Feeling all the (Partisan) Feels

Exploring the Drivers of Affective Polarization at the Individual Level

Thesis submitted in partial fulfilment of the requirements for the degree of DPhil in Information, Communication and the Social Sciences, Oxford Internet Institute, University of Oxford.

Nahema Marchal
Oxford Internet Institute, University of Oxford

Supervisors:
Dr. Jonathan Bright, Prof. Philip N. Howard
2021

Word count: 64,337

Abstract

Affective polarization—hostility and aversion between rival political groups—is an increasingly salient feature of Western democratic politics. While scholars agree that this trend raises serious concerns for public life, its underlying drivers are still debated. This thesis enriches this line of research by clarifying some of the mechanisms shaping affective polarization at the individual level. My overarching research question is: *What drives affective polarization at the individual level*? I tackle this question through three distinct empirical studies that engage with separate aspect of this phenomenon. Study I leverages survey data to show that, among UK partisans, perceived conflict with their in-party on policy issues curbs affective polarization. Moving on to the digital sphere, Study II deploys computational methods to reveal the existence of affective biases in online political discussions between ideologically opposed Reddit users, and shows how the tone of their exchanges can both mitigate and worsen these patterns. Study III, finally, provides experimental evidence that only exposure to certain forms of online partisan messages have downstream consequences for affective polarization: those that explicitly threaten and devalue an individual's own political party. Taken together, these findings reveal the drivers of affective polarization to be contingent on both situational and dispositional factors, including how ideologically aligned partisans are with their own political party, how their perceived political rivals engage with them in online conversations, and whether they are made to feel threatened by them.

Table of Contents

List of Tables 7

List of Figures 9

Acknowledgments 10

Chapter 1. Feeling all the (Partisan) Feels 12

1.1 Introduction 12

1.2 Polarization: A Tale of Two Narratives 16

1.2.1 Politics in a World of 'Tribes' 16

1.2.2 'Social Media Is Driving Us Apart' 18

1.3 Research Questions 20

1.4 Key Definitions 22

1.4.1 Affective Polarization 22

1.4.2 Digital Political Communication 23

1.5 Outline of Thesis 25

1.6 Significance of Research 27

Chapter 2. Bringing 'Affect' Back Into the Study of Affective Polarization and Its Drivers 29

2.1 Introduction 29

2.2 Affective Polarization in Context 30

2.2.1 Mass Polarization: from Ideology to Affect 31

2.2.2 Affective Polarization: Manifestations 36

2.2.3 Political and Social Consequences of Affective Polarization 39

2.3 Locating the Drivers of Partisan Animosity 41

2.3.1 Two Explanatory Models: Policy Versus Identity 42

2.3.2 Beyond Identity Versus Ideology: Re-centering the Role of Affect 45

2.4 Towards an Affective Perspective on the Drivers of Affective Polarization 48
2.4.1 Attitudinal Factors 49
2.4.2 Online Factors 51
2.4.3 The Missing Piece: Bringing 'Affect' Back In 55
2.5 Conclusion 59
Chapter 3. The Paradox of Poor Representation: How Voter-Party Incongruence Curbs Affective Polarization 60
3.1 Abstract 60
3.2 Introduction 61
3.3 Ideological Disagreement and Affective Polarization 63
3.4 The Consequences of Voter-Party Incongruence for Affective Evaluations 66
3.5 Data & Methods 70
3.6 Measures 74
3.7 Results 76
3.8 Discussion 80
3.9 Conclusion 83
3.10 Appendix A 84
Chapter 4. 'Be Nice or Leave Me Alone': An Intergroup Perspective on Affective Polarization in Online Political Discussions 91
4.1 Abstract 91
4.2 Introduction 91
4.3 Affective Polarization and Political Group Identity 93
4.4 Affective Dynamics and Intergroup Communication 95
4.5 Data and Methodology 100
4.5.1 Estimating Users' Ideological Leaning 102

4.5.2 Sentiment Scores 105
4.6 Measures 106
4.6.1 Dependent Variable 106
4.6.2 Predictors 107
4.6.3 Controls 107
4.7 Models 108
4.8 Results 109
4.9 Discussion 114
4.10 Conclusion 116
4.11 Appendix B 117
Chapter 5. 'Who Do You Think You Are?' The Differential Effect of In-Party Threats and Reassurances on Affective Polarization 125
5.1 Abstract 125
5.2 Introduction 125
5.3 Online Political Talk and Affective Polarization 127
5.4 Partisan Prejudice and Threat Perceptions 129
5.5 Methodology 134
5.5.1 Sample Size & Recruitment 134
5.5.2 Study Design 135
5.6 Measures 139
5.7 Results 140
5.8 Discussion 147
5.9 Conclusion 149
5.10 Appendix C 150
Chapter 6. Conclusion 158
6.1 Introduction 158

6.2 Summary of Findings.....159
6.2.1 Is Partisan Identity One Hell of a Drug? 159
6.2.2 The Limits of Partisan Hostility 161
6.2.3 On the Role of Emotions 163
6.3 Strengths and Limitations 165
6.4 Lessons and Future Directions 171
6.4.1 Looking Beyond the Echo Chamber..... 172
6.4.2 Reducing Affective Polarization Through Targeted Intervention 172
6.5 Concluding Remarks 174
References 175

List of Tables

Table 3.2. Impact of voter-party incongruence on affective evaluations of political parties and candidates 77

Table 3.3. Impact of voter-party incongruence on affective evaluations of political parties and candidates, by strength of partisanship 78

Table 3.4. Impact of voter-party incongruence on salient issues on affective evaluations of political parties and candidates 79

Table 4.1. Distribution of interaction types 105

Table 4.2. Effects of interaction type on average sentiment 109

Table 4.3. Effect of interactions features on likelihood of follow-up crosscutting negative response or conversation end 110

Table 4.4. Effect of interaction features on the likelihood of follow-up crosscutting positive response 112

Table 4.5. Effect of likeminded interaction features on intragroup affective bias 113

Table B.1. Coding scheme for the classification of subreddits 118

Table B.2. Confusion matrix for validation of estimated ideological leanings .. 120

Table B.4. Complete list of political entities 122

Table B.5. Performance of VADER sentiment analysis 122

Table B.6. Descriptive statistics for comment sentiment 123

Table B.7. Likelihood of a crosscutting interaction to be followed up by another crosscutting one 124

Table 5.1. Treatment design 136

Table 5.2. Impact of exposure to partisan praise and partisan criticism on affective polarization measures 142

Table 5.3. Impact of exposure to partisan praise and criticism on affective polarization measures, by strength of partisanship 144

Table 5.4. Impact of exposure to partisan criticism on affective polarization measures, by party affiliation 146

Table C.1. Power analysis 152

Table C.2. Demographic distribution of experimental participants 153

Table C.3. Summary statistics for outcome variables 154

Table C.4. Additional OLS regression models 156

Table C.5. Additional logistic regression models 157

List of Figures

Figure 4.1. Model of Reddit discussion thread 101

Figure 4.2. Ideological alignment of Reddit users 104

Figure B.1. Example of 'ideology-based' liberal leaning subreddit 119

Figure B.2. Example of 'figure-based' conservative leaning subreddit 119

Figure 5.1. Examples of treatment stimuli 137

Figure C.1. Average knowledge evaluations and out-party ratings per treatment and partisan group 155

Acknowledgments

I must start by acknowledging my supervisors Jonathan Bright and Philip Howard. I could not have completed this doctoral project without your mentorship and patient guidance over the last three years. Thank you for encouraging me to believe in myself, even when I was filled with doubts, and for helping me develop as a scholar. Furthermore, I am grateful to Victoria Nash, Yptach Lelkes and Claes de Vreese for agreeing to serve as examiners on this thesis. Your valuable comments and insights have helped shaped this work into what it is today.

This has been a long journey and one I would probably not have embarked on had it not been for Jonathan Levin. Thank you for pushing me to apply to the Oxford Internet Institute (OII) and for being the kindest friend one could ever ask for. I am also grateful to Odysseas Sclavounis for guiding me through the early stages of the application process, when my doctoral ambitions still revolved around the democratic potential of blockchain technology (who would have thought!). Thank you, finally, to my old friends Meghan Robinson and Joseph Dodd for being critical readers of my research proposal.

I owe an enormous amount of gratitude to my academic community at the OII. I have met some of the most inspiring and generous people in the corridors of 41 St Giles and watching my peers grow around me has been an immensely rewarding experience. I could write entire paragraphs about each person who has lent me a helping hand, but space constraints force me to only name a few. I am first thankful to David Watson—my longer-lasting (and most faithful) co-author. I have learned so much from working with you and will remember fondly our late-night chats in the pubs of Oxford.

I would also like to acknowledge all the wonderful ladies who I have had the privilege to work with (and learn from!) over the years, many of whom offered me time and tea, and provided thoughtful feedback on various parts of my work: Kate Sim, Corinne Cath-Speth, Mona Elswah, Lisa-Maria Neudert, Nicole Au, Cindy Ma, Amelia Hassoun, Erin Simpson, Samantha Bradshaw, and Rutendo Chabikwa. You are all absolute legends, fantastic scholars and allies, and I feel incredibly lucky to count so many of you as friends.

I would also like to thank Bertie Vidgen, Sian Brooke and Jack LaViolette for their help and guidance as I navigated the choppy waters of 'Big Data' analytics for the first time. Thanks also go to my colleagues and friends from the Programme for Democracy and Technology, especially Hubert Au and Bence Kollyani, for being patient guides in the dark arts of computational social science and for always putting a smile on my face.

To my Oxonian friends, Andrew McLean, Namratha Rao, Lily Bussman, Rachel Benoit and Barclay Bram: thank you for nourishing me (in every sense of the word) and for bringing so much joy into my life as a doctoral student.

The greatest appreciation of all goes to my husband Moritz Kraemer. Without your unwavering love and support, this thesis would never have seen the light of day. You are an endless source of inspiration—both as an academic and as a life partner. Thank you for standing beside me through thick and thin. I joked to you many moons ago that if I ever completed a doctorate and we stayed together long enough, you would make it into my acknowledgements section. I could not be happier to be typing these words today.

Last but not least, I would like to thank my parents, Adiaratou Sarr and Jean-Pierre Marchal. Thank you for instilling in me the love of learning, always encouraging me in my pursuits, and teaching me the value of patience and determination. Without your sacrifices, I would not be where I am today. I love you both dearly.

Finally, and because God only knows how hard it has been to wrap up this thesis in the most tumultuous of times, I would like to end these acknowledgements with a quote from the 20th century wordsmith Calvin Cordozar Broadus Jr., who once dared to say:
'I wanna thank me.
I wanna thank me for believing in me.
I wanna thank me for doing all this hard work.
I wanna thank me for having no days off;
I wanna thank me for never quitting.'

1 Feeling all the (Partisan) Feels

1.1 Introduction

On a sunny Tuesday afternoon in October 2020, Democratic presidential candidate Joe Biden made his way to the stage where he was due to address a small crowd of journalists and voters, and took his mask off. Overlooking the battlefield in Gettysburg, Pa. where Union and Confederate forces had fought centuries earlier—altering forever the course of the American Civil war—Biden deplored the cost of division in contemporary America and made a plea for unity. The country, he warned, was 'is in a dangerous place' (Biden, 2020):

> Our trust in each other is ebbing. Hope seems elusive. Too many Americans see our public life not as an arena for mediation of our difference but, rather, they see it as an occasion for total, unrelenting partisan warfare. Instead of treating each other's party as the opposition, we treat them as the enemy. This must end.

This quote illustrates a widespread sentiment among the political and academic class. American society, we are told, is deeply polarized (Drutman, 2020; Klein, 2020). The last four years of US politics have brought into sharp relief the extent of the ideological chasm between Republicans and Democrats on a wide range of issues, from immigration to national security, healthcare and climate policy. So profound is the perceptual gap between members of both parties that Donald Trump's presidency, in fact, set the record for the largest difference in job approval ratings ever recorded by Gallup (79-percentage points), topping the already polarized ratings from Obama's years in the White House (Jones, 2019).

US partisans not only disagree over the best way to run the country. Today, many also despise each other—to a level hitherto unprecedented—and that political scientists warn poses serious threats to democratic cohesion (Iyengar et al., 2019). This phenomenon, commonly referred to as *affective polarization*, manifests in a number of ways, including intense dislike and animosity between partisans, and greater stigmatization (Iyengar et al., 2012). Since the 1980s, for example, Democrats and Republicans have expressed increasingly hostile attitudes towards the opposite party and its members (Iyengar & Krupenkin, 2018). Rank-and-file partisans are also more likely to hold unfavorable opinions and negative

stereotypes of their political rivals (Iyengar et al., 2012; Kalmoe & Mason, 2019). In a telling video filmed at the height of the 2016 US presidential election, reporters from Jimmy Kimmel Live asked attendees of the Democratic and Republican Conventions what they thought about members of the opposite party. At the RNC, a self-identified Republican derided Democrats as 'brainwashed,' while another one referred to the entire Democratic Party as a 'crime syndicate.' The tone was equally acrimonious at the DNC. When asked about their opinions of Republicans, Democrats described them in turn as 'close-minded', 'racist', and 'mean' (Jimmy Kimmel Live, 2016).

The level of partisan animosity captured in these comments has only intensified over the last four years. In 2019, 55% of Republicans and 47% of Democrats said members of the opposite party were more 'immoral' than the average American—up a respective 8 and 12 percentage points since 2016 (Doherty et al., 2019a). Seventy-five percent of Democrats also said that Republicans were more 'close-minded' than other citizens, while 64% of Republicans said the same about Democrats (Doherty et al., 2019a). The 2020 presidential election further underscored the pervasiveness of these divisions: around the time that Biden gave his Gettysburg speech, approximately eight-in-ten voters from both parties said their differences with the other side were not just about facts, but about core American values (Dimock & Wike, 2020).

The United States is far from the only country facing these polarizing pressures. According to recent research, affective polarization among ordinary citizens is just as high, if not higher, in some parts of Europe as in the US (Gidron et al., 2019b; Reiljan, 2019; Wagner, 2021). In comparative terms, partisans' tendency to express dislike towards their political opponents is most pronounced in Eastern and Southern European countries, such as Bulgaria, Portugal, Spain and Greece (Reiljan, 2019). Still, French and British partisans are about as affectively polarized as their American counterparts (Gidron et al., 2020; Hobolt et al., 2020). In fact, in the UK, Belgium, and Spain, partisan animosity is so strong that voters are now more likely to discriminate against each other on the basis of partisanship than on the basis of religious, linguistic or ethnic differences (Westwood et al., 2018).

Beyond traditional partisan divides, in recent years the rise of populism across Europe and a polarizing Brexit vote have also generated new affective cleavages. Over the last decades, many European democracies including Austria, Switzerland, and Germany for example, have seen the ascent of anti-establishment parties that use hostile rhetoric against mainstream political actors and their supporters, prompting in turn intense dislike from their opponents (Gidron et al., 2019b; Helbling & Jungkunz, 2020). Today, voters' willingness to interact with and trust each other in daily life is thus strongly determined by where they see others falling on the populism spectrum (Gidron et al., 2019b; Helbling & Jungkunz, 2020). Social identities formed in the wake of the 2016's referendum on European Union membership have led to similar patterns in the UK. In the aftermath of the vote, citizens' perceptions of each other became strongly colored by whether they were a 'Leave' or a 'Remain' voter (Hobolt et al., 2020). Ordinary citizens who voted for 'Leave' for example were more likely to view likeminded voters in a favorable light—calling them 'intelligent', 'honest' and 'open-minded'—while attributing negative characteristics to their opponents, and vice versa (Hobolt et al., 2020).

As I detail at length in the next chapter, there are good reasons to be concerned about the rise of affective polarization in the general public. Partisan animosity is already linked to a wide range of negative social and economic outcomes, including lowered interpersonal trust and cooperation. Affectively polarized citizens are, for example, less likely to trust their political opponents, interact with them socially (Carlin & Love, 2018; Martini & Torcal, 2019) and compromise on matters of public affairs (Levendusky, 2017). Negative partisan affect is also a strong predictor of discrimination in hiring and of preferential treatment of co-partisans in resource allocation (Gift & Gift, 2015; McConnell et al., 2018). Finally, there are also suggestions that affective polarization could have damaging consequences for democracy in the long-run. In a prescient paper, McCoy and Somer (2019) note that the greatest danger of affective polarization is that it contributes to growing perceptions among citizens that their political opponents 'pose a threat to the nation' or to their 'way of life.' More worryingly for democracy, they argue, these perceptions of threat 'open the door to undemocratic

behavior by an incumbent and his/her supporters to stay in power, or by opponents to remove the incumbent from power' (p. 258).[1]

With so much at stake for social and democratic cohesion, understanding what drives affective polarization emerges as a critical research task. While scholars agree that affective polarization is on the rise in several Western democracies and have advanced several explanations for this pattern, as I detail in my literature review, there is still little consensus about the mechanisms underlying this phenomenon. This thesis contributes to this emerging literature by clarifying some of these mechanisms at the individual level. Doing so is important for a number of reasons. First, getting a better sense of the circumstances under which partisans develop hostile attitudes towards their political counterparts can help us develop better explanatory frameworks for this phenomenon at the mass level. Furthermore, identifying what factors accentuate people's tendency to discriminate against their political rivals can teach us a lot about potential ways to mitigate polarization.

My thesis engages with these debates through three distinct studies that each touch on a separate aspect of my overarching research question: *What drives affective polarization?* Through this work, I demonstrate that the individual drivers of affective polarization are contingent upon both situational and dispositional factors, including how ideologically aligned someone feels with their own political party, how their perceived political rivals engage with them, and whether or not they are made to feel threatened by them. A more detailed outline of my research questions and empirical studies is provided in section 1.3.

The remainder of this chapter is structured as follows. In the next section, I review two popular narratives that have recently emerged to explain current levels of partisan hostility—the rise of 'tribalism' in public life and the divisiveness of social media—and show how their blind spots inform and motivate my inquiry. In 1.3., I then outline the main goals of the thesis and the principal questions that my empirical studies seek to answer. After defining key concepts that appear

[1]Though no formal link has yet been established between these events in the literature, the storming of the Capitol Building in Washington D.C. on January 6th, 2021 by thousands of Trump supporters who refused to accept the outcome of the 2020 presidential election was widely framed worldwide as the culmination of years of partisan rancor and animosity in the US (Bremmer, 2021).

throughout subsequent chapters in 1.4, I conclude by presenting the overall structure of the thesis and its contributions to the field.

1.2 Polarization: A Tale of Two Narratives

1.2.1 Politics in a World of 'Tribes'

In recent years, several popular narratives have begun to emerge about how and why Western societies have become polarized in this more affective and visceral manner. For some scholars and pundits, today's political fissures are symptomatic of a shift to a new form of 'sectarianism': a religious-like devotion to one's political identity, an excessive belief in the moral superiority of its ideals, and a profound disdain for those that appear to threaten it (Finkel et al., 2020). Another popular concept frequently used to describe these dynamics is that of 'tribalism' (Chua, 2018; Mason & Wronski, 2018; Whitt et al., 2020). Following the UK's decision to leave the European Union in 2016 for example, newspaper headlines frequently called upon politicians to 'bring the country back together' to heal a 'bitterly divided' and more 'tribalized nation' (Duffy et al., 2019). Crucially, in this framework, political tribalism is often presented as an innate and natural tendency, to be contrasted with the learned behavior of democratic compromise.[2]

For supporters of this view, the deterioration of partisan relations in the US reflects profound, structural changes to the US political system and to American society at large. One such change is the rise of identity politics (Sides et al., 2018) and the crystallization of the cultural wars along party and religious lines (Castle, 2019). In the US, political affiliation has not only become increasingly aligned with cultural values (Mason, 2018a), but is also closely matched with geographical location, economic status, and religious preferences: affluent socially liberal folks tend to live in large, coastal cities while more conservative folks tend to live in smaller towns (Drutman, 2020). The convergence of these factors, for these authors, has led to the emergence of politicized supra-identities, pitting ever larger sections of society against each other and fostering a climate of 'zero-sum

[2] In a widely watched Ted Talk entitled 'How Partisan Tribalism is Killing Democracy', for example, Brian Klass confidently states that: 'We, humans, are not naturally democratic animals. We are naturally tribal. Democracy isn't hard-wired into our brains' (TED, 2018). Amy Chua's book *Political Tribes: Group Instinct and the Fate of Nations,* begins by the sentence: 'Humans are tribal' (Chua, 2018, p.2).

competition, in which parties succeed by stoking voters' fears and appealing to their ugliest us-versus-them instincts' (Chua & Rubenfeld, 2018). While the perspective summarized above is deeply rooted in the socio-political context of the US, a similar narrative has begun to take hold in other parts of the Western world. In France, as in the UK, prominent politicians and intellectuals increasingly lament the encroachment of American-style politics and the growing focus on questions of identity (notably gender, religion and race) in public discourse, which they say fuels communitarianism and erodes national unity (Casalicchio, 2021; Onishi, 2021).

Offering the rise of tribal politics and sectarianism as an explanation for current levels of affective polarization is intuitive. By relying on universalizing assumptions about people's political behavior and psychology, however, this type of overarching narrative runs the risk of glossing over important contextual factors and alternative theories. Psychology is an evolving field and advances in neuroscience have pushed scholars to continuously update and refine their assumptions about how people come to see themselves in the way they do, how they develop the opinions that they have, and how these processes relate or differ from what we can measure in the observable world (Maloney, 2018). For that reason and others that I discuss at length in the literature chapter, it is actually far from clear that 'tribalism' is the most appropriate construct to explain affective polarization, especially at the individual level. Other scholars, for example, suggest that partisan hostility is not so much an expression of people's loyalty to their political 'tribes', as a reflection of their disdain for their opponents' beliefs (Bougher, 2017; Webster & Abramowitz, 2017).

Before drawing conclusions about these mechanisms at the mass-level, it is therefore essential to further investigate whether antipathy between partisans truly stems from an innate tendency to dislike outgroups, or whether it has a rational component. Doing so will not only inform current theoretical understandings of affective polarization, but is also an important step towards identifying the types of interventions that may be most effective at curbing it. At the heart of it also lies a fundamental question about the nature of democracy itself, and whether it can survive this moment. The 'totalizing, regulatory and 'othering' characteristics of identity' (W. Brown, 1993) are often seen as incompatible with the level of recognition and crosscutting engagement required in functioning democracies. If

partisans held each other in contempt based on tribal tendencies alone (whether or not they form part of the same 'kin'), this means we may struggle to ever truly transcend our differences. If, on the other hand, partisan hostility had a strategic component—in that it was indicative of a desire to defend one's political ideas—then there might be a possibility for us to 'compromise' our way out of hatred and find mutually agreeable solutions. I discuss this and related ideas in Chapter 2 and 3.

1.2.2 'Social Media Is Driving Us Apart'

With the rapid expansion of social media as vehicles for political discourse, another narrative has begun to emerge. Following a series of unprecedented political events, starting with the 2016 US presidential election, journalists and scholars have increasingly pointed to digital technology—and particularly social media—as the cause for our growing divisions. Algorithms have been accused of 'destroying' democracy by pushing people to retreat into ideological bubbles, fomenting hatred, precipitating the UK's exit from the European Union (Hern & Sabbagh, 2018) and bringing Donald Trump to power (Rosenberg et al., 2018).[3] This narrative has become so popular and is so frequently amplified in the press that it is rarely goes questioned. Yet, scholars are only just beginning to fathom how complex socio-technological systems like social media platforms impact us, our psychology, and our politics.

Digitalization has fundamentally altered the way people access and communicate political information. A growing number of people access information about politics and public affairs through digital means, making social media a primary gateway to political information (Newman et al., 2020). Whether through platforms like Twitter and Facebook, or discussion forums like Reddit, individuals are increasingly taking to social media to voice their opinions and engage in new forms of civic expression and participation (Hill, 2018; Margetts et

[3]President Barack Obama's long-term advisor and legal scholar Cass Sunstein was instrumental in popularizing the view that social media bore a large responsibility for growing divisions in the American electorate. Discussing the role of social media in the 2016 US elections, Obama said 'the capacity to disseminate misinformation, wild conspiracy theories, to paint the opposition in wildly negative light without any rebuttal—that has accelerated in ways that much more sharply polarize the electorate and make it very difficult to have a common conversation' (Remnick, 2016).

al., 2015; Vaccari, 2017). Social media networks empower large segments of society to coordinate political action and are widely recognized as having played an integral role in the mobilization of mass grassroots movements, from the Arab Spring to the Gilets Jaunes protests (Gerbaudo, 2012; Howard & Hussain, 2011). By lowering the costs of information dissemination, finally, these platforms have also emerged as essential vehicles of communication between civil society and political actors.

Yet the initial optimism that accompanied the mass uptake of information and communication technologies has recently been replaced by a 'techlash'—widespread public concerns about their negative impact on society. The Cambridge Analytica data privacy scandal and the meteoric rise of the disinformation-for-hire industry (Silverman et al., 2020) have both shown how easily these digital tools can be manipulated for questionable political ends.[4] Theoretically, there are also good reasons to believe that digital media could be exacerbating affective polarization. Partisan incivility and divisive messaging are common on some corners of the internet and could likely contribute to negative attitudes, while the homogeneity of some users' networks could warp their perceptions of political reality (Garrett et al., 2014; Levendusky, 2013). Today, for example, two-thirds of US adults say they find it 'stressful and frustrating' to talk about politics on social media with people they disagree with, often leaving them feeling like they have 'less in common' with others (M. Anderson & Auxier, 2020).

Whether internet and social media use does have polarizing effects on the electorate, however, is still somewhat unclear and, as I demonstrate in later chapters, the field is far from settled on the matter. Scholars in particular do not agree on what precisely is polarizing about the experience of being online: whether it is the content that people come across, the frequency with which they do so, the nature of their social interactions with politically similar and different others, or some combination of the all the above (Barberá, 2020; Tucker et al., 2017). Answering these questions in any conclusive way also entails overcoming significant methodological limitations. Even accurately measuring exposure to

[4] In the aftermath of the election, a series of government, journalistic, and academic inquiries uncovered incriminating evidence that the Internet Research Agency (IRA) had orchestrated a sophisticated social media operation, posting polarizing content across multiple platforms with the aim of disfavouring Hillary Clinton (Howard et al., 2018; *United States of America v. Internet Research Agency*, 2018).

political messaging in a multimedia environment for example is, in itself, a thorny task (de Vreese & Neijens, 2016). The rapidly changing nature of these technologies also complicates any efforts to draw causal conclusions. Access to reliable social media data for researchers, finally, is still sparse, unevenly distributed, and reliant on the whims of technology companies (Freelon, 2018).

Bringing empirical clarity to these questions, however, is not only a worthwhile academic endeavor but also a matter of public concern. In the current climate, misguided or exaggerated beliefs about the impact of social media on politics can lead to rushed and ill-conceived policies that impinge on individual freedoms and digital rights (The Economist, 2021). Recent work also shows that it could seriously hamper trust in institutions and satisfaction with democracy at large (Nisbet et al., 2021). I come back to these debates in Chapter 4 and in the concluding chapter of the thesis.

1.3 Research Questions

As the above discussion makes clear, how and why several Western democracies have become more affectively polarized over the last decades is still an unsolved puzzle. Scholars have already advanced several explanations for these patterns at the mass-level—from growing 'tribalism' in public life to the rise of social and digital media. At this point, however, we still lack a sophisticated understanding of the mechanisms underlying this phenomenon at the *individual-level*. It is still unclear, for example, whether one's allegiance to a political party is more important than one's policy preferences in shaping affective evaluations. Likewise, empirical work has so far failed to specify the conditions under which partisans may be more or less motivated to act in a hostile manner towards their political opponents. This is where this doctoral thesis seeks to make a contribution. Drawing on theories that center the role of affect, emotions and motivational processes in political thinking and behavior, in the following chapters I explore how individuals' characteristics, on the one hand, and the nature of their online political interactions, on the other hand, impact affective polarization.

My first empirical study focuses on the role of individuals' political attitudes versus their political identity in driving affective polarization. As I detail at length in the literature chapter, whether partisan hostility is best explained by political

loyalty or policy conflict is a source of ongoing debate. The goal of this study is to bring clarity to this question by examining what happens when these two motives stand in tension with each other within a single individual. Do partisans who disagree with their party's policies also report negative affect towards the opposition? And does this incongruence challenges how they feel towards their own party? The main research question guiding this study is thus:

- *How does incongruence between voter and party on policy issues impact affective evaluations of political parties and their leaders?*

In the second half of the thesis, I then turn my attention to the nature of people's political interactions in digital settings, with a particular emphasis on the content of their communications. As digital technology gets more and more embedded into our everyday lives—mediating what we see and how we communicate with one another—it is imperative to further elucidate which aspects of online political talk have the capacity to exacerbate affective polarization. It is still unclear, for example, whether certain types of exchanges have worse consequences for partisan affect than others. My second and third empirical studies speak to this goal, by addressing the following research questions:

- *Are online political discussions between different users affectively polarized? And if so, what communicative factors drive these affective biases?*
- *What types of online partisan comments drive affective polarization?*

Taken together, all three studies form an integrated whole. While each looks at affective polarization from a slightly different dimension, they all take the individual as a unit of analysis and draw on theories of intergroup relations and social identity to their build hypotheses.

The analyses presented in the following chapters are restricted to two countries, namely the United Kingdom (Chapter 3) and the United States (Chapter 4 and 5). These two countries present important similarities that allow for generalization across cases. Yet, they also present meaningful differences that allow for interesting comparative conclusions. The UK, just like the US, is a

multicultural, advanced democracy with a history of rancorous politics between two dominant political parties: The Labour and the Conservative Party. The UK, in fact, is arguably the closest to having a two-party system in Europe (Gidron, 2020). In the UK, traditional cleavages along the left-right ideological spectrum and across party lines still dominate on multiple issues including the UK's relationship with Europe, education, and the economy (Schumacher, 2019). Both countries also exhibit equivalent levels of affective polarization (Gidron et al., 2019a). Contrary to the US, however, UK voters have shown signs of high electoral volatility in recent elections and, although partisanship is high, it is lower on average than in the US (British Election Study, 2019).

Having set out my goals and objectives, in the section below I turn to defining some key terms and concepts that will be used throughout the thesis. I conclude this section by providing an outline of the upcoming chapters.

1.4 Key Definitions

1.4.1 Affective Polarization

In political science, polarization traditionally refers to the process whereby two groups with contrasting views become more extreme in their respective opinions, thus diverging on the ideological spectrum (Isenberg, 1986). Though related, *affective polarization* draws on the sociological concept of affective social distance, which centers on individuals' psychological orientations towards other groups of people. Following this logic, affective polarization is commonly defined as a form of intergroup prejudice: partisans' tendency to evaluate and behave more positively towards their co-partisans, while discriminating and holding more hostile attitudes towards their political opponents (Iyengar et al., 2012; Iyengar & Westwood, 2015). Affective polarization has thus two distinct components—ingroup favoritism and outgroup hostility—which can be captured along two dimensions: a behavioral and an evaluative one.

In the literature, the term is frequently used to refer to both a state and a process. As a state, affective polarization refers to the extent to which partisans display favoritism towards their fellow partisans and prejudice towards the opposition (Iyengar et al., 2019). Accordingly, the more affectively polarized someone is, the more differently she will behave towards her political friends and

foes. However, the term is also often employed to describe the process whereby partisans in Western mass publics have become increasingly hostile towards each other since the mid twenty-first-century (Iyengar et al., 2012). Unless specified otherwise, throughout the thesis I will be using the terms *affective polarization, partisan prejudice* and *partisan animosity* interchangeably to refer to the former: partisans' differential treatment of their co-partisans and those they perceive as political rivals.

As we will see in the next chapter, until now affective polarization has mostly been examined through the prism of partisanship. Most studies in the literature thus document affective biases between members of opposing political parties. There are a couple of exceptions to this rule, however. Recently, researchers have started to show that people also treat differently those who they see as belonging to rival ideological groups, such as 'conservative' or 'liberal', 'left-wing' or 'right-wing' (Mason, 2018b), rival political coalitions (González et al., 2008) and rival opinion based-groups, such as pro-'Leave' and pro-'Remain' (Hobolt et al., 2020). For the sake of clarity and given that the majority of my empirical studies are concerned with party identifiers, throughout the remainder of the thesis I refer to all these other forms of intergroup prejudice as *partisan* prejudice.

1.4.2 Digital Political Communication

There are many definitions of political communication, grounded in different disciplinary traditions and shifting understandings of what the most relevant object of study is (Jamieson & Kenski, 2017). If one were to synthesize these definitions, however, the term *political communication* would generally describe the process of creation, exchange, and dissemination of political information between different sets of actors, which include citizens, political and government officials, and the media (Jamieson & Kenski, 2017). In the last two empirical studies of this thesis, I am particularly interested in the interpersonal dimensions of political communication, that is the more informal exchange of political information between citizens themselves.

Following this definition, how should one then understand *digital political communication*? Traditional face-to-face communication with friends, family, and

colleagues is still the most dominant form of interpersonal interaction. As digital technology increasingly permeates our lives, the boundaries between digital and analog social spaces are becoming more and more porous. Computer-mediated communication, however, differs from face-to-face communication in important ways. While the latter is oral, synchronous, and very rich in visual and social cues, up until recently the former was mostly text-based and asynchronous (although this is now changing with the advent of video-conferencing software, and image- or audio-centric platforms such as Instagram and Clubhouse). This form of communication can be particularly conducive to political expression and participation (Shah, 2016; Vaccari & Valeriani, 2018). The architecture of online platforms also affords users the option to disclose or hide their identity: while anonymity is possible on Twitter and Reddit, for example, Facebook profiles tend to reveal much more information about its users, which serve as contextual cues (Valkenburg et al., 2016).

Digitalization also enables users to interact with a much broader network of individuals than they would in real life. Every day, millions of people can go online and broadcast their opinions to large audiences of known and unknown others or choose to engage with a selected few based on topical interests, forming ephemeral 'network publics' (boyd, 2010). Another important feature of digital political communication is the highly social and networked nature of information exchange. On the internet or over social media, users can share their own take on political events, comment on each other's posts, share their reactions to a news article, and amplify that content to their own social network. This creates new dynamics of social influence that impact the reception of political messages (Anspach, 2017). In that sense, 'the online social world is simultaneously a reflection of the offline world of politics, a facilitator of offline politics, and a space in which politics occurs' (Carpini, 2020, p. 317). Lastly, political interactions in the digital space tend to be quite frequent and incidental. The social and algorithmic logic that powers most digital platforms leads habitual users—even the least politically interested ones—to regularly come across political content even when they do not actively seek it (Jungherr et al., 2017; Settle, 2018a).

In this thesis, I define and conceptualize *digital political communication* as online political talk; that is the informal exchange of messages or posts of political nature in a digital setting, characterized by the features described above.

1.5 Outline of Thesis

In the chapters that follow, I return to the key debates presented in this introduction. In Chapter 2, the literature review, I begin by critically examining existing scholarship on affective polarization and its drivers. I divide the current literature on this question into two distinct and competing schools of thoughts—one which explains partisan prejudice through the lens of policy preferences, and another one which explains it through the lens of identity conflict. Throughout the chapter, I argue that, while these explanatory models have generated useful findings on what shapes affective polarization, their overreliance on cognitive factors makes them ultimately quite reductive. I suggest instead that better incorporating the role of affect into models of political thinking and judgment will work to enrich theory on the underlying drivers of partisan prejudice.

The study presented in Chapter 3, 'The Paradox of Poor Representation: How Voter-Party Incongruence Curbs Affective Polarization', first engages with the question of the micro-foundations of affective polarization.[5] Drawing on secondary survey data obtained from the British Election Studies in 2019, the paper investigates how perceived incongruence between party and voter on policy issues impacts evaluations of that party and the opposition. Estimating causal effects using observational data is notoriously tricky, especially in the absence of a randomized experimental set up that would allow for controlled manipulation of the independent variables. To circumvent this issue, an instrumental variable (IV) approach was implemented, using real voter-party congruence as the instrument. Using this technique, this paper demonstrates that perceived incongruence over policy issues is, in fact, causally linked to lower levels of affective polarization. By showing that, at least among UK partisans, perceived ideological conflict with the in-party actually improves feelings towards the opposition, these findings thus

[5] This chapter is an extended version of a paper I co-authored with my colleague David Watson. As lead author, I conceived and designed the study, performed most of the data analysis and drafted the manuscript. A statement from my co-author detailing both of our contributions is attached in *Appendix A*.

problematize the notion that blind partisan allegiance is the driving force of partisan prejudice.

In Chapter 4, I turn my attention to digital political communication, and explore how various aspects of discussing or being exposed to political talk online impact users' affective dispositions towards each other. My second empirical study 'Be Nice or Leave Me Alone: An Intergroup Perspective on Affective Polarization in Online Political Discussions' is an observational study of conversation dynamics on one of the most active forums for discussion of US politics: Reddit's *r/politics.* In this paper, I investigate the extent to which interactions between politically different users are characterized by affective prejudice, and explore what communicative features mitigate or worsen these conversational dynamics. My analyses demonstrate that online users indeed display affective biases when interacting with likeminded users and with users who are opposed to them. By examining not only the ideological relationships between users on the forum, but also the tone and target of their exchanges, I demonstrate that the intersection of these two elements provide valuable information about what does and does not trigger intergroup hostility. Results from my analyses show that interactions between users who leaned towards opposite sides of the ideological spectrum were also more likely to be cut short than sustained if one user referred negatively to the other's political ingroup. Conversely, crosscutting interactions in which one of the users expressed positive sentiment towards the other one's group were more likely to attract a positive than a negative response, thus mitigating intergroup prejudice.

Chapter 5 builds on the insights generated in the previous chapter and further explores the relationship between social media use and affective polarization through an experimental design. In this chapter, 'Who Do You Think You Are? The Differential Effect of In-Party Threats and Reassurances on Affective Polarization', I present evidence from an online survey experiment conducted on 499 US partisans in December 2020 to test the effects of exposure to partisan praise and criticism in online comments on affective polarization. The findings from this study reveal that, when it comes to online partisan commentary, not all messages have the same effects. Indeed, in this experiment only certain forms of partisan comments had an impact on affective polarization: those that threatened or devalued individuals' own political affiliation. By specifying the

conditions under which exposure to online partisan talk is most polarizing, this paper thus brings nuance to ongoing discussions about the effects of social media on political attitudes.

In the concluding chapter, I summarize the findings from my three empirical studies and reflect on the lessons that can be drawn from them in light of the initial research questions posed. This conclusion chapter also serves as a discussion of this thesis' limitations and outlines remaining tasks and possible avenues for future research.

1.6 Significance of Research

The contributions of my thesis to the literature are three-fold. Whereas dominant explanations of affective polarization tend to rest on universalizing assumptions about human psychology and partisanship (for example, that out-party hostility is ingrained in partisans' minds), my thesis investigates and brings to light specific mechanisms that shape the expression of partisan animosity at the individual level. Through this work, I provide empirical evidence that partisan relations do not amount to a zero-sum game; in fact, the extent to which partisans are willing to act favorably towards their co-partisans and in a hostile way towards their political rivals is highly context-dependent. Specifically, I show that this hinges on both situational and dispositional factors, including how aligned partisans feel with their own political party on policy issues, what happens when they interact with politically-different others online, and whether these interactions are threatening or amicable.

By bringing insights from political communication and social psychology—and specifically from the intergroup relations literature—to bear on the study of affective polarization, my work also advances current debates about the potential role of social media in furthering the affective divide between US partisans. Specifically, this thesis shifts the focus away from concerns about lack of ideological diversity or 'echo chambers' as the driving forces of polarization to hone in instead on the nature of people's political encounters with others online. At a time of growing concern about the effects of digital technology on democratic politics, these findings will be of prime relevance to the academic and policy communities.

From a methodological perspective, finally, my thesis offers one of the first investigations of how affective polarization manifests in the digital context, and the communicative factors that underpin it. As I detail in the next chapter, affective polarization has thus far been largely measured through survey instruments and behavioral measures of discrimination in trust games and resource allocations. By combining the analysis of social media comment data with computational inferences of user features to examine affective polarization in online political discussions, my work provides a template that other researchers can build on. I hope it will serve to further scholars' current understanding of how users' ideological relationships, the content of the messages they exchange, and the power dynamics this creates, affect their attitudes and behavior.

2 Bringing 'Affect' Back Into the Study of Affective Polarization and Its Drivers

2.1 Introduction

Few topics have captured the attention of political scientists and opinion scholars as much as political polarization. Whether and to what extent our societies have become polarized has also emerged as a prominent subject of journalists and commentators in recent years (Carothers & O'Donohue, 2019; Cohn, 2014; Naim, 2019). While polarization was, for a long time, largely conceived as ideological distance, a new type of division has emerged in the mass public over the last two decades. Citizens across Western democracies increasingly distrust and dislike members of the opposition, while showing reluctance to associate or even socialize with them (Gidron et al., 2020; Iyengar et al., 2019). This phenomenon of growing avoidance and animosity between political opponents is referred to as affective polarization (Iyengar et al., 2012).

Understanding what drives affective polarization at the individual-level, and how to reduce it, requires knowledge of its foundations. In this chapter, I offer a review of existing scholarship on this topic, which I divide into three parts. In Part I, I begin by providing some key definitions and contextualize the study of affective polarization within the broader political science literature on political polarization. Following this contextualization, I lay out the contours of contemporary affective polarization, as well as its consequences both within and outside the political realm. In Part II, I divide existing explanations for this phenomenon at the individual-level into two categories: a *policy-based* model, which argues that partisan hostility reflects parties and voters' disagreements about policy issues, and an *identity-based* model, which sees partisan animosity as a corollary of partisanship's role as a group identity. I argue that while both models have some currency, they primarily rely on cognitive factors to explain affective polarization. In doing so, ironically, they tend to overlook the centrality of affective and motivational processes to the development of intergroup prejudice. In Part III, I suggest instead that bringing affect 'back' into the study of affective polarization and its drivers has the potential to enrich existing explanatory models, and may

even provide a better framework for understanding how aspects of our digital communication environment impact this phenomenon.

2.2 Affective Polarization in Context

Polarization has long been a topic of interest to political scientists and public opinion scholars. In the political science literature, some degree of political polarization is often regarded as beneficial to democracy: distinctiveness between political parties encourages citizens to think through their electoral choices, provides them with clear programmatic options, and motivates political engagement (Lupu, 2015; Mutz, 2002). Over the last decade, however, academics, journalists, and political observers concerned about the health of democratic societies have increasingly pointed to political and societal polarization as a threat to democratic stability. Growing levels of polarization have been cited as a leading cause for a wide range of issues, from legislative gridlock to lowered trust in institutions, elite incivility, and democratic backsliding (Fiorina & Abrams, 2008; M. J. Hetherington & Rudolph, 2015; Layman et al., 2006; F. E. Lee, 2015).

This lack of consensus over the purported effects of polarization is explained, in part, by a certain definitional catholicity. Polarization is defined and measured in varying ways by political scientists and journalists, which can often lead to confusion (Lelkes, 2016). For a long time, American political science scholars used the term to describe the growing ideological divide between the Republican and Democratic party leaders on social issues such as abortion and immigration (McCarty et al., 2006; Theriault, 2008), as measured through roll call voting (McCarty et al., 2006; Poole & Rosenthal, 2001), party vote percentages and interest group ratings (Bond & Fleisher, 2000). As scholars turned their attention to the question of whether citizens themselves had become polarized in the same fashion, the term was increasingly used to describe division at the mass level (Abramowitz & Saunders, 2008; Fiorina & Abrams, 2008).[6] In this context,

[6] This question has engaged prominent political scientists for several years, generating a fruitful, if inconclusive, scholarly discussion. Having reviewed existing evidence in support of what they call the 'polarization narrative,' Fiorina and Abrams contend that the idea that American voters are becoming more divided is largely a myth (Fiorina & Abrams, 2008). Taking the opposite view, Abramowitz and his colleagues argue that the gap between American partisans has widened dramatically, pointing to a decrease in the number of moderate voters and a 'dramatic increase in the correlation between party identification and stated ideology since 1972' (Abramowitz & Saunders, 2008, p.547).

polarization is often discussed as both a state and a process (DiMaggio et al., 1996): the process whereby a population splits into distinct and homogenous groups opposed to one another (Esteban & Ray, 1994) as well as the measure of the gap or divide between them. Before moving on with the remainder of the chapter, it is therefore important to clear out any potential definitional confusion. In the section below, I review three definitions of polarization at the mass level that frequently appear in the literature and that I will often come back to in later chapters. I then introduce *affective polarization* as a clearly distinct construct from other forms of polarization, and survey the different ways in which it is measured in the literature.

2.2.1 Mass Polarization: from Ideology to Affect

Determining exactly *what* divides the electorate politically, and how to measure it has been a core question for political polarization research. For a long time, political scientists assumed that divisions at the mass level were best captured through the lens of ideology. Over time, this led to the development of three distinct measures of mass polarization in the political science and political opinions scholarship: ideological divergence, ideological consistency, and false polarization (Lelkes, 2016).

Ideological divergence denotes the extent to which citizens' opinions on policy issues fall on opposite sides of the ideological spectrum (Fiorina & Abrams, 2008; Levendusky, 2010). In this view, the more partisans adopt extreme policy positions, the more polarized the electorate. A related, but distinct, measure is the degree to which individuals' policy positions are consistent with that of the party with which they identify. This is what some political scientists refer to as *ideological consistency* (Abramowitz & Saunders, 2005). To these authors, the existence of political conflict within society—in the form of non-overlapping interest groups and identities—is typically seen as a useful 'balancing mechanism' (Coser, 1964) that prevents unilateral power grabs and secures the 'openness of the democratic process' (Baldassarri & Gelman, 2008, p.2). In this perspective, the real threat to democratic politics is thus not so much that citizens' opinions differ, but rather that they crystallize into opposite political factions, built around exclusive identities and internally consistent interests (Baldassari & Gelman, 2008). In this ideology-oriented perspective, finally, mass polarization is also operationalized as *perceived*

or *false* polarization (Levendusky & Malhotra, 2016b; Westfall et al., 2015), that is the extent to which the electorate perceives political parties and other voters to be more divided on social and political issues than they really are. This last measure is of interest to opinion researchers as it has been shown to impact partisan attitudes independently of other measures of ideological polarization: perceived polarization, for instance, moderates voters' issue positions while driving dislike of the opposition (Levendusky & Malhotra, 2016a).

Other scholars, however, have argued that this focus on ideology is misguided and insufficient, as it may only provide a partial picture of mass polarization, based on 'empty signifiers' (Lauka et al., 2018). Crucial to this argument is the fact that multiple studies show that partisans are not necessarily ideologues (Abramowitz & Webster, 2018; Mason, 2015). The average US citizen, for example, struggles to correctly identify parties' positions on issues of national importance (Carpini & Keeter, 1996; Kohut et al., 2012). The same is true in the UK: today, a majority of Britons struggle to place political parties on the ideological spectrum and very few hold consistent left- and right-wing outlooks (M. Smith, 2019). Recently, political scientists have thus paid growing attention to a fourth type of mass polarization: *affective* polarization (Iyengar et al., 2012). Unlike previous measures, affective polarization is not operationalized in terms of ideological distinctiveness, but in terms of voters' dispositions towards other individuals and groups—how much they like or dislike them, what mental image they develop of them, and how they behave towards them. Specifically, affective polarization has two components: the extent to which partisans express dislike and aversion towards the opposition and strong preferences for their own political party (Iyengar & Westwood, 2015).[7] In that regard, affective polarization is conceptually similar to the notion of intergroup prejudice from the social-psychology literature: the tendency for people to evaluate and behave positively towards members of their ingroup and negatively towards outgroup members (Brewer, 2001).

This view of polarization as rooted in affect draws heavily on established theories of partisanship as a powerful, and relatively stable form of group identity

[7] A related concept is that of 'negative partisanship' which refers to voters' tendency to align themselves with candidates and form opinions mainly in opposition to the party they dislike (Abramowitz & Webster, 2018; Mayer, 2017).

(Converse, 1964; Green et al., 2002; Sears & Funk, 1999). Since the publication of *The American Voter* (Campbell et al., 1960), a predominant view among political scientists is that most voters develop psychological attachments to a political party or ideology at an early age, which tend to remain fairly consistent throughout their life and exert strong influence on their political attitudes and behaviors (Bafumi & Shapiro, 2009; M. J. Hetherington & Rudolph, 2015; Huddy et al., 2015; Theodoridis, 2017).[8] This holds true regardless of economic fluctuations or changes in party policies (Adams et al., 2011). This process of party identification also involves a cognitive representation of the party's composition in terms of various socio-political groups, with evaluations of the party reflecting an individual's feelings towards those groups (Ahler, 2018; Green et al., 2002). The extent of voters' allegiance to a political party is quite impressive in the US: less than 1% of the electorate says it feels no sense of attachment to either the Republican or the Democratic Party (Abramowitz and Webster, 2018). Recent scholarship even indicates a 'strengthening' of partisan identity in the US. Today, married couples and intergenerational families are even more aligned in terms of partisanship than they were in the mid-1960s (Iyengar et al., 2018). This stands in contrast to European democracies, where party identification tends to be more modest and where party-switching has been on the rise since the turn of the decade (Bischoff, 2013).

2.2.2 Measurements

Affective polarization is measured in a number of ways. On the evaluative side, three types of survey measurements are common: feeling thermometers of various political parties and their supporters, social distance measures, and trait ratings (Druckman & Levendusky, 2019). The first instrument, which is by far the most widely used in the literature, aims to capture the extent to which partisans view each other as a disliked out-group. This is generally achieved by asking respondents to rate how 'warm' or 'cold' they feel towards specific parties and candidates on a

[8] In fact, according to Theodoridis (2017), this visceral sense of association with a political party is sometimes even more pronounced than voters themselves realize or report through explicit measures of party identification.

scale from 0-100, though like-dislike scales are also common (Lelkes & Westwood, 2017).

Alternative measures of affective polarization also include trait ratings and social distance measures (Garrett et al., 2014; Iyengar et al., 2019; Settle, 2018a). These items are mostly concerned with partisans' appraisals of one another. In their respective studies, Levendusky (2017), Garrett et al. (2014), and Settle (2018a), all measure partisan animosity by asking voters what traits they most associate with different party figures and party supporters. Positive traits generally include intelligence, open-mindedness, and generosity, while negative traits often include selfishness, meanness, and intolerance. Another popular and less intrusive set of measures are social distance items, which gauge the extent to which individuals feel comfortable interacting with their political rivals in different social settings (Iyengar et al., 2019). The most common survey question asks partisans how they would feel if their child married someone from the other party[9], while another set of questions asks how comfortable partisans would feel having friends from the other party and how comfortable they would feel having them as neighbors (Druckman & Levendusky, 2019).

An important criticism that has been levelled against measures based on affective evaluations of political parties is that they primarily capture voters' attitudes towards *party elites* rather than their attitudes towards *other citizens* (Druckman & Levendusky, 2019). This would be especially problematic if the mechanisms that governed citizens' relationship to political parties were totally different from those that governed their relationship to other partisans. In reality, however, studies show that there is considerable overlap between those two measures (Iyengar et al. 2012, Druckman & Levendusky, 2019). This makes sense from a logical perspective: all else being equal, animosity directed at supporters of a political party necessarily reflects some degree of disdain towards that party as a whole, and what it represents. In the absence of survey instruments that explicitly tap voters' dispositions towards other partisans, party-based measures are thus frequently used as proxies.

[9] Though some scholars suggest that this measure exaggerates the extent of partisan animosity, by conflating identity-based hostility with a disdain for partisanship altogether (Klar et al., 2018).

A third, and final, way of measuring affective polarization is to gauge how much party affiliation biases people's behavior towards each other (Iyengar & Westwood, 2015; Suhay et al., 2018). In this view, strong partisan bias makes citizens more likely to give preferential treatment to their co-partisans, and more likely to discriminate against their political rivals. Behavioral measures of affective polarization are often deployed in economic trust games that assess the extent to which partisans would discriminate against each other in resource allocation, in evaluations of job applicants, and in labor markets (Gift & Gift, 2015; McConnell et al., 2018). Though the different sets of measures presented above all closely relate to one another, they can tell a slightly different story about the nature of partisan affective polarization. Whenever possible, scholars should thus carefully consider which type of item best matches the underlying concept and research question driving their study (Druckman & Levendusky, 2019).

In a two-party system like the US, capturing differential patterns of affect between competing partisan groups is relatively straightforward. However, moving to multi-party settings raises a number of conceptual and methodological challenges. First, democracies in which multiple parties can compete for power and form coalitions across the political spectrum rarely exhibit clear-cut 'ingroup/out-group' dynamics (Harteveld, 2021). The fragmented nature of the political landscape also means that citizens in multi-party systems tend to form weaker attachments to political parties (Bankert et al, 2017). By the same logic, research shows that they can also develop positive attachments to more than one party (Garry, 2007). Consequently, in this context, negative affect is unlikely to be restricted to a single party or group of supporters and voters might display varying levels of animosity towards these different groups (Wagner, 2020).

Deciding how to draw the line between political 'ingroup' and 'outgroup', and whether to account for the fact that voters may experience a sense of allegiance to multiple parties, are thus central tasks for scholars of affective polarization in multi-party systems. Existing empirical studies have tackled these issues in slightly different ways. Some scholars either restrict their observations to respondents who identify with the top two political parties (by size or vote share) (see for example, Hobolt et al., 2020), or broaden their unit of analysis to compare affect towards the two largest party coalitions (Huddy et al. 2018). Other scholars, such as Reiljan

(2019), Boxell et al. (2020) and Mayer (2017) embrace a more traditional view of partisanship as a social identity by assuming that positive affect can only be exclusive to one party. In this view, affective polarization is operationalized as the averaged difference between someone's feelings towards their own party versus *every other party* in the system, though other weighted variants of this measure also exist (Harteveld, 2021).

As this brief review shows, to this day, there remains some degree of contention as to what measure of polarization most meaningfully captures divisions at the mass level. Notably, the extent to which affective evaluations can be divorced from ideological considerations has been called into question. At the aggregate level, however, ideological polarization—as measured through growing ideological divergence—and affective polarization are far from congruent (Gaertner et al., 1993; Gidron et al., 2019). Moreover, while the latter has steadily increased in the US and elsewhere in recent years, there has been no comparable increase in ideological divergence on most issues (Iyengar et al., 2012). In the remainder of the thesis, I therefore treat them both as separate constructs and phenomena. Having traced the contours of affective polarization, and explained the ways in which it differs from other measures in the literature, I now review existing evidence for this form of polarization in the United States and Europe.

2.2.3 Manifestations

Over the last 25 years, affective polarization has been on the rise in a number of Western polities (Gidron et al., 2019b). In comparative terms, nowhere has this increased more rapidly since the mid-1990s in the US (Gidron et al., 2020). The clearest manifestation of this trend can be seen in the growing levels of inter-partisan animosity expressed in US public surveys. Between 1994 and 2016, the number of Democrats and Republicans with unfavorable views of the opposite party grew by 37 and 32 percentage points, respectively (Doherty et al., 2016). Today, the extent of interparty animosity has only deepened. Since 2016, the percentage of Republicans who rated Democrats negatively and expressed 'cold' feelings towards them has grown by 14 percentage points. Democrats' perception of Republicans has equally worsened: in 2019, 57% gave a 'very cold' rating of their political rivals (and 83% either 'cold' or 'very cold')—up from 41% three years prior

(Doherty et al., 2019a). Importantly, while Democrats and Republicans' feelings towards their own party have mostly remained positive and stable, it is feelings *against* the out-party that have gotten much worse (Iyengar & Krupenkin, 2018; West & Iyengar, 2020). Over the past 50 years, both groups have grown increasingly hostile towards each other. Partisans denigrate their political rivals as 'close-minded', 'lazier' and 'more dishonest' while holding an inflated sense of trust for members of their own party (Iyengar et al., 2012; Mason, 2015). In 2019, large swaths of partisans in the US expressed alienation from members of the opposite party—with the vast majority stating that they did not share non-political values and goals (Doherty et al., 2019)

Recent studies also underscore the fact that partisanship goes hand-in-hand with deep misconceptions about the political opposition. When thinking about rank-and-file party supporters, most people tend to hold stereotypical and often widely inaccurate views about their personality and demographic characteristics. They tend to perceive Democrats as progressive young urbanites, often union members, and Republicans as old, southern, white folks, for example (Drutman, 2020). Most Americans, for example, believe that nearly 40% of Republicans belong to the top 1% earners in the country (Ahler, 2018). This is true across the board, regardless of party, but particularly pronounced among voters who display high levels of affective polarization. Less than 15% of individuals see the opposition and its supporters as having core moral traits (Miller & Conover, 2015), with majorities of Republican and Democrats qualifying their rivals as 'immoral' (Doherty et al., 2019a), 'prejudiced', 'lazy', and 'closed minded' (J. E. Rothschild et al., 2019) while ascribing positive traits to their own party members. Partisans are also particularly willing to dehumanize members of the opposition, castigating them as animals that lack essential human traits (Martherus et al., 2019).

By some measures, discrimination based on party affiliation in the US now even trumps other forms of racial and social bias (Iyengar & Westwood, 2015; Mason, 2015). Following a series of survey experiments using measures of implicit partisan prejudice, Iyengar & Westwood (2015) find that unconscious bias towards the out-party is just as significant as comparable bias directed at racial outgroups. Furthermore, unlike with religious or racial biases, people do not shy away from expressing their contempt for political rivals. Where explicit measures of bias

normally underestimate the extent of prejudice against other groups due to social desirability bias, it is the exact opposite with partisan bias. Direct measures of people's animus towards political rivals reveal just as much if not more bias than implicit measures (Iyengar & Westwood, 2015).

Similar trends have been observed across several European democracies, despite large differences in their political systems and in the strength of their social and religious cleavages (Bankert et al., 2017; Reiljan, 2019; Wagner, 2021; Gidron et al. 2020, Boxell et al. 2020). Using economic trust games, for example, Westwood et al., (2018) show that partisan prejudice—operationalized as unwillingness to trust and cooperate with other partisans—now even exceeds other forms of religious, linguistic and ethnic bias in the UK, Belgium and Spain. Drawing on CSES data, Reiljan (2019) also finds hostility towards out-partisans to be widespread across Europe. It is especially acute in Central Eastern and Southern European states such as Bulgaria, Portugal, Czech Republic and Spain. In contrast, Northwestern, coalition-based countries such as Germany, Finland and the Netherlands present more moderate levels of affective polarization. While affective polarization has intensified sharply in the US over the last two decades, studies show that it is not that extreme in comparative terms. Today, by some measures, many European citizens (including in the UK) are just as much if not more polarized than their American counterparts (Gidron et al., 2020).

These effects also extend beyond traditional partisan divides. Research has shown that ideologies and opinion-based groups—as distinct from the substantive set of policies with which these are attached—also function as symbolic, social identities (Ellis & Stimson, 2012; Mason, 2018b). These are referred to in turn as 'ideological social identity' (Devine, 2015) and 'ideological identity' (Malka & Lelkes, 2010). Ideological and opinion-based identities are rooted in the same basic psychological mechanisms as partisan identities, meaning that they are just as powerful in motivating political activism (Huddy et al., 2015), distorting information processing (Malka & Lelkes, 2010), and generating bias against ideological opponents (Hobolt et al., 2020; McGarty et al., 2009). This is particularly relevant for European democracies, where party identification tends to be more complex and where ideological divisions between parties are less clear cut. González et al., (2008) and Huddy et al. (2018), for example, document significant

levels of affective polarization between voters who identify with competing political coalitions in Chile and Sweden, respectively. In both cases, however, participants still showed more positive affect towards members of their in-party than towards other coalitions members. Other studies have also found hostility and prejudice to take place across concrete issue-positions, such as vote choice in the wake of Britain's 2016 referendum on EU membership (Hobolt et al., 2020; Murray et al., 2017).

Political conflict is a constitutive part of democracy. As this review shows, however, the roots of today's discord extend far beyond policy disagreements and are becoming increasingly affective and personal in nature. This not only carries major consequences for our politics, but it also bodes poorly for social cohesion. In the next section, I summarize evidence on how partisan animosity affects behaviors and attitudes both within and outside the political realm, with damaging implications for social and economic life.

2.2.4 Political and Social Consequences of Affective Polarization

Affective polarization has wide-ranging political, social, and economic implications. In the political sphere, a nascent body of literature suggests that animosity towards the opposition biases partisans' policy attitudes. A recent study finds strong evidence that US citizens' levels of partisan animosity determined their attitudes towards the COVID-19 pandemic and towards the Trump administration's response to the crisis. Republicans who harbored negative feelings towards Democrats, for example, were much less concerned about the virus, and less likely to see actions such as mask-wearing and social distancing as beneficial, than their less polarized peers (Druckman et al., 2020). Partisans who are affectively polarized also tend to be less trusting of government, especially when their party is out of power (M. J. Hetherington & Rudolph, 2015), and to believe that the opposition and their policies threaten the nation's well-being (McCoy & Somer, 2019). They are also more likely to see themselves as more distant than they really are from the 'other side' on a wide range of issues, thus promoting misperceptions of widespread polarization (Armaly & Enders, 2020).

Rising partisan hostility also has ramifications for a number of social processes. Partisanship is now a defining factor in people's personal and romantic

relationships in the US for example, with co-partisans often being perceived as more physically attractive than members of rival parties (Klar & Krupnikov, 2016). Longitudinal survey data shows that party affiliation also influences social distance, such as desirability and willingness to associate with and even live close to people from rival political groups (M. K. Chen & Rohla, 2018; Gimpel & Hui, 2015). Large proportions of both Republicans and Democrats report, for example, being troubled by the prospect of having their child marry a supporter of the opposite party (Iyengar & Westwood, 2015; Westwood et al., 2018) and many recoil at the idea of living too close to political foes (Huber & Malhotra, 2017). Partisan labels, finally, have also been shown to dictate how willing citizens are to trust and cooperate with one another (Carlin & Love, 2018; Martini & Torcal, 2013).

These dynamics also extend to social relationships in the digital sphere. In a series of three experiments, Settle & Carlson (2019) show that partisan preferences influence how much someone would request to be paid to participate in a political discussion with a stranger. When given the option, partisans request significantly more compensation to discuss politics with out-partisans than in-partisans online, no matter the topic of conversation. During key moments in public life with heightened levels of partisan discussions, such as elections or protests, users also take active steps to disengage from political rivals, by unfollowing or unfriending them (John & Dvir-Gvirsman, 2015; Skoric et al., 2018).

Beyond impacting the kind of social relationship individuals enter into and seek to maintain, partisanship can also distort economic behavior. Work in this area has shown that taxi drivers in Ghana tend to offer preferential prices to people they recognize as co-partisans while overcharging out-partisans (Michelitch, 2015). Through a series of experiments in real-world settings conducted in the US, partisanship also had similar consequences on various everyday economic choices, such as one's preferences over who to make deals with, callbacks on job interviews, and hiring decisions (Gift & Gift, 2015; McConnell et al., 2018). Worryingly, similar dynamics apply to altruistic behavior and the distribution of public goods, with partisans consistently favoring their peers in a way that 'verges on ethnic cleavages […] observed in conflict-ridden societies' (Whitt et al., 2020) These

effects are particularly robust among strong partisans, and are not just restricted to ingroup favoritism. In several experimental tests of consumer reactions to corporate giving, for instance, Panagopoulos et al. (2016) find that participants were less likely to accept gift cards coming from organizations that supported a rival political party.

Taken together, this evidence clearly shows that the effects of partisanship extend far beyond the political realm. Negative sentiment not only impacts core aspect of sociability, such as interpersonal trust, but also shapes citizens' willingness to engage with one another beyond political difference. Such unwillingness to interact or even socially associate with people who hold different values poses serious concerns for democracy and social cohesion, as it has long been argued that healthy polities require citizens to be exposed to, interact with and tolerate a diverse range of political viewpoints (Downs, 1957). At a time of heightened political tensions across several geographies, it is therefore imperative to shed more light on the underlying mechanisms behind this phenomenon. Yet, while scholars agree that partisan animosity has intensified in recent years, there is far less consensus on what factors drive this process. In the second part of this chapter, I now critically engage with existing scholarship on this question.

2.3 Locating the Drivers of Partisan Animosity

As we saw, partisanship has real and far-reaching consequences for how voters think and behave towards each other. However, there is still plenty of debate as to what actually drives these effects. Scholars have advanced a number of competing explanations in this regard, which can be broadly summarized into two distinct theoretical camps: a *policy-based* model and an *identity-based* model (Huddy & Yair, 2021; Lelkes, 2021; Orr & Huber, 2020). In the section below, I outline the theoretical foundations of these two models and review the insights they have generated with respect to individual and online drivers of affective polarization. Throughout, I argue that while they offer some valuable insights into factors that may worsen partisan animosity, both rest on cognitive assumptions that have been severely challenged and apply to a limited range of situations. Notably, as I argue in the second half of the section, both models are ill-suited to explain the effects of computer-mediated communication on partisan animosity. I instead propose an

extension of the identity-based model beyond its strictly cognitive focus that attends to the role of affect in shaping the effects of political identification.

2.3.1 Two Explanatory Models: Policy Versus Identity

2.3.1.1 Policy Model

Proponents of the *policy-based* model believe that there is a rational basis to voters' negative appraisals of the opposition. In this view, partisanship not only conveys critical information about political elites' positions on salient issues but also serves as a proxy for party members' social and political characteristics, beliefs, and values. Partisans who dislike members of opposite parties thus do so because they see their political rivals as having fundamentally incompatible views on policy issues, such as immigration or social welfare (Abramowitz & Webster, 2018; Enders & Lupton, 2020; Rogowski & Sutherland, 2016; Webster & Abramowitz, 2017).

This assumption has various psychological bases. Early research from social psychology has produced extensive evidence that similarity in belief is a strong determinant of interpersonal attraction (Rokeach & Rothman, 1965). Attitude theory, for example, posits that individuals' evaluations of others are strongly determined by their own beliefs about other people's attributes (such as their opinions, for instance) and whether they find themselves in alignment or opposition to them (Festinger, 1954, 1957; Fishbein & Ajzen, 1975; Heider, 1958). Put differently, people tend to be attracted to other people who think similarly to them. Studies have shown that individuals tend to hold more favorable views towards parties and candidates whose issue stance they agree with, whilst negatively evaluating those with whom they disagree (K. Hetherington, 1996; Ottati et al., 2002). Based on these findings, some scholars have argued that the main reason why people hold negative attitudes towards particular individuals and groups is they have different systems of beliefs and issues positions than their own (Bougher, 2017), and not because of the value ascribed to their belonging to a particular ethnic or social group, for instance.

Drawing on belief congruence theory, Bougher (2017) embraces this view, arguing that partisan hostility is best explained by perceived 'dissimilarity in beliefs' and issue positions between oneself and members of rival party. Through cross-

sectional and panel analyses of survey data, she provides evidence that issue alignment—falling on the same side of the ideological spectrum as one's own party on multiple issues—is a leading cause of hostility towards the out-party, regardless of party affiliation. By the same logic, scholars have shown that individuals who perceive high levels of ideological divergence within the electorate and between political elites themselves—what I defined earlier as false polarization—are more likely to be affectively polarized (Levendusky & Malhotra, 2016b).

Another psychological mechanism underlying the policy-based model has to do with the nature of electoral choice and its impact on partisans' attitudes and behaviors. According to the rational choice model of voting behavior, partisans become more psychologically invested in picking the right candidate when political parties adopt increasingly divergent policy stances, as there are higher stakes associated with their choice (Downs, 1957; Stokes, 1963). Building on spatial models of party competition, Rogowski & Sutherland (2016) argue that these dynamics also extend to how citizens evaluate political leaders, with voters responding to large ideological differences between political officials by providing more favorable evaluations of their own candidate than of rival politicians. Confirming these findings, Webster & Abramowitz (2017) linked perceptions of political parties as ideologically distant with higher affective polarization and show that manipulating the extent to which a candidate appears to be more liberal or conservative also boosted participants' negative evaluation of that candidate.

2.3.1.2 Identity Model

In contrast to the policy-based model, the *identity-based model* locates the origins of out-party hostility to partisanship's role as a social identity. Proponents of this view largely draw on Tajfel's social identity and social categorization theory. In this view, individuals develop a strong value and emotional attachment to the social and political groups they identify with, which become internalized as an important component of their sense of self (West & Iyengar, 2020). Much of the contemporary affective polarization literature is grounded in the more cognitive strand of social identity theory, which posits that mere categorization—even on the basis of arbitrary and insignificant differences between groups—is itself enough to elicit intergroup discrimination (Billig & Tajfel, 1973). Extended to the political context,

scholars have interpreted this to imply that simply identifying with a political group is enough to trigger biased attitudes and behaviors (Iyengar et al., 2012). In this view, there is something intrinsic to group membership that automatically drives these effects. Mason makes this point clearly in the opening of her book *Uncivil Agreement* (2018), citing a famous study of intergroup conflict—the Robbers Cave experiment—in which a group of boys were randomly sorted into two teams and pitted against each other for several weeks:

> The boys at Robbers Cave needed nothing but isolation and competition to almost *instantaneously* consider the other team to be "dirty bums", to hold negative stereotypes about them, to avoid social contact with them, and to overestimate their own group's abilities. In very basic ways, group identification and conflict change the way we think and feel about ourselves and our opponents (p.2, emphasis added)

For proponents of the identity-based approach to affective polarization, the mechanism that form the basis of any of these effects is *identity salience* (Gaertner et al., 1993)—that is how prominent an identity is to someone's sense of self. The more salient one's political identity is in any context, the more likely it is to shape one's attitudes and behaviors towards perceived others. Salience itself, however, hinges on different factors, such as the degree to which an identity is central or peripheral to someone's self-concept, how loyal they feel towards a particular social group, how often they are reminded of their affiliation to that group, and the information environment they are in (Stryker & Serpe, 1994). Citizens with the strongest allegiance to their political party thus typically display more biased views towards the opposition, even when they themselves do not hold extreme policy positions (Huddy et al., 2018; Mason, 2015). For these authors, party identity plays a more powerful role in shaping out-party animosity than ideological extremity. Similar findings apply to ideology-based identities: the more strongly someone identifies with an ideological group (such as liberal), the more biased they are against perceived ideological rivals, even when attitude extremity is low (Mason, 2018b).

According to Mason and others, partisan attachment is a powerful enough predictor in isolation, but its effects are even stronger when working in concert with other social identities. Whereas the existence of cross-pressures works to reduce

perceived differences between social groups and mitigate social conflict (Dahl, 1981; Nordlinger, 1972), the more alignment there is between partisans' cultural, ideological and social preferences—a phenomenon termed *sorting* (Levendusky, 2009; Mason, 2015)—the less tolerant and more negatively they feel against perceived outgroups (Roccas & Brewer, 2002). Voters who identify with the party that most closely reflects their worldview have been shown to perceive their political opponents as more socially distant and extreme than they really are (Levendusky & Malhotra, 2016b), and display higher levels of inter-party bias and anger (Mason, 2015, 2018c; Mason & Wronski, 2018). The percentage of individuals whose partisan identity directly aligns with their ideology has greatly increased in the US since the 1970s (Levendusky, 2009). Some scholars have thus argued that the intensification of affective polarization in the US over the last 50 years is a direct corollary of the progressive decline of crosscutting political identities (Mason, 2018a).

2.3.2 Beyond Identity Versus Ideology: Re-centering the Role of Affect

Which of those two competing perspectives best explains partisan hostility? Providing a robust answer to this question is an area of ongoing research. For a long time, both strands of the literature evolved in parallel, as most studies did not explicitly seek to test the effects of partisanship and ideological disagreement in isolation. Most studies in this line of work have examined differences in individual characteristics to infer whether affective polarization was best explained by policy considerations or partisanship (for example, Bougher, 2017; Halberstam & Knight, 2016; Mason, 2015; Webster & Abramowitz, 2017). At the time of writing, only two studies have directly compared the predictive power of party versus ideology on affective polarization. Using a survey experiment, Lelkes (2021) demonstrates that when asked to evaluate a fictional political candidate, ideologically extreme partisans who were given information about the candidate's policy preferences were more likely to give them a harsh rating than those who were given information about the candidate's political party. Confirming these findings, Orr & Huber (2020) also provide experimental evidence that partisanship's effects on interpersonal evaluations is more limited than that of policy positions, and so among both weak and strong partisans. Also relevant to this question are findings

from West & Iyengar (2020) showing that reducing the salience of someone's partisan identity though a self-affirmation treatment does not make them less inclined to express out-party hostility. This suggests that although partisanship may function as a strong social identity, affective polarization itself may have its origins elsewhere than in the 'classic ingroup versus outgroup distinction' (p. 1).

On the surface, it therefore appears as though the policy-based model has more empirical currency than the identity-based one in explaining partisan animosity. However, I would argue that neither of them provides exhaustive nor mutually exclusive explanations for this phenomenon. In reality, it is difficult to truly disentangle these two mechanisms at the individual level (Goggin et al., 2020). There are several reasons for this. One, as Orr and Huber (2020) poignantly remark, is that the two are often confounded observationally, which complicates any attempt to tell their effects apart. Furthermore, the two may not be distinguishable constructs in any meaningful way. Individuals make a number of inferences about someone's characteristics based on partisan orientation (Green et al., 2002). Partisanship itself could thus be a proxy for several factors, from an individual's policy position to a wide range of social features including ethnicity, social class, and religion (Ahler & Sood, 2018), all of which could impact how affectively disposed a voter is towards that person. However, as we saw earlier, certain policy preferences (such as being pro or against immigration) and ideological labels (such as liberal and conservative) can also confer a sense of group identity (Devine, 2015). In Lelkes's study, it is thus entirely possible that respondents reacted negatively to a candidate's policy stance because it conveyed a rival psychological team orientation. Without breaking into the 'black box' that is the human mind, it is thus practically impossible to meaningfully distinguish between these two processes.

Furthermore, by foregrounding a primarily cognitive view of human psychology, both models are in fact quite reductive as they neglect the centrality of affect in political thinking and behavior (Cassino & Lodge, 2007). Indeed, a core assumption running through much of the identity-oriented perspective is that the mere existence of divisions between partisan groups suffices to foster biased behavior—a notion inferred from Tajfel's experiments ran in 'minimal group' settings. A key implication of this assumption, Bernadette Park and Judd (2005)

note is that partisan prejudice may simply be an unavoidable consequence of the cognitive process of categorizing the social world into different groups.[10] Likewise, the theoretical premises of the policy-based model largely draw on a cognitive understanding of opinion formation, according to which people develop attitudes towards political objects based on the information they are provided with. In this perspective, voters are regarded as rational actors and partisanship as merely an 'expression of beliefs about which party will best pursue one's interests' (Green et al., 2002). A voter endorses a politician during an election, for example, because she knows that this politician's policy agenda will be better for her.

Experimental evidence, however, calls into question the validity of this assumption. Drawing on a wealth of research from the social, cognitive and, neurosciences, Lodge and Taber (2013) demonstrably show in *The Rationalizing Voter* that, when it comes to how citizens evaluate political candidates, parties, and issues, how they *feel* mostly takes precedence over what they know (see also Abelson et al., 1982; Ottati et al., 1992). Decades of studies in psychology have clearly indicated that affect is not only a key organizing principle of the mind that enables us to make *prima facie* sense of the world around us, but that affective processes also strongly impact our cognition (Abelson, 1959). Emotions distort information processing by motivating certain forms of reasoning (MacKuen et al., 2007) and are themselves used as relevant information when making political judgements (Cassino & Lodge, 2007), often leading to ex-post rationalization of our beliefs and attitudes. In other words, evaluations are less the product of careful, rational considerations and cognitive calculus, than they are of a complex system that integrates both affective and cognitive components (Mackie & Hamilton, 1993). Similarly, scholars of intergroup prejudice have long warned against an excessive focus on cognitive factors when studying discrimination processes, as they only apply to a limited range of situations and bypass the importance of affective forces in motivating our thoughts and actions (Bernadette Park & Judd, 2005).

Emotions, in other words, play a central role in shaping social and political cognition—how people make sense of themselves and others, how they interpret

[10] Referring to Tajfel's work, Hamilton suggest that his experiments advance 'the more general principle that, at least under some circumstance, aspects of our cognitive functioning may, by themselves, constitute the bases for stereotyping and intergroup discrimination' (Hamilton, 2015).

political events, and make decisions. Hence, beyond the primarily cognitive explanations of affective polarization outlined above, I argue that there is some value in better integrating the role of affective states, and their impact on the formation of political judgments, into our models of the drivers of partisan prejudice. This is not to say that extant approaches are not valuable or insightful, but rather that they may only provide a partial understanding of underlying processes too complex and elusive for only one model to account for satisfactorily. As I show in the section below, this is particularly the case when thinking about the personal and communicative drivers of affective polarization.

2.4 Towards an Affective Perspective on the Drivers of Affective Polarization

Research into the antecedents of political attitudes and behavior have long been dominated by cognitive approaches, and as we saw in the previous section, affective polarization is no exception. Cognitive approaches are mostly based on a mechanistic understanding of how citizens form impressions of the political world and develop attitudes towards it. In contrast, a more affective orientation to social and political psychology emphasizes the role of emotions, motivations, and affective states in shaping information processing and evaluations of political objects, especially in intergroup contexts (Mackie & Hamilton, 1993).

Important elaborations of both social identity and attitude-formation theories, which I will draw on throughout subsequent chapters, include intergroup emotions theory (Mackie et al., 2008), the affective intelligence model of political judgement (Marcus et al., 2000), and integrated threat theory (Stephan & Stephan, 2000). I will spell out each of these in greater detail in the following chapters. Broadly, however, all three theories depart from the aforementioned frameworks in three significant ways. First, all three posit that the rational and the emotional—thinking and feeling—are not oppositional systems, but instead interact and mutually influence each other: affect impacts cognitive processes and cognition provides the normative and socio-cultural framework to help individuals make sense of their experienced emotions. Consequently, as Hamilton and Trolier (1986) remark:

> Any particular form of stereotyping or prejudice [...] is in all likelihood multiply determined by cognitive, motivational, and social learning processes, whose effects combine in a given social context to produce specific judgmental and behavioral manifestations. Therefore, any attempt to understand such phenomena as a product of one process alone is probably misguided (1986, p. 153)

Second, they are all concerned with the way in which individuals' emotional reactions to particular situations or stimuli—for example, when placed in an intergroup context or a new informational environment—influence their action tendencies, attitudes and preferences. Finally, all three posit that specific, discreet emotions act as motivational factors for different attitudinal and behavioral outcomes. Negatively-valenced emotions like anger and shame, for example, produce different outcomes than enthusiasm or pride (Brader & Marcus, 2013). I argue that integrating these insights into existing theoretical models of affective polarization may help to address some of the shortcomings in the literature, and offer new explanatory avenues that ultimately enrich our understanding of the phenomenon at the individual level.

2.4.1 Attitudinal Factors

As we saw above, whether affective polarization is driven by partisan allegiances or by disagreements over policy preferences is an ongoing source of debate in the literature. From this latter policy-based standpoint, partisans hold each other in contempt because they believe that their positions differ broadly from that of out-partisans. The problem with such a framing is that it wrongly assumes that individuals' partisanship and issue preferences are always consistently aligned.[11] While this may be true at the elite level, evidence shows that this is far from the case within the mass public: to this day, many voters still struggle to place the parties on policy issues (Achen & Bartels, 2017) and recent longitudinal studies show that, although sorting has increased in the US since the 1980s, that sorting is very limited in absolute terms: in other words, there is still more ideological agreement than disagreement on issues across party lines (Barum Park, 2018).

[11] At Orr and Huber (2020)'s own admission, their aforementioned analysis of the effects of policy and partisan cues on affective ratings did not cover issues around which there is a wide consensus or strong intraparty conflict.

Importantly, there are also clear indications that partisans themselves are ideologically cross-pressured. Analyzing data from the American National Election Studies survey, for example, Hillygus & Shields (2009) find that a substantive share of US partisans disagree with their own party's candidate on at least one policy issue. Several studies identify similar trends in Western European electorates, with voters reporting both right-wing and left-wing opinions on different sets of issues (Gidron, 2020).

While research on electoral choice already shows cross-pressures to impact candidate choice (R. Bakker et al., 2020), it is less clear whether these dynamics transfer to affective evaluations of party and candidates. To my knowledge, no study has yet examined how experiencing such a disjuncture between party identification and policy preferences impacts partisans' dispositions towards their own and rival political parties. Yet, studying such intraparty conflict can offer insights into the micro-foundations of affective polarization as it provides a natural test of the explanatory power of both policy and identity-based models at the individual level. Indeed, if considerations of partisanship really mattered more for evaluations, than even cross-pressured partisans who disagreed with their own party on policy issues would express more hostility towards the out-party than the in-party. If, on the contrary, policy conflict was the real driver of prejudice, then cross-pressured partisans should feel more negatively towards their own party than towards the opposition.

Here, affective theories such as the theory of affective intelligence offer some useful insights into the potential consequences of such ambivalence. Affective intelligence theory (Marcus et al., 2000) posits that 'affective systems manage both our response to novelty or threat and our reliance on established habits [..] and also influence when and how we think about such things' (p. 9). According to Marcus et al. (2000) ordinary people switch between two emotional systems when dealing with information: a dispositional system and a surveillance system, which are linked to different sets of emotions. Under habitual circumstances, voters tend to rely on established dispositions and familiar heuristics to process information from the political world. Novel and unpredictable circumstances on the other hand—such as finding out that their party's representative embraces an issue position that conflict with their own values—trigger feelings of anxiety and uncertainty, which

provokes thoughtful and deliberate thinking of partisan's choices and evaluations. Put differently, the negative emotions associated with realizing that their own political party stepped out of line with their expectations may drive partisans to be more critical of or break away from their partisan allegiance (MacKuen et al., 2007). I come back to this debate and test this hypothesis in Chapter 3.

2.4.2 Online Factors

Another topic of ongoing discussion within the literature is the extent to which various aspects of our contemporary media and communication landscape could be exacerbating affective polarization (Iyengar et al., 2019). At the macro-level, the relationship between internet access, social media use, and affective polarization is still a contested one. To date, only few studies have investigated this question, with mixed results: while some scholars find a small but positive correlation between broadband internet access and negative partisan affect (Lelkes et al., 2017), others argue that if digital media played a role in the rise of affective polarization in multiple OECD countries post-2000, it is unlikely to be the most important factor, given that the most polarized segments of the population are not those most likely to be active internet and social media users (Boxell et al., 2017, 2020). These inconclusive results clearly highlight the need for further research. Although internet and social media use may not be the biggest contributing factor to existing levels of partisan hostility, however, a bourgeoning literature argues that it still has a non-negligible impact on users.

2.4.2.1 Social Media as an 'Affective' Communication Space

Digital media is increasingly interwoven with our social and political lives. Today, around 85% of US citizens say they go online on a daily basis, with 31% saying they are 'almost constantly' online (Perrin & Atske, 2021). The mass uptake of digital technology has also facilitated the speed and ease with which people can get access to news and political information: more than 60% of US adults report getting news from digital platforms 'sometimes' or 'very often—up from 47% in 2018 (Shearer & Grieco, 2019). These numbers are similar in the UK, with over half of the population reporting using social media to keep up with public affairs. After

television, social media is the second most popular vehicle for news (Ofcom, 2019). Digital platforms have also emerged as key venues for discussing politics and engaging in a wide range of civic activities: in 2020, 45% of users said that social media was important to them when it comes to getting involved with political and social issues and expressing their political opinions (Auxier, 2020).

But what makes social media any different from other informational and communicative spaces? I identify several distinguishing features. One is the ease and frequency with which individuals come across politically-charged content online. Most social media users do not go online with the explicit purpose of discussing politics or looking for political information (Bode, 2016). However, they are very likely to stumble upon it incidentally through their social contacts (Fletcher & Nielsen, 2018; Settle, 2018a), often in a repeated and habitual fashion. Majorities of social media users say they come across political content, in the form of comments, posts, and news articles, on their feeds on a daily basis—to the point of fatigue (M. Anderson & Auxier, 2020; Shearer & Gottfried, 2017). Crucially, research shows that these explicit and implicit political cues are informative enough to lead users to infer the views of the person who posted them (Settle, 2018a), despite their interpretations being 'neither *perfectly accurate* nor *particularly well-honed* [emphasis added] in making fine grained differentiations between people' (2018, p.161).

Information sharing and political communication on social media is also quintessentially social. Though digital platforms vary in their degree of public-facingness, one of the distinguishing features of political expression online is that it is highly visible and interactive, in that a plurality of others can readily provide feedback, express their support or dismay at something another user said, and, in most cases, share that content to their own network. As I mentioned in the previous chapter, it is also increasingly easy for users to connect with varied audiences as well as likeminded others who may share similar interests and opinions.

One final distinctive feature is the nature and quality of online political discourse. As Jungherr et al., (2020) put it, 'political talk online is emotional, phatic, humorous, and personal' (p.117). In hybrid media systems (Chadwick, 2013), the boundaries between facts and opinions can become blurred, as users provide their own takes on social and political events that reflect deeply subjective

frames and interpretations (Reeve, 2015). Public appraisals and displays of emotions are also central aspects of our digital cultures (Döveling et al., 2018; Tettegah, 2016). While online speech can be positive, empathetic and joyful (Lu & Steele, 2019); incivility, ad hominem attacks, and vitriolic language sometimes referred to as 'flaming' (Coe et al., 2014; da Silva, 2013) are also pervasive online (Rossini, 2020).[12] Incivility is not necessarily undesirable in itself (G. M. Chen et al., 2019). However, its pervasiveness underscores the fact that political discussions over social media rarely meet the conditions for constructive political deliberation.

2.4.2.2 Theorizing the Effects of Social Media on Partisan Affect

All of these factors have led some scholars to suggest that the widespread use of digital technologies could be to blame for exacerbating negative partisan affect (Bail et al., 2018; Sunstein, 2017). Scholars have identified three broad ways in which the internet in general and social media in particular may be contributing to these effects. These causes loosely map themselves onto the theoretical frameworks presented above: (1) by driving exposure to political content that accentuates user's misperceptions about the other side; (2) by fostering a social environment in which users are more likely to develop extreme ideological beliefs, and (3) by increasing exposure to content that activates their sense of partisanship. I review each of these proposed explanations in turn, before highlighting their conceptual limitations, and proposing a new way forward.

A first prominent hypothesis in the literature is that social media use drives exposure to news and information that reinforces people's misperceptions about their political rivals. The high-choice nature of the digital media environment has, theoretically, made it easier for internet users to access partisan and ideologically slanted news that reflects their pre-existing beliefs (Gentzkow & Shapiro, 2011; van Aelst et al., 2017). Exposure to these pro-attitudinal messages, scholars argue, not only makes users more extreme (Levendusky, 2013), but also drives negative affective evaluations of the opposition (Garrett et al., 2014, 2016; Lau et al., 2017). The reporting of online news outlets has also been flagged as an aggravating factor. Since the turn of the century, hyper-partisan blogs and media outlets—many of

[12] To only take one platform, political speech on Twitter has been described as 'simple, impetuous, and frequently denigrating and dehumanizing' (Ott, 2017, p.59).

which thrive on outrage and divisive rhetoric (Sobieraj & Berry, 2011)—have proliferated on the open web (Faris et al., 2017). Not only do these outlets overwhelmingly focusing on scandals (Puglisi & Snyder, 2011), but they often portray political opponents in negative terms by questioning their integrity and intentions (Sood & Iyengar, 2017), which could confirm existing (and engender new) stereotypes towards the other side (Puglisi & Snyder, 2011).

A second hypothesis advanced by scholars of online communication is that the internet is driving us apart by allowing citizens to pursue social relationships that reinforce their pre-existing beliefs, thus making them more hostile towards their political opponents. The bulk of the literature on this topic focuses on the diversity of individuals' social networks online—that is the extent to which they encase themselves in homogenous political bubbles (Eady et al., 2019; Sunstein, 2017) or to which they interact with dissimilar others and viewpoints. Variants of this 'echo chamber' hypothesis have existed since the early days of the internet (Negroponte, 1995; Papacharissi, 2004; Sunstein, 2001) and draw on previous research into face-to-face interpersonal communication which suggest that discussions among likeminded groups make participants' attitudes more extreme (R. Brown, 2010; Myers & Lamm, 1976; Sunstein, 2001) by boosting their confidence in their own opinions (Brauer et al., 1995). In her investigations of 'group polarization' in the online context, Wojcieszak (2010, 2008) did find that members of radical and homogenous online forums displayed more extreme attitudes, and were restricted in their ability to reach consensus with others which may, as we have seen, impact affective polarization down the line. Thus far, however, only one study has established a clear link between homogeneity of political discussion and affective polarization. Analyzing panel data across two different elections cycles, Hutchens et al. (2019) find that discussing politics with likeminded others online increases negative affect towards the opposition—though that relationship is reciprocal.

The third and final hypothesis is that social media use drives partisan animosity by boosting exposure to political cues that activate users' sense of partisanship. According to social identity theorists, exposure to symbols, images, and words containing political group cues heightens the psychological centrality of one's political identity and the degree to which that person will behave in a manner

consistent with the meanings attached to it (Carr et al., 2013). There is already solid evidence that being exposed to partisan media strengthens audiences' sense of partisanship, making them more prone to develop ill-feelings towards the 'other side'(Horwitz & Nir, 2015; Knobloch-Westerwick & Kleinman, 2012; Levendusky, 2013). Given the ubiquity of politically informative cues online, some scholars argue that our social media ecosystem provides the perfect conditions to activate users' political allegiances and thus shape their subsequent evaluations of others. In her 2018 book *Frenemies,* Settle demonstrates through a series of survey and experimental studies that substantial amounts of content posted on Facebook—from comments to status updates—contain cues that clue users about their contacts' political affiliation. Exposure to these cues, in turn, triggers a host of polarizing effects, including negative judgement towards the other party, perception of greater social distance between political groups; and desire to preferentially connect with members of one's perceived ingroup (2018, p.224). Considering that many online discussion spaces, such as forums, blogs, and newspaper comment sections, are still characterized by relative anonymity or pseudonymity, however, testing whether these effects travel to different and more naturalistic settings will also be an important step towards generalizing any claims about the polarizing impact of social media use.

2.4.3 The Missing Piece: Bringing 'Affect' Back In

As this review shows, investigations into the effects of digital media on affective polarization have produced rich insights. Yet, they also still leave a number of questions unanswered, raising several interesting conceptual and theoretical issues.

Several of these explanations are based on assumptions about user behavior that have increasingly called into question in the digital context. Take, for example, the echo chamber hypothesis. Despite initial concerns, evidence is mounting that insulation from outside perspectives on social media is, in fact, far from the norm. Most social media users come across ideologically diverse news and information through weak social ties, including distant friends and colleagues (Bakshy et al., 2015; Barnidge, 2017; Messing & Westwood, 2014). Studies leveraging behavioral data and recent self-reports confirm this. Partisans sometimes seek likeminded news, however, the average user's media repertoire is actually much more diverse

than previously thought (Dubois & Blank, 2018; Dvir-Gvirsman, 2017; Fletcher & Nielsen, 2017). Finally, while early studies show that people tend to gravitate towards consonant information sources (Garrett, 2009; Iyengar & Hahn, 2009), they rarely avoid counter-attitudinal information altogether when it is presented to them (Knobloch-Westerwick & Meng, 2009). Further challenging the echo chamber narrative, empirical evidence for political homophily online is also rather mixed. While some studies of online blogs and social networks find a degree of political homophily among content creators (Adamic & Glance, 2005; Hargittai et al., 2008) and Twitter users (Miller & Conover, 2015) through their linking and sharing patterns for example (Colleoni et al., 2014)—overall ideological segregation on social media appears to be 'overestimated' (Barberá et al., 2015, p.1531). Importantly, the bulk of this ideological clustering appears limited to the most partisan and engaged users (Barberá & Alvarez, 2015; Halberstam & Knight, 2016), while the broader population is embedded in politically diverse networks and are often exposed to diverse political information (Eady et al., 2019). All in all, this goes on to show that it would be mistaken to picture our social media environment as a collection of isolated, political bubbles. Instead, I join Carr (2017) and Settle (2018) in arguing that it should primarily be seen as a space for *intergroup* communication: an environment in which individuals frequently come into contact with people and ideas that are different to them (Barnidge, 2017). These encounters, in turn, activate cognitive and emotional processes that have the potential to shape users' political thinking and behavior.

Furthermore, the majority of studies in the literature tend to overwhelmingly focus on only one aspect of digital political communications: their content, on the one hand, or the extent to which they take place across ideological lines, on the other hand. In doing so, they often fail to investigate how the two intersect in producing specific outcomes. Political interactions do not happen in a vacuum, however; they take place within a specific social and communicative context, and one that, as we saw above is deeply phatic and emotional. How a political message is interpreted itself depends on the context within which it is shared, its sender (e.g., whether it originates from a peer or a foe) and its content (e.g., whether adversarial or amicable). In my mind, this is a critical elaboration that is currently missing from the literature. Both of these dimensions—whether

interlocutors believe to be part of the same 'team' and how they themselves behave and interact with each other—are key in shaping our political attitudes and perceptions. Carefully studying the impact of online political communication thus requires that we not only pay close attention to the ideological relationships between users, but also to the affective nature of their communication.

Unfortunately, we still lack a sophisticated understanding of the role of emotions in political discussions taking place outside of face-to-face settings; not so much from a lack of recognition that they are important (Papacharissi, 2014), but rather from a relative absence of studies that develop specific hypotheses about 'the way that the type, level, and dynamic of emotion in a political conversation amplify or alter the patterns of influence we expect to see between discussants' (Settle, 2018b) especially when those discussants are from opposite sides of the political spectrum. Integrating these perspectives is one of the goals of this thesis, and I will come back to this question in Chapter 4.

Lastly, current explanations fail to specify what types of online messages are most impactful for partisan affect. Extant research on exposure to incivility and partisan media already provides clear indications that some aspects of our communicative landscape have the potential to exacerbate negative sentiment. Individuals exposed to incivility tend to react with anger and aversion, engage in copycat aggression (Hmielowski et al., 2014), develop biased risk perceptions (A. Anderson et al., 2014), and be less openminded (Borah, 2013). Recent findings also show that exposure to uncivil and disagreeable comments on Facebook reduces willingness to read more comments and leads to negative emotions and more polarized attitudes (Y. Kim & Kim, 2019). As we saw in the previous section, exposure to partisan news is also linked to out-party animosity. However, because the treatment oftentimes consists in exposure to a partisan outlet, it has thus far been difficult to tease out the precise mechanism—priming of partisan identity, exposure to extreme viewpoints, or denigration of political opponents—through which partisan media drives these effects. Recent experimental studies have begun to shine light on this question. Suhay et al. (2018), for example, show that participants who are exposed to partisan criticism—online comments that explicitly demean their in-party—become more prejudiced against the opposition as a result. Yet, extant research has mostly focused on the negative effects of online

communication on deliberation at the expense of other dynamics. It is unclear, for example, whether partisan incivility always bolsters animosity or whether expressions of interparty respect and camaraderie could reverse these effects. Such questions still remain to be explored, and I will do so in greater detail in Chapter 5.

By centering the role of specific emotions in group contexts, elaborations of social identity theory such as intergroup emotions theory, intergroup contact, and intergroup threat theory, all provide useful theoretical frameworks to address these limitations. These models are based on the social identity perspective in that they also view the social groups with which individuals identify as key catalysts for particular psychological processes. Unlike the strictly cognitive paradigm of self-categorization theory, however, they place a greater emphasis on the context in which people become motivated to defend their ingroup. This approach suggests that identification with a group motivates individuals to positively distinguish one's group from others to maintain a positive social identity—and that it is these motivational processes that are responsible for intergroup conflict and prejudice (Brewer, 1999). Intergroup scholars have long underscored the importance of affective mechanisms to intergroup relations: in this view, discrimination and prejudicial behavior are best understood 'through the lens of group-based emotional reactions' (E. R. Smith & Mackie, 2015). Intergroup emotions theory in particular posits that individuals can experience emotions on behalf of the social groups they identify with, and so even in the absence of other group members (E. R. Smith et al., 2007). This has far-reaching consequences with regards to the digital context, as it entails that the large proportion of the population who might only get *exposed* to politically salient content or people who typically observe online conversations without participating in them—rather than directly experiencing it, may still experience negative feelings on behalf of their political ingroup.

These emotional reactions are triggered by situations or events that ostensibly impact the group with which they identify. They can be either aimed at someone's own group (such as patriotic pride) or at an outgroup (such as feelings of anger or disgust towards ethnic minorities), and have direct and tangible consequences for a range of behaviors including outgroup discrimination (E. R. Smith & Mackie, 2015). Situations that might accentuate an individual's perceptions of an outgroup as a threat to the ingroup, for example, typically

accentuate negative appraisals towards its members (Stephan & Stephan, 2000). In contrast, positive contact with members of an outgroup has been shown to decrease such threat perceptions, which can ultimately improve attitudes towards them (N. Kim & Wojcieszak, 2018). Applying these insights to the digital context, thus offers an opportunity to investigate a wider range of situations and interactions that, by virtue of eliciting different affective states, may produce differential outcomes—both prejudice-enhancing and prejudice-reducing—for partisan evaluations.

2.5 Conclusion

Affective polarization, its nature, manifestations, and consequences, have been the subject of extensive scholarship in recent years. Yet, as we saw in this review, identifying its root causes is still an area of ongoing work. Proposed explanations for this phenomenon bring together valuable insights from both the political science and social psychology literature, but largely draw on cognitive models of political thinking and behavior. Throughout this chapter I have argued that these existing approaches are quite limiting and that taking into account the emotional states that people experience when encountering new information or interacting with politically different others may offer a more nuanced understanding of the drivers of partisan prejudice. The papers presented in the next three chapters do so in different contexts.

At a time when scholars still debate whether affective polarization is best explained by considerations of identity or ideology, Study I, presented in the next chapter, integrates insights from affective intelligence theory to investigate what happens when voters feel poorly represented by their own political party on issues they care about, and how this disconnect impacts their evaluations of their own and rival parties.

In Study II, I then draw on theories of intergroup emotions to investigate the extent to which digital interactions between politically different users are affectively polarized. I then go one layer deeper to explore how the content of their communications can mitigate or worsen these conversational dynamics.

In Study III, finally, I mobilize intergroup threat theories to test whether different types of partisan messages—especially those who praise or deride specific political parties—have different consequences for affective polarization.

3 The Paradox of Poor Representation: How Voter-Party Incongruence Curbs Affective Polarization[13]

3.1 Abstract

Research on the relationship between ideology and affective polarization suggests that ideological disagreement is a key driver of animosity between partisan groups. By operationalizing disagreement on the left-right ideological dimension, however, existing studies mostly overlook incongruence between voter and party on specific policy issues as a potential predictor of affective polarization. How does voter-party

[13] An abridged version of this paper has been revised and resubmitted to *The British Journal of Politics and International Relations*. As lead author, I conceptualized and designed the paper, cleaned up the data, performed most of the analysis and drafted the manuscript. A statement of my co-author's contributions is included in *Appendix A*.

incongruence impact affective evaluations of mainstream political parties and their leaders? This paper explores this question by deploying an instrumental variable approach on survey data from the British Election Study collected in the lead up to the 2019 UK General Election. In line with initial expectations, findings reveal that voter-party incongruence has a significant causal impact on affective evaluations. Perceived representational gaps between voter and party drive negative evaluations of the in-party and its leader and positive evaluations of the opposition, thus lowering affective polarization overall. These effects on evaluations of the in-party and the opposition leader are even stronger when that gap concerns an issue that partisans personally care about. The results offer a more nuanced perspective on the role of ideological conflict in driving affective polarization.

3.2 Introduction

Over the last two decades, several Western democracies have been plagued by rising animosity and dislike across political lines—a phenomenon referred to as 'affective polarization' (Boxell et al., 2020; Reiljan, 2019; Wagner, 2021). This is a worrying trend that poses known threats to democracy. Previous studies suggest that negative partisan affect may have negative consequences on interpersonal trust (Carlin & Love, 2018), makes citizens less likely to seek diverse perspectives on sensitive issues or compromise in political debates (M. J. Hetherington & Rudolph, 2015) and more likely to discriminate against out-partisans in hiring (Gift & Gift, 2015) and economic exchanges (McConnell et al., 2018). Partisan rivalry is also a key driver of prejudice—even surpassing other forms of racial and social bias (Westwood et al., 2018)—which, at its most extreme, can lead to severe forms of dehumanization (Martherus et al., 2019) and aggression (Kalmoe & Mason, 2019). Finally, animosity towards party elites has been shown to lower trust in political institutions, thereby fostering greater political cynicism and apathy (Layman et al., 2006). Given the implications of this phenomenon for social and political cohesion, understanding its foundations has emerged as a central research task.

While its exact origins are still disputed (Iyengar et al., 2019), a growing number of scholars suggest that interparty animosity is largely explained by ideology. Specifically, partisans hold their political opponents in contempt, because they perceive large ideological differences between themselves and their rivals on a

wide range of issues (Abramowitz & Webster, 2018; Rogowski & Sutherland, 2016). Studies in this line of work tend to treat ideology as one-dimensional—measuring congruence between political parties and voters or between political elites themselves along the left-right ideological dimension (Bougher, 2017; Gidron et al., 2019a; Rogowski & Sutherland, 2016). Political scientists, however, have long underscored the multi-dimensional nature of the political space; one which can also incorporates social, cultural and value-based cleavages (Albright, 2010; Bornschier, 2010; Kriesi, 2010; van Der Brug & van Spanje, 2009). For that reason, despite being mostly congruent at the ideological level, Western European political parties and party members can often find themselves at odds on policy issues that are poorly captured by the general left-right dimension, such as the question of European integration (de Vries & Marks, 2012; Thomassen, 2012), with deleterious consequences for traditional relationships between parties and voters (Hobolt & Rodon, 2020).

While the impact of ideological disagreement with the out-party on affective polarization is well-established empirically, in contrast, we lack a clear understanding of how perceived conflict with one's *own* party impacts affective evaluations. The research question guiding this paper is thus the following: How does voter-party incongruence impact affective evaluations of mainstream political parties and their leaders? Shedding light on this question is important for a number of reasons. First, a more careful look at the relationship between partisanship, evaluations and policy preferences will help bring nuance to ongoing debates about the micro-foundations of affective polarization. If ideological disputes were indeed the main driver of negative political affect—as recent work suggests (Lelkes, 2021; Orr & Huber, 2020)—by the same logic, not seeing eye-to-eye with one's party or political leader on policy should elicit negative evaluations of said party or leader, regardless of partisan allegiance. If the opposite were true, however, these findings would underscore the primacy of partisan allegiance in shaping affective political evaluations. Second, teasing out these mechanisms will also inform the most effective strategy to curb political animus within the electorate. Should perceived voter-party conflict on policy issues be a strong predictor of positive evaluations of the opposition, for example, then reducing affective polarization may be best achieved by emphasizing areas of compromise between political parties. If, on the

other hand, even being at odds with the in-party does not impact voters' dispositions, then efforts to depolarize the electorate should probably put a greater emphasis on moderating partisan attachment.

The current study contributes to these debates by examining how voter-party incongruence impacts affective evaluations of parties and candidates in the United Kingdom. To do so, it leverages data from the 2019 British Election Study (BES) and uses an instrumental variable design intended to isolate causal effects. The paper is structured as follows. The first section provides a theoretical overview of affective polarization, and formulate three hypotheses about the role that policy incongruence might play in shaping citizens' affective evaluations. After outlining the research design, the second section tests how policy incongruence impacts affective evaluations of parties and candidates alongside other demographic factors. The analysis shows that voter-party incongruence over policy preferences is, in fact, causally linked to lower levels of affective polarization. Perceived representational gaps between a voter and her own party have a significantly negative impact on her evaluations of that party and a positive impact on evaluations of the opposition. This negative effect on evaluations of the in-party and positive effect on evaluations of the opposition leader are even stronger when that gap concerns issues that are personally salient to voters. The results offer a more nuanced perspective on ideological conflict and its role in driving affective polarization, and suggest some avenues for how to reduce partisan animosity within the electorate.

3.3 Ideological Disagreement and Affective Polarization

Over the last two decades, levels of inter-partisan hostility have reached new highs in the US (Gidron et al., 2020; Iyengar et al., 2012) and across several European democracies (Huddy et al., 2018; Lauka et al., 2018; Reiljan, 2019). In two-party and multi-party systems alike, partisans are increasingly expressing animosity and distrust towards the opposition while holding favorable views of their political ingroup (Knudsen, 2021; Wagner, 2021). Recent studies show comparatively similar levels of affective polarization among British and US voters (Gidron et al., 2020). Labour and Conservative identifiers in particular hold widely divergent perceptions of one another, with Conservatives scoring their co-partisans 3.5 points

higher in terms of 'intelligence', 'honesty' and 'open-mindedness' than they score their Labour counterparts. The reverse is true for Labour party supporters, who are more likely to view themselves in a positive light while categorizing Conservatives as 'selfish', 'closed-minded', and 'hypocritical' (Hobolt et al., 2020).

Two significant theories have been advanced to explain this phenomenon thus far: one grounded in considerations of *identity*, and another grounded in considerations of *ideology*. Supporters of the former argue that animosity across party lines is rooted in partisanship's role as a social identity (Iyengar et al., 2012; West & Iyengar, 2020). In this view, partisans are more like football fans, rooting for their own team and vilifying competitors (Mason, 2018a; Miller & Conover, 2015). According to social identity theory, most people define themselves through their membership to various social groups, such as their family or local football club, which form a core part of their own identity (Tajfel, 1974; Tajfel & Turner, 1979). This subjective sense of belonging generates powerful attachments to those groups and their members, with distinct and predictable behavioral effects. Once identified, for example, members will generally strive to distinguish their own group positively from others by elevating their ingroup's status ("ingroup favoritism"), and developing biased perceptions of the outgroup ("outgroup denigration") (Brewer & Kramer, 1985; Miller & Conover, 2015). Partisanship itself has long been identified as a key social identity that predicts a range of political behaviors, from involvement in political campaigns (Huddy et al., 2015) to out-party evaluations (Iyengar & Krupenkin, 2018).[14] Several empirical studies confirm this: strong partisans tend to be more hostile and prejudiced towards the opposition, even when their issue positions are moderate. This is not just true of partisanship, but also of other subjective political identities, such as those formed around 'ideological groups'[15] (Mason, 2015), 'coalitions' (González et al., 2008), and 'opinion-based groups', including those of 'Leaver' and 'Remainer' formed in the wake of 2016 referendum on European Union membership (Hobolt et al., 2020).

[14] As Achen & Bartels, (2017) point out, 'in the political sphere, the most salient groups are parties and the self-justifications that sustain group life are primarily grounded in—and constructed to maintain—partisan loyalties' (p. 296).

[15] Crucially in this regard scholars have shown that a person's ideological identity (such as 'liberal' or 'conservative') and their policy preferences may be related, but are not necessarily concurrent (Malka & Lelkes, 2010; Mason, 2015) and should therefore treated as distinct phenomena.

Supporters of the latter challenge this position, arguing that there is an irreducible ideological component to affective polarization. Studies in this line of work posit that there is a rational, and even strategic, basis for partisans' negative evaluations of their political rivals. In this view, perceiving large ideological differences between major parties and candidates drives out-party animosity because it raises the stakes associated with partisans' electoral choice (Rogowski & Sutherland, 2016; Webster & Abramowitz, 2017). Put differently, the cost of picking the wrong party or candidate increases when the two major parties hold extreme and ideologically distinct positions. And this, according to spatial theories of voting, incentivizes negative evaluations of the opposition (Stokes, 1963). Several empirical studies provide evidence for this: manipulating the extent to which candidates appear to be more liberal or conservative substantially boosts negative evaluations of that candidate (Rogowski & Sutherland, 2016). Likewise, Webster and Abramowitz (2017) demonstrate that people who perceive political parties to be ideologically far apart are the most likely to be affectively polarized, and that priming participants about a candidate's extreme political preferences drives negative evaluations towards that candidate. When being presented with information about a political candidate's policy preferences or about their party affiliation in experimental vignettes, voters also tend to react more strongly to ideology—especially if the policy positions shown were extreme (Lelkes, 2021). Taken together this body of evidence thus suggests that, although relevant, partisan identity might only provide a '(noisy) cue' (Orr & Huber, 2020, p. 48) for the greater source of partisan animosity: perceived ideological differences (Abramowitz & Webster, 2018; Lelkes, 2021; Orr & Huber, 2020). This is especially significant given that partisans tend to widely exaggerate the ideological extremity of their political opponents (Ahler & Sood, 2018; Levendusky & Malhotra, 2016b). However, this line of reasoning heavily rests on the assumption that people's partisan identity and ideological preferences are always in reasonable alignment. While this is mostly true in the US (Abramowitz & Webster, 2018), the next section demonstrates that these conditions are far from the being norm, especially in Europe where political parties many voters tend to be cross-pressured and hold policy preferences that do not always match their stated party identity.

3.4 The Consequences of Voter-Party Incongruence for Affective Evaluations

Reasonable alignment between political parties and voters on policy issues is considered fundamental to good democratic representation (Pitkin, 2004). In the US, where most of the literature on the drivers of affective polarization is situated, the last 50 years have seen a steady polarization of the major political parties along the liberal-conservative continuum (Bougher, 2017). As political elites increasingly drifted towards the extremes, so did party supporters, with voters progressively sorting themselves into political parties on the basis of their ideological stances, and adopting increasingly consistent policy preferences (Dimock et al., 2014; Finkel et al., 2020; Levendusky, 2009, 2010).[16] During that time, American partisans have also come to perceive a growing ideological gap between themselves and the opposition (Levendusky & Malhotra, 2016b), while in contrast perceived distance with their own party on key issues such as abortion, social welfare and gay rights has remained fairly stable (Webster & Abramowitz, 2017).

In Europe, the picture is somewhat different. While parties and voters are reasonably well-aligned in *ideological* terms (Costello et al., 2012; Mattila & Raunio, 2006) this alignment often breaks down on issues that are high on the political agenda but do not neatly fall on the left-right ideological continuum, such as immigration and European politics (Dalton, 2017). Studies find that parties demonstrate some degree of responsiveness to voters' attitudes on the question of European integration (Williams & Spoon, 2015). Nevertheless, many scholars find that mainstream European parties do a fairly poor job of representing their voting base on the EU dimension, especially since the 2015 Euro crisis (Mattila & Raunio, 2006, 2012; Rohrschneider & Whitefield, 2016; Thomassen & Schmitt, 1997). After decades of ideological convergence on major policy issues, there are clear indications that the two British major political parties—the Labour and Conservative Party—are re-polarizing again (G. Cohen & Cohen, 2019). Unlike in the US, however, this has not swung voters towards the extreme or made them more ideologically consistent. If anything, British voters are more electorally volatile and cross-pressured than their American counterparts (British Election

[16]An important caveat is that only a small minority of US voters hold *fully* consistent liberal or conservative views (G. Cohen & Cohen, 2019). However, as Dimock et al. (2014) point out those voters have 'disproportionate influence on the political process' as they are likely than cross-pressured ones to vote and be actively engaged in politics.

Study, 2019; Gidron, 2020). Today, high proportions of British voters say they feel poorly represented by the major political parties on a wide range of issues (Dommett & Temple, 2019; Hansard Society, 2019)—a sentiment exacerbated by a bitter referendum on European Union membership that sparked sharp cleavages within the political class and the country as a whole (Phillips et al., 2018; H. Schmitt & Loughran, 2017).

There is already substantial evidence that representational gaps between parties and their supporters on multiple issue dimensions can destabilize party systems by disrupting traditional patterns of electoral behavior. Perceived incongruence on issues of immigration, economic redistribution, and European integration, for instance, have all been linked to stronger political disaffection with democracy (Stecker & Tausendpfund, 2016) and a higher propensity to switch to anti-establishment and challenger parties during elections (R. Bakker et al., 2020; Hobolt & Tilley, 2016; Hong, 2015). Despite the prevalence of voter-party incongruence, however, at this point it is still unclear whether it also impacts affective evaluations of political parties and leaders.

In that respect, theoretical models from political and social psychology, such as attitude and affective intelligence theories, offers some interesting insights on how attitude incongruence might shape political judgements. The basic tenet of attitude theory is that individuals' attitudes towards any social objects, groups or events are strongly determined by their own beliefs and affective orientation towards the attributes of these objects, and whether they are themselves in alignment or opposition to them (Fishbein & Ajzen, 1975; Heider, 1958; Ottati et al., 1992). Adapted to the context of this study, this signals that if someone believes that a political figure has a certain opinion on an issue, and feels negatively towards that opinion, then they are more likely to evaluate this person unfavorably. Providing evidence for this, studies have shown that voters tend to hold more favorable views towards parties and candidates whose issue stance they agree with, whilst having more negative views of those with whom they disagree, regardless of shared ideology (K. Hetherington, 1996; Ottati et al., 2002). This is explained by the fact that sharing the same party or ideological label does not necessarily lead voters and candidates to exhibit congruent patterns of policy preferences (Converse, 1964).

In the same vein, proponents of affective intelligence theory (AIT) also argue that individuals' emotional reactions to informational stimuli are key in shaping social evaluations, and the formation of political judgements in particular. In this view, if a voter holds a generally positive disposition towards a particular party or candidate—by virtue of being part of the same political ingroup, for instance (Campbell et al., 1960)—being exposed to new information that contravenes this prior expectation will trigger a negative affective response (MacKuen et al., 2007). By the same logic, if good representation is considered normative among most partisans, then failing to meet that standard might be experienced as a violation of that norm and therefore result in negative affect, feelings of disappointment, and diminished trust in one' party. Survey data indeed shows that respondents who perceive themselves to be at odds with their political party or representative are more likely to report feeling angry at them (MacKuen et al., 2007; Steenbergen et al., 2001; Wolak et al., 2003). This ambivalence—a disjuncture between one's partisan identity and one's feelings towards the party (Lavine, Johnston, Steenbergen, et al., 2013)—also weakens party identification, which may ameliorate feelings towards the opposition by lessening partisans' tendency to pit themselves again the opposition (Brewer, 2001). Building on this logic, in the context of this study one should therefore expect that high levels of incongruence between voter and party would drive more negative attitudes towards a respondent's own party and more positive attitudes towards the rival party. This leads to the two following hypotheses:

H1a: Voter-party incongruence increases negative perceptions of the in-party and its leader.

H1b: Voter-party incongruence increases positive perceptions of the out-party and its leader.

There are also good reasons to believe that the effect of incongruence on affective polarization is moderated by partisanship. Indeed, a vast literature underscores the power of partisanship to shape voters' understanding of political reality (Bartels, 2002). However, existing scholarship on this topic points in two different directions. Partisan-motivated reasoning has been shown to distort how

individuals process informational cues (Gaines et al., 2007; Goren, 2002; Groenendyk, 2013) and drive biased political evaluations, for example (Jilke, 2018). This could imply that strong partisans will be more reluctant to express dislike towards their own party—despite not seeing eye-to-eye with them on policy. However, even the staunchest supporters of the 'socio-psychological' model of party identification—which view partisanship as a deeply-rooted psychological attachment (Campbell et al., 1960)—concede that the primacy of partisanship in shaping individuals' political attitudes is not entirely independent from considerations of party performance and ideological compatibility (Bartels, 2002; Campbell et al., 1960; Fiorina, 1981). Writing about the possibility of incongruence between voter and party in *The American Voter,* Campbell et al. (1960) noted that when individuals develop attitudes inconsistent with the party line, these attitudes 'exert some pressure on the individual's basic partisan commitment' acting as 'potential agents of change' in their behavior (p. 135). This expectation is consistent with spatial theories of electoral behavior, which have long argued that voters readily update their partisan allegiance based on policy shifts (Downs, 1957) and notably whether they think the party they identify with represents their interests in a satisfactory way (Stokes, 1963). In the context of the British public, one might therefore expect voters to be sensitive to perceived disjuncture with their own party's policy, and adjust their evaluations and preferences accordingly. This leads to the following hypothesis:

H2: Partisanship positively moderates the relationships between voter-party incongruence and affective evaluations.

Finally, this paper turns its attention to the role of issue salience—the subjective degree of concern that individuals ascribe to specific issues—in shaping the relationship between congruence and evaluations (Dennison, 2019). Different issues matter to different people, and how personally important an issue is to someone can exerts powerful effects on their cognition. Studies shows that citizens are not only more motivated to inform themselves about issues that they personally care about (Jerit, 2008) but that they are also more likely to pick up on policy-relevant information (Bolsen & Leeper, 2013; Ciuk & Yost, 2016) and form stable

opinions about them over time (Boninger et al., 1995; Krosnick, 1990). Beyond increasing opinion stability, salience also drives awareness and consideration of party positions on these topics (Plescia & Staniek, 2017; Walgrave & Lefevere, 2013). For that reason, as Krosnick notes, salient attitudes are 'especially powerful anchors that cause individuals to see others as falling primarily into one of two groups: those with whom they agree […] and those with whom they disagree' (Krosnick, 1990, p.63). Of most relevance to this study, research also indicates that ideological proximity between voter and party tends to be higher on personally salient issues (Giger & Lefkofridi, 2014) and on issues emphasized by the party during elections (party-level salience), with niche parties performing better in both instances (Costello et al., 2020). These findings imply that the importance citizens attach to single issues play a fundamental role in the representational relationship between parties and voters. A large body of research already shows ideological proximity on salient issues to be a powerful determinant of candidate preferences and voting behavior (Krosnick, 1988, 1990). Similarly, several studies find that voter-party congruence on salient issues is a strong predictor of how positively citizens evaluate government performance, and whether they view political candidates as suitable leaders (Fournier et al., 2003; Rabinowitz et al., 1982). Extending these findings about candidate evaluations, finally, some scholars have shown that voters who ascribe a lot of importance to a policy issue readily switch their partisan allegiance if they perceive the party they identify with to be at odds with them on that issue (Carsey & Layman, 2006). Taken together, these findings thus lead to the following final hypothesis:

H3: Issue salience positively moderates the relationships between voter-party incongruence and affective evaluations.

3.5 Data & Methods

To investigate the relationship between voter-party incongruence and affective polarization, this study makes use of survey data from the 2019 pre-election wave of the British Election Study (BES) 2014-2023 panel (Fieldhouse et al., 2020). The BES has historically conducted post- and pre-election surveys of the British population on a number of demographic, political, and social items. Since 2014,

the BES has also conducted several panel studies with additional survey items that follow respondents over time through multiple waves, with each wave containing a 'core' representative sample of respondents established for the purposes of cross-sectional analysis. This weighted, nationally representative sample is used for subsequent analyses. Specifically, this study focuses on the representative subset of respondents who were asked to answer the personal issue salience item as part of the November 2019 survey of the British electorate. In line with other works in the literature, analyses are restricted to voters who identify with one of the two major parties in the UK—namely the Labour Party and the Conservative Party—including party 'leaners', that is voters who do not strongly support but instead lean towards one of the two major parties (Petrocik, 2009). This brings the study's total sample size to 4,748, covering 63.36% of all respondents and a full 74.21% of all party identifiers in the sample.

Model

One of the disadvantages of using cross-sectional, observational data from surveys is the difficulty to establish that the variables of interests are not endogenous to the phenomenon being investigated. This is particularly problematic in this case given the potential for projection effects—i.e. the fact that respondents could be projecting their own issue preferences onto parties and leaders they like, while ideologically distancing themselves from parties and leaders they dislike (Merrill et al., 2001). To circumvent this issue, this study employs and instrumental variable (IV) approach—specifically, multivariate two-stage least squares (2SLS) regression—to assess the causal relationship between the predictors of interest and affective evaluations of political parties and their leaders. The instrument in all cases is partisans' *actual* incongruence with their in- party (i.e., the gap between their reported issue positions and expert-identified party positions) as opposed to their *perceived* policy incongruence with the in-party (i.e., the gap between their own issue positions and perceived party positions). Details of these measures are provided in the section below.

Since actual incongruence can be expected to only impact affective evaluations through respondent's own perceptions, this instrument satisfies the exclusion restriction requirement for IV regression (Angrist et al., 1996). Another requirement of IV models is that the instrument is a strong predictor of the endogenous variable in the initial regression model: in this case, perceived incongruence. As shown in *Table 3.1.* below, actual incongruence is a strong predictor of perceived incongruence with one's party ($\beta = 0.49$, $p < 0.001$), easily exceeding conventional thresholds for instrumental validity (Staiger & Stock, 1997). The models adjust for ideological extremity as well as demographic covariates (age, gender, and education) at both stages of the IV design to control for unwanted variation. This IV design, similar to the ones deployed by Ansolabehere and Jones (2010) and Jessee and Malhotra (2013), addresses issues of projection among endogenous predictors, which would otherwise complicate the interpretation of regression outputs.

Table 3.1. OLS estimates of perceived incongruence with the in-party

	Perceived Incongruence	*Perceived Incongruence, Salient Issues*
	Estimate (SE)	Estimate (SE)
Intercept	0.50*** (0.11)	0.90*** (0.31)
Age	-0.001 (0.001)	0.003 (0.002)
Male	-0.02 (0.04)	0.09 (0.07)
Education	0.08*** (0.02)	0.02 (0.03)
Labour	-0.08* (0.04)	1.07*** (0.19)
Extremity	-0.03** (0.01)	-0.11*** (0.02)
Actual Incongruence	0.49*** (0.02)	
Actual Incongruence, Salient Issues		0.15*** (0.04)
Observations	3,365	2,634
Adjusted R^2	0.22	0.02
Residual Std. Error	1.15 (df = 3358)	1.71 (df = 2627)
F Statistic	154.80*** (df = 6; 3358)	10.63*** (df = 6; 2627)

*Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE = Standard Errors*

As if often the case with survey data, a number of cases presented missing information at the individual level. The percentage of missing values in the dataset ranged from 0% for some demographic covariates including age and gender, to as high as 14% for measures of educational attainment and ideological self-placement. Overall, however, only 2.5% of the dataset was missing, primarily due to non-response to certain items by participants. A Little's test was performed to ascertain that the missing data was indeed missing completely at random (F = 2091, p < 0.001). Given the low risk of bias introduction of the parameters estimates and its standard errors (Little, 1992), all IV regression models were conducted with listwise deletion of missing data. All variables are standardized to unit variance in order to facilitate the interpretation of coefficient estimates. Standard diagnostic tests were also performed for all reported models (see *Appendix A*), which did not provide any

reason to doubt the results presented. Results for the conditional association tests and interaction effects are reported in the main text, with the remainder of the regressions in *Appendix A.*

3.6 Measures

Party Identification: The BES party identification item asked respondents whether they generally think of themselves as 'Labour, Conservative, Liberal Democrat or what?' Depending on their answers, respondents were further asked if they were 'a little closer to one of the parties than the others.' Following Bankert et al. (2017), Gidron et al. (2020), Hobolt et al. (2020) and others, subsequent analyses are restricted to Labour and Conservative identifiers and include party leaners.

Strength of Partisanship: Strength of party identification was determined in the BES by asking respondents to rank how strongly they felt attached to the party they identify with on a 3-point scale ranging from weak (1) to strong (3) (M= 1.89, SD = 0.69).

Perceived Voter-Party Incongruence: Perceived voter-party incongruence is operationalized as the average distance between a respondent's own position and their estimation of their in-party's position on all possible sets of issues provided in the BES survey, namely economic redistribution, immigration, the environment, and the issue of the UK's membership in the European Union. For each issue, respondents were asked to evaluate their own as well as the Labour and Conservative parties' positions on a 0-10 ideological scale (see *Table A.2.* in *Appendix A* for exact wording). Perceived incongruence consists in the absolute difference between respondents' estimations of their own party's position and their self-position (M = 1.43, SD = 1.36). Out of all 4,748 respondents, 97.9% perceive some degree of incongruence between their averaged policy preferences and that of the party with which they identify.

Actual Voter-Party Incongruence: Actual ideological congruence is calculated by taking the average distance between respondents' stated positions and their in-party's actual positions on the same four aforementioned policy issues, as measured

by the BES through expert surveys conducted that same year. These independent surveys gather information on 'objective' party policy positions by asking members of the Elections Parties and Public Opinion (EPOP) specialist group of the Political Studies Association to place British political parties on a wide range of issues (Prosser, 2020). The issues that experts are asked to place parties on, as well as the question wording and response scales all matched those fielded to BES survey respondents. This combined data can thus be used to make party-voter comparisons (Prosser, 2020). Descriptive statistics show that actual incongruence with the in-party on these four policy issues is slightly higher on average than perceived incongruence based on respondents' estimates (M = 1.69, SD = 1.32).

Issue Salience: To identify which policy issues respondents selected as personally salient, responses to the BES survey item asking participants to indicate 'which policy is most important' to them among a set of multiple, non-mutually exclusive issues that include the economy, Brexit, immigration and the environment were collected. A full 94.4% of respondents in the sample picked at least one of the aforementioned issues as their most personally important, with the average respondent selecting 2.32 (SD = 1.14) (see *Table A.1.* in *Appendix A* for full descriptive statistics).

Affective Evaluations: Affective polarization is commonly operationalized as the absolute difference between respondent's affective evaluations of their own party and the out-party (Iyengar et al., 2019). Early research from social psychology, however, shows that positive feelings towards the in-party and negative feelings towards the out-party do not always go hand-in-hand (Allport, 1954; Brewer, 1999) and that they can, in fact, often evolve independently of one another (Webster & Abramowitz, 2017). To gain a more fine-grained understanding of how incongruence impacts affective evaluations, subsequent analyses thus disambiguate between evaluations of the in- and out-parties, and their respective leaders. Affective scores are based on the BES 10-point survey item, asking respondents to evaluate political parties and leaders on a scale ranging from 0 for 'strong dislike' to 10 for 'strong like.' On average, respondents in the sample gave more favorable evaluations of their own-party (M =7.26, SD = 2.18) and its leader (M = 6.34, SD

= 2.96) than of the opposition (*M* = 1.51, *SD* =2.03) and its leader (*M* = 1.31, *SD* = 2.28) (see *Table A.1* in *Appendix A* for full details).

Ideological Extremity: Respondents' ideological extremity, finally, is operationalized by taking the midpoint of the 10-point ideological self-placement item ranging from 'left' to 'right' and convert it to a 5-point range where 0 corresponds to the most neutral ideological point and 5 denotes the most ideologically extreme point (*M* = 2.14, *SD* = 1.39). High scores on this ideological extremity scale indicate strongly left-wing or strongly right-wing views on policy issues, while low scores indicate ideologically moderate positions.

3.7 Results

The first set of hypotheses were concerned with the effect of voter-party incongruence on perceptions of the in-party, the out-party and their respective leaders. Specifically, H1a predicted that incongruence would drive negative evaluations of the in-party and its leader, while H2b predicted that it would drive positive evaluations of the rival party and its leader. To address it, four separate 2SLS models were fitted to estimate effects of voter-party incongruence on affective evaluations of one's (a) in-party, (b) in-party leader, (c) out-party, and (d) out-party leader, respectively. Results are summarized in models M1.0 to M1.3 in *Table 3.2.*

Similar trends are evident across each pair of models. After controlling for all covariates, the analyses show that ideological incongruence leads to negative evaluations of one's in-party ($\beta = -1.10$, $p < 0.001$) and its leader ($\beta = -1.40$, $p < 0.001$), while increasing evaluations of one's out-party ($\beta = 1.07$, $p < 0.001$) and its leader ($\beta = 1.77$, $p < 0.001$). It is interesting to note that, in both cases, effects are stronger at the leader level than the party level, suggesting that ideological disagreements with one's own party may inspire some personal animus toward salient party representatives. Given the uniformity of results across these models, it appears that both H1a and H1b are corroborated by the evidence.

Table 3.2. Impact of voter-party incongruence on affective evaluations of political parties and candidates

	In-Party	*In-Party Leader*	*Out-Party*	*Out-Party Leader*
	M1.0	M1.1	M1.2	M1.3
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	7.57*** (0.06)	6.95*** (0.08)	1.43*** (0.06)	0.85*** (0.07)
Age	-0.07** (0.03)	-0.13*** (0.04)	-0.18*** (0.03)	-0.09** (0.04)
Male	-0.12* (0.07)	0.09 (0.09)	0.20*** (0.07)	0.13 (0.08)
Education	-0.05 (0.04)	-0.06 (0.05)	-0.04 (0.04)	-0.25*** (0.05)
Labour	-0.45*** (0.07)	-1.31*** (0.09)	-0.05 (0.07)	0.94*** (0.09)
Extremity	0.60*** (0.03)	0.82*** (0.04)	-0.52*** (0.04)	-0.34*** (0.04)
Incongruence	-1.10*** (0.08)	-1.40*** (0.10)	1.07*** (0.08)	1.77*** (0.09)
Observations	3,357	3,355	3,358	3,356
Residual Std. Error	1.90 (df = 3350)	2.44 (df = 3348)	1.97 (df = 3351)	2.33 (df = 3349)

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE = Standard Errors.

Table 3.3. Impact of voter-party incongruence on affective evaluations of political parties and candidates, by strength of partisanship

	In-Party	*In-Party Leader*	*Out-Party*	*Out-Party Leader*
	M2.0	M2.1	M2.2	M2.3
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	7.64*** (0.06)	7.01*** (0.08)	1.41*** (0.06)	0.87*** (0.08)
Age	-0.10*** (0.03)	-0.16*** (0.04)	-0.19*** (0.03)	-0.09** (0.04)
Male	-0.20*** (0.06)	0.02 (0.08)	0.21*** (0.07)	0.13 (0.08)
Education	-0.03 (0.03)	-0.06 (0.05)	-0.05 (0.04)	-0.26*** (0.05)
Labour	-0.53*** (0.07)	-1.36*** (0.09)	-0.01 (0.07)	0.97*** (0.09)
Extremity	0.31*** (0.03)	0.57*** (0.05)	-0.41*** (0.04)	-0.30*** (0.05)
Incongruence	-0.99*** (0.08)	-1.26*** (0.11)	1.01*** (0.09)	1.81*** (0.11)
Party Id Strength	0.75*** (0.04)	0.65*** (0.05)	-0.30*** (0.04)	-0.09* (0.05)
Incongruence * Party Id Strength	0.01 (0.08)	0.16 (0.10)	-0.04 (0.09)	0.20± (0.10)
Observations	3,326	3,325	3,327	3,325
Residual Std. Error	1.74 (df = 3317)	2.34 (df = 3316)	1.93 (df = 3318)	2.35 (df = 3316)

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE = Standard Errors.

The second hypothesis posits that the impact of voter-party incongruence on affective evaluations of parties and their representatives is moderated by strength of partisanship. To test this hypothesis, models M2.0 to M2.3 build upon the models presented above, adding a covariate for partisan strength as well as an interaction term between this variable and perceived ideological incongruence. The focus here on is the interaction term, which, after controlling for covariates including main effects, represents the moderating impact of partisan strength on the relationship between voter-party incongruence and affective evaluations. Results this time were less uniform and effect sizes generally smaller, which makes some sense given that ideological incongruence and partisan strength are inversely correlated ($\rho = -0.134$, $p < 0.001$). Results are summarized in *Table 3.3*. Only a single interaction term borders the significance threshold: voter-party

incongruence, even among strong partisans, still leads to positive evaluations of one's out-party leader ($\beta = 0.20$, $p = 0.055$). Therefore, one cannot reject the null hypothesis that partisan strength positively moderates on the relationship between voter-party incongruence and affective evaluations.

The final hypothesis concerns the moderating effect of issue salience on the relationship between voter-party incongruence and affective evaluations. Models M3.0 to M3.3 adopt a similar design as the four M1 models, except this time real and perceived voter-party incongruence are replaced with weighted versions of those variables, where weight 1 is given to issues a voter identifies as highly important, and 0 otherwise. Results are summarized in *Table 3.4.* There is some interesting variation between these results and those for the M1 models. Salience-weighted effects of incongruence on evaluations of one's in-party are similar to the

Table 3.4. Impact of voter-party incongruence on affective evaluations of political parties and candidates, by issue salience

	In-Party	*In-Party Leader*	*Out-Party*	*Out-Party Leader*
	M3.0	M3.1	M3.2	M3.3
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	7.58*** (0.10)	7.11*** (0.13)	1.43*** (0.10)	1.19*** (0.19)
Age	-0.07* (0.04)	-0.17*** (0.05)	-0.21*** (0.04)	-0.19** (0.08)
Male	-0.18** (0.08)	-0.01 (0.11)	0.19** (0.08)	0.01 (0.16)
Education	-0.06 (0.04)	-0.07 (0.05)	-0.04 (0.04)	-0.26*** (0.08)
Labour	-0.35** (0.14)	-1.46*** (0.18)	-0.08 (0.14)	0.33 (0.27)
Extremity	0.57*** (0.06)	0.93*** (0.07)	-0.56*** (0.06)	-0.20* (0.11)
Incongruence, Salient Issue	-1.12** (0.54)	-0.11 (0.70)	0.74 (0.53)	3.84*** (1.03)
Observations	2,629	2,628	2,629	2,629
Residual Std. Error	1.98 (df = 2622)	2.53 (df = 2621)	1.91 (df = 2622)	3.77 (df = 2622)

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE = Standard Errors.

unweighted results ($\beta = -1.12$, $p < 0.05$), the effects are much *stronger* on positive evaluations of the out-party leader ($\beta = 3.84$, $p < 0.001$). Meanwhile, model M3.2 shows insignificant effects of salience-weighted incongruence on out-party evaluations ($\beta = 0.74$, $p = 0.161$).

3.8 Discussion

The present research set out to examine the effect of voter-party incongruence on affective evaluations of political parties and their leaders. Drawing on a large-scale survey of British voters conducted before the 2019 UK General Election, this study finds that voters who feel poorly represented by the party they identity with are more likely to give negative evaluations of the in-party and its leader, as well as more positive evaluations of the opposition. This finding validates previous work on partisan ambivalence (Lavine, Johnston, & Steenbergen, 2013), which shows that experiencing a disconnect between one's partisan identity and one's belief in party representation generates 'ambivalence' that challenges traditional structures of political thinking. These findings also confirm the notion that personal issue salience *matters* for political evaluations, and can function as an extension of individuals' moral values and sense of self (Duncan & Stewart, 2007) in a way that drives support for parties who 'own' these issues (Neundorf & Adams, 2018).

Several findings stand out from these analyses. Across all models, the effects of voter-party incongruence were especially pronounced on evaluations of in- and out-party leaders. British partisans who felt at odds with their own party on issues they had singled out as important, were much more likely to view the leader of the opposition favorably, while expressing negative feelings towards their own party leader. There could be several explanations for this pattern. One is that feeling poorly represented by the leader of one's own party—the very incarnation of the ideals that the party is supposed to uphold—was lived as a personal affront or a betrayal by many partisans. Indeed, as politics in general and political campaigning in particular has become more personalized (Bennett, 2012), party leaders have increasingly come to replace political parties themselves as key heuristics for policy positions in partisans' minds (Berz, 2020). This is particularly true of British political elites, which have increasingly capitalized on and strategically positioned themselves on Brexit-related standpoints since the 2016 EU referendum (Ridge-

Newman, 2019). Recall for example, Boris Johnson's evocative promise to 'Get Brexit Done' or the Liberal Democrat leader Jo Swinson's campaign to 'stop Brexit' during the last UK General Election. If party leaders themselves have become proxies for party positions, they could therefore be bearing the brunt of the discontent when voters disagree with them on these issues.

Another plausible explanation for this pattern could be that as European political systems have become increasingly de-aligned (Dalton & Wattenberg, 2003), political parties have had to take stronger stances on wedge issues to compete against rising single-issue parties such as UKIP (Usherwood, 2008). Indeed, while the decoupling of class and partisan identity in post-war Britain brought with it perceptions of ideological 'decline' among UK party leaders, over the last few years there has been a resurgence of ideology in political discourse (G. Cohen & Cohen, 2019). [17] Shifts in the electoral landscape and the introduction of new societal issues such as growing concerns over immigration and the meaning of 'Britishness', the environment, and the UK's place within the European Union (Sanders, 2017) have all led political elites to adopt more internally homogenous, programmatic, and ideological stances to distinguish themselves from the opposition (Dahlgreen, 2014).[18] This realignment of partisanship and ideology at the elite level could have thus increased respondents' association of political figures with strong policy stances. Given that voters' attitudes on issues like Brexit and the environment themselves shifted during that time (Hobolt et al., 2021), this gap could have, in turn, resulted in higher perceptions of ideological disconnect with party leaders.

Interestingly, this study also shows the relationship between partisanship and party evaluations to be more heterogenous than commonly thought. Prior studies in the literature have tried to single out either partisanship or ideological conflict as the true cause of affective polarization (Orr & Huber, 2020; Rogowski & Sutherland, 2016; Webster & Abramowitz, 2017). What the analyses presented here make clear, however, is that doing so might be creating a false dichotomy,

[17] Katharine Dommett (2016), for example, contends that between 1947 and 1997 Labour and Conservative Party leaders engaged in a certain 'ideological quietism', moving away from referencing ideological concepts such as socialism and conservativism and instead putting a stronger emphasis on values and traditions.

[18] Jeremy Corbyn's leadership presents a notable exception to this, as the Labour Party's public stance on Brexit famously shifted multiple times between the 2016 and 2019 General Election (G. Lee, 2019)

because policy preferences and partisan attachment *both* impact affective evaluations in subtle ways. Ideological misalignment with one's own party—especially on wedge issues such as the UK's place in the European Union—does exert some pressure on partisan attachments by generating identity-challenging evaluations. The negative in-party evaluations provided by cross-pressured voters in this study are a clear example of this dynamic. Yet, this does not mean that partisanship does not matter at all. While even the strongest partisans were still likely to 'betray' their party by endorsing the leader of the opposition, they were not as prone to criticize their own party. Taken together, these findings thus reinforce the notion that partisans' evaluations of political parties and candidates are far from static, but instead contingent upon perceptions of parties' issue stances and performance (Greene, 2004). From a theoretical standpoint, they also inject some nuance into ongoing debates about the drivers of affective polarization by highlighting that, when it comes to affective evaluations, partisanship is neither a purely affective, uncritical bond nor a mere shortcut for congenial views, but instead a dynamic process with both emotional and evaluative elements.

This paper also enriches the nascent literature on strategies to mitigate negative partisan affect in the electorate (Huddy & Yair, 2021; Wojcieszak & Warner, 2020). Some scholars, including Orr and Huber (2020), have already suggested that 'if issue conflict is at the heart of [partisan] animosity, emphasizing… areas of common policy agreement may reduce animosity' (p. 571). While this study does not experimentally manipulate parties' policy positions—something which is difficult to execute in a convincing way (Huddy & Yair, 2021)—its instrumental variable design still provides a robust, causal test of the power of perceived intra-party conflict in reducing affective polarization. This suggests that emphasizing partisan ambivalence—through primes strong enough to make voters question their allegiance to their own party—may produce positive results in alleviating partisan discord. [19]

[19] Although these might be in itself quite difficult to achieve: for example, Levendusky (2018) shows that priming partisan ambivalence in a survey experiment only reduced affective polarization for the weakest party identifiers, despite working well in other contexts. To the author's own admission, however, 'one concern with null results is that the manipulation failed (i.e., the treatment did not increase ambivalence or affirm apolitical dimensions of individuals' identities)'

3.9 Conclusion

Poor voter-party representation is not normally regarded as a desirable, let alone normative feature, in a well-functioning democracy. Indeed, for many political theorists, democracy rests on an ideal of reciprocity and responsiveness between voters and the 'representatives and political parties who speak and act in their name' (Urbinati & Warren, 2008). Yet, somewhat counter-intuitively, the findings from this paper suggest that the existence of cross-pressures or perceptions of misalignment with the one's own party may actually be good for democracy, by lessening affective polarization overall. Taken together, the findings reported in this paper show a strong causal relationship between voter-party incongruence and affective evaluations. When voters feel that their party's position does not align with their own—especially on the issues that matter most to them—they display much more *positive affect* towards out-party leaders, and more hostility towards their own party. The former is true even among strong partisans, after controlling for all other variables. These effects confirm existing findings suggesting that emphasizing partisan ambivalence may help reduce negative partisan affect.

3.10 Appendix A

Table A.1 Descriptive statistics

Statistic	*N*	*Mean*	*St. Dev*	*Min*	*Max*
Age (years)	4,748	54.83	16.07	18	92
Gender (1 = male)	4,747	0.46	0.50	0	1
Party (1 = Labour)	4,748	0.45	0.50	0	1
Partisanship Strength	4,662	1.89	0.69	1	3
Ideological Extremity	4,083	2.14	1.39	0	5
Education Level	4,052	2.94	1.35	0	2
Most salient issue					
Brexit (1 = yes)	3,365	0.71	0.45	0	1
Immigration (1 = yes)	2,571	0.53	0.50	0	1
Environment (1 = yes)	1,979	0.42	0.49	0	1
Economy (1 = yes)	3,193	0.67	0.47	0	1
Issues selected (number)	4,748	2.32	1.14	0	4
Out-party ratings	4,703	1.51	2.03	0	10
Out-leader ratings	4,701	1.31	2.28	0	10
In-party ratings	4,709	7.26	2.18	0	10
In-leader ratings	4,699	6.34	2.96	0	10
Perceived Incongruence (out-party)	4,471	4.05	2.49	0	10
Perceived Incongruence (in-party)	4,515	1.43	1.36	0	10
Actual Incongruence (out-party)	4,719	3.63	1.86	0	10
Actual Incongruence (in-party)	4,719	1.69	1.32	0	10

Note: St. Dev stands for standard deviation.

Table A.2. BES survey items measuring self and estimated party policy positions

Policy issue	*Survey item*
Immigration	Some people think that the UK should allow many more immigrants to come to the UK to live and others think that the UK should allow many fewer immigrants. Where would you place yourself and the parties on this scale?
Economy	Some people feel that government should make much greater efforts to make people's incomes more equal. Other people feel that government should be much less concerned about how equal people's incomes are. Where would you place yourself and the political parties on this scale?
Environment	Some believe that protecting the environment should have priority even if that reduces economic growth. Others believe that economic growth should have priority even if that hinders protecting the environment. Where would you place yourself and the political parties on this scale?
Brexit	Some people feel that Britain should do all it can to unite fully with the European Union. Other people feel that Britain should do all it can to protect its independence from the European Union. Where would you place yourself and the political parties on this scale?

Source: Fieldhouse, E., J. Green, G. Evans, J. Mellon & C. Prosser (2020) British Election Study Internet Panel Wave 17.

Diagnostics

The following diagnostic checks were performed on the instrumental variable regressions. Residual plots were checked and indicated no significant deviation from normality. The linearity of the partial relationship between the response variables and all predictors were also inspected through visualizations. Finally, collinearity diagnostics were performed by examining generalized variance inflation factors. None of these checks provided any reason to doubt the validity of the results.

In addition to the instrumental variable model reported in the main text, below ordinary least squares regressions for all variables and interactions of interest are provided below. These models estimate the effects of voter-party incongruence on evaluations of the in-party, the out-party and their respective leaders, controlling for the same set of variables: age, gender, education, as well as party identification and ideological extremity.

Table A.4. OLS regression, effect of voter-party incongruence

	In-Party	*In-Party Leader*	*Out-Party*	*Out-Party Leader*
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	1.34*** (0.06)	0.69*** (0.06)	7.64*** (0.06)	7.02*** (0.08)
Age	-0.16*** (0.03)	-0.05 (0.04)	-0.08*** (0.03)	-0.15*** (0.04)
Male	0.22*** (0.07)	0.18** (0.07)	-0.14** (0.07)	0.07 (0.09)
Education	-0.03 (0.04)	-0.23*** (0.04)	-0.06* (0.04)	-0.07 (0.05)
Extremity	-0.57*** (0.03)	-0.44*** (0.04)	0.64*** (0.03)	0.86*** (0.04)
Labour	0.09 (0.07)	1.18*** (0.07)	-0.54*** (0.07)	-1.41*** (0.09)
Incongruence	0.32*** (0.03)	0.51*** (0.04)	-0.63*** (0.03)	-0.85*** (0.04)
Observations	3,358	3,356	3,357	3,355
Adjusted R^2	0.12	0.17	0.21	0.26
Residual Std. Error	1.84 (df = 3351)	2.00 (df = 3349)	1.84 (df = 3350)	2.38 (df = 3348)

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE = Standard Errors.

Table A.5. OLS regression, interaction of voter-party incongruence and partisanship

	In-Party	*In-Party Leader*	*Out-Party*	*Out-Party Leader*
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	1.31*** (0.06)	0.67*** (0.06)	7.71*** (0.05)	7.08*** (0.07)
Age	-0.17*** (0.03)	-0.05 (0.03)	-0.11*** (0.03)	-0.17*** (0.04)
Male	0.25*** (0.06)	0.19*** (0.07)	-0.23*** (0.06)	-0.003 (0.08)
Education	-0.04 (0.04)	-0.23*** (0.04)	-0.04 (0.03)	-0.06 (0.04)
Extremity	-0.43*** (0.03)	-0.34*** (0.04)	0.33*** (0.03)	0.59*** (0.04)
Labour	0.13* (0.07)	1.20*** (0.07)	-0.61*** (0.06)	-1.47*** (0.09)
Party Id Strength	0.25*** (0.03)	0.45*** (0.04)	-0.52*** (0.03)	-0.74*** (0.04)
Incongruence	-0.40*** (0.04)	-0.27*** (0.04)	0.80*** (0.03)	0.71*** (0.04)
Incongruence * Party Id Strength	-0.09*** (0.03)	-0.11*** (0.03)	0.09*** (0.03)	0.12*** (0.04)
Observations	3,327	3,325	3,326	3,325
Adjusted R^2	0.16	0.19	0.33	0.31
Residual Std. Error	1.80 (df = 3318)	1.98 (df = 3316)	1.68 (df = 3317)	2.29 (df = 3316)

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE = Standard Errors.

Table A.6. OLS regression, interaction of voter-party incongruence and partisanship

	In-Party	*In-Party Leader*	*Out-Party*	*Out-Party Leader*
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	1.38*** (0.07)	0.76*** (0.08)	7.65*** (0.07)	7.03*** (0.09)
Age	-0.20*** (0.04)	-0.09** (0.04)	-0.09** (0.04)	-0.15*** (0.05)
Male	0.21*** (0.08)	0.16* (0.08)	-0.21*** (0.08)	0.02 (0.10)
Education	-0.04 (0.04)	-0.26*** (0.05)	-0.06 (0.04)	-0.07 (0.05)
Extremity	-0.58*** (0.04)	-0.46*** (0.04)	0.62*** (0.04)	0.88*** (0.05)
Labour	0.001 (0.08)	1.04*** (0.09)	-0.47*** (0.08)	-1.33*** (0.10)
Incongruence In-Party, Salient Issues	0.36*** (0.04)	0.47*** (0.04)	-0.56*** (0.04)	-0.73*** (0.05)
Observations	2,629	2,629	2,629	2,628
Adjusted R^2	0.13	0.15	0.18	0.23
Residual Std. Error	1.88 (df = 2622)	2.09 (df = 2622)	1.91 (df = 2622)	2.46 (df = 2621)

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE = Standard Errors.

LONDON'S GLOBAL UNIVERSITY

UCL

London, 4 May 2021

Subject: Collaboration with Nahema Marchal

Dear Oxford Graduate Studies Committee,

This letter is to certify that Nahema Marchal was the primary author of our article, "The paradox of poor representation: how voter-party incongruence curbs affective polarization." Nahema conceived of the piece, conducted a thorough literature review, and drafted the manuscript. We split the computational burden of the experiments, such that she cleaned the data and performed multiple imputation, while I implemented the two-stage least squares regression models and associated statistical tests. A complete version control history of the project's code can be found at our dedicated GitHub repository: https://github.com/dswatson/Affective_Polarisation.

In summary, the work is substantially her own, and I have no objections to her submitting a revised version of the article as a chapter of her doctoral thesis. Please do not hesitate to reach out if you have any questions.

Sincerely,

David Watson

Postdoctoral Research Fellow
University College London
Department of Statistical Science
1-19 Torrington Place, London WC1E 7HB
Telephone: +44 (0)7478 730102
Email: david.watson@ucl.ac.uk

4 'Be Nice or Leave Me Alone': An Intergroup Perspective on Affective Polarization in Online Political Discussions[20]

4.1 Abstract

Affective polarization—growing animosity and hostility between political rivals—has become increasingly characteristic of Western politics. While this phenomenon is well-documented through surveys, few studies investigate whether and how it manifests in the digital context, and what mechanisms underpin it. Drawing on social identity and intergroup theories, this study employs computational methods to explore to what extent political discussions on Reddit's *r/politics* are affectively polarized, and how intergroup communicative factors shape these affective biases. Results show that interactions between ideologically opposed users were significantly more negative than likeminded ones. These interactions were also more likely to be cut short than sustained if one user referred negatively to the other's political ingroup. Conversely, crosscutting interactions in which one of the users expressed positive sentiment towards the outgroup were more likely to attract a positive than a negative response, thus mitigating intergroup affective bias. Implications for the study of online political communication dynamics are discussed.

4.2 Introduction

As digital spaces have grown as arenas of political communication, scholars have become increasingly concerned with the nature and diversity of online political talk. Digital communication has strong affective dimensions (Papacharissi, 2014; Stark, 2020) and there is a bourgeoning literature exploring how emotions and sentiment on social media impact various political outcomes, from political

[20] An abridged version of this paper has been revised and resubmitted to *Communication Research.*

misperceptions to engagement (Eberl et al., 2020; Papacharissi, 2014; Papacharissi & De Fatima Oliveira, 2012; Weeks & Garrett, 2019). Over the last two decades, the question of whether political discussions take place in ideological silos has also emerged as a major source of scholarly inquiry (for reviews of the literature see Barberá, 2020; Tucker et al., 2018). Individuals' tendency to associate with similar others and to seek out content that reinforces their pre-existing beliefs are well-established precepts of communication research (Knobloch-Westerwick et al., 2015). Evidence for them in the online context, however, is mixed. While some degree of ideological homophily is commonly observed in online social networks (Bright, 2018; Halberstam & Knight, 2016), a growing body of studies challenge the 'echo chamber' hypothesis (Bruns, 2019; Zuiderveen Borgesius et al., 2016), arguing that Internet and social media are key drivers of exposure to oppositional viewpoints (Bakshy et al., 2015; Barberá et al., 2015; Barnidge, 2017; Dubois & Blank, 2018).

Despite the fact that individuals frequently come into contact with others they disagree with online, however, we still know relatively little about the *affective dimension* of these interactions. In recent years, public opinion scholars have noted the rise of 'affective polarization' across the US and Europe: growing animus and aversion between members of rival parties and ideological groups (Iyengar et al., 2019; Mason, 2018a; Reiljan, 2019; Wagner, 2021). Yet, thus far few studies have explored whether online political discussions are characterized by similar dynamics and how intergroup communicative factors might be shaping these patterns. This is especially pertinent, as social media continues to provide exposure to politically-charged content and communications with the potential to distort social attitudes towards others (Settle, 2018a; Zhuravskaya et al., 2020). While users may engage in frequent crosscutting talk, the nature of these exchanges (especially if adversarial) might carry severe consequences for political trust and tolerance.

Using data from Reddit's *r/politics*—one of the largest and most active political discussion forums online—in this paper, I leverage computational methods to show that ideological rivals are more likely to engage in negative political exchanges with one another than they are towards likeminded peers. Results further indicate that interactions between opposed users are more likely to end than be followed up by another exchange if they refer negatively to the other's

political ingroup. Conversely, expressions of positive sentiment towards the outgroup in crosscutting talk are more likely to drive further positive engagement, thus mitigating affective polarization. The contributions of this study to the literature are twofold. First, it makes an important methodological contribution by suggesting a way to identify the ideological leanings of Reddit users based on their prior engagement in other political communities on the site. Second, these findings expand the current literature on affective polarization by showing that crosscutting online political discussions are characterized by affective biases—with users treating their political ingroup more favorably than the outgroup—and that these interactions are significantly mediated by the direction and the sentiment of the messages exchanged between users.

The following sections are structured as follows. In the next section, I first provide a theoretical overview of affective polarization as a process driven by cognitive and affective responses to other people's political group identity, and show how particular features of communication in an anonymous online setting can activate these political identities. Following this contextualization, in the second section I build on social identity and intergroup relations theories to generate hypotheses about how intergroup communicative factors might shape affective polarization in political discussions. The third section details the data collection, methodology and operationalization of key variables, and reports results. Findings are then discussed in light of their implications for future research on online polarization.

4.3 Affective Polarization and Political Group Identity

Political identity is a powerful predictor of social and political behavior. Decades of research show that partisanship and ideological affiliation are key in determining our perceptions, evaluations and treatment of others. Partisans are not only less inclined to trust and interact with political opponents, but also overwhelmingly discriminate against them in resource allocation, online labor markets (McConnell et al., 2018), and in their dating preferences (Huber & Malhotra, 2017). For social identity theorists, these biases are the direct corollary of partisanship's role as a social identity. According to this view, most people develop a sense of self through their membership in various social groups, such as their family or football club

(Tajfel, 1972). This sense of belonging, in turn, generates powerful allegiances to the group ('us') that, in turn, shape our actions and social attitudes towards perceived outgroup members ('them') in predictable ways (Brewer & Kramer, 1985). Individuals tend to preferentially engage with ingroup members, for instance, and generally strive to protect their own group's status and standing, while developing negative perceptions and hostile feelings towards perceived rivals: a pattern described as 'ingroup favoritism' and 'outgroup hostility' (Brewer & Kramer, 1985). Scholars have also suggested that people can form identities around ideologies and opinion-based groups (Devine, 2015; Malka & Lelkes, 2010; Mason, 2018b) that are just as potent as partisanship in generating intergroup biases (McGarty et al., 2009). Following from this idea, in this paper I focus specifically on the 'liberal' and 'conservative' ideological identities as drivers of social behavior.

The core mechanism behind the effects of group identity on social behavior is identity salience (Mullen et al., 1992)—how significant an identity is to one's self-concept and to their perception of others—which itself, hinges on other factors such as how frequently one is reminded of their affiliation to the group (Gaertner et al., 1993). In situations of intense intergroup competition, individuals are acutely aware of their group membership and therefore primed to feel greater animosity and aversion towards members of the opposite team. There is mounting evidence that several features of online communication environment make political identities highly salient, with direct consequences for social behavior. According to social identity model of deindividuation effects (SIDE) scholars, the relative absence of personally identifiable information in anonymous digital settings compared to face-to-face settings, coupled with the abundance of partisan language and cues denoting political group membership, including 'we' talk and other ways to discursively invoke one's ingroup or distance oneself from outgroups (Carr, 2017; Hinck & Carr, 2020; Scott, 2007), all make political labels more salient, giving them more weight during interactions (Lea et al., 2001). [21] Confirming this pattern, SIDE research has shown that internet users operating under conditions of anonymity tend to form impressions of each other based on social group affiliations

[21] In digital settings in particular, others tend be identified as part of the 'outgroup' by virtue of the fact that they do not present features that make them recognizably part of the ingroup. In other words, anyone is not clearly an ingrouper is more likely to be seen as part of the outgroup (Douglas & McGarty, 2001).

rather than individual differences (Li & Zhang, 2018; Postmes et al., 1998; Spears et al., 2002). Recent scholarship also shows that Facebook users readily infer the political affiliation of their interlocutors based on the content they post (Settle, 2018a) and modify their online behavior accordingly. Salient political identities have been shown to determine people's word choices (Tamburrini et al., 2015), their propensity to be uncivil in online conversations (Gervais, 2015; Rains et al., 2017), their evaluations of users (Settle, 2018a; Suhay et al., 2018), as well as their responses to moral suasion from perceived ingroup members (Munger, 2020). While none of these studies specifically focus on Reddit, given the pseudonymity affordances and community-based structure of the site (which I present in greater detail below), it is reasonable to assume that participation on the platform will trigger similar processes.

4.4 Affective Dynamics and Intergroup Communication

Emotional expression is a core feature of political discourse on social media (Stark, 2020; Tettegah, 2016). From sharing live reactions to political events, to engaging in public displays of mourning or unabashed joy, affective statements 'define modalities of belonging that are articulated as strangers connect and attach to each other' (Papacharissi, 2014, p.117). The centrality of emotion in online discourse is also exemplified by the widespread use of emojis, and the ubiquity of emotional 'reaction' features (e.g., like, upvote or heart buttons) across social media platforms like Facebook, Twitter, Instagram or Reddit. Research shows that such affective information is easily transferred from person to person within online social networks (Ferrara & Yang, 2015; Harris & Paradice, 2007; Kramer et al., 2014). Experimental evidence has indeed suggested that individuals on the receiving end of a message can easily detect the emotional state of a sender through linguistic markers and emotional cues (Harris & Paradice, 2007). Despite the fact that most Internet and social media users regularly come across individuals they disagree with online (Barnidge, 2017), however, studies that investigate the affective dynamics of interactions between politically opposed users are still scant. This is partly explained by the difficulty of inferring the party or ideological affiliations of large numbers of users (Barberá & Alvarez, 2015) and of adequately linking message sentiment to different types of interactions (Hillmann & Trier, 2012).There

are a handful of notable exceptions, however. Comparing the incidence of outgroup hostility in discussion of a political controversy in Israel across different platforms, Yarchi et al. (2020) find that interactions across political camps on Twitter and WhatsApp tend to be more negative than the ones taking place between supporters. Using network-based methods to analyze Twitter discourse on climate change, Tyagi et al. (2020) similarly find climate deniers to be more hostile towards climate believers than towards likeminded users and vice versa. Based on these prior findings, I therefore hypothesize that crosscutting interactions on Reddit will be more negative than likeminded ones:

H1: Interactions between ideologically opposed users will be more negative than likeminded ones.

Beyond efforts to document the phenomenon, thus far little attention has been paid to the communicative factors that could be reinforcing or mitigating these biases between users. Decades of psychological research have established that individuals respond differently to positive and negative emotional stimuli—a phenomenon termed 'negativity bias' (Soroka, 2014)—with negative information typically generating stronger affective and behavioral responses than positive information (for a review, see Skowronski & Carlston, 1989; Soroka et al., 2019). In the digital context, however, studies that explore the role of emotions in shaping online conversation dynamics have thus far yielded conflicting results. Leveraging digital trace data, several large-scale studies of emotional contagion and reaction on social media appear to confirm the 'negativity bias' hypothesis. Emotionally-charged posts—especially negative ones—have been shown to travel faster, elicit more feedback and more shares than neutral ones (Hornik et al., 2015; Stieglitz & Dang-Xuan, 2013). Contradicting these findings, a number of recent studies indicate that positive content actually attracts more attention from online users (Nave et al., 2018), generating more feedback and reciprocity (Stieglitz & Dang-Xuan, 2012) and spreading wider than negative content (Berger & Milkman, 2012; Ferrara & Yang, 2015; Kramer et al., 2014). Communication scholars have also shown that the presence of incivility in online political talk elicits negative emotions in those who encounter it (Gervais, 2015; Han & Brazeal, 2015), which can, in

turn, trigger copycat aggression (G. M. Chen & Lu, 2017). Incivility is still a contested concept, however, and although uncivil speech in the form of profanities or derogatory comments is mostly negatively valenced, not all negative utterances amount to incivility (G. M. Chen, 2017). It is thus important to distinguish between both concepts. Moreover, several scholars have argued that incivility alone does not preclude meaningful dialogue across political lines and that forms of expressions such as hate speech carry far worse consequences for political tolerance (Bilewicz & Soral, 2020; Rossini, 2020).

Some of the conflicting patterns identified above, I would argue, can be explained by the fact that the role of intergroup dynamics in shaping behavioral responses to message sentiment is rarely considered in studies of social media communication. Yet, how a user experiences and responds to emotionally-charged messages is itself contingent upon the social context within which these are shared (Mackie & Silver, 2004). Research shows, for example, that partisans are more likely to perceive their political rivals as more impolite than their peers during an exchange, even if the two engage in exactly the same behaviors (Muddiman, 2017; Mutz, 2015). Studying how online political talk impacts users' behavior towards one another thus requires one to look beyond message sentiment in isolation, and examine instead how the context within which a message is shared and who it is aimed at impacts its reception. Here, work from intergroup emotion theory (IET) in particular offers some useful avenues.

First, it is important to specify that communication need not be taking place between groups as units to be regarded as 'intergroup'; rather the term applies to situations in which 'the transmission or reception of message is influenced by the group memberships of the individuals involved' (Harwood et al., 2005, p.3). Scholars of intergroup emotions have shown that individuals can experience emotions on behalf of their ingroup and so even in the absence of other group members; a finding that has been replicated in a wide range of settings (Hogg et al., 2004). IET likewise postulates that individuals may experience different group-level emotions (such as fear, anger and pride) depending on the intergroup context in which they are operating (Iyer & Leach, 2008). These are generally triggered by events or appraisals that can be interpreted as harming the individual's in- group. Exposure to hostile attacks, criticisms or threats directed at one's ingroup, for

example, tend to be lived as personal attacks—especially among strong partisans (Mackie et al., 2008). Besides being taken personally, derogatory comments of this sort can, in turn, provoke feelings of anger, humiliation and hate, thus motivating a desire for revenge by attacking 'the other side' (Lickel, 2012). Extending these findings to the context of crosscutting online discussions between users from opposing ideological inclinations, one would therefore expect that negative sentiment aimed at a sender's ingroup would spur others to retaliate:

H2a: In crosscutting interactions, a negative reply that mentions the sender's ingroup will be more likely to be followed by a negative response than a positive one.

Other studies in the literature point in a slightly different direction. Another branch of social psychology concerned with the potential for communication to stoke or mitigate intergroup rivalries is intergroup contact theory (Hewstone et al., 2014). At odds with the first proposition, work in this area has found that negative encounters with members of an outgroup can increase individuals' perceptions that the group is threatening, therefore reinforcing intergroup anger (Hayward et al., 2017). In these situations, negative contact may actually hamper intentions to engage in contact with said outgroup in the future (Barlow et al., 2012)—a phenomenon called the 'avoidance generalization effect' (Meleady & Forder, 2019). These contrasting findings leads me to formulate the following competing hypothesis:

H2b: In crosscutting interactions, a negative reply that mentions the sender's ingroup will be more likely to end the discussion than be followed up by another crosscutting response.

Even though individuals are inevitably going to come across members of political communities they are opposed to online, as conflict scholars often underline the mere existence of incompatibility 'does not always result in a confrontation' (Knapp & Daly, 2011). To the contrary, it is reasonable to assume that under certain conditions amicable exchanges between dissimilar users could have positive effects on future contact between those groups, thus decreasing affective bias. Touching on this possibility, the 'intergroup contact hypothesis', which was first

developed by Allport (1954), suggests that positive and meaningful interactions between members of different social groups can ameliorate intergroup relations and reduce intergroup prejudice (Hewstone et al., 2014). Studies of interpersonal behavior already indicate that associating other individuals or groups with positive events or experiences strengthens attraction and produces positive feelings towards them (Veitch & Griffitt, 1976). Online communication has expanded possibilities for such contact and scholars suggest that the specific affordances of computer-mediated communication—disembodied, text-based exchanges coupled with a lack of interpersonal cues—make digital discussion environments particularly fitting venues for effective intergroup interactions (Amichai-Hamburger, 2008; N. Kim & Wojcieszak, 2018). Indeed, recent studies show that positive direct online contact with outgroup members through online comments improve user attitudes toward them and predicted intentions to interact again in the future (N. Kim & Wojcieszak, 2018). In the context of this study, I therefore expect that positive references of the outgroup will be followed by a positive response.

H3: In crosscutting interactions, a positive reply that mentions the sender's ingroup will be more likely to be followed by a positive response than a negative one.

Affective polarization is not limited to derogation of and negative affect towards the political outgroup, however. In fact, scholars of intergroup relations have long argued that intergroup conflict is largely driven by 'ingroup favoritism' rather than 'outgroup hate' (Halevy et al., 2012). Put differently, when given the choice, people are generally more motivated to support their ingroup (rewarding them in resource allocation, for example) and avoid the company of outgroupers (Iyengar et al., 2019; Mason, 2015) than to directly harm them (Amira et al., 2019; Brewer, 1999; Iyengar & Krupenkin, 2018). Several factors could be reinforcing this tendency towards ingroup favoritism. Research from the IET literature demonstrates that, in intragroup settings, expressing positive feelings such as pride towards the ingroup has positive functional effects on strength of identification, cooperation, and motivation for future action, regardless of whether the group is a long standing or fleeting one (Kessler & Hollbach, 2005; Knight & Eisenkraft, 2015). Coupled with findings suggesting that positive discussions among

likeminded peers strengthens group identity (Yardi & boyd, 2010) and encourage continued participation (Joyce & Kraut, 2006), one would therefore expect that positive references to the ingroup in a likeminded conversation would drive further contact between group members.

H4: In likeminded interactions, a positive reply that mentions the sender's ingroup is more likely to be followed by a positive response than a negative one.

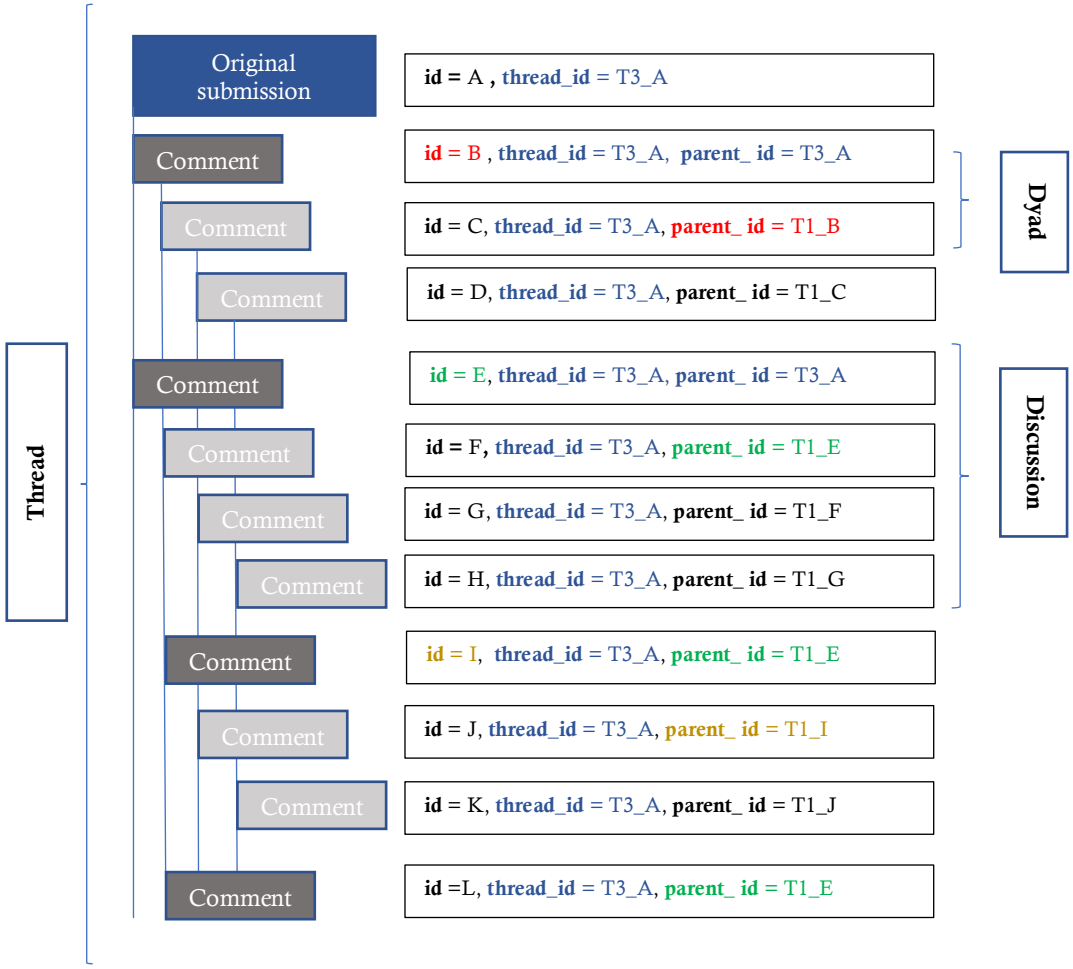
4.5 Data and Methodology

Testing the hypotheses outlined above requires first a careful examination of users' ideological affinities. Moreover, it requires identifying the emotional valence and content of their exchanges, including whether these are more positive or negative, and whether they mention specific political entities. To this end, textual and behavioral digital traces can provide valuable insights, provided they are considered as bounded by the social practices and affordances of the medium through which they arose (Jungherr et al., 2017). In this study, I turn to such traces of commenting activity to map out instances of crosscutting and likeminded interactions in my dataset, and extract the sentiment of messages exchanged to determine the nature of these interactions.

Data for this study were collected from the Reddit's *r/politics* subreddit between October 15th, 2018 and November 5th, 2018—the day prior to the 2018 US midterm elections—as this coincided with a period of heightened political discourse. Data were collected using Jason Baumgartner's Pushshift.io API (Baumgartner et al., 2020). Despite its large user base, Reddit remains surprisingly under-studied as a platform of political communication. A bibliometric search for titles containing the word 'Reddit' in Web of Science's publication database performed on April 2nd, 2021 only yielded 188 results, compared to 8,191 entries for Twitter and 6,729 for Facebook. Reddit is a compelling platform to study social and political behavior for a number of reasons. Unlike micro-blogging sites like Twitter that limits posts to 280 characters, Reddit does not impose length restrictions on comments and messages, allowing users to engage in both short and reactive exchanges, as well as long and more sophisticated discussions of political topics. Besides, Reddit's pseudonymity affordances and open access ethos—

allowing anyone to participate in message boards—has also created a strong political culture on the site making it a fertile ground for the study of political exchanges.

Figure 4.1. Model of Reddit discussion thread



Note: This figure breaks down the elements of a Reddit discussion thread. Thread_id corresponds to the unique ID of a thread, parent_id is the ID of the comment a user is responding to and id is the unique ID of a comment.

With over 5 million subscribers, *r/politics* in particular is one of the largest and most active political forums on Reddit—itself the third most visited social networking site in the US (Statista, 2021). Conversations on Reddit are organized, bulletin-board style, around original submissions that other users can comment on. Every comment posted on Reddit is identifiable by a series of parameters, which make it easy to identify if a comment is posted in response to an original submission or in response to another user's comment. As can be seen in *Figure 4.1.* above

comments and replies nested under an original submission together make up a discussion thread, which can span anywhere between a minute to several months for the most popular ones.

After filtering out comments that had either been removed, posted by moderator bots or by authors who had deleted their accounts since the data collection, 1.2 million valid comments from 141,171 unique authors, nested in 20,623 discussion threads remained in the dataset. Since the units of interest for this analysis are interactions between pairs of users (a comment followed by at least one response), only discussion threads (12,447 in total) that contain more than one interaction are included in the final analysis. Following all transformations, the final dataset contains a total of 176,339 comment-response dyads.

4.5.1 Estimating Users' Ideological Leaning

Research shows that people tend to engage in online communities that reflect their latent ideological preferences (Barberá & Alvarez, 2015; Shi et al., 2019). Here, I follow Shi et al. (2019) in estimating the ideological position of users in my sample based on their total number of contributions to liberal-leaning and conservative-leaning subreddits outside *r/politics* over a 6-month period ranging from June 2018 to November 2018. For this task, I downloaded a publicly available list of 200 politically-relevant subreddits, known as the Reddit Politosphere, from the *r/politics* wiki in August 2019. I then refined and augmented this list by searching for 'politics' in Reddit search bar and collecting the URLs for all suggested subreddits not already listed, culminating in a final list of 258 subreddits. A team of two coders then proceeded to classify each subreddit by ideology based on a pre-defined coding scheme. Ideological orientation is, of course, not unidimensional and may be denoted in a number of ways, from stated social values and positions on policy issues to one's attitudes towards political foes (Hanson et al., 2019). Linguistic choices in political conversations are also regarded as a reliable indicator of an individual's ideological leaning: what liberals refer to an 'estate tax', for example, is typically called the 'death tax' by a conservative, with no ideologically neutral alternative (Lakoff, 2002). With this in mind, a holistic approach was taken to the classification scheme (the full codebook, confusion matrix and example codes for this task are all available in Table B.1. to B.3., and in Figures B.1. and B.2. of

Appendix B). In each case, the subreddit's title, its description, guidelines as well as the content posted in it were all taken as potential indicators of its users' ideological orientation. Both coders reached high inter-rater reliability (Krippendorf's α = 0.77) during the test coding round (n = 20) and any disagreements were resolved by the author.

Having collected information about users' commenting activity in liberal-leaning and conservative-leaning subreddits, I then proceed to estimate their ideological alignment using a conservative Bayesian estimation framework (Shi et al. 2019). A 'neutral' beta prior $\text{Beta}(1/3, 1/3)$ was selected (Kerman, 2011), which assumes that each author is as likely to contribute to a liberal as they are to contribute to a conservative subreddit. This is designed to prevent 'lurkers' and occasional commenters from introducing too much uncertainty in the overall estimation of Redditors' ideological alignment.[22] If p is assumed to have a prior distribution $p \sim Beta(a, b)$ before any observations, then the distribution is updated using Bayes' law with observations X (number of comments contributed to liberal subreddits) and Y (comments contributed to both liberal and conservative subreddits), such that: $p|X \sim Beta(a + X, b + Y - X)$. Finally, political alignment is defined as the posterior mean of p:

$$E[p|X] = \frac{(X + a)}{(Y + a + b)}$$

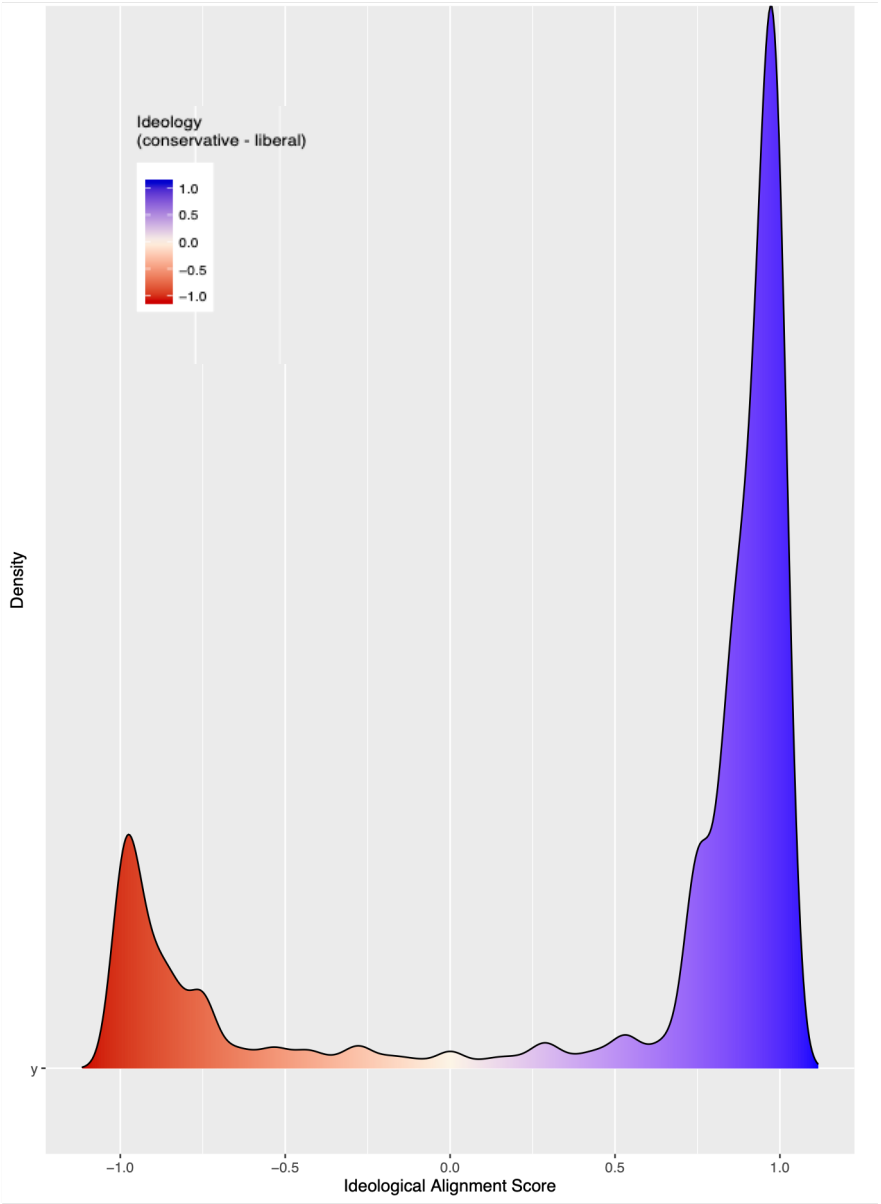
Following this method, I was able to infer the ideological leaning of 35,794 authors. To validate these estimations, a random sample of 100 comments posted by authors who had been classified as either 'liberal-leaning' or 'conservative-leaning' through this method was manually coded following the same coding scheme used to classify subreddits. Comparing the classification generated by the Bayesian framework with that obtained through manual coding resulted in a 0.78 accuracy score (95% CI: 0.68, 0.86). *Figure 4.2* below shows the distribution of ideological alignment for *r/politics* users in my sample.

One may note that the distribution is quite skewed towards liberals (77.8%), with the peaks at both extrema corresponding to users who have made the highest number of contributions to liberal or conservative subreddits. As shown in *Table 4.1.,* the vast majority of interactions (69%) over the data collection period appeared

[22] Authors were included if they had at least contributed one comment to either type of subreddit.

to be taking place between likeminded users, especially liberal-leaning ones, while crosscutting interactions were more marginal (31% of total).

Figure 4.2. Ideological alignment of Reddit users



Note: The distribution of ideological alignment of Reddit users, ranging from -1 (most conservative, where users have contributed exclusively and significantly to conservative-leaning subreddits) to +1 (most liberal, where users have contributed exclusively and significantly to liberal-leaning subreddits). N = 35,794. Y-axis represents the proportion of authors who were assigned a particular ideological alignment score.

Table 4.1. Distribution of interaction types

Type of interaction	N	%
Likeminded	121,939	69
Conservative-Conservative	6,463	
Liberal-Liberal	115,476	
Crosscutting	54,400	31
Total	176,339	100

4.5.2 Sentiment Scores

Having predicted the ideological leanings of users in my sample, I then move on to extracting sentiment scores for each comment in the dataset. Sentiment scores are compiled using NLTK's Vader Sentiment Analyzer—a human-validated sentiment analysis package developed for micro-blogging and social media. VADER is a rule-based sentiment model that has both a dictionary and associated intensity measures that returns a compound sentiment score in the [-1, +1] range, where -1 is the most negative sentiment and +1 is the most positive sentiment (M = -0.07, SD = 0.53). Although prediction performance for off-the-shelf sentiment analysis techniques can vary considerably from one dataset to another, VADER has been shown to be the best performing method to classify positive, negative, and neutral sentences (3-class tasks) in comments when compared with twenty-four other popular tools (Ribeiro et al., 2016). To test VADER's estimations, I manually annotated a random sample of 300 comments from the dataset as positive, negative or neutral, resulting in overall 0.70 accuracy score (see *Appendix B* for confusion matrix).

Only comments that scored either highly negative with compound scores in the 25th% percentile, and comments that scored highly positive with scores in the 75th% percentile were classified as respectively 'negative' or 'positive', while all others were classified as 'neutral'. While this technique is robust to predict the emotional valence of comments, it should be acknowledged that sentiment analysis

is limited in its ability to identify and deal with instances of sarcastic or non-literal comments (Muresan et al., 2016).

4.6 Measures

4.6.1 Dependent Variable

Affective polarization is defined as partisans' tendency to feel negatively towards and discriminate against their political opponents, while feeling greater affinity and acting more positively towards their co-partisans (Iyengar et al., 2012). The most common measure of affective polarization in the literature is based on survey self-reports. Specifically, it is computed as the absolute difference between the evaluation score given to a respondent's own party and the score given to the rival party (Iyengar et al., 2019). Other popular measures include trait ratings of party supporters and measures of social distance, such as willingness to socially interact or marry individuals from a rival political group (Garrett et al., 2014; Iyengar et al., 2012). Beyond these evaluative measures, scholars have also investigated various behavioral manifestations of partisan prejudice in economic games, such as partisans' tendency to financially withhold or reward other people based on whether they share the same partisan label (Carlin & Love, 2013; Iyengar & Westwood, 2015).

In line with this last set of measures, in this study, I operationalize affective polarization as the tendency for users to engage positively with likeminded peers and negatively with ideological opponents. This is consistent with other studies in this line of work and theories of intergroup behavior that see intergroup prejudice as characterized by both ingroup favoritism and outgroup hostility (Brewer, 1999). My primary research task is therefore to determine the affective nature of interactions between crosscutting and likeminded users. Averaging the sentiment scores of both original and response comments in each dyad indicates that likeminded interactions are slightly less negative (n = 121,939, M = -0.057, SD = 0.39) than crosscutting ones (n = 54,400, M = -0.060, SD = 0.39).

This study also seeks to uncover what communicative factors play in driving or mitigating affective bias. Here, I am particularly interested in how the tone and direction of the last reply in a conversation dyad impacts future interactions. I distinguish between seven future interaction types, depending on (1) the ideological

affinity between past and future commenter (likeminded or crosscutting), (2) whether a future comment has been classified negative, positive or neutral, including the possibility that an interaction is not followed by another one. A transition to a different interaction is recorded when an initial comment-response dyad elicits another reply, forming another dyad between either the same or a different pair of users, and so on until the exchange ends.

4.6.2 Predictors

To identify the presence of ingroup or outgroup political entities in the comment corpus, I used spaCy's named entity recognition feature with a custom dictionary of common nouns and abbreviations closely associated with the words 'liberal' and 'conservative', such as 'libs', 'dems' and 'cons' (see *Appendix B* for complete list). Users may, of course, refer to liberal or conservative ideologies in comments without explicitly mentioning these groups by name (through discussion of specific policies or political leaders for example). Given that the presence of explicit social group cues accentuates identity salience, however, in the analysis I opt to focus primarily on these identifiers. When a comment references multiple political entities, it is impossible to properly determine the direction in which the sentiment of the comment is expressed—the presence of a liberal (n = 12,582) or conservative entity (n = 14,678) was thus only recorded when a reply strictly mentioned one or the other. Following from this, I measure the presence of an ingroup or outgroup political entity in a response by matching the ideology of the sender and the political entity mentioned. The tone of the last reply in a conversation dyad, finally, is determined based on the sentiment score previously allocated to each comment (negative or positive).

4.6.3 Controls

Several social factors might influence the tone and nature of users' responses. Here, I first control for *a response's popularity score*, or the number of upvotes minus the number of downvotes a comment has received at the time of data collection, as highest scoring comments have a higher chance of ranking at the top of a discussion thread. Finally, given that the presence of other ingroup and outgroup members

can influence individual behavior in intergroup settings (Coe et al., 2014), I also control for the proportion of *liberal or conservative-leaning users* in a discussion thread.

4.7 Models

There are several sources of dependencies in the data: interactions are nested in pairs of users, in that two users can have more than one interaction in the same discussion. Pairs of users are then also separately nested in individual discussions, which form part of larger threads. To determine what proportion of the variance in the sentiment of user interactions can be explained by these levels, I first calculate intra-class correlations. Nesting within thread only accounted for 14% of the variance in message sentiment (ICC = 0.14) while nesting within discussion and pair of users explained between 41.7% and 55.1% of the variance in the outcome variable (ICC = 0.42 and ICC = 0.55, respectively). To account for these dependencies in the data, I thus adopt a series of multilevel models to address my hypotheses.

The first part of the analysis addresses the question of whether users of different ideological camps display affective biases in their interactions. To answer this, Model 1 examines the effect of ideological affinity between two users on the sentiment of their interactions, including the identification of the user pair, the discussion in which the interaction took place and the thread as random effects. Estimates of statistical significance were computed using the Satterthwaite's degrees of freedom method (Kuznetsova et al., 2017). To calculate marginal and condition R^2, I follow the methods proposed by Nakagawa and Schielzeth (2013).

The second part of the analysis is concerned with how the tone and target of the last comment in a conversation dyad impact the follow-up interaction, including the possibility that a conversation ends. To explore the mechanisms outlined in Hypotheses 2a to 4, I run a series of separate multilevel logit models, which allows me to specify the model equation for different binomial contrasts of interest in the outcome variable. Model 2a and Model 3 evaluates the odds of future negative engagement between ideologically opposed users compared to any other types of interactions. Model 2b contrasts the likelihood that a crosscutting interaction ends or is followed by another crosscutting interaction of any type. Model 4, finally, is concerned with the probability of positive future interactions

between likeminded users compared to any other types of interactions. There are some limitations associated with this model choice. Notably, individualized models can lead to efficiency loss (Agresti, 2002). However, in each case this is mitigated by a large sample size in the reference category.

4.8 Results

Hypothesis 1 predicted that interactions between crosscutting users would be more negative than those between likeminded users. Results from this model (Model 1) are summarized in *Table 4.2*. Model 1 clearly shows that ideologically opposed users tend to engage in significantly more negative interactions than likeminded ones, though the effect size is modest (β = -0.02, $p < 0.01$). Crosscutting interactions have, on average, a 0.021 lower sentiment score than likeminded ones. Hypothesis 1 is thus supported.

Table 4.2. Effects of interaction type on average sentiment

	Avg. Sentiment of Interaction		
	Model 1		
Random effects	Variance	Std. Dev	
User pair	0.47	0.68	
Discussion	0.01	0.12	
Thread	0.09	0.29	
Residual	0.44	0.66	
Fixed effects	Estimate	Std. Error	t
Intercept	0.05***	0.007	6.56
Crosscutting (Yes = 1)	-0.02**	0.007	-3.15
Marginal R^2		0.001	
Conditional R^2		0.56	

Note: *** p <0.001, **p<0.01, * p<0.05. N = 176,339.

Before addressing the second part of the analysis, I run an initial model with the same sets of random effects to assess the baseline likelihood of transitioning from crosscutting to likeminded interactions (see Model 0 in *Appendix B*). This indicates

that users rarely transition between these two states: interactions between opposed users are significantly more likely to be followed by another crosscutting exchange than a likeminded one (OR = 61.05, p <0.001).

Moving on to the hypotheses of interest, Hypotheses 2a and 2b suggested that negative crosscutting interactions that referenced the sender's ingroup (for example, if a liberal user made a negative comment about conservatives as a response to something a conservative user wrote) would be more likely to prompt another negative crosscutting exchange (H2a) or to halt discussion altogether (H2b). Model 2a and 2b in *Table 4.3* address this set of hypotheses.

Table 4.3. Effect of interactions features on likelihood of follow-up crosscutting negative response or conversation end

	'Crosscutting Negative'		*'End'*	
	Model 2a		Model 2b	
Fixed effects	Odds Ratio	95% CI	Odds Ratio	95% CI
Intercept	0.27***	0.25-0.28	0.70***	0.67-0.73
Crosscutting negative (1 = Yes)	0.12***	0.11-0.12	0.92***	0.89-0.96
Refers sender's ingroup (1 = Yes)	1.16**	1.06-1.27	0.69***	0.63-0.74
Comment score	1.08***	1.05-1.11	0.13***	0.12-0.15
Proportion of ingroup members in thread	40.43***	27.10-60.33	0.08***	0.07-0.09
Proportion of outgroup members in thread	0.83***	0.79-0.86	1.12***	1.08-1.15
Crosscutting negative * Refers sender's ingroup	0.93	0.70-1.23	1.37***	1.18-1.59
Marginal R^2	0.16		0.25	
Conditional R^2	0.95		0.83	
AIC	70323.7		100758.0	

Note: *** p <0.001, **p<0.01, * p<0.05. N = 176,339 for Model2a and N = 98,886 for Model 2b. All models summarize fixed effects.

The results in *Table 4.3* show that negative interactions between opposed users are slightly less likely to be followed up by a negative crosscutting response (OR = 0.12, p <0.001). Examining the interaction term, negative exchanges in which the respondent mentions another user's political identity are not significantly associated with a higher chance of negative response. Hypothesis 2a is therefore not supported. Interestingly, in this scenario, large proportions of political ingroupers in the thread also significantly drive the odds of a negative crosscutting response, and so to a high degree (OR = 40.43, p <0.001) while the negative effect of the presence of outgroup members is marginal (OR = 0.83, p <0.001). In the same type of interactions, a high score comment is also slightly more conducive to a negative follow-up (OR = 1.08, p <0.001).

Turning now to Model 2b, there is clear evidence that the mention of a user's political identity moderates the relationship between previous and future crosscutting interaction. Negative crosscutting interactions in which the reply negatively references the sender's ingroup are indeed significantly more likely to end, than to be followed up by any type of future crosscutting exchange (OR = 1.37, p <0.001). Hypothesis 2b is thus confirmed. Among control variables, the effect of high proportions of ingroup and outgroup members in the thread are both significant, though their effect is small (OR = 0.08 and OR = 1.12 respectively).

The third hypothesis dealt with the possibility that positive intergroup contact would enhance the likelihood of further positive engagement between ideologically opposed users. As is evident from Model 3 in *Table 4.4*, interactions between different users in which the respondent positively references the sender's ingroup (for example, a conservative expressing positive sentiment towards liberals in an exchange with a liberal user) are significantly more likely to be followed up by another positive exchange than any other types of interactions (OR = 2.06, p <0.001). This confirms Hypothesis 3. Among control variables, the presence of fellow ingroup members is also a significant driver of crosscutting positive engagement (OR = 53.5, p <0.001). In contrast, high proportions of outgroup members in the thread only have a small impact on the outcome variable (OR = 1.07, p <0.05).

Table 4.4. Effect of interaction features on the likelihood of follow-up crosscutting positive response

	'Crosscutting Positive'	
	Model 3	
Fixed effects	Estimate	95% CI
Intercept	0.05***	0.05-0.06
Crosscutting positive (1 = Yes)	0.19***	0.15-0.23
Refers sender's ingroup (1 = Yes)	0.99	0.86-1.13
Comment score	0.97	0.92-1.02
Proportion of ingroup members in thread	53.55***	27.71-103.50
Proportion of outgroup members in thread	1.07*	1.01-1.13
Crosscutting positive * Refers sender's ingroup	2.06*	1.13-3.75
Marginal R^2	0.11	
Conditional R^2	0.96	
AIC	34119.8	

Note: *** p <0.001, **p<0.01, * p<0.05. N = 176,339 for Model 3. Model 3 summarizes fixed effects.

Having reviewed the drivers of intergroup affective bias, I now turn to intragroup dynamics. Hypothesis 4 made predictions about the effect of positive references to the ingroup in likeminded exchanges. Model 4 in *Table 4.5* shows that, controlling for other factors, positive interactions between similar users are actually less likely to be followed up by a positive likeminded response (OR = 0.23, $p < 0.001$) than other interaction types. Furthermore, references to the political ingroup do not moderate that relationship. Hypothesis 4 is therefore not supported.

Table 4.5. Effect of interaction features on likelihood of likeminded positive response

	'Likeminded Positive'	
	Model 4	
Fixed effects	Estimate	95% CI
Intercept	0.05***	0.05-0.06
Likeminded positive (1 = Yes)	0.23***	0.20-0.27
Refers sender's ingroup (1 = Yes)	1.01	0.90-1.13
Comment score	0.99***	0.96-1.03
Proportion of ingroup members in thread	1.06**	1.02-1.09
Proportion of outgroup members in thread	8.24***	6.07-11.19
Likeminded positive * Refers sender's ingroup	1.29	0.75-2.21
Marginal R^2	0.12	
Conditional R^2	0.72	
AIC	61176.4	

Note: *** p <0.001, **p<0.01, * p<0.05. N = 176,339 for Model 4. Model 4 summarizes fixed effects.

4.9 Discussion

The present study set out to determine whether discussions between liberal and conservative users on Reddit's *r/politics* were affectively polarized, and what intergroup factors shaped these communicative dynamics. Drawing on theories of intergroup relations and leveraging computational methods to infer the political leanings of forum members, I find that interactions between ideologically-opposed users are indeed polarized along affective lines. In other words, Reddit users express more negative sentiment when interacting with people they disagree with than with likeminded peers. This finding joins a nascent body of studies that identify similar patterns on other platforms (Yarchi et al., 2020), suggesting the need to expand current definitions of online polarization to include affective dimensions.

The analyses presented here also indicate that the contents of these conversations and the group context within which they take place are significant determinants of affective polarization. Confirming initial expectations about 'generalized avoidance', negativity aimed at the opposition in crosscutting exchanges discourages further interaction of the same kind. This held true even after controlling for comment popularity and proportion of ingroup members in the conversation, underscoring that, even in the face of threats, retaliation is often a less preferable option than opting out from a chat. This outcome may also reflect the fact that most individuals are quite adverse to uncivil and excessively partisan behavior (Klar, 2018; Shafranek, 2020), and generally prefer cooperating with their political ingroup than engaging in overt intergroup conflict (Halevy et al., 2012).

Conversely, I find that positive crosscutting engagement has a *depolarizing* effect: when users made positive references to the opposition, they were more likely to engage in further positive crosscutting interactions. This finding enriches current research on the effects of online intergroup contact by showing that positive encounters with ideologically different 'others' not only improve individuals' feelings towards that outgroup (N. Kim & Wojcieszak, 2018), but also their disposition towards them. In line with findings from Wojcieszak and Warner (2020), Huddy and Yair (2021), and Skytte (2020), it also provides support for warm group relations and positive contact as effective strategies for reducing partisan prejudice in online discussions.

Finally, and contrary to initial expectations, ingroup praise did not have a significant impact on affective polarization among likeminded users. One possible explanation for this has to do with the fact that group norms tend to vary with situational factors: the extent to which groups will engage in discriminatory behavior, for example, depends heavily on group members' perception of threat from rivals (Hogg et al., 1984). In situations when the power differential between dominant and minority groups is strong, as is the case on *r/politics*—a predominantly liberal enclave—the need for positive distinctiveness is reduced (Rains et al., 2017). Ingroup praise might have therefore not carried the same normative weight in this particular context as it would in other situations.

This study presents several limitations that must be acknowledged. However, each presents clear opportunities for future research. Given the disproportionately high number of liberal users in my sample, the opportunity for crosscutting interactions in these discussions is inherently restricted. Moreover, as I highlight above, the power imbalance between liberals and conservative users may have affected the way in which the two groups interacted with one another. It is entirely possible, for example, that the minority of conservative users engaged on the forum would behave differently in a space where they had a dominant presence. It will therefore be interesting to follow up this study with analyses of other subreddits that are either more ideologically segregated or where the power dynamic is reversed.

Second, the present analysis only focused on a restricted number of communicative and group factors, namely the presence of other group members, tone, and salience of political identities in exchanges between users. The strategy proposed for the coding of political entities in particular might have excluded comments that involved clear intergroup dynamics without making explicit references to 'liberals' or 'conservatives.'[23] Likewise, the reduction of the range of message sentiment scores to a binary distinction might have accentuated the effects observed. Future studies could thus usefully expand on these findings by analyzing the effects of more granular communicative features and constructs on affective

[23] As one reviewer noted, only about 27,000 comments or 25% of comments were identified as mentioning a liberal or conservative entity, meaning that great majority of comments did not have an ingroup or outgroup label.

polarization, such as discussion of policy issues or the presence of incivility, name-calling or dismissal.

A third limitation is the potential loss of relevant data through community moderation. Through Reddit is characterized by a strong free speech ethos, in practice subreddit moderators are empowered to remove or delete messages that violate their community guidelines if they contain personal attacks, hate speech or flaming. While only a small proportion of comments had been deleted by moderator bots at the time of data collection, some of these omissions might have constituted the most extreme displays of negative sentiment on the forum.

Finally, it will be important for future studies to test whether the findings presented here hold with alternative sentiment analysis techniques, including entity-level sentiment analysis and supervised approaches. Previous work in this area has mostly shown group identification to take place primarily with textual cues and language use. However, many social media platforms are also rich in visual cues that denote one's political affiliation and views. In such environments, it could thus be interesting to explore how political cues contained in users' real names (including emojis), self-descriptions and even profile pictures affect the dynamics outlined above.

4.10 Conclusion

Affective polarization is seldom researched in the digital context. In this paper, I advanced an intergroup perspective to demonstrate that online political discussions are polarized along affective lines. Taken together, my findings show that whether and how users engage with one each other online is informed by complex communicative factors— including the tone and direction of their messages. While the literature on online polarization is rich, the vast majority of studies still operationalize polarization as a *state*—namely, the presence or absence of homogenous clusters of users with little connectivity between them—and provide largely descriptive snapshots of its existence at a given point in time. In contrast, the effects presented above show that integrating emotions in the study of intergroup communicative dynamics can tell a more revealing story about the nature of online polarization than simply looking at how frequently users interact with each other.

4.11 Appendix B

Subreddit classification task

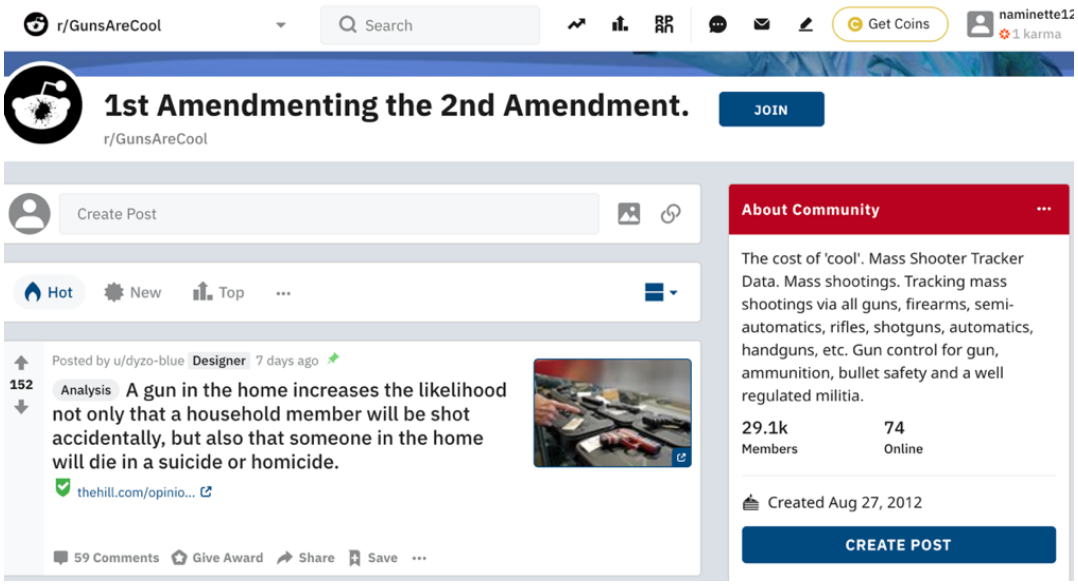
The coding of the subreddits was performed by two coders, including the author. Each coder was assigned all 258 subreddits to classify. Each coder was instructed to read a subreddit's title, description, as well as some of its members' most recent posts before assigning it to one of three categories: 'conservative', 'liberal' or 'not applicable' (N/A).

Subreddits that explicitly referenced Republican party membership or adherence to conservative ideology and clearly espoused conservative or right-wing social and economic views—including support for the cultural and social status quo, anti-immigration attitudes (Gil de Zúñiga et al., 2012), support for guns (Wozniak, 2017), and concern for mainstream media bias (Domke et al., 1999)—were categorized as 'conservative.' This code was also applied to subreddits that expressed favorable views toward economic elites and free market capitalism (Haidt et al., 2009). Conversely, subreddits that explicitly stated their support for the Democratic party, promoted liberal ideology, and espoused progressive political views such as support of LGBTQ views (Herek et al., 2010), promotion of social change, egalitarian ideals and redistributive justice (Lakoff, 2002), were categorized as 'liberal.' Subreddits where no clear ideological hues were discernible were coded as 'neutral.' Both coders reached a high inter-rater reliability (Krippendorf's $\alpha = 0.772$) during the test coding round (N=20) and any disagreements were resolved by the author.

Table B.1. Coding scheme for the classification of subreddits

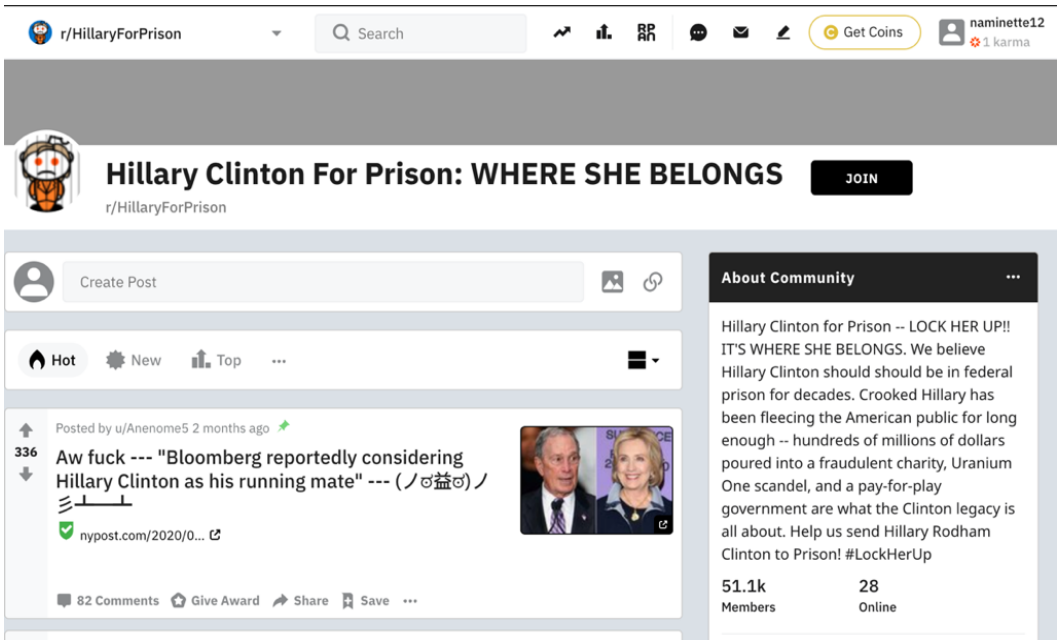
	Content Codes		
	Conservative	Liberal	Not Applicable
Figure-based	Subreddits that espouse Republican and conservative political figures or explicitly target and disparage liberal ones, as indicated by the subreddit's title, description or community guidelines.	Subreddits that espouse Democrat and liberal political figures or explicitly target and disparage conservative ones, as indicated by the subreddit's title, description or community guidelines.	Subreddits that do not fall into either the 'conservative' or 'liberal' category.
Ideology-based	Subreddits that espouse conservative social and economic views, as stated either in the subreddit's title, description or community guidelines. This includes policies that uphold individual liberties, support the social/cultural status quo, and embraces the moral values of authority/respect, ingroup loyalty and hierarchy. The label also applies to subreddits that promote religious traditions, authorities and hold favourable attitudes toward the free market, the capitalist system in general, the sacredness of human life, individual character and strong family values.	Subreddits that espouse progressive social and economic views, as stated either in the subreddit's title, description or community guidelines. This includes progressive policies that promote social change egalitarian ideals, economic redistribution and social justice, and embrace the moral values of care and fairness. The label also applies to subreddits that challenge the supremacy of structure they regard as sources of oppression, such as capitalism and the patriarchy.	

Figure B.1. Example of 'ideology-based' liberal leaning subreddit



Note: Screenshot of a subreddit advocating for stricter gun control. *Source:* Reddit. (May 11, 2019) *GunsAreCool Subreddit* [Screenshot]. Retrieved from: https://www.reddit.com/r/GunsAreCool/

Figure B.2. Example of 'figure-based' conservative leaning subreddit



Note: Screenshot of an anti-Hillary Clinton subreddit. *Source:* Reddit. (May 12, 2019) *HillaryForPrison Subreddit* [Screenshot]. Retrieved from: https://www.reddit.com/r/HillaryForPrison/

Bayesian estimation of ideological leaning

To validate estimations from the Bayesian framework outlined in the paper, a random sample of 100 comments posted by authors who had been classified as either 'liberal-leaning' or 'conservative-leaning' through this method was manually coded following the same coding scheme used to classify subreddits. The confusion matrix is presented in *Table B.2*.

Table B.2. Confusion matrix for validation of estimated ideological leanings

Bayesian Estimation	Conservative	Liberal
Conservative	30	16
Liberal	6	47

Note: Accuracy : 0.778, Sensitivity : 0.833, Specificity : 0.746, Pos Pred Value : 0.652, Neg Pred Value : 0.887

Table B.3. Complete list of coded subreddits

Liberal Subreddits	*Conservative Subreddits*
accidentallycommunist;Activism;AgainstHateSubreddits;AllTheLeft;anarcha;AnarchaFeminism;Anarcho_Communist;anarcho_socialism;anarcho_syndicalism;anarchocommunism;ani_communism;antifascism;antifascistsofreddit;AntiTrumpAlliance;AOC;BannedFromThe_Donald;SandersForPresident;beto2020;bidenbro;bluemidterm2018;breadtube;BullMooseParty;centerleftpolitics;Chapo_Trap_House;chapotraphouse2;chomsky;Communalists;Communism;communism101;Communist;CornbreadLiberals;dankleft;debateacommunist;debatecommunism;Delete_The_Donald;democrat;demsocialists;Democrats;Demsocialist;Drumpf;Egalitarian;Egalitarianism;elizabethwarren;enlightendedcentrism;esist;evolutionReddit;Feminism;Feminisms;FirstLook;fragilewhiteredditor;fuckthealtright;FULLCOMMUNISM;GamerGhazi;GenderEgaliatarian;GenderCritical;gendercynical;Green;GreenParty;GunsAreCool;HillaryForAmerica;HillaryClinton;impeach_trump;IWW;ironfrontusa;JillStein;Justice Democrats;keep_track;Kossacks_for_Sanders;Labor;Labour;LateStageCapitalism;Leftcommunism;LeftWithoutEdge;Leninism;LGBT;Liberal;liberalgunowners;LibertarianLeft;LibertarianSocialism;MarchAgainstTrump;MarchForScience;MartinO'Malley;Marxism;MocktheAltRight;moretankiechapo;NaziHunting;ndp;neoprogs;neoliberal;NeverTrump;newpatriotism;Obama;OccupyWallStreet;onguardforthee;OurPresident;Political_Revolution;pragerurine;PrisonReform;ProChoice;Progressive;PutinsBitch;RadicalFeminism;russialago;selfawarewolves;shitthe_donaldSays;SocialDemocracy;socialanarchism;SocialEcology;socialism;Socialism_101;socialistra;socialjustice101;StrikeAction;stupidpol;taxmarch;thanksobama;The_Mueller;TinyhandsInc;TinyTrumps;toiletpaperusa;Trump_Watch;TrumpCriticizesTrump;trumpgret;Union;voteblue;WayOfTheBern	AltRightChristian; Anarcho Capitalism;AskThe_Donald;AsktRP;BannedFromThe_Donald;BenCarson;BobbyJindal;CarlyFiorina;Censorship;ChrisChristie;Christian_Ancaps;Classical_Liberals;Conservative;Conservatives;Conservatives Only;DNCLeaks;Enough_Sanders_Spam;EnoughObamaSpam;EnoughPaulSpam;FlushTheTPP;Hillary ForPrison;JohnKasich;KotakuInAction;MarcoRubio;MarriedRedPill;MartinOMalley;MensRights;Monarchism;neoliberal;New_Right;Objectivism;Paleoconservative;presidentbannon;PresidentKushner;ProGun;RandPaul;Republican;Republicanism;Republicans;RickSantorum;Romney;SargonofAkkad;ShitLiberalsSay;Sorosforprison;TeaParty;TedCruz;the_donald;TheNewRight;TheRedPill;TheRightBoycott;trump;Tuesday;TrueObjectivism;White Rights;Wikileaks

Sentiment analysis and entity recognition in the text corpus

Table B.4. Complete list of political entities

Liberal Entities	*Conservative Entities*
communists,Communists,communist,Communist,commies,Commies,democratic_party,Democratic_party,democrats,Democrats,democrat,Democrat,dem,Dem,dems,Dems,the_Left,the_left,left_wing,Left_winger,left_winger,Left_wingers,leftist,Leftist,leftists,Leftists,lefties,Lefties,lefty,Lefty,liberal,Liberal,liberals,Liberals,libruls,Libruls,lib,Lib,libs,Libs,libtard,Libtard,progressive,Progressive,progressives,Progressives,socialist,Socialist,socialists,Socialists	Cons,cons,Conservatives,conservatives,Conservative,conservative,conservatard,Conservatard,far_right,gop,GOP,Gop,neocon,Neocon,neocons,Neocons,republican,Republican,republicans,Republicans,the_right,the_Right,right_wing,Right_wing,right_winger,right_wingers,,alt_right,Alt_Right,alt_lite,new_right,New_Right,trumpster,Trumpster,trumpsters,Trumpsters,tory,Tory,tories,Tories,trad,Trad,trads

Table B.5. Performance of VADER sentiment analysis

Manual Coding								
Negative			Neutral			Positive		
Pr.	Re.	F1	Pr.	Re.	F1	Pr.	Re.	F1
0.736	0.609	0.666	0.766	0.756	0.761	0.535	0.703	0.608

Note: Overall Accuracy: 0.704 [0.648, 0.755], Kappa = 0.517, Pr. = Precision, Re. = Recall

Table B.6. Descriptive statistics for comment sentiment

Mean	-0.030
Std. Deviation	0.508
Min	-1.000
25%	-0.440
50%	0.000
75%	0.366
Max	0.999

Diagnostics

In additional to the intra-class correlation analysis detailed in the main manuscript, the following diagnostic checks were performed. Multilevel regression models assume that the residuals of the analysis are normally distributed. The normality of residuals was checked and the plots indicated no significant deviation from normality. Secondly, the linearity of continuous covariates was also inspected through visualizations. These did not suggest any issue, except in the case of comment scores and proportions of in and outgroup members, which exhibited strong skews. This was addressed through a cube and two log transformations, respectively. No issues of multicollinearity were detected after computing variance inflation factors.

Table B.7. Likelihood of a crosscutting interaction to be followed up by another crosscutting one

	'Future Crosscutting'	
	Model 0	
Fixed effects	Odds Ratio	95% CI
Intercept	0.02***	0.02-0.02
Crosscutting (Yes = 1)	61.06***	57.79-64.51
Comment score	1.02*	1.00-1.04
Proportion of ingroup members in thread	1.01***	1.22-1.38
Proportion of outgroup members in thread	1.30	1.00-1.03
Marginal R^2	0.50	
Conditional R^2	0.55	

Note: *** p <0.001, **p<0.01, * p<0.05. N = 176,339.

5 ‘Who Do You Think You Are?’ The Differential Effect of In-Party Threats and Reassurances on Affective Polarization

5.1 Abstract

Evidence is mounting that several aspects of our online communication landscape have the potential to drive animosity between partisan groups. A common explanation in the literature is that these effects are driven by the power of digital media and partisan incivility to activate individual’s partisan identities, thus accentuating an ‘us’ versus ‘them’ mentality. Yet, it has so far failed to specify what aspects of partisan communication are the most likely to drive prejudice. In this paper, I tackle this question by investigating how exposure to online partisan criticism and partisan praise in online comments differentially impacts affective polarization in an experimental setting. Drawing on intergroup threat theory, I demonstrate that not all forms of partisan comments drive affective polarization: while exposure to comments threatening someone’s in-party’s status significantly increases negative evaluations of political opponents, exposure to comments derogating the out-party and praising the in-party mostly do not. Furthermore, findings show that Democrats are more likely to react adversely to viewing comments praising Republicans than the reverse. Implications of these findings for the study of affective polarization are discussed.

5.2 Introduction

Rising inter-partisan hostility and distrust is a growing source of concern for political scientists (Iyengar et al., 2019). Over the last three decades, partisans across Western democracies have become increasingly prejudiced against one another (Hobolt et al., 2020; Iyengar & Krupenkin, 2018; Martini & Torcal, 2019; Murray et al., 2017), with negative consequences for interpersonal trust and social

cohesion (Carlin & Love, 2018; Klar, 2018). In comparative terms, nowhere has this spike in partisan animosity been more pronounced than in the US (Boxell et al., 2020; Gidron et al., 2020). Over the last two decades, the share of US partisans expressing negative views of the out-party has more than doubled, while Democrats and Republicans' feelings towards their own party have remained fairly stable (West & Iyengar, 2020).

Evidence is mounting that certain aspects of our online communication environment have the potential to exacerbate this phenomenon at the individual level. Research shows, for example, that our digital news and social media ecosystems are rife with partisan cues that not only activate users' sense of partisanship, but can also promote negative attitudes towards the political opposition (Settle, 2018a; Suhay et al., 2018). A bourgeoning literature also investigates the effects of partisan incivility on affective polarization—though it has thus far produced mixed results (Druckman et al., 2019; Gervais, 2015, 2019; Suhay et al., 2018). While existing studies have predominantly focused on the more *negative* aspect of online political talk by examining how partisan 'criticism' or 'incivility' impacts interparty hostility, in comparison other dimensions of online political talk have received far less attention. Notably, it is unclear whether other forms of political expression, such as partisan praise, have the same effects on affective polarization and whether these effects differ when a message is directed towards the in-party or the out-party. Yet, identifying which aspects of partisan communication have the strongest impact on animosity is important as it informs strategies to mitigate polarization.

In this paper, I address this gap by exploring how exposure to partisan criticism and partisan praise in online comments differentially impacts affective polarization. My main research question is thus: What types of online partisan comments drive affective polarization? Drawing on intergroup threat theory and using data from an online survey experiment conducted in December 2020 on a sample of 499 US partisans, I demonstrate that exposure to laudatory comments about the opposition is as strong a driver of affective polarization among partisans as exposure to attacks on the in-party. While this relationship is not moderated by the strength of one's partisan identity, Republicans and Democrats exhibit different sensibilities to partisan praise and partisan criticism: the latter were more likely to

react negatively to positive statements about the opposition. This study enriches existing scholarship on the determinants of affective polarization by positing distinct mechanisms for exposure to in-party status threats, both in the form of ingroup attacks and outgroup praise, and in-party reassurances in the form in-party praise and outgroup derogation. These findings also underscore that studying the effects of partisan communication on affective polarization requires looking beyond partisan incivility, and paying greater attention to both a message's rhetorical function and its target.

The paper is structured as follows. In the first section, I provide an overview of existing literature on the effects of online political talk on affective polarization, and highlight its limitations. Following this contextualization, I draw on findings from integrated threat and intergroup emotions theories to derive specific hypotheses about the factors that may shape how people react to different types of partisan comments. The third section outlines the survey experiment's design, explaining how participants were recruited, how the experimental treatments were built, and how key variables of interest are operationalized. I conclude the chapter by reporting the study's main findings and discussing their implications for the study of digital media and their effects on political attitudes.

5.3 Online Political Talk and Affective Polarization

Many scholars converge in their assessment that American society is polarized, both socially and affectively. Since the 1980s, rank-and-file partisans have expressed growing animosity and ill-will towards their political opponents while harboring mostly positive feelings towards fellow partisans (Iyengar et al., 2019; Iyengar & Krupenkin, 2018). Data from longitudinal surveys shows that partisans are increasingly keen to distance themselves socially from their political rivals and have become less and less trusting of them (Huber & Malhotra, 2017; Westwood et al., 2018). So profound are these divisions that, by some measures, partisan prejudice in America today exceeds comparable bias based on racial differences (Iyengar & Westwood, 2015). [24]

[24] Though recent scholarship suggests that race and partisanship are deeply intertwined in contemporary American politics, in such a way that shifts in racial affect are mirrored by shifts in partisan affect and vice-versa (Westwood & Peterson, 2020).

This intensification of affective partisan polarization has worrying implications. Although the literature exploring its effects in explicitly political contexts is still nascent, existing findings are already cause for concern. [25] Negative partisan affect has been shown to undermine trust in the opposition (M. J. Hetherington & Rudolph, 2015), reinforce racial hostility (Westwood & Peterson, 2020), and bias adherence to public health guidelines during the COVID-19 pandemic (Druckman et al., 2020). Beyond these political outcomes, scholars have also documented several social consequences of affective polarization. Partisanship has been linked to discriminatory practices in domains as diverse as hiring (Gift & Gift, 2015), health care provisions (Hersh & Goldenberg, 2016), and economic exchanges (McConnell et al., 2018). Bitter partisan relations also 'permeate[s] judgements about interpersonal relations' (Iyengar et al., 2012), with partisans being less likely to date out-partisans (Huber & Malhotra, 2017) and to engage in cross-party dialogue with family members (M. K. Chen & Rohla, 2018).

Recently, communication scholars have begun investigating how various features of our digital media environment could be exacerbating affective polarization. The dominant perspective in this literature is that internet and social media reinforce out-party animosity by driving exposure to political content that activates users' political allegiances and reinforces negative judgments of the political opposition. Work in this perspective mostly draws on social identity and social categorization theory (Tajfel, 1981) to explain this effects. In this view, partisanship is an influential, internalized social identity (Iyengar et al., 2012; West & Iyengar, 2020) that shapes people's attitudes and behavior towards others. When partisan identity salience is high, such as during elections, individuals are more likely to see the world as a battle between 'us' versus 'them' (Campbell et al., 1960; Greene, 2004; Huddy & Bankert, 2017) and discriminate towards out-partisans (Gaertner et al., 1993), even when they do not feel strongly attached to a political party (West & Iyengar, 2020).

Prior studies have shown that people frequently come into contact with political content online in the form of news articles, comments or posts that can

[25] Druckman et al. (2020) and Iyengar et al. (2019) highlight this lacuna, the latter noting that 'little has been written on this topic, as most studies have focused on the more surprising apolitical ramifications' (p. 139).

drive partisan identity salience (Bode, 2016; Fletcher & Nielsen, 2018; Settle, 2018a). It is well-established that exposure to partisan media, for example, strengthens individuals' partisan identity, which can in turn reinforce negative attitudes towards the opposition (Horwitz & Nir, 2015; Levendusky, 2013). The more partisan cues there are in the informational environment, the more individuals' group membership influences how they interpret and respond to political stimuli (Carr, 2017; Diamond, 2020; Hinck & Carr, 2020). Partisan outlets also routinely denigrate the out-party, using abrasive language and ad-hominem attacks, all of which cultivates hostility and negative affect towards the 'other side' (Hasell & Weeks, 2016; Jamieson & Capella, 2010; Mutz, 2015; Sobieraj & Berry, 2011). Beyond partisan news, however, other scholars argue that even the most trivial, non-political content posted on social media is rich in cues that can signal political affiliation, and that identifying other users as potential 'rivals' drives negative judgments towards them. In a series of survey-based studies and experiments, Settle (2018) demonstrates that a substantial share of political content on the Facebook News Feed is informative enough for users to (rightly or wrongly) infer the political views of those who posted it. Once people start engaging in this process of social categorization, by distinguishing others as either one of 'us' or one of 'them', they express more critical views towards Facebook users they regard as political opponents, and tend to overestimate differences between themselves and the outgroup.

5.4 Partisan Prejudice and Threat Perceptions

This partisan identity salience mechanism is an intuitive explanation of social media's effects on affective polarization. Yet, it falls short of specifying the conditions under which partisan identity priming leads (or not) to biased judgements. Indeed, implicit in most of the literature is the idea that 'mere' categorization (i.e., simply calling attention to group boundaries) is a necessary and *sufficient* condition for individuals to engage in ingroup favoritism and outgroup discrimination. If this were true then being exposed to messages that contain explicit partisan cues should be enough to trigger one's partisan identity and generate out-party animus. However, experimental work on this question has produced contradictory findings. While some studies show that exposure to

messages that derogate out-partisans drives affective polarization (Sood & Iyengar, 2017; Suhay et al., 2018), for instance, others point in the opposite direction. Recent studies show that exposure to hostile attacks on the out-party from likeminded sources has limited effects on partisan affect, and might even improve favorability towards the out-party (Druckman et al., 2019; Gervais, 2016; Rains et al., 2017). One common explanation for this pattern is that incivility towards the 'other side' is not offensive, and thus does not generate defensive anger (Gervais, 2016). Instead, it reassures partisans about the superiority of their beliefs, which boosts positive ingroup feelings and can drive tolerance and even empathy for outgroups (Vanbeselaere et al., 2006). These findings cast doubt on the idea that partisan identity salience alone is responsible for these polarizing effects, and suggests instead some degree of specificity to the ways in which partisan communication impacts partisan prejudice.

In this regard, it is useful to go back to social identity theory (SIT), and particularly to the motivational processes that underpin intergroup behavior. A central, though often overlooked, tenet of SIT is that since individuals' social identities are such an important part of their self-concept, in order to maintain a positive self-image they will generally strive to achieve positive distinctiveness through favorable comparisons with other groups (Abrams & Hogg, 1988; Brewer, 2001). While favoring one's ingroup is considered normatively appropriate (Blanz et al., 1997; Platow et al., 1995), inflicting harm on others simply because they are outgroup members is costly and less justifiable (Blanz et al., 1997)—unless one's group status, beliefs or image is perceived to be threatened (Brewer, 1999). This *positive-negative asymmetry* has been repeatedly observed in studies of intergroup relations: absent explicit threats, individuals are more inclined to help the ingroup than derogate the outgroup (Buhl, 1999; Mummendey et al., 2000). In other words, outgroup animosity is primarily driven by a desire to protect the status of the ingroup ('ingroup love') from threats than an explicit intention to harm the outgroup ('outgroup hate').

Threats themselves come in different flavors. Research from the intergroup threat literature shows that any action on the part of the outgroup that can be interpreted as jeopardizing the ingroup's status, beliefs and collective self-esteem tends to generate hostility and a desire to avoid contact with outgroup members

(Branscombe et al., 2002; Steele et al., 2002; Stephan & Stephan, 2000; Tajfel & Turner, 1979). Confirming this, communication scholars have shown that disagreeable incivility—especially directed insults, ad-hominem attacks, name-calling or any other intentional expressions of disdain or contempt for the group and its members (Muddiman, 2017)—triggers feelings of anger and anxiety in those exposed to them, which motivates defensive hostility and a desire to attack the 'other side' (Gervais, 2015, 2016, 2019; Z. K. Rothschild et al., 2021; van Zomeren et al., 2007). Likewise, other studies have found that individuals exposed to attacks on the in-party were also less keen to socially interact with political rivals (Suhay et al., 2018). This leads me to formulate the following hypothesis:

H1: Exposure to ingroup criticism increases affective polarization.

Beyond hostile attacks, however, perceptions of competition between rival groups have also been shown to accentuate negative outgroup attitudes (Riek et al., 2006). When two groups are in competition for political power or other resources such as money for example, the potential success of one group with respect to influence or prestige directly threatens the chances of the other group, resulting in growing resentment and outgroup hostility (Bobo, 1988). Beliefs that the outgroup is being rewarded for pushing goals that are seen as harmful to the ingroup drive similar effects (Bobo, 1988). Following this logic, in the context of digital political communication, comments that could potentially be quite aggravating for partisans are those that boast the greatness or superiority of a rival party. Some studies show that this kind of 'collective narcissism'—an excessive idealization of someone's ingroup (Golec de Zavala, 2011)—is a common form of identity performance on social media (Papacharassi & Mendelson, 2010). Based on the existing literature, one might thus reasonably expect that being exposed to statements praising the opposition will trigger a desire to reaffirm the superiority of one's own group by correcting the record, or engaging in negative evaluations of the outgroup. My second hypothesis is therefore:

H2: Exposure to outgroup praise increases affective polarization.

If perceived threats to the ingroup are linked to higher intergroup bias, as we saw above, absent explicit threats, individuals are generally less inclined to engage in discriminatory behavior. While they may not feel the need antagonize the outgroup, however, scholars of intergroup relations have shown that people's desire to achieve a positive self-image in society often leads them to adopt various strategies to enhance the status of the groups they identify with (Hogg, 2003; Tajfel & Turner, 1979). One of these strategies—social creativity—involves creatively re-defining the terms of the intergroup relationship in a favorable way, by portraying the ingroup in a more positive light, and emphasizing its positive attributes (Tajfel & Turner, 1979). For that reason, one might expect that partisans will be particularly receptive to messages that praise their political party's actions, as it will also boost their own self-esteem. Furthermore, research from intergroup emotions theory has shown that experiencing positive emotions towards one's ingroup, such as pride and joy, enhances ingroup favoritism and increases individuals' desire to be close to other groups members (Iyer & Leach, 2008; E. R. Smith & Mackie, 2015). I therefore hypothesize that being exposed to comments praising one's in-party will increase positive disposition towards the in-party, thus reinforcing affective polarization.

H3: Exposure to ingroup praise increases affective polarization.

Scholars have also speculated that exposure to criticisms of the opposition might spur further hostility towards the out-party by reinforcing the idea that they are an inferior outgroup and that attacking them is warranted or legitimate (Iyengar et al., 2012; Iyengar & Westwood, 2015; Jamieson & Capella, 2010). As I briefly summarized above, however, the literature on partisan incivility has so far produced inconclusive findings regarding its effects on affective polarization. On the one hand, some studies find that exposure to outgroup derogation increases negative affect (Suhay et al., 2018). Others cast doubt on this assumption, showing that individuals are not as reactive to out-party incivility as commonly thought—in fact, it might even improve positive feeling towards the opposition (Gervais, 2015, 2019; Rains et al., 2017). Keeping an intergroup threat framework in mind, one could suspect that since partisans are not directly antagonized and threatened

by messages derogating the outgroup, they might therefore not feel the necessity to react in a hostile way. Given the lack of clear and consistent direction from existing work about this relationship, however, I therefore pose the following as another research question:

RQ1: How does exposure to outgroup criticism impact affective polarization?

Several factors might shape how partisans interpret messages they are exposed to. One of the factors most frequently highlighted in the literature is the strength of someone's party identification. Studies have indeed shown that strong attachment to the ingroup mediates the relationship between group esteem threats and intergroup bias: specifically, the more someone identifies with an ingroup, the more likely they are to react to threats to the group's status (Branscombe et al., 1999). Conversely, individuals who are only weak identifiers may not feel the same sense of allegiance to their political party, and may therefore not be as willing to respond to an attack to defend their own party's status (Branscombe & Wann, 1994). This leads me to formulate the following hypothesis:

H4: Strength of partisanship mediates the relationships between exposure to (a) ingroup criticism, (b) outgroup praise, (c) ingroup praise, (d) outgroup criticism, and affective polarization.

Another mediating factor to consider is political orientation. Indeed, several studies from political psychology reveal some asymmetries in the ways that partisans respond to information and opinions they are exposed: scholars have argued that conservatives, in particular, might be more sensitive to threats than their liberal counterparts, as they tend to react more strongly to negative information and are more averse towards uncertainty (Haas & Cunningham, 2014; Jost et al., 2017; van Leeuwen & Park, 2009). In contrast, scholars find that liberals, by virtue of their concern for fairness and harm avoidance, are less prone to outgroup hostility and aggression than conservatives (Kugler et al., 2014). In the context of this study, I would therefore expect that Republicans react in a more hostile way towards the outgroup when exposed to ingroup threats (i.e., statements

attacking their own party and praising the opposition) and are more receptive to messages elevating their own party's status (i.e., in-party praise and criticisms of the opposition). My final hypothesis is thus:

> *H5: Republicans will react more strongly to being exposed to (a) ingroup criticism, (b) outgroup praise, (c) ingroup praise, and (d) outgroup criticism than Democrats.*

5.5 Methodology

5.5.1 Sample Size & Recruitment

To test these hypotheses, I designed and conducted a 2x3 factorial survey-experiment in December 2020, in which participants were randomly assigned to one out of a total of 6 possible vignettes. One key advantage of factorial designs compared to other types of experimental designs is that they allow for analysis of interactions between factor variables (Collins et al., 2009) and require fewer experimental subjects per condition to reach the same level of statistical power as alternative experimental designs (Wu & Hamada, 2009). Participants for the experiment were recruited on Amazon Mechanical Turk (MTurk). Since the experiment was designed to focus on US voters, all participants were required to be at least 18 years-old and to be residents of the United States. To ensure the quality of the subject pool, I also restricted participation to individuals who had at least a 90% approval rates on MTurk for previous tasks (Berinsky et al., 2012).

Several studies attest to the quality of MTurk as a recruitment tool and source of experimental data sample. Research shows, for example, that liberals and conservatives in MTurk samples closely mirror the psychological divisions of liberals and conservatives in the mass public compared to two benchmark national samples, confirming its suitability as a tool for research on political ideology (Clifford et al., 2015). MTurk's subject pool tends to be much more geographically and ethnically diverse than a typical student convenience sample (Paolacci et al., 2010) and respondents tend to be older and more experienced than undergraduate students (Berinsky et al., 2012).

To calculate the appropriate sample size for the experiment, I performed a statistical power analysis using GPower (Erdfelder et al., 1996). I first determined the approximate effect size by estimating the total variance in the outcome variable

that would be accounted for by a predictor. A conventional partial *η2* for small effect sizes—which are common in media effects literature—using Cohen (1988) criteria is 0.02. This corresponds to an approximate effect size of 0.14. With an alpha of $\alpha = 0.05$ and power set at 0.80, I therefore projected that the minimum sample size needed for this effect would be 476. This sample size allows for expected attrition. The full details for the power analysis are included in *Table C.1.* of *Appendix C*.

The final sample consisted of 499 participants, who were equally split between Republicans (49.9%) and Democrats (50.1%). Consistent with other studies in the literature (see for example. Gervais, 2019; Levendusky, 2013), and to better hone in on the effects of partisan comments on party identifiers, pure independents—those who do not lean towards either political party—were excluded from the sample. In each partisan group, around 22% of respondents identified as strong partisans (see *Appendix C* for complete descriptive statistics). The sample skewed more male ($n = 327$, 65.5%), and younger than the general population, with a mean of 38 years old ($SD = 11.3$) and the modal category being 30-34-year-olds. The majority of respondents were also highly educated, having received either an undergraduate (38.8%) or graduate education (35.1%). While this sample is not perfectly representative of the US population, studies have shown considerable similarity in treatment effects between nationally representative population-based and convenience samples like this one (Berinsky et al., 2012; Mullinix et al., 2015).

5.5.2 Study Design

In the pre-test, all participants began by answering demographic questions about their age, education, party affiliation and the frequency of their social media use. Based on their stated political affiliation, Republican and Democrat identifiers (including party leaners, but excluding pure Independents) were randomly assigned to one out of six possible vignettes. As can be seen in *Table 5.1*, the key factors manipulated in each condition were (1) the tone of a comment (either positive or negative) and (2) its target (either aimed at a respondents' in-party, out-party or no one in particular). Including negative comments that were not targeted at a specific political entity was critical to test my hypotheses, which aim to zero in on the effects

of *partisan* praise and criticism on participants' responses and not simply the effects of negative or positive comments. To lend the study greater external validity, the vignettes were designed to emulate the type of content a Facebook user would come across on their News Feed. Instead of being shown a comment in isolation, after completing the pre-test questionnaire, participants across all conditions were presented with the same Facebook post linking to an article from *MarketWatch* about the disappearance of the middle class (Suhay et al., 2018)—a topic that broadly unites voters across party lines[26]— followed by a single comment. Participants were not required to read the article, but were instructed to focus on the comment underneath the post.

Table 5.1. Treatment design

Experimental Condition	*Tone of comment*	*Target of comment*	*N*	*%*
Generic Positive	Positive	None	83	16.6%
Pro-Ingroup	Positive	In-party	76	15.2%
Pro-Outgroup	Positive	Out-party	93	18.6%
Generic Negative	Negative	None	85	17.1%
Anti-Ingroup	Negative	In-party	84	16.8%
Anti-Outgroup	Negative	Out-party	78	15.6%

The choice of Facebook as an appropriate platform for this study is justified in three ways. First, Facebook is the second most popular social media platform in the US (used by 69% of US adults), right after YouTube, which lends the study greater generalizability among US social media users (Auxier & Anderson, 2021). Second, exposure to political content on Facebook is habitual and recurrent: seven-out-of-ten Facebook users currently say they use the site every day (Auxier & Anderson, 2021), and even users who are not interested in politics report stumbling across political content on Facebook when browsing the site on a regular basis (M. Anderson & Quinn, 2019; Anspach, 2017; Vaccari et al., 2016). Finally, there is already research indicating that the Facebook News Feed ecosystem in particular—with its cue-rich blend of political expression, news sharing and discussion—activates users' political identities, with downstream effects on political attitudes (Settle, 2018a).

[26] Survey show majorities of US adults agreeing that current economic conditions are hurting those in middle class (Igielnik & Parker, 2019).

Figure 5.1. Examples of treatment stimuli



Note: From the top left corner to bottom right, examples of the 'anti-Democrats', 'generic negative', 'pro-Democrats' and 'generic positive' comments. *Source:* Author.

Participants assigned to the 'Pro-Ingroup' (n = 76) and 'Pro-Outgroup' (n = 93) conditions saw a comment praising either Republicans or Democrats for supporting the middle class underneath the Facebook post, while participants treated with the 'Anti-Ingroup' (n = 84) and 'Anti-Outgroup' (n = 78) stimuli saw a comment criticizing the same groups for destroying the US economy (see *Figure 5.1*). These four treatments mirrored each other, with the name of the political party being the only changing feature across conditions. Participants who were assigned to the 'Generic Negative' (n = 85) condition read a negative comment deploring the end of the American dream, while those who were assigned to the 'General Positive' (n = 83) condition saw a positive comment in which a user expressed gratitude for being employed.

To avoid influencing participants' responses, I purposefully chose gender neutral names and abstract profiles pictures for both the poster and commenter (see *Figure 5.1.*), as visual cues in pictures typically lead to inferences about an individual's personal traits. Gendered profile pictures and selfies, for example, have been shown to have differential effects on impression formation (Krämer et al., 2017). Likewise, I also refrained from not including a picture altogether, as most people consider 'egg profiles'—or picture-less profiles—as suspicious, often associating them with bots (Hunt, 2017).

To maximize participant's attention and ensure the quality of responses, I followed Qualtrics procedure and restricted respondents' ability to skip through the survey: each vignette was displayed for a minimum of 15 seconds and all post-treatment questions had a 5 second minimum time display. To check that respondents had been appropriately treated, I included a manipulation check in the post-test survey asking respondents whether they thought the person who had posted the comment in response to the Facebook post was a 'Republican', 'Democrat' or 'Independent' (see *Appendix C* for the complete pre- and post- survey questionnaire).[27] No additional information was supplied. Of all the participants exposed to the partisan comments, 95.9% classified the Facebook user behind the

[27] In this case, successful treatment constitutes identifying the Facebook user behind the comment they were exposed to as a member of the political ingroup (when treated with the pro-ingroup or anti-outgroup comment) or as a member of the political outgroup (when treated with the pro-outgroup or anti-ingroup comment).

anti-Republican and the pro-Democrat message as a Democrat, and the one behind the pro-Republican and anti-Democrat message as a Republican. Only 1.15% said they didn't know and about 2.96% of respondents across all conditions classified the commenter as an Independent. Given the high percentage of accurate inferences, all participants were retained in subsequent analyses.

5.6 Measures

Affective Polarization: Affective polarization has been defined as partisans' tendency to feel negatively towards and discriminate against their political opponents, while feeling greater affinity and acting more positively towards members of their own party (Iyengar et al., 2012, 2019). In this sense, affective polarization has two distinct components—in-party favoritism and out-party hostility—which can be measured along two dimensions: a behavioral and an evaluative one. On the evaluative side, scholars have measured affective polarization in at least three ways: through survey self-reports of affective ratings towards the in-party and the out-party, social distance measures and trait ratings (Druckman & Levendusky, 2019). The first approach, which is the most common in the literature, measures how differently respondents feel towards their own party versus the out-party through a feeling thermometer instrument (Lelkes & Westwood, 2017). Scholars also employ alternative measures of partisan animosity, including trait ratings of party supporters (Garrett et al., 2014; Iyengar et al., 2019; Settle, 2018a). Levendusky (2017), Levendusky and Malhotra (2016), and Settle (2018a) for example, all document that partisans overwhelmingly associate negative traits—such as lack of morality, open-mindedness, intelligence and political knowledge—with their political opponents. Finally, a third approach engages with behavioral manifestations of affective polarization, such as the extent to which partisans prefer to interact with members of the ingroup over the outgroup, or give preferential treatments to their co-partisans (Iyengar & Westwood, 2015; Suhay et al., 2018).

Although these measures strongly relate to one another, they can paint a slightly different picture of affective partisan polarization depending on whether an instrument taps into partisans' attitudes towards other voters or towards political elites (Druckman & Levendusky, 2019). To enrich my analysis and better separate between these two mechanisms, in this study I thus choose to measure affective

polarization in three distinct ways: through knowledge evaluations of ingroup and outgroup members, affective ratings of the out-party, and a behavioral engagement item.

Knowledge Evaluation: The first item takes the form of a trait rating measure. Immediately after the seeing the stimuli, I first asked participants to evaluate the level of political knowledge of the person who had posted the comment they were assigned to on a 4-point scale ranging from 'Not very knowledgeable' to 'Very knowledgeable' (M = 2.71, SD = 0.96).

Out-Party Hostility: Following the knowledge question, I collected affective evaluations of the participant's out-party on a 10-point scale ranging from 0 for 'Strong dislike' to 10 for 'Strong like'. The average out-party score is 4.72 (SD = 3.07), with Democrats displaying slightly stronger out-party hostility (M = 4.05, SD = 2.92) than Republicans (M = 5.39, SD = 3.09). Unsurprisingly, strong identifiers of both parties provided slightly lower ratings of the opposition (M = 4.75, SD = 3.51) than other partisans (M = 4.83, SD = 2.76). The measure was reverse coded for the analyses so that high scores correspond to strong dislike and low scores correspond to strong like.

Engagement: Finally, to estimate subjects' willingness to engage with ingroup and outgroup members, I asked participants to indicate what action they would take if they saw the comment they were assigned to on their social media feed. They were given the choice to either 'Like it' (n = 193, 38%), 'Reply to it' (n = 66, 13.2%) or 'Do nothing' (n = 240, 48.1%). Since I am interested in exploring whether a particular stimulus would trigger participants to engage with the comment or not, this was later recoded as a binary variable where 1 corresponds to either liking or responding to the comment, and 0 corresponds to no engagement at all.

5.7 Results

The main research question guiding this study is what types of partisan online comments are the most consequential for affective polarization. To test this, I begin by running two separate Ordinary Least Squares (OLS) models and regress the first

two measures of affective polarization—knowledge evaluations and out-party hostility—on the treatment variables, with the 'Generic Negative' and 'Generic Positive' groups used as comparison categories. I control for party affiliation and strength of partisanship in these models, as stronger partisans have been shown to display significantly higher levels of affective polarization than weak identifiers (Iyengar et al., 2012; Mason, 2018a).[28] I also run logistic regressions to determine the odds that a participant would engage with a social media comment following treatment. Results from all three of these models are reported in *Table 5.2*. Following this step, I analyze the interactions between the treatment variables, strength of partisanship and party affiliation. These results are reported in *Tables 5.3* and *5.4.*

All outcome variables were rescaled between 0 and 1 to allow for comparison. Diagnostic tests for the OLS models showed the relationship between the dependent and independent variables to be linear. These models also indicated no issue with multicollinearity: for all predictors, the variance inflation factor (VIF) fell between 1 and 2. However, in both cases the assumption of constant variance is violated, indicating heteroskedasticity. To provide unbiased standard errors, I therefore followed the method proposed by MacKinnon and White (1985) and Samii and Aronow (2012), and report heteroskedasticity-robust (HC2) standard errors throughout the analyses. I also checked for the presence of multicollinearity between the predictor variables for the logistic regression models. Here again, VIF values did not suggest any issues.

The first two hypotheses suggested that exposure to ingroup criticism (H1) and outgroup praise (H2) would increase affective polarization. As can be seen from models M1.2 and M1.3 in *Table 5.2*, exposure to the pro- outgroup and anti-ingroup treatment both had quite a strong negative effect on political knowledge evaluations. Participants who read a comment praising the opposition attributed a significantly lower knowledge score to the person who posted it ($\beta = -0.14$, $p < 0.05$), while those who read a comment attacking their in-party were even more likely to

[28] As a robustness check, I also ran separate logistic regressions on the likelihood that the treatment would return a negative out-party rating (1 = yes, 0 = no) and a negative knowledge evaluation (1 = yes, 0 = no). The results are presented in *Appendix C* and are consistent with those reported in the main manuscript.

mark the knowledge level of the commenter as low (β = -0.20, p <0.05). Note, however, that neither exposure to in-party criticism nor out-party praise had a significant effect on out-party ratings (see M1.0 and M1.1) or participants' readiness to engage with the online comment (see M1.4 and M1.5). Hypotheses 1 and 2 are thus only confirmed in the case of knowledge evaluations.

Table 5.2. Impact of exposure to partisan praise and partisan criticism on affective polarization measures

	Out-Party Ratings		*Knowledge Evaluation*		*Engagement*	
	M1.0	M1.1	M1.2	M1.3	M1.4	M1.5
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Odds Ratio [CI]	Odds Ratio [CI]
Intercept	0.50*** (0.06)	0.45*** (0.05)	0.54*** (0.05)	0.50*** (0.05)	0.36** [0.16-0.75]	0.42* [0.18-0.90]
Pro-Ingroup	0.04 (0.05)		0.03 (0.04)		1.93± [0.98-3.85]	
Pro-Outgroup	0.02 (0.05)		-0.14** (0.05)		1.62 [0.85-3.10]	
Anti-Ingroup		-0.01 (0.05)		-0.20*** (0.05)		1.08 [0.57-2.05]
Anti-Outgroup		-0.02 (0.05)		0.04 (0.04)		0.77 [0.40-1.47]
Party ID Strength	0.04 (0.06)	-0.02 (0.06)	0.02*** (0.06)	0.22*** (0.06)	7.17*** [3.26-16.55]	5.73*** [2.56-13.41]
Democrat (Yes = 1)	-0.17*** (0.04)	-0.13*** (0.04)	-0.10** (0.04)	-0.10* (0.04)	0.39*** [0.22-0.66]	0.50* [0.29-0.85]
Observations	252	247	252	247	252	247
Adjusted R^2	0.05	0.03	0.11	0.18		
AIC					318.7	327.4

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE stands for standard error and CI indicates the 95% confidence interval. N = 247 for models testing the effects of partisan criticism and N = 252 for those testing the effects of partisan praise.

Hypothesis 3 suggested that exposure to comments praising participants' own party would increase affective polarization by driving favorable evaluations of the in-party and boosting social engagement. As can be seen in models M1.0, M1.2 and

M1.4 of *Table 5.2*, while the pro-ingroup treatment had no effect on participants' ratings of the out-party or their knowledge evaluations of the person who posted the comment, it has a strong effect on their readiness to behaviorally engage with it (OR = 1.93, p <0.1). Specifically, participants assigned the pro-ingroup condition had a much higher probability of saying they would either like it or respond to the comment they saw than participants in all other conditions. Thus, Hypothesis 3 is partially confirmed for this outcome variable.

One of the research questions also formulated at the onset of the paper was whether exposure to out-party criticism would have any effect on affective polarization. As is evident from the results reported in models M1.1, M1.3 and M1.5 of *Table 5.2.*, however, this does not seem to be the case. Indeed, being exposed to comments that criticized the out-party did not have a statistically significant effect on any of the affective polarization measures.

Moving on to Hypothesis 4a to 4d, I now turn my attention to whether the strength of participants' party identification moderates the effects of the treatment variables (in-party criticism, out-party praise, in-party praise and out-party criticism) on affective polarization. The models for these hypotheses are reported in *Table 5.3.* Most interaction terms are not statistically significant, with the exception of two: as can be seen in model M2.1 and M2.5, partisanship negatively moderates the relationship between exposure to the out-party criticism and affective polarization (β = -0.27, p <0.1) and positively moderates the relationship between exposure to anti-ingroup stimuli and affective polarization (OR = 0.12, p <0.1). Specifically, strong partisans who were assigned to the anti-outgroup condition were more likely to provide negative ratings of the opposition. And interestingly, when exposed to a comment criticizing their own party, strong partisans had a higher probability of saying they would engage with the comment. Thus, only Hypotheses 4a and 4d are confirmed. Here, it is noteworthy that after adding an interaction term, exposure to criticism of the in-party has a strong positive effect on the probability of engaging with the comment (OR = 4.83, p <0.1), suggesting that the joint effect of this treatment variable and strength of partisanship are more influential in explaining variation in this outcome variable than the treatment variable alone.

Table 5.3. Impact of exposure to partisan praise and criticism on affective polarization measures, by strength of partisanship

	Out-Party Ratings		*Knowledge Evaluation*		*Engagement*	
	M2.0	M2.1	M2.2	M2.3	M2.4	M2.5
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Odds Ratio [CI]	Odds Ratio [CI]
Intercept	0.55*** (0.07)	0.35*** (0.07)	0.58*** (0.06)	0.48*** (0.07)	0.47 [0.16-1.28]	0.14** [0.03-0.53
Pro-Ingroup	-0.06 (0.11)		-0.01 (0.10)		1.03 [0.27-5.39]	
Pro-Outgroup	-0.03 (0.10)		-0.20± (0.11)		1.20 [0.13-8.51]	
Anti-Ingroup		0.08 (0.09)		-0.16 (0.10)		4.83± [0.95-28.3]
Anti-Outgroup		0.17± (0.10)		0.04 (0.10)		3.13 [0.52-20.4]
Party ID Strength	-0.02 (0.10)	0.11 (0.11)	0.15* (0.08)	0.24* (0.09)	4.76* [1.37-18.4]	25.1*** [4.87-166]
Democrat (Yes = 1)	-0.16*** (0.04)	-0.13** (0.04)	-0.10** (0.04)	-0.10* (0.04)	0.39*** [0.22-0.67]	0.47** [0.27-0.81]
Pro-Ingroup * Party ID Strength	0.16 (0.15)		0.04 (0.12)		2.54 [0.36-19.0]	
Pro-Outgroup * Party ID Strength	0.06 (0.14)		0.10 (0.15)		1.50 [0.22-10.1]	
Anti-Ingroup * Party ID Strength		-0.13 (0.14)		-0.06 (0.14)		0.12± [0.01-1.01]
Anti-Outgroup * Party ID Strength		-0.27± (0.15)		-0.01 (0.13)		0.15 [0.01-1.40]
Observations	252	247	252	247	252	247
Adjusted R^2	0.04	0.03	0.11	0.18		
AIC					321.9	327.2

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE stands for standard error and CI indicates the 95% confidence interval. N = 247 for models testing the effects of partisan criticism and N = 252 for those testing the effects of partisan praise.

Finally, Hypotheses 5a to 5d were concerned with whether party affiliation would mediate the relationships between my predictors of interest and affective polarization. Specifically, I anticipated that Republican identifiers would react more strongly to being exposed to in-party criticism, out-party praise, in-party praise, and out-party criticism than Democrats. As can be seen in model M3.2 of *Table 5.4*, only one interaction term is significant in this set of models: identifying with the Democratic Party negatively impacts the relationship between exposure to pro-outgroup comments and knowledge evaluations (β = -0.18, p = 0.06). Specifically, and contrary to my initial expectations, the output shows that Democrats who saw a social media comment praising the opposition were more likely to say that it had been written by someone who was not knowledgeable about politics, compared to Republicans. Based on the data analyzed here, I therefore cannot reject the null hypothesis that Republicans are most sensitive to in-party threats than Democrats.

Table 5.4. Impact of exposure to partisan criticism on affective polarization measures, by party affiliation

	Out-Party Ratings		*Knowledge Evaluation*		*Engagement*	
	M3.0	M3.1	M3.2	M3.3	M3.4	M3.5
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Odds Ratio [CI]	Odds Ratio [CI]
Intercept	0.50*** (0.06)	0.43*** (0.06)	0.51*** (0.05)	0.46*** (0.06)	0.31*** [0.13-0.71]	0.45± [0.19-1.05]
Pro-Ingroup	0.06 (0.07)		0.06 (0.06)		2.98* [1.10-8.56]	
Pro-Outgroup	-0.01 (0.07)		-0.05 (0.07)		1.60 [0.64-4.02]	
Anti-Ingroup		0.06 (0.08)		-0.12± (0.08)		0.95 [0.37-2.49]
Anti-Outgroup		-0.03 (0.07)		0.05 (0.06)		0.70 [0.28-1.72]
Party ID Strength	0.05 (0.06)	-0.02 (0.06)	0.19*** (0.06)	0.22*** (0.06)	7.36*** [3.31-17.25]	5.79*** [2.59-13.58]
Democrat (Yes =1)	-0.17* (0.08)	-0.10 (0.07)	-0.01 (0.06)	-0.04 (0.06)	0.49 [0.19-1.22]	0.44± [0.17-1.07]
Pro-Ingroup * Democrat	-0.04 (0.11)		-0.06 (0.08)		0.45 [0.11-1.78]	
Pro-Outgroup * Democrat	0.04 (0.10)		-0.18± (0.10)		0.99 [0.27-3.58]	
Anti-Ingroup * Democrat		-0.12 (0.10)		-0.15 (0.10)		1.26 [0.35-4.55]
Anti-Outgroup * Democrat		0.04 (0.10)		-0.02 (0.08)		1.19 [0.32-4.42]
Observations	252	247	252	247	252	247
Adjusted R^2	0.04	0.03	0.12	0.18		
AIC					321.1	331.4

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. SE stands for standard error and CI indicates the 95% confidence interval. N = 247 for models testing the effects of partisan criticism and N = 252 for those testing the effects of partisan praise.

5.8 Discussion

Partisan animosity is a growing problem in the United States, and remedying it requires a nuanced understanding of its driving factors. It is becoming increasingly apparent that our online communication environment exposes us to a variety of political messages that have downstream effects on how we view the world and our fellow citizens (Rösner et al., 2016). Political appraisals in particular, from partisan jabs traded in comment sections, to formed opinions about the state of public affairs, form a core part of our digital culture. Yet studies that analyze which of these types of messages are the most consequential for partisan hostility are still scant.

In this paper, I set out to investigate how exposure to social media comments that make positive or negative value judgements about Republicans and Democrats would impact affective polarization. Taken together, findings from this experiment clearly indicate that the comments with the most significant effects on behavioral and evaluative measures of partisan prejudice were those that constituted symbolic threats to the image of participants' partisan ingroup. Exposure to comments that either praised the political opposition or criticized a participant's own party were not only seen as coming from members of the opposition, but led participants to express negative judgement about the knowledge of the person who had posted it. Conversely, exposure to derogatory comments about the opposition did not spur partisans to engage in copy-cat derogation. Finally, contrary to their political counterparts, Democrats were particularly likely to react in an adverse manner to positive statements about the opposition. However, in both cases strength of partisanship did not moderate treatment effects.

Results from this study join others in suggesting that exposure to pro-attitudinal political incivility does not always aggravate hostile feelings towards the opposition (Druckman et al., 2019; Gervais, 2019). Furthermore, they challenge the assumption underwriting some parts of the literature that making partisan affiliation salient is enough to trigger intergroup biases (Levendusky & Malhotra, 2016a; Settle, 2018a). Study participants did not become more affectively polarized after being exposed to any type of partisan comment. Instead, these findings posit distinct mechanisms for exposure to messages that symbolically *threaten* an individual's partisan ingroup status versus those that elevate it (through outgroup

derogation, for example). This is an important distinction that opens up avenues for further research. While most of the political communication literature has in recent years focused on the more toxic side on online political speech, this study shows that references to a political party's greatness could be as consequential for driving animosity towards the 'other side.' This is particularly relevant at a time when political figures regularly invoke marginalized identity status while extolling the virtues and superiority of their own political faction—the hallmarks of collective narcissism—on public platforms like Twitter (Ott & Dickinson, 2019; Sclafani, 2018).

This study is not without limitations. In order to focus on the impact of partisan commentary on affective partisan polarization, I elected to restrict the experiment to 'pure' Democrat or Republican identifiers. However, it will be important for future work to test whether these findings hold with a more representative sample of the US population that includes Independents. Secondly, the mechanisms outlined here rest on the assumption that individual users operate in a social online environment with politically different others. As political conversations increasingly migrate to 'closed', interest-based groups (Wilson, 2020), in which users often engage in ingroup praise and outgroup derogation as a form of identity performance or to gain greater acceptance from other ingroups members (Bright et al., 2020; Postmes et al., 2000, 2005; Rains et al., 2017), it would be worth exploring how group contextual factors moderate these effects.

It should also be noted that the explained variation for some of models presented in this paper is quite small, meaning that the vast majority of the variation in the sample is explained by factors that are not included in the model. Formally, however, the models are a statistically significant improvement over having no model at all.

Social media effects on political attitude and behavior are inherently tied to a platform's specific affordances, that is the set of features, designs and policies that constrain what users can do, and the norms that develop around its use. Consequently, it will be important to test whether these effects are replicated in a different digital setting. Online users are also known to oscillate between individualistic and group-based modes of cognition and communication in digital environments (Hinck & Carr, 2020). While testing the extent to which different

messages actually activate users' sense of partisanship was outside the scope of this paper, finally, it will also be interesting for future studies to tackle this question and address the role that identifiability plays in moderating the relationship between partisan commentary and affective polarization.

5.9 Conclusion

There is still a paucity of research exploring the impact of exposure to online political talk on affective polarization. In particular, it has thus far been difficult to tease out what types of partisan messages are the most consequential for partisan affect. In this paper, I argued that our understanding of these effects is enhanced by looking beyond the presence of partisan cues in online comments. Instead, the results reveal that how a message is received and whether it triggers partisan animosity and aversion hinges on a complex interplay between the tone—whether critical or complementary—and the target of a given message. Statements that constitute either a direct or symbolic threat to an individual's in-party were more likely to trigger a negative reaction, and further out-party hostility, while reassurances did not have the same effects.

5.10 Appendix C

Survey questionnaire

Pre-test

1) Generally speaking, do you think of yourself as a Democrat, a Republican, an Independent, or something else?
    - o Strong Democrat
    - o Weak Democrat
    - o Lean Democrat
    - o Independent
    - o Lean Republican
    - o Weak Republican
    - o Strong Republican

2) If Independent, do you think of yourself as closer to the Republican Party or the Democratic party?
    - o Democratic Party
    - o Republican Party
    - o Neither

3) Please enter your age:

4) What is your gender identity?
    - o Male
    - o Female
    - o Other

5) What is the highest level of education you have completed?
    - o None
    - o High School diploma
    - o Technical/vocational training
    - o Undergraduate
    - o Postgraduate
    - o Other

6) Are you a social media user?
    - o Yes
    - o No

7) Which social media platform do you use the most?
    - o Facebook
    - o Twitter
    - o Instagram
    - o YouTube
    - o TikTok
    - o Other

8) On a typical day, how much time would you say you spend on social media (either on a mobile or a computer)?
    - Less than 30 minutes
    - 30-60 minutes
    - 1-2 hours
    - 2-3 hours
    - 3+ hours

Imagine you are scrolling down social media and come across this post.
Please pay attention to the comment underneath the post.

Post-test

1) How knowledgeable about politics is the person who posted this comment
    - Very knowledgeable
    - Somewhat knowledgeable
    - Not very knowledgeable
    - Not at all knowledgeable
    - I don't know

2) Let's think about political parties and their supporters for a moment. Using a scale from 0 to 10, where 0 means strongly dislike and 10 means strongly like, how you do feel about…? [Scale 0 to 10]

- Republicans
- Democrats

3) What action would you take if you saw the above social media comment?
    - Like
    - Reply
    - Do nothing

4) In your opinion, is the person who posted this comment a…?
    - Democrat
    - Republican
    - Independent
    - I don't know

5) In your opinion, was this comment written by a real person?
    - Yes
    - No

6) How interested are you in politics and public affairs?
    - Very interested
    - Somewhat interested
    - Not very interested
      Not at all interested

Table C.1. Power analysis

Analysis:	A priori: Compute required sample size		
Input:	Effect size f	=	0.14
	α err prob	=	0.05
	Power (1-β err prob)	=	0.80
	Numerator df	=	2
	Number of groups	=	6
Output:	Noncentrality parameter λ	=	9.71
	Critical F	=	3.01
	Denominator df	=	470
	Total sample size	=	476
	Actual power	=	0.80

To test the effectiveness of the treatments, I ran a pilot study on a convenience sample (n = 15). Qualitative insights from this pilot indicated that treated participants tended to focus more on the Facebook post as a whole, rather than on the comment underneath. This was remedied by displaying the following prompt before allocating participants to the different vignettes: 'Imagine that you are scrolling down social media and you come across the following post on your timeline. Please pay attention to **the comment** under the post**.**'

Table C.2. Demographic distribution of experimental participants

Age Groups	*N*	*%*
18-24	18	3.61
25-29	96	19.2
30-34	126	25.2
35-39	71	14.2
40-44	52	10.4
45-49	45	9.02
50-54	28	5.61
55-59	27	5.41
60-64	23	4.61
65-69	7	1.40
70-74	4	0.80
75+	1	0.20
Missing	1	0.20
Party Affiliation		
Strong Democrat	111	22.2
Moderate Democrat	79	15.8
Weak Democrat	33	6.61
Lean Democrat	27	5.41
Lean Republican	34	6.81
Weak Republican	38	7.62
Moderate Republican	63	12.6
Strong Republican	114	22.8
Education		
None	3	0.60
Technical/Vocational	29	5.81
High School Diploma	98	19.6
Undergraduate	194	38.8
Graduate	175	35.1
Total	499	100

Table C.3. Summary statistics for outcome variables

		Knowledge Evaluation		*Out-Party Rating*	
Experimental Condition	N	Mean	St.Dev	Mean	St.Dev
Anti-Ingroup	84	0.94	0.24	4.17	3.07
Anti-Outgroup	78	0.95	0.22	4.19	2.77
Generic Negative	85	0.90	0.31	4.24	3.03
Pro-Ingroup	76	0.96	0.20	5.34	3.09
Pro-Outgroup	93	0.97	0.18	5.24	3.07
Generic Positive	83	0.92	0.26	5.13	3.23

Note: St. Dev stands for standard deviation.

Figure C.1. Average knowledge evaluations and out-party ratings per treatment and partisan group

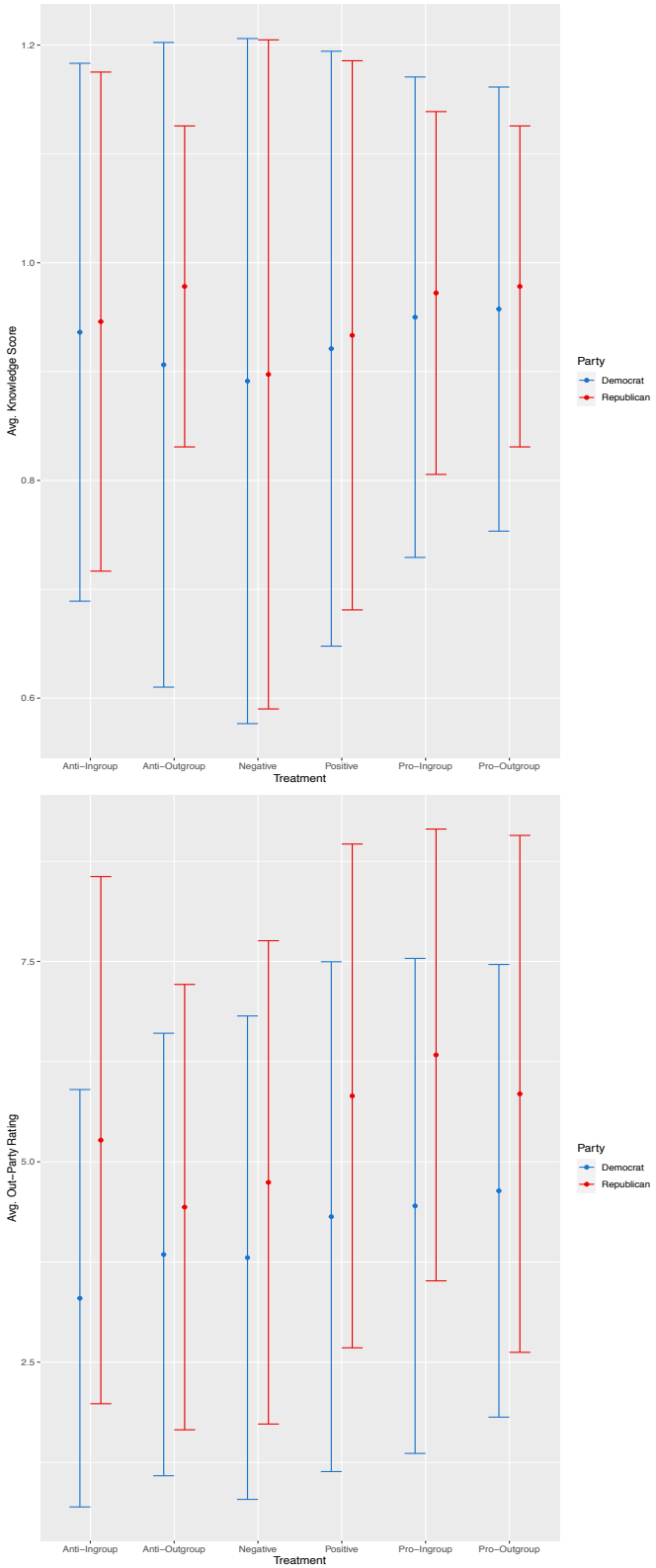


Table C.4. Additional OLS regression models

	Out-Party Rating		*Knowledge Evaluation*		*Behavioural Engagement*	
	Estimate	Std. Error	Estimate	Std. Error	Odds Ratio	95% CI
Intercept	0.52***	0.05	0.534***	0.04	0.36***	0.19-0.67
Pro-Ingroup	0.03	0.05	0.03	0.04	1.89±	0.97 - 3.72
Pro-Outgroup	0.02	0.05	-0.14**	0.05	1.59	0.85-3.02
Anti-Ingroup	-0.09±	0.05	-0.23***	0.05	1.24	0.65-2.39
Anti-Outgroup	-0.11*	0.05	0.01	0.04	0.86	0.44-1.66
No Target Negative	-0.09±	0.05	-0.03	0.04	1.13	0.60-2.16
Democrat (Yes = 1)	-0.15***	0.03	-0.10***	0.03	0.44***	0.30-0.64
Party ID Strength	0.01	0.04	0.21***	0.04	6.43***	3.65-11.6
Observations	499		499		499	
Adjusted R-squared	0.06		0.15		AIC: 642.77	

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1.

Table C.5. Additional logistic regression models

	Negative Out-Party Rating		*Negative Knowledge Evaluation*	
	Odds Ratio [95% CI]		Odds Ratio [95% CI]	
Intercept	0.63 [0.31-1.28]	1.18 [0.55-2.55]	0.57 [0.25-1.24]	0.73 [0.32-1.65]
Pro-Ingroup	0.91 [0.48-1.73]		0.84 [0.38-1.84]	
Pro-Outgroup	1.05 [0.57-1.92]		3.24*** [1.65-6.58]	
Anti-Ingroup		1.10 [0.57-2.11]		3.53*** [1.80-7.12]
Anti-Outgroup		1.20 [0.62-2.36])		1.09 [0.53-2.26])
Party ID Strength	1.04 [0.50-2.15]	1.22 [0.55-2.67]	0.28** [0.12-0.64]	0.29** [0.12-0.66]
Democrat (Yes = 1)	2.39*** [1.44-3.99]	1.97* [1.15-3.44]	1.64± [0.92-2.95]	1.61 [0.91-2.88]
Observations	252	252	252	252
AIC	347.4	316.1	283.5	287.12

Note: *** p <0.001, **p<0.01, * p<0.05, ± p<0.1. N = 247 for all models.

6 Conclusion

6.1 Introduction

What drives affective polarization at the individual-level? Throughout this thesis, I addressed this question by exploring some of the mechanisms that shape how partisans think and behave towards their political counterparts. I did so through two distinct lines of inquiry. In Chapter 3, I first attended to the 'micro-foundations' of affective polarization by examining how individual's political attitudes impact their affective evaluations. Drawing on survey data from the British Election Study, I showed that perceived gaps between voters' personal beliefs and those of the party they identify with lessen affective polarization, by improving their feelings towards the opposition and worsening evaluations of their own party. In other words, internalized conflict between one's partisan identity and preferences on key issues has a depolarizing effect on voters. In Chapter 4 and 5, I then turned my attention to the nature of individuals' online political interactions. Analyzing digital trace data from Reddit, my second study showed that online discussions between politically opposed users are also characterized by affective biases, with likeminded users being much more likely to engage positively with one another than with dissimilar users. Moreover, these patterns are contingent upon the tone and target of user messages. Positive exchanges across party lines, for example, can foster more positive engagement over time. Finally, in my last empirical study, I tested the effects of exposure to various types of online partisan comments on affective polarization. Results from this study indicated that not all partisan messages have downstream effects on partisan prejudice. Only when exposed to comments that criticized their own party or commended the opposition, did US partisans react with greater animosity towards their political opponents. Through this work, I thus demonstrate that affective polarization at the individual level is contingent on a number of situational and dispositional factors. Specifically, the extent to which partisans treat their political opponents as a dislike outgroup depends on how ideologically aligned they feel with their own political party, how their political rivals engage towards them in online conversations, and whether they are made to feel threatened by them.

Tying together the findings from these studies, what do they reveal about the drivers of partisan animosity, and how does this inform the study of affective polarization more broadly? I identify two main take-aways, which I address in turn in the next section. They first suggest—contrary to a common assumption in the literature (Bishop, 2009; Iyengar et al., 2012; Iyengar & Westwood, 2015)—that even people who care deeply about politics, such as party identifiers and people who discuss politics with others online, may not be intrinsically motivated to discriminate or express negative views against their political rivals. In fact, unless provoked or threatened, most people seem to mostly shy away from this type of behavior. Furthermore, all three of my studies underscore the fact that the *extent* to which partisans express prejudice towards their political rivals is not fixed or immutable. Whenever partisans were in situations that could make them feel ambivalent towards their co-partisans or develop increased sympathy for the other side, they actually became less inclined to behave in a prejudicial way. Attending to these individual-level variations is vital for the study of partisan prejudice.

In the sections below, I explore these conclusions in more depth by revisiting the findings from my empirical studies. I highlight the links and logic between them, and their broader implications for the study of affective polarization. Throughout, I also reflect on the limitations of my methodological design choices—what they illuminate and what they leave in the dark—and consider how future studies may compliment and improve on the presented research designs. In conclusion, I draw out several recommendations for future research and interventions to reduce political hostility.

6.2 Summary of Findings

6.2.1 Is Partisan Identity One Hell of a Drug?

As we saw in Chapter 2, several propositions have been advanced to explain the emergence and strengthening of affective polarization. Whether rooted in the policy-based or identity-based model, however, all seem to operate under a similar assumption: that identifying with a political party or set of ideological beliefs *automatically* begets hostility and negative attitudes towards those who we perceive as our political or ideological 'enemies' (Bishop, 2009; Iyengar et al., 2012; Iyengar & Westwood, 2015).

For proponents of the identity-based model, in particularly, affective polarization 'stems directly' from this process of differentiation (Levendusky, 2018).[29] Summarizing the canonical position of scholars in the literature, Levendusky (2018) writes:

> Affective polarization is an outgrowth of this classification: because the other party is an outgroup, they take on a negative connotation and individuals dislike and distrust them' (p. 584).

In their seminal 2015 paper, Iyengar and Westwood even go as far as suggesting that 'hostile feelings for the opposing party are *ingrained* or *automatic* in voters' minds' (Iyengar & Westwood, 2015). This characterization of out-party hostility as an automatic rather than reflexive process is reminiscent of a popular meme among political scientists likening partisanship to a 'psychoactive drug' with an 'intoxicating' grip and influence on people's behavior (Fowler, 2020). Whereas the phrase is typically used in reference to the predictive power of partisanship on voting choice, the analogy easily extends to its purported impact on social attitudes.

The pervasiveness of this assumption in the literature makes sense, as variants of it can be found across many cognitive, political and social theories. There is a wealth of evidence suggesting, for instance, that political judgment is subject to a negativity bias. Put differently, human beings tend to privilege negative over positive information when forming impressions of others, which would imply that affective polarization is driven primarily by outgroup hostility (Skowronski & Carlston, 1989; Soroka, 2014; Soroka et al., 2019). This negativity bias is also reflected in the notion of 'negative partisanship': the tendency for voters to form political opinions and justify their affiliation with a political party in oppositional or negational terms (Abramowitz & Webster, 2018; Bankert, 2020). Partisans do so not because they believe in their party's platform per se, but mainly because they are averse to the consequences of having the opposition in power. Epitomizing this view, in his 1932 book *The Concept of the Political*, the German political theorist Carl Schmitt famously argued that 'the specific political distinction to which political actions and motives can be reduced is that between friend and enemy' (C. Schmitt,

[29] Although as I have argued in previous chapters, specific policy positions or ideological beliefs can themselves be internalized as social identities, and therefore generate a similar categorization process.

1932). For Schmitt and others, political enmity is an inevitable consequence of distinguishing between those who are part of one's community and those who stand outside of it. Beyond serving as one of its essential functions, enmity is also the very essence of the political, and politics itself is reducible to a perpetual struggle between those we view as 'allies' and as those we construe as 'enemies.'

This coupling of political identification—and therefore categorization—with unfettered hostility and prejudice is also present in much of the literature on social media's effects on affective polarization. As we saw in earlier chapters, supporters of social categorization theory largely explain the effects of digital media on partisan hostility through the power of partisan cues to activate people's tendency to sort the world into liked ingroups and disliked outgroups. Concomitant with the proliferation of partisan news and political vitriol online, many scholars, journalists, and pundits have come to argue that incivility and partisan rhetoric on social media are largely responsible for rousing partisan combativeness at the expense of political tolerance (for summaries see example Barberá, 2020; Jungherr et al., 2020; Tucker et al., 2017). Others yet again believe that partisan media and commentary perpetuate misleading stereotypes about the opposition or reinforce ideological differences between parties. In both cases, however, the implicit assumption is that partisans and politically interested citizens are so 'intoxicated' by partisan heuristics that they hold little critical agency over their actions and motives, and are instead automatically prompted to respond in a biased way (Fowler, 2020).

6.2.2 The Limits of Partisan Hostility

The findings from my thesis, however, paint a more nuanced picture of the context and manner in which partisan hostility manifests. Throughout all my studies, I find consistent evidence that affective polarization does not, in fact, automatically follow from partisan identification. Indeed, partisans only express prejudice towards their political opponents under certain circumstances. They do so, for example, when having an unpleasant exchange with someone who holds different opinions to them, or when exposed to messages that threaten the status of their partisan ingroup. As we saw in Chapter 3 and 4, when their own views do not line up with that of the party they identify with, or when the actions of their supposed

opponents do not conform with expectations, however, many partisans refrain from hostile confrontation. My analysis of Reddit conversation dynamics reveals that while political interactions can be rancorous, online users are more amenable to positive crosscutting engagement than outright conflict. In line with the literature on conflict avoidance, when confronted with a negative comment from a political opponent for example, users were simply more likely to leave a conversation than to retaliate in a hostile way (Meleady & Forder, 2019). Likewise, and at odds with the accepted wisdom on enmity, negative mentions of the opposition in conversations between likeminded users did not encourage further outgroup hostility or derogation. These results present an interesting contrast to those exhibited in Chapter 5. Even in the survey experiment, study participants did not respond to partisan cues in online comments in a uniform manner. While they were not triggered by statements denigrating the opposition, in this case, however, they were more likely to react in a hostile manner when exposed to comments that overtly criticized their own party.

It is worth exploring potential explanations for these different outcomes. One is simply the extent to which people are willing to express hostility in public versus in private settings. The set of norms that rule a particular communication space impact how people discuss politics. Within the confines of an online experiment, with less expectation of social desirability, participants may have been more at ease with expressing negative attitudes towards other users. However, as we saw in that chapter, when offered the option to engage with a disagreeable comment to counter it or correct the record, few participants were actually willing to respond. In contrast, we know from the literature on the online disinhibition effect that anonymous environments can oftentimes encourage anti-deliberative or combative behavior (Nithyanand et al., 2017). Reddit users operating in a pseudonymous environment were more likely to be hostile towards their political opponents, on average. When zooming in on particular conversation dynamics, however, it became clear that it was not outgroup hostility that incentivized this type of behavior. Although exploring these patterns was beyond the scope of my analysis, the visibility of an interlocutor's political affiliation and group factors linked to the size and dominance of a particular group in a conversation space could also be important moderators of this relationship. It will thus be interesting

for future work to test whether individuals behave differently when members of their own ideological or political group outnumber the opposition (whether it emboldens them to show prejudicial behavior, for example). Likewise, it will be pertinent to assess whether people behave differently on a platform like Facebook, where there is a lower expectation of anonymity and where social interactions primarily take place with friends, acquaintances, and weak ties.

Whichever way we look at it, however, the take-away is clear: partisans do not express *unconditional* hatred towards the 'other side' simply by virtue of identifying with a political party or being put in situations in which their political identity is activated. Assuming otherwise by-passes important contextual factors. Taken together, my work thus lends support to decades of previous scholarship from social psychology that sees attachment to the partisan ingroup ('ingroup love') as having psychological primacy over dissimilarity with the outgroup ('outgroup hate') when forming social judgments. Extended to the political domain, this suggests that partisan prejudice as we currently measure it may be more an expression of individuals' desire to protect the status, beliefs and principles of their own group, and ultimately themselves, than of gratuitous hatred towards the other side (Brewer, 1979; Mummendey et al., 2000).

6.2.3 On the Role of Emotions

The findings from my empirical studies also seem to indicate that the *extent* to which partisans express hostility towards one another is shaped, at least in part, by the emotions they associate with their ingroup and outgroup in a given situation. This is not to dismiss the power of differentiations: there is no question that accentuating differences between partisan groups can have negative outcomes on people's attitudes. Throughout this thesis, however, I have demonstrated that the existence of categorical boundaries in itself does not predict the magnitude of affective partisan polarization. Instead, what my work suggests is that people's emotional reactions to situations and events in which their political identity is brought to the fore play a key role in shaping the contours of these category boundaries, and that it is this shaping process that may be the most consequential for partisan prejudice.

In her 2004 volume *The Cultural Politics of Emotion*, Sara Ahmed foregrounds a similar model of emotions as mediators between oneself and one's social environment. Emotions, she argues, are inherently relational. Emotions are not simply an expression of how individuals feel but also the way through which they respond to objects and others and through which 'surfaces or boundaries are made. The "I" and the "we" are shaped by, and even take the shape of contact with others' (2014, p. 10). Writing nearly a century prior, Walter Lippman (1922) paints a similarly evocative picture of the affective nature of social categorization. He writes:

> [Stereotyping] is not merely a way of substituting order for the great blooming, buzzing confusion of reality. It is not merely a short cut. It is all these things and more. It is the guarantee of our self-respect; it is the projection upon the world of our own sense of our own value, our own position and our own rights. The stereotypes are, therefore, *highly charged with the feelings that are attached to them*. They are the fortress of our tradition, and behind its defenses we can continue to feel ourselves safe in the position we occupy (pp. 63-64, emphasis added)

For Lippman, as for Ahmed, emotions are foundational to the process of ordering and typifying the mass of information we receive from our environment. They are also central in helping us make sense of our place in the world and that of others within it. As I suggested above, and as Lippman argues in this passage, our affective dispositions towards others also form a core part of our *own* political identity—our inner narrative about our political values, preferences, and the individuals we view as allies on that basis—and are also a way for us to uphold the integrity of that political identity. We might attach negative stereotypes to outgroups, for instance, in order to feel better about ourselves; and we might think of them as lazy or immoral as a way to highlight our own sense of moral superiority.

However, as I have demonstrated throughout this thesis, and as the work of intergroup emotion scholars clearly underscores, the nature of the feelings we attach to these groups is not impervious to change. In fact, they are highly dependent on situational and dispositional factors. Participants to my survey experiment, for example, did not automatically react to all partisan comments they were exposed to on social media, even with clear partisan cues. They only reacted to the ones that somehow called into question or jeopardized the perceived

superiority of their own party. Likewise, in Chapter 3, we learned that British partisans whose policy preferences were inconsistent with that of the party they identify with held more negative attitudes towards that party and more favorable attitudes towards the opposition. In fact, it seems that whenever partisans were put in situations that could make them feel ambivalent towards their co-partisans or develop increased sympathy for the other side, they became less inclined to behave in a prejudicial way. This was also apparent in the study of Reddit political discussion dynamics, where positive engagement between users across ideological lines had a clear depolarizing effect.

Abstracted to the broader scholarly conversation about the dynamics of partisan prejudice, these findings imply that the 'otherness' of our political rivals, and the negative connotations we associate with them as a result, may fluctuate as our affective dispositions towards them shift. Some of these patterns were clearly evident in my empirical studies. Scholars of intergroup emotions have long established that hostility towards the outgroup generally manifests when strong negative emotions, such as anger or fear, are associated with that group (Mackie et al., 2000). In the survey experiment presented in Chapter 5, it is entirely possible that study participants ascribed negative traits to their political rivals because they felt anger or resentment after being exposed to comments praising the political opposition. Likewise, it is equally plausible that British partisans' perceptions of disconnect with their own party instilled in them feelings of anger or disappointment towards that party, and generated greater enthusiasm for the opposition, thereby improving their overall affective disposition towards them (McLaughlin et al., 2020). Of course, without the ability to probe respondents' inner feelings and emotions towards their own party and its leader, these conclusions are necessarily conjectural. However, recent work that identifies negative affect as a direct response to counter-attitudinal political rhetoric suggests that this could be a valid explanatory pathway and one that is worth investigating in greater detail in the future (B. Bakker et al., 2021).

6.3 Contributions and Limitations

My thesis presents a number of theoretical and methodological strengths. First, it advances existing debates about the mechanisms underlying affective

polarization by challenging the notion that policy conflict and partisan loyalty are mutually exclusive, or indeed sufficient, explanations for partisan hostility. Debates about the relative importance of party identification versus policy preferences in driving individuals' attitudes have long been central to the political science literature (Campbell et al. 1960) on performance evaluations and voting behavior. Though both theories offer competing explanations for the source of voters' behavior, they each assume that one takes precedence over the other. In other words, individuals are either motivated by a desire to defend their policy interests or by a desire to maintain their partisan identity. Findings from my thesis complicate this assumption, however, by showing that partisans not only report some degree of mismatch between their affiliation and policy preferences, but that this tension brings them closer to the political opposition affectively, while moving them away from their own party. In line with the dual motivation theory (Groenendyk, 2013), this may suggest that partisans in fact frequently 'wrestle' between these two competing motives and strive to reconcile them by updating their affective allegiances. Likewise, however, it is entirely possible that voters do not interpret policy preferences through the lens of their own rational motivations, but in terms of what they represent for the party's identity. In this case, for example, a pro-Remain Labour voter may express dislike towards a Labour candidate or supporter they perceive as pro-Leave, not necessarily because they disagree with them on policy, but because they care about the integrity of the party (and may therefore regard them as disloyal). Taken together, the analyses presented in previous chapters thus suggest that further research, including new models and methodological approaches, are needed to identify what factors that truly motivate affective polarization.

Second, this thesis enriches theory about the effects of exposure to online political content online on affective polarization by positing separate mechanisms for messages that threaten the status of an individual's political ingroup and those that simply comfort its status. Research into the polarizing potential of partisan news and online political talk has thus far been largely dominated by studies of online hate and incivility. What my thesis underscores, however, is that this approach is somewhat limiting, in that it ignores other modes of communication —such as collective narcissism—that carry worse consequences for partisan affect.

There is currently a lack of comprehensive theoretical model to explain how interacting with political content in digital settings impact users' attitudes and behavior towards their political counterparts. To reach a sophisticated understanding of these mechanisms, it is therefore imperative that we continue to broaden our theoretical expectations about the possible types of content, affordances, and situations that could foster hostile feelings between groups. Furthermore, the analyses presented in this thesis also underscore the usefulness of integrating the role of emotions and group dynamics to existing models of online prejudice formation. In doing so, my work joins that of Settle (2018) and Suhay et al. (2018), as well as Kim & Wojcieszak, (2018), Wojcieszak & Warner, (2020) and others who have already done so for prejudice reduction.

Finally, my research contributes to the burgeoning literature on the reduction of affective polarization by offering several promising avenues for intervention. Results from my first empirical study, for example, suggest that increasing partisans' ambivalence towards their own political party, and emphasizing areas of common policy agreement between rival partisans (especially on issues they care about) may both be fruitful in reducing inter-party animosity. Moreover, by showing expressions of positivity across political lines to have a depolarizing effect on users, my work also highlights warm group relations and positive contact as effective strategies for reducing partisan prejudice in online discussions.

From a methodological standpoint, by combining multiple sources of data (namely, digital trace and survey data), methodological approaches (experimental, observational, and statistical) and measurement strategies (evaluative and behavioral), my thesis provides a nuanced and holistic account of the factors driving individual-level affective polarization across multiple settings. There is already a long and rich tradition of experimental studies testing the effects of digital communication on social and political attitudes. The survey experimental design presented in Chapter 5 contributes to this tradition, by allowing me to reveal strong causal relationships between exposure to in-party threats and negative interpersonal affect. One of the limitations of these types of designs, however, is that they can sometimes lack ecological validity, as they do not fully reflect the way in which individuals interact with content on social media. Studies that harness

social media data, like the one I presented in Chapter 4, counter these limitations by allowing researchers to analyze patterns and relationships between variables in a naturalistic setting. Taking a computational approach to the analysis of conversational dynamics in Chapter 4 has not only enabled me to analyze behavioral data at scale, but also to observe people's behavior, as well as the types of messages and communicative strategies that were the most emotionally activating for them, in a non-obstructive way. Juxtaposing these two approaches provides a powerful way to offset the limitations of using either method on its own.

Beyond these strengths, my methodological choices also come with limitations, which must be acknowledged. However, in most cases, these limitations provide clear opportunities for future research. First, my empirical studies largely focused on the British and US electorate. This choice was justified from a theoretical standpoint in that political life in both countries is dominated by two leading parties, creating the perfect conditions for the development of strong outgroup-ingroup dynamics among the electorate. However, this inevitably limits the transportability of my findings to European multi-party contexts, where partisan identification is more complex and political parties can form coalition groups that cut across ideological lines. In recent years, a growing number of scholars have started investigating how affective polarization is expressed in multi-party systems, and what country-level factors explain this phenomenon (Knudsen, 2021; Wagner, 2021). However, there is currently little to no research exploring the mechanisms driving affective polarization in individuals with varying degrees of attachments to one or more party. One possibility could be that citizens' affective evaluations towards different partisan groups reflects these parties' underlying ideological structure: Socialists would exhibit more positive affect towards members of a Green party than a Liberal party, for example. If that were the case, ideology could be functioning as a 'supra' identity that determines voters' affective dispositions. Equally, partisan affect may reveal voters' latent economic or cultural preferences. It is therefore important for future studies to test whether the mechanisms outlined in this thesis can be transported to other political contexts, where partisans can hold multiple, non-overlapping, partisan identities. This will be especially critical in the digital context, as social identity theory has thus far

fallen short of specifying the conditions under which one person's identity becomes more salient than another online.

Secondly, one of the major criticisms that could be levelled at my empirical studies is that they did not explicitly capture the underlying affective states of cross-pressured British voters, Reddit users, and US partisans exposed to social media comments, only how these may have manifested through their attitudes and behaviors. Accurately measuring a person's emotional state is, in fact, quite complex (Mauss & Robinson, 2009). Self-reports of emotions via survey instruments are known to be quite limiting and unreliable, as they often depend on participants accurately recalling their emotions and sorting them out into discreet units, instead of capturing them as experienced. Measuring individuals' affective responses to a stimulus in an experimental setting can help bypass some of these limitations, but has also been shown to introduce desirability biases. Analyzing text-based social media data, finally, as I did in Chapter 4, offers clear advantages over other techniques but is not without its flaws either (Settle, 2018b). What individuals post on platforms like Facebook or Twitter—from their opinions, to status updates—offer researchers a valuable glimpse into their emotional responses to specific events or messages as they happen, and without intrusion. Yet, it is unclear whether these effusions always reflect users' underlying emotional states. Consider, for example, how social media users readily post a 'crying' emoji online while not necessarily feeling that emotion on a physiological level. Despite these limitations, the advent of new measurements techniques beyond classic measures of arousal or discreet emotions such as electromyography and skin conductance (B. Bakker et al., 2021), and methodological advances in sentimental analysis techniques bode well for further investigations in this area. While the studies presented in this thesis barely began to explore emotions' role as potential mediators between party identification and affective polarization, they provide a clear path for future research. It will be interesting for subsequent research designs to build on the hypotheses and frameworks introduced in this thesis and further investigate the role of emotions in shaping how individuals form political judgements, and how they can be manipulated to mitigate interparty hostility.

Another fundamental question concerns the generalizability of my findings to the wider population. The design of some of my empirical studies has allowed

me to draw robust causal conclusions about the power of political cross-pressures to curb affective polarization, and the power of exposure to in-party threats in comments to drive affective polarization. One limitation of my work as a whole, however, is that it mostly focuses on the most politically interested and engaged members of society: namely, partisans and people who actively participate in online political discussions. More experienced or frequent social media users, such as members of Reddit's r/politics, may behave differently than those who do not go online as often, or who cannot afford regular access. As I have argued in previous chapters, there are plenty of reasons to believe that even casual users and 'lurkers' who do not engage in rancorous political conversations or actively seek political information may still be impacted by the confluence of political stimuli present in our modern social media ecosystems. Nevertheless, it will be important for future studies to explore whether the mechanisms outlined in Chapter 4 and 5 are also found in more representative samples of the population, and to test how they are impacted by frequency of social media use.

Determining whether internet and social media have broader negative consequences for users' social and political attitudes (on not simply on the most engaged ones) is an area of ongoing work (Allcott et al., 2020; Mosquera et al., 2020). At the heart of it lies the question of whether social media effects are persuasive at all, and whether our traditional methodological and theoretical frameworks are well-suited to studying the complex interplay between media exposure and political attitudes in new socio-technological environments (Holbert et al., 2010). While the likelihood that exposure to political information on social media has meaningful and lasting effects on users' attitudes is high, that link is notoriously difficult to establish due to measurement issues. People inhabit complex, multi-modal informational environments and can often shift between multiple platforms (Dubois & Blank, 2018; Dutton et al., 2017). Accurately capturing the content users expose themselves to, disentangling which elements of that content are most pertinent for polarization, and testing their effects are challenging tasks. Much of the work in my thesis, for example, is predicated on the idea that people infer what political 'camps' other users belong in based on what they post, what they say, and how they present themselves. While this assumption is well-supported in the literature, what is less clear, however, is what aspect of the

heap of politically informative content users come across online actually leads to this process of categorization. Today, for example, it is very common for users to express their political allegiance through self-descriptions, emojis or political filters. Yet, we still lack a proper understanding of whether these alternative ways of signaling one's political identity accentuate any of the effects identified throughout my studies. Are political exchanges on Facebook as affectively polarized as they are on Reddit, both of which have their own rules of engagement and unique discussion cultures? And if so, is it attributable to the fact that Facebook users can more easily identify the ideology of their friends and close acquaintances? These remain open questions that future studies could usually explore.

Even assuming that we have accurately captured the most relevant content people see online and identified important mechanisms for how it may impact their behavior, there is also the question of how 'sticky' that information is. In the experimental study of Chapter 5, for example, I measured participants' affective polarization responses immediately after they were exposed to the stimuli. However, it is entirely possible that the polarizing effects of being exposed to a threatening partisan comment vanished shortly after treatment. An appropriate test of the long-term impact of these effects on participants would have necessitated either a repeated measures or delayed post-test design. These designs might not necessarily prove that social media exposure incites a *change* in people's attitudes or leads them to develop new attitudes, as opposed to only reinforcing their pre-existing dispositions. However, such designs could go a long way towards providing a more sophisticated explanation of how social media use impacts people's psychology and behavior over time.

6.4 Lessons and Future Direction

What lessons can be drawn from these studies? And how do these inform nascent efforts to reduce polarization both online and off? In the below section, I reflect on current debates about the role of social media in furthering polarization in society, and offer recommendation on possible interventions that could be tested and rolled out to foster healthier digital environments.

6.4.1 Looking Beyond the Echo Chamber

In recent years, the public conversation around social media's impact on polarization has mostly been dominated by fears around 'echo chambers' and filter bubbles. Social media algorithms and selective exposure, the narrative goes, are tearing citizens apart by pushing them into spaces where they are only served attitude-reinforcing content and only engage with likeminded others. By and large, these concerns are motivated by a Habermasian conception of a healthy public sphere, against which online political discussions are evaluated. Accordingly, citizens should engage in reasoned and civil deliberation across party lines. The danger is that, by focusing on features mostly relevant to this deliberative ideal, and despite the fact that the conditions for it are rarely met online, researchers and technologists alike might ignore other relevant dimensions of online political communication that may be more consequential for polarization. These fears also rest on assumptions about user behavior that are being increasingly challenged in the digital context. Indeed, as I detailed in the literature chapter, a growing body of studies is starting to refute the notion that people either go online to shelter themselves from disagreeable information or pursue social relationships that reinforce their pre-existing political beliefs. So, what does this mean for the future? The studies I presented in this doctoral thesis suggest that there is much to be gained by looking beyond 'echo chambers' as the sole, or indeed the greatest ill, of our digital existence. To solve the puzzle of social media's impact on polarization, future efforts will have to focus not only on the diversity or toxicity of online conversations, but also on examining the nature of people's political encounters with (potentially more vocal, more extreme, and more influential) others online, as well as the psychological consequences of these interactions.

6.4.2 Reducing Affective Polarization Through Targeted Intervention

Secondly, part of the work presented in this thesis underscores the importance of stronger collaborations between researchers and technologists. Today, social media platforms such as Facebook, Instagram, and Twitter act as powerful intermediaries between citizens, political elites, civil society and news producers. Properly elucidating the social and political impacts of these digital platforms requires us to

think deeply about 'how' and 'why' users seek or interact with the content and people that they do; about the specificity of these platforms' affordances and how they shape the nature of these interactions' and the ways in which information circulates. To this end, and at a time of restricted data access, close collaborations between technologists and social scientists are needed more than ever before. Strong partnerships between industry and academia can not only ensure that technologists rigorously think through, anticipate, and avoid the unintended consequences of policy or feature changes that have the potential to impact millions of people.

Findings from my studies also show that there is still plenty of work to be done in identifying exactly what types of messages and engagements have the most negative consequences for social and political outcomes. Partisan language is common online, and so is incivility. Yet, as I have made clear in preceding chapters, not all types of partisan comments carry implications for polarization. The better researchers become at disentangling these complex dynamics, and quantifying their effects, the better technologists will be able to adapt their own policies to combat the issue-at-hand and devise effective interventions.

Shining light on these nuances also means that social media companies may be able to strike a better balance between content moderation and optimization of user experience moving forward. One of the most pressing debates about the power of technology firms today concerns their role in moderating speech online. In recent years, critics have frequently decried social media companies' tendency to over-moderate content that may have deleterious effects on users. These efforts make sense from a business perspective: few users want to navigate or spend time in an online environment that is replete with toxic or harmful speech, or where they might be antagonized by political rivals. This is detrimental to social media platforms' business model as well as their reputations among users and advertisers.

Findings from my studies, however, make me cautiously optimistic about the potential for transparent and collaboratively-designed interventions to reduce affective polarization online. As I showed throughout my thesis, a gesture as simple as breaking away from expectations of hostility by engaging in a positive way with a political outgroup can have depolarizing effects. In my studies, this was specifically the case under conditions of anonymity and when the political

affiliation of users was not immediately apparent. Tools such as bots are already frequently provided to community administrators on social media platforms like Reddit to aid their content moderation efforts. In most cases, their primary function is to remind users of a subreddit's rules and take down any content that violates them because it is uncivil or toxic. However, recent studies have shown that bots can also be leveraged to nudge users' behavior in a more democratically desirable direction through moral appeals, lessening racism and partisan incivility in online comments (Munger, 2020). Moving forward, one could thus imagine that encouraging or surfacing positive cross-partisan engagement might offer a promising way to reduce partisan hostility on social media. These interventions are, of course, no panacea for deep-seated political divisions. Moreover, they require thorough ethical considerations before being tested or deployed at scale. Nonetheless, they present new policy avenues—beyond content take-downs—that could be further explored by researchers and trust and safety teams to promote healthier digital conversations.

6.5 Concluding Remarks

Affective polarization is a complex and multifaceted phenomenon, with worrying social, economic, and political implications. What drives this phenomenon at the individual level? Throughout this thesis, I have shone light on some of the factors that can spur partisans to aggression and animosity, and uncovered several mechanisms through which these outcomes could be reversed. Partisan prejudice, I have argued, is not immutable but shaped by a multitude of factors, including how congruent partisans are with their own political party on policy issues, as well as the nature of their interactions with others online. I hope that these findings, and that the thesis as a whole, serve as a first step towards a better understanding of the individual-level drivers of affective polarization in the digital age, and provide useful guidelines for policy-makers and technologists seeking to address these challenges to social and democratic life.

References

Abelson, R. P. (1959). Modes of Resolution of Belief Dilemmas. *Journal of Conflict Resolution*, *3*(4), 343–352.

Abelson, R. P., Kinder, D. R., Peters, M. D., & Fiske, S. T. (1982). Affective and Semantic Components in Political Person Perception. *Journal of Personality and Social Psychology*, *42*(4), 619–630. https://doi.org/10.1037/0022-3514.42.4.619

Abramowitz, A., & Saunders, K. (2005). Why Can't We All Just Get Along? The Reality of a Polarized America. *The Forum*, *3*(2). https://doi.org/10.2202/1540-8884.1076

Abramowitz, A., & Saunders, K. (2008). Is Polarization a Myth? *Journal of Politics*, *70*(2), 542–555. https://doi.org/10.1017/S0022381608080493

Abramowitz, A., & Webster, S. (2018). Negative Partisanship: Why Americans Dislike Parties But Behave Like Rabid Partisans. *Political Psychology*, *39*(S1), 119–135. https://doi.org/10.1111/pops.12479

Abrams, D., & Hogg, M. A. (1988). Comments on the Motivational Status of Self-Esteem in Social Identity and Intergroup Discrimination. *European Journal of Social Psychology*, *18*(4), 317–334. https://doi.org/10.1002/ejsp.2420180403

Achen, C., & Bartels, L. M. (2017). *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton University Press.

Adamic, L. A., & Glance, N. (2005). The Political Blogosphere and the 2004 U.S. Election. *Proceedings of the 3rd International Workshop on Link Discovery*, 36–43. https://doi.org/10.1145/1134271.1134277

Adams, J., Ezrow, L., & Somer-Topcu, Z. (2011). Is Anybody Listening? Evidence That Voters Do Not Respond to European Parties' Policy Statements During Elections. *American Journal of Political Science*, *55*(2), 370–382. https://doi.org/10.1111/j.1540-5907.2010.00489.x

Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, Ltd.

Ahler, D. J. (2018). The Group Theory of Parties: Identity Politics, Party Stereotypes, and Polarization in the 21st Century. *The Forum*, *16*(1), 3–22. https://doi.org/10.1515/for-2018-0002

Ahler, D. J., & Sood, G. (2018). The Parties in Our Heads: Misperceptions about Party Composition and Their Consequences. *The Journal of Politics*, *80*(3), 964–981. https://doi.org/10.1086/697253

Ahmed, S. (2014). *The Cultural Politics of Emotion*. Edinburgh University Press.

Albright, J. J. (2010). The Multidimensional Nature of Party Competition. *Party Politics*, *16*(6), 699–719. https://doi.org/10.1177/1354068809345856

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The Welfare Effects of Social Media. *American Economic Review*, *110*(3), 629–676. https://doi.org/10.1257/aer.20190658

Allport, G. (1954). *The Nature Of Prejudice*. Addison-Wesley.

Amichai-Hamburger, Y. (2008). The Contact Hypothesis Reconsidered: Interacting via Internet: Theoretical and Practical Aspects. *Psychological Aspects of Cyberspace: Theory, Research, Applications*, *11*(3), 209–227. https://doi.org/10.1017/CBO9780511813740.010

Amira, K., Wright, J. C., & Goya-Tocchetto, D. (2019). In-Group Love Versus Out-Group Hate: Which Is More Important to Partisans and When? *Political Behavior*. https://doi.org/10.1007/s11109-019-09557-6

Anderson, A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "Nasty Effect:" Online Incivility and Risk Perceptions of Emerging Technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373–387. https://doi.org/10.1111/jcc4.12009

Anderson, M., & Auxier, B. (2020, August 19). *55% of U.S. Social Media Users Say They Are 'Worn Out' by Political Posts and Discussions*. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/08/19/55-of-u-s-social-media-users-say-they-are-worn-out-by-political-posts-and-discussions/

Anderson, M., & Quinn, D. (2019, August 8). *46% of Social Media Users Worn out by Political Posts, Discussions*. Pew Research Center. https://www.pewresearch.org/fact-tank/2019/08/08/46-of-u-s-social-media-users-say-they-are-worn-out-by-political-posts-and-discussions/

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.2307/2291629

Ansolabehere, S., & Jones, P. E. (2010). Constituents' Responses to Congressional Roll-Call Voting. *American Journal of Political Science*, *54*(3), 583–597. https://doi.org/10.1111/j.1540-5907.2010.00448.x

Anspach, N. M. (2017). The New Personal Influence: How Our Facebook Friends Influence the News We Read. *Political Communication*, *34*(4), 590–606. https://doi.org/10.1080/10584609.2017.1316329

Armaly, M. T., & Enders, A. M. (2020). The Role of Affective Orientations in Promoting Perceived Polarization. *Political Science Research and Methods*, 1–12. https://doi.org/10.1017/psrm.2020.24

Auxier, B. (2020, July 13). *Activism on Social Media Varies by Race and Ethnicity, Age, Political Party*. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/07/13/activism-on-social-media-varies-by-race-and-ethnicity-age-political-party/

Auxier, B., & Anderson, M. (2021). *Social Media Use in 2021*. Pew Research Center. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/04/PI_2021.04.07_Social-Media-Use_FINAL.pdf

Bafumi, J., & Shapiro, R. Y. (2009). A New Partisan Voter. *Journal of Politics*, *71*(1), 1–24. https://doi.org/10.1017/S0022381608090014

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Fallin Hunzaker, M. B., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to Opposing Views on Social Media Can Increase Political Polarization. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Bakker, B., Schumacher, G., & Rooduijn, M. (2021). Hot Politics? Affective Responses to Political Rhetoric. *American Political Science Review*, *115*(1), 150–164. https://doi.org/10.1017/S0003055420000519

Bakker, R., Jolly, S., & Polk, J. (2020). Multidimensional Incongruence, Political Disaffection, and Support for Anti-Establishment Parties. *Journal of European Public Policy*, *27*(2), 292–309. https://doi.org/10.1080/13501763.2019.1701534

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Baldassarri, D., & Gelman, A. (2008). Partisans Without Constraint: Political Polarization and Trends in American Public Opinion. *American Journal of Sociology*, *114*(2), 408–446. https://doi.org/10.1086/590649

Bankert, A. (2020). Negative and Positive Partisanship in the 2016 U.S. Presidential Elections. *Political Behavior*. https://doi.org/10.1007/s11109-020-09599-1

Bankert, A., Huddy, L., & Rosema, M. (2017). Measuring Partisanship as a Social Identity in Multi-Party Systems. *Political Behavior*, *39*, 103–132. https://doi.org/10.1007/s11109-016-9349-5

Barberá, P. (2020). Social Media, Echo Chambers, and Political Polarization. In N. Persily & J. A. Tucker (Eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform* (pp. 34–56). Cambridge University Press.

Barberá, P., & Alvarez, R. M. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, *23*(1), 76–91. https://doi.org/10.1093/pan/mpu011

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting From Left to Right. *Psychological Science*, *26*(10), 1531–1542. https://doi.org/10.1177/0956797615594620

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R. M., Harwood, J., Rubin, M., & Sibley, C. G. (2012). The Contact Caveat: Negative Contact Predicts Increased Prejudice More Than Positive

Contact Predicts Reduced Prejudice. *Personality and Social Psychology Bulletin*, *38*(12), 1629–1643. https://doi.org/10.1177/0146167212457953

Barnidge, M. (2017). Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings. *Political Communication*, *34*(2), 302–321. https://doi.org/10.1080/10584609.2016.1235639

Bartels, L. M. (2002). Beyond the Running Tally: Partisan Bias in Political Perceptions. *Political Behavior*, *24*(2), 117–150. https://doi.org/10.1023/A:1021226224601

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). *The Pushshift Reddit Dataset*. arXiv. http://arxiv.org/abs/2001.08435

Bennett, W. L. (2012). The Personalization of Politics: Political Identity, Social Media, and Changing Patterns of Participation. *Annals of the American Academy of Political and Social Science*, *644*(1), 20–39. https://doi.org/10.1177/0002716212451428

Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? *Journal of Marketing Research*, *49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(03), 351–368. https://doi.org/10.1093/pan/mpr057

Berz, J. (2020). The Effect of Voters' Economic Perception, Brexit and Campaigns on the Evaluation of Party Leaders Over Time. *British Journal of Politics and International Relations*, *22*(2), 202–219. https://doi.org/10.1177/1369148119886220

Biden, J. (2020, October 6). *Remarks by Vice President Joe Bien in Gettysburg, Pennsylvania [Speech Transcript]*. Biden For President. https://joebiden.com/2020/10/06/remarks-by-vice-president-joe-biden-in-gettysburg-pennsylvania/

Bilewicz, M., & Soral, W. (2020). Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization. *Political Psychology*, *41*(S1), 3–33. https://doi.org/10.1111/pops.12670

Billig, M., & Tajfel, H. (1973). Social Categorization and Similarity in Intergroup Behaviour. *European Journal of Social Psychology*, *3*(1), 27–52. https://doi.org/10.1002/ejsp.2420030103

Bischoff, C. S. (2013). Electorally Unstable by Supply or Demand? An Examination of the Causes of Electoral Volatility in Advanced Industrial Democracies. *Public Choice*, *156*(3), 537–561. https://doi.org/10.1007/s11127-011-9910-z

Bishop, B. (2009). *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart*. Mariner Books.

Blanz, M., Mummendey, A., & Otten, S. (1997). Normative Evaluations and Frequency Expectations Regarding Positive Versus Negative Outcome Allocations Between Groups. *European Journal of Social Psychology*, *27*(2), 165–176. https://doi.org/10.1002/(SICI)1099-0992(199703)27:2<165::AID-EJSP812>3.0.CO;2-3

Bobo, L. (1988). Attitudes Toward the Black Political Movement: Trends, Meaning, and Effects on Racial Policy Preferences. *Social Psychology Quarterly*, *51*(4), 287–302. https://doi.org/10.2307/2786757

Bode, L. (2016). Political News in the News Feed: Learning Politics from Social Media. *Mass Communication and Society*, *19*(1), 24–48. https://doi.org/10.1080/15205436.2015.1045149

Bolsen, T., & Leeper, T. J. (2013). Self-Interest and Attention to News Among Issue Publics. *Political Communication*, *30*(3), 329–348. https://doi.org/10.1080/10584609.2012.737428

Bond, J. R., & Fleisher, Richard. (2000). *Polarized Politics: Congress and the President in a Partisan Era*. CQ Press. https://doi.org/10.5860/choice.38-2423

Boninger, D. S., Krosnick, J. A., & Berent, M. K. (1995). Origins of Attitude Importance: Self-Interest, Social Identification, and Value Relevance. *Journal of Personality and Social Psychology*, *68*(1), 61–80. https://doi.org/10.1037/0022-3514.68.1.61

Borah, P. (2013). Interactions of News Frames and Incivility in the Political Blogosphere: Examining Perceptual Outcomes. *Political Communication*, *30*(3), 456–473. https://doi.org/10.1080/10584609.2012.737426

Bornschier, S. (2010). The New Cultural Divide and the Two-Dimensional Political Space in Western Europe. *West European Politics*, *33*(3), 419–444. https://doi.org/10.1080/01402381003654387

Bougher, L. D. (2017). The Correlates of Discord: Identity, Issue Alignment, and Political Hostility in Polarized America. *Political Behavior*, *39*(3), 731–762. https://doi.org/10.1007/s11109-016-9377-1

Boxell, L., Gentzkow, M., & Shapiro, J. (2020). Cross-Country Trends in Affective Polarization. *NBER Working Paper Series*, *26669*. https://doi.org/10.3386/w26669

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater Internet Use Is Not Associated With Faster Growth in Political Polarization Among Us Demographic Groups. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(40), 10612–10617. https://doi.org/10.1073/pnas.1706588114

boyd, danah. (2010). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In Z. Papacharissi (Ed.), *Networked Self: Identity, Community, and Culture on Social Network Sites* (pp. 39–58). Routledge. https://doi.org/10.1162/dmal.9780262524834.119

Brader, T., & Marcus, G. E. (2013). *Emotion and Political Psychology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199760107.013.0006

Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The Context and Content of Social Identity Threat. In *Social Identity: Context, Commitment, Content* (pp. 35–38). Blackwell Publishers.

Branscombe, N. R., Spears, R., Ellemers, N., & Doosje, B. (2002). Intragroup and Intergroup Evaluation Effects on Group Behavior. *Personality and Social Psychology Bulletin*, *28*(6), 744–753. https://doi.org/10.1177/0146167202289004

Branscombe, N. R., & Wann, D. L. (1994). Collective Self-Esteem Consequences of Outgroup Derogation When a Valued Social Identity Is on Trial. *European Journal of Social Psychology*, *24*(6), 641–657. https://doi.org/10.1002/ejsp.2420240603

Brauer, M., Judd, C. M., & Gliner, M. D. (1995). The Effects of Repeated Expressions on Attitude Polarization During Group Discussions. *Journal*

of Personality and Social Psychology, *68*(6), 1014–1029. https://doi.org/10.1037/0022-3514.68.6.1014

Bremmer, I. (2021, January 16). The U.S. Capitol Riot Was Years in the Making. Here's Why America Is So Divided. *Time*. https://time.com/5929978/the-u-s-capitol-riot-was-years-in-the-making-heres-why-america-is-so-divided/

Brewer, M. B. (1979). In-Group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis. *Psychological Bulletin*, *86*(2), 307–324. https://doi.org/10.1037/0033-2909.86.2.307

Brewer, M. B. (1999). The Psychology of Prejudice: Ingroup Love or Outgroup Hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brewer, M. B. (2001). The Many Faces of Social Identity: Implications for Political Psychology. *Political Psychology*, *22*(1), 115–125. https://doi.org/10.1111/0162-895X.00229

Brewer, M. B., & Kramer, R. M. (1985). The Psychology of Intergroup Attitudes and Behavior. *Annual Review of Psychology*, *36*(1), 219–243. https://doi.org/10.1146/annurev.ps.36.020185.001251

Bright, J. (2018). Explaining the Emergence of Political Fragmentation on Social Media: The Role of Ideology and Extremism. *Journal of Computer-Mediated Communication*, *23*(1), 17–33. https://doi.org/10.2139/ssrn.2839728

Bright, J., Marchal, N., Ganesh, B., & Rudinac, S. (2020). *Echo Chambers Exist! (But They're Full of Opposing Views)*. arXiv. http://arxiv.org/abs/2001.11461

British Election Study. (2019). *Most Volatile British Electorate in Modern Times*. https://www.britishelectionstudy.com/bes-resources/press-release-most-volatile-british-electorate-in-modern-times/

Brown, R. (2010). *Prejudice: Its Social Psychology*. John Wiley & Sons, Ltd.

Brown, W. (1993). Wounded Attachments. *Political Theory*, *21*(3), 390–410.

Bruns, A. (2019). *Are Filter Bubbles Real?* Polity.

Buhl, T. (1999). Positive-Negative Asymmetry in Social Discrimination: Meta-Analytical Evidence. *Group Processes & Intergroup Relations*, *2*(1), 51–58. https://doi.org/10.1177/1368430299021004

Campbell, A., Converse, P., Warren, M., & Stokes, D. (1960). *The American Voter*. University of Chicago Press.

Carlin, R., & Love, G. (2013). The Politics of Interpersonal Trust and Reciprocity: An Experimental Approach. *Political Behavior*, *35*(1), 43–63. https://doi.org/10.1007/s11109-011-9181-x

Carlin, R., & Love, G. (2018). Political Competition, Partisanship and Interpersonal Trust in Electoral Democracies. *British Journal of Political Science*, *48*(1), 115–139. https://doi.org/10.1017/S0007123415000526

Carothers, T., & O'Donohue, A. (2019). *How to Understand the Global Spread of Political Polarization*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2019/10/01/how-to-understand-global-spread-of-political-polarization-pub-79893

Carpini, M. X. D. (2020). Political Communication Research in a Networked World. In B. Foucault Welles & S. González-Bailón (Eds.), *The Oxford Handbook of Networked Communication* (pp. 310–320). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190460518.013.19

Carpini, M. X. D., & Keeter, S. (1996). *What Americans Know about Politics and Why It Matters*. Yale University Press.

Carr, C. T. (2017). Social Media and Intergroup Communication. In *Oxford Research Encyclopedia of Communication*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.460

Carr, C. T., Vitak, J., & McLaughlin, C. (2013). Strength of Social Cues in Online Impression Formation. *Communication Research*, *40*(2), 261–281. https://doi.org/10.1177/0093650211430687

Carsey, T. M., & Layman, G. C. (2006). Changing Sides or Changing Minds? Party Identification and Policy Preferences in the American Electorate. *American Journal of Political Science*, *50*(2), 464–477. https://doi.org/10.1111/j.1540-5907.2006.00196.x

Casalicchio, E. (2021, January 8). Britain's Culture War Extends Beyond Brexit. *Politico*. https://www.politico.eu/article/uk-tories-plan-assault-woke-culture-war/

Cassino, D., & Lodge, M. (2007). The Primacy of Affect in Political Evaluations. In W. R. Neuman, G. E. Marcus, A. Crigler, & M. MacKuen (Eds.), *The Affect Effect: Dynamics of Emotion in Thinking and Political Behavior* (pp. 101–124). University of Chicago Press.

Castle, J. (2019). New Fronts in the Culture Wars? Religion, Partisanship, and Polarization on Religious Liberty and Transgender Rights in the United States. *American Politics Research*, *47*(3), 650–679. https://doi.org/10.1177/1532673X18818169

Chadwick, A. (2013). *The Hybrid Media System*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199759477.001.0001

Chen, G. M. (2017). *Online Incivility and Public Debate*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-56273-5

Chen, G. M., & Lu, S. (2017). Online Political Discourse: Exploring Differences in Effects of Civil and Uncivil Disagreement in News Website Comments. *Journal of Broadcasting and Electronic Media*, *61*(1), 108–125. https://doi.org/10.1080/08838151.2016.1273922

Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We Should Not Get Rid of Incivility Online. *Social Media + Society*, 1–5. https://doi.org/10.1177/2056305119862641

Chen, M. K., & Rohla, R. (2018). The Effect of Partisanship and Political Advertising on Close Family Ties. *Science*, *360*(6392), 1020–1024. https://doi.org/10.1126/science.aaq1433

Chua, A. (2018). *Political Tribes: Group Instinct and the Fate of Nations*. Bloomsbury.

Chua, A., & Rubenfeld, J. (2018, October). The Threat of Tribalism. *The Atlantic*. https://www.theatlantic.com/magazine/archive/2018/10/the-threat-of-tribalism/568342/

Ciuk, D. J., & Yost, B. A. (2016). The Effects of Issue Salience, Elite Influence, and Policy Content on Public Opinion. *Political Communication*, *33*(2), 328–345. https://doi.org/10.1080/10584609.2015.1017629

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are Samples Drawn From Mechanical Turk Valid for Research on Political Ideology? *Research and Politics*, *2*(4), 1–9. https://doi.org/10.1177/2053168015622072

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10.1111/jcom.12104

Cohen, G., & Cohen, S. (2019). Depolarization, Repolarization and Redistributive Ideological Change in Britain, 1983–2016. *British Journal of Political Science*, 1–22. https://doi.org/10.1017/S0007123419000486

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Inc.

Cohn, N. (2014, June 12). Polarization Is Dividing American Society, Not Just Politics. *The New York Times*. https://www.nytimes.com/2014/06/12/upshot/polarization-is-dividing-american-society-not-just-politics.html

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication*, *64*(2), 317–332. https://doi.org/10.1111/jcom.12084

Collins, L. M., Dziak, J. J., & Li, R. (2009). Design of Experiments With Multiple Independent Variables: A Resource Management Perspective on Complete and Reduced Factorial Designs. *Psychological Methods*, *14*(3), 202–224. https://doi.org/10.1037/a0015826

Converse, P. E. (1964). The Nature of Belief Systems in Mass Publics. *Critical Review*, *18*(1), 1–74. https://doi.org/10.1080/08913810608443650

Coser, L. A. (1964). *The Functions of Social Conflict*. Simon and Schuster.

Costello, R., Thomassen, J., & Rosema, M. (2012). European Parliament Elections and Political Representation: Policy Congruence between Voters and Parties. *West European Politics*, *35*(6), 1226–1248. https://doi.org/10.1080/01402382.2012.713744

Costello, R., Toshkov, D., Bos, B., & Krouwel, A. (2020). Congruence Between Voters and Parties: The Role of Party-Level Issue Salience. *European Journal of Political Research*, *60*(1), 92–113. https://doi.org/10.1111/1475-6765.12388

da Silva, M. T. (2013). Online Forums, Audience Participation and Modes of Political Discussion: Readers' Comments on the Brazilian Presidential Election as a Case Study. *Comunicación y Sociedad*, *26*(4), 175–193.

Dahl, R. A. (1981). *Democracy in the United States: Promise and Performance* (4th ed.). Houghton Mifflin Co.

Dahlgreen, W. (2014, July 23). *Britain's Changing Political Spectrum*. YouGov. https://yougov.co.uk/topics/politics/articles-reports/2014/07/23/britains-changing-political-spectrum

Dalton, R. J. (2017). Party Representation Across Multiple Issue Dimensions. *Party Politics*, *23*(6), 609–622. https://doi.org/10.1177/1354068815614515

Dalton, R. J., & Wattenberg, M. P. (2003). *Parties Without Partisans: Political Change in Advanced Industrial Democracies*. Oxford University Press. https://doi.org/10.1093/0199253099.001.0001

de Vreese, C. H., & Neijens, P. (2016). Measuring Media Exposure in a Changing Communications Environment. *Communication Methods and Measures*, *10*(2–3), 69–80. https://doi.org/10.1080/19312458.2016.1150441

de Vries, C. E., & Marks, G. (2012). The Struggle Over Dimensionality: A Note on Theory and Empirics. *European Union Politics*, *13*(2), 185–193. https://doi.org/10.1177/1465116511435712

Dennison, J. (2019). A Review of Public Issue Salience: Concepts, Determinants and Effects on Voting. *Political Studies Review*, *17*(4), 436–446. https://doi.org/10.1177/1478929918819264

Devine, C. J. (2015). Ideological Social Identity: Psychological Attachment to Ideological In-Groups as a Political Phenomenon and a Behavioral Influence. *Political Behavior*, *37*(3), 509–535. https://doi.org/10.1007/s11109-014-9280-6

Diamond, E. P. (2020). The Influence of Identity Salience on Framing Effectiveness: An Experiment. *Political Psychology*, *41*(6), 1133–1150. https://doi.org/10.1111/pops.12669

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's Social Attitudes Become More Polarized? *American Journal of Sociology*, *102*(3), 690–755. https://doi.org/10.1086/230995

Dimock, M., Doherty, C., Kiley, J., & Oates, R. (2014). *Political Polarization in the American Public*. Pew Research Center. https://www.pewresearch.org/politics/wp-content/uploads/sites/4/2014/06/6-12-2014-Political-Polarization-Release.pdf

Dimock, M., & Wike, R. (2020, November 13). *America Is Exceptional in the Nature of Its Political Divide*. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/11/13/america-is-exceptional-in-the-nature-of-its-political-divide/

Doherty, C., Kiley, J., & Asheer, N. (2019a). *Partisan Antipathy: More Intense, More Personal*. Pew Research Center.

https://www.pewresearch.org/politics/wp-content/uploads/sites/4/2019/10/10-10-19-Parties-report.pdf

Doherty, C., Kiley, J., & Asheer, N. (2019b). *In a Politically Polarized Era, Sharp Divides in Both Partisan Coalitions*. Pew Research Center. https://www.pewresearch.org/politics/wp-content/uploads/sites/4/2019/12/PP_2019.12.17_Political-Values_FINAL.pdf

Doherty, C., Kiley, J., & Jameson, B. (2016). *Partisanship and Political Animosity in 2016*. Pew Research Center. https://www.pewresearch.org/politics/wp-content/uploads/sites/4/2016/06/06-22-16-Partisanship-and-animosity-release.pdf

Domke, D., Watts, M. D., Shah, D. V., & Fan, D. P. (1999). The Politics of Conservative Elites and the "Liberal Media" Argument. *Journal of Communication*, *49*(4), 35–58. https://doi.org/10.1111/j.1460-2466.1999.tb02816.x

Dommett, K. (2016). Ideological Quietism? Ideology and Party Politics in Britain. *Political Studies*, *64*(1), 106–122. https://doi.org/10.1111/1467-9248.12160

Dommett, K., & Temple, L. (2019). *What People Want To See In Parties Today*. Involve. https://www.involve.org.uk/sites/default/files/uploads/final_-_what_people_want_from_parties_today.pdf

Douglas, K. M., & McGarty, C. (2001). Identifiability and Self-Presentation: Computer-Mediated Communication and Intergroup Interaction. *British Journal of Social Psychology*, *40*(3), 399–416. https://doi.org/10.1348/014466601164894

Döveling, K., Harju, A. A., & Sommer, D. (2018). From Mediatized Emotion to Digital Affect Cultures: New Technologies and Global Flows of Emotion. *Social Media + Society*, 1–11. https://doi.org/10.1177/2056305117743141

Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, *65*(2), 135–150. https://doi.org/10.2307/1827369

Druckman, J. N., Gubitz, S. R., Levendusky, M. S., & Lloyd, A. M. (2019). How Incivility on Partisan Media (De)polarizes the Electorate. *Journal of Politics*, *81*(1), 291–295. https://doi.org/10.1086/699912

Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2020). How Affective Polarization Shapes Americans' Political Beliefs: A Study of Response to the COVID-19 Pandemic. *Journal of Experimental Political Science*, 1–12. https://doi.org/10.1017/XPS.2020.28

Druckman, J. N., & Levendusky, M. S. (2019). What Do We Measure When We Measure Affective Polarization? *Public Opinion Quarterly*, *83*(1), 114–122. https://doi.org/10.1093/poq/nfz003

Drutman, L. (2020, October 5). *How Hatred Came to Dominate American Politics*. FiveThirtyEight. https://fivethirtyeight.com/features/how-hatred-negative-partisanship-came-to-dominate-american-politics/

Dubois, E., & Blank, G. (2018). The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media. *Information, Communication & Society*, *21*(5), 729–745. https://doi.org/10.1080/1369118X.2018.1428656

Duffy, B., Hewlett, K., McCrae, J., & Hall, J. (2019). *Divided Britain? Polarisation and Fragmentation Trends in the UK*. The Policy Institute, King's College

London. https://www.kcl.ac.uk/policy-institute/assets/divided-britain.pdf

Duncan, L. E., & Stewart, A. J. (2007). Personal Political Salience: The Role of Personality in Collective Identity and Action. *Political Psychology*, *28*(2), 143–164. https://doi.org/10.1111/j.1467-9221.2007.00560.x

Dutton, W. H., Reisdorf, B. C., Dubois, E., & Blank, G. (2017). *Search and Politics: The Uses and Impacts of Search in Britain, France, Germany, Italy, Poland, Spain, and the United States*. SSRN. https://doi.org/10.2139/ssrn.2960697

Dvir-Gvirsman, S. (2017). Media Audience Homophily: Partisan Websites, Audience Identity and Polarization Processes. *New Media and Society*, *19*(7), 1072–1091. https://doi.org/10.1177/1461444815625945

Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data. *SAGE Open*, 1–21. https://doi.org/10.1177/2158244019832705

Eberl, J.-M., Tolochko, P., Jost, P., Heidenreich, T., & Boomgaarden, H. G. (2020). What's in a Post? How Sentiment and Issue Salience Affect Users' Emotional Reactions on Facebook. *Journal of Information Technology & Politics*, *17*(1), 48–65. https://doi.org/10.1080/19331681.2019.1710318

Ellis, C., & Stimson, J. A. (2012). *Ideology in America*. Cambridge University Press. https://doi.org/10.1017/CBO9781139094009

Enders, A. M., & Lupton, R. N. (2020). Value Extremity Contributes to Affective Polarization in the US. *Political Science Research and Methods*, 1–10. https://doi.org/10.1017/psrm.2020.27

Erdfelder, E., Faul, F., & Buchner, A. (1996). Gpower: A General Power Analysis Program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Esteban, J.-M., & Ray, D. (1994). On the Measurement of Polarization. *Econometrica*, *62*(4), 819–851. https://doi.org/10.2307/2951734

Faris, R. M., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., Benkler, Y., Roberts, H., Faris, R. M., Etling, B., Zuckerman, E., & Bourassa, N. (2017). *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. Berkman Klein Center for Internet & Society. https://dash.harvard.edu/bitstream/handle/1/33759251/2017-08_electionReport_0.pdf?sequence=9&isAllowed=y

Ferrara, E., & Yang, Z. (2015). Measuring Emotional Contagion in Social Media. *PLOS ONE*, *10*(11), e0142390. https://doi.org/10.1371/journal.pone.0142390

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Fieldhouse, E., Green, J., Evans, G., Mellon, J., & Prosser, C. (2020). *British Election Study Internet Panel Wave 17*. British Election Study. https://www.britishelectionstudy.com/data-object/wave-17-of-the-2014-2023-british-election-study-internet-panel/

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political

Sectarianism in America. *Science*, *370*(6516), 533–536. https://doi.org/10.1126/science.abe1715

Fiorina, M. (1981). *Retrospective Voting in American National Elections*. Yale University Press.

Fiorina, M., & Abrams, S. J. (2008). Political Polarization in the American Public. *Annual Review of Political Science*, *11*(1), 563–588. https://doi.org/10.1146/annurev.polisci.11.053106.153836

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley.

Fletcher, R., & Nielsen, R. K. (2017). Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. *Journal of Communication*, *67*(4), 476–498. https://doi.org/10.1111/jcom.12315

Fletcher, R., & Nielsen, R. K. (2018). Are People Incidentally Exposed to News on Social Media? A Comparative Analysis. *New Media and Society*, *20*(7), 2450–2468. https://doi.org/10.1177/1461444817724170

Fournier, P., Blais, A., Nadeau, R., Gidengil, E., & Nevitte, N. (2003). Issue Importance and Performance Voting. *Political Behavior*, *25*(1), 51–67. https://doi.org/10.1023/A:1022952311518

Fowler, A. (2020). Partisan Intoxication or Policy Voting? *Quarterly Journal of Political Science*, *15*(2), 141–179. https://doi.org/10.1561/100.00018027a

Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, *35*(4), 665–668. https://doi.org/10.1080/10584609.2018.1477506

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The Common Ingroup Identity Model: Recategorization and the Reduction of Intergroup Bias. *European Review of Social Psychology*, *4*(1), 1–26. https://doi.org/10.1080/14792779343000004

Gaines, B. J., Kuklinski, J. H., Quirk, P. J., Peyton, B., & Verkuilen, J. (2007). Same Facts, Different Interpretations: Partisan Motivation and Opinion on Iraq. *The Journal of Politics*, *69*(4), 957–974. https://doi.org/10.1111/j.1468-2508.2007.00601.x

Garrett, K. (2009). Echo Chambers Online? Politically Motivated Selective Exposure Among Internet News Users. *Journal of Computer-Mediated Communication*, *14*(2), 265–285. https://doi.org/10.1111/j.1083-6101.2009.01440.x

Garrett, K., Gvirsman, S. D., Johnson, B., Tsfati, Y., Neo, R., & Dal, A. (2014). Implications of Pro- and Counterattitudinal Information Exposure for Affective Polarization. *Human Communication Research*, *40*(3), 309–332. https://doi.org/10.1111/hcre.12028

Garrett, K., Weeks, B., & Neo, R. (2016). Driving a Wedge Between Evidence and Beliefs: How Online Ideological News Exposure Promotes Political Misperceptions. *Journal of Computer-Mediated Communication*, *21*(5), 331–348. https://doi.org/10.1111/jcc4.12164

Gentzkow, M., & Shapiro, J. M. (2011). Ideological Segregation Online and Offline. *Quarterly Journal of Economics*, *126*(4), 1799–1839. https://doi.org/10.1093/qje/qjr044

Gerbaudo, P. (2012). *Tweets and the Streets: Social Media and Contemporary Activism*. Pluto Press. https://doi.org/10.2307/j.ctt183pdzs

Gervais, B. T. (2015). Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-Based Experiment. *Journal of Information Technology and Politics*, *12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416

Gervais, B. T. (2016). More than Mimicry? The Role of Anger in Uncivil Reactions to Elite Political Incivility. *International Journal of Public Opinion Research*, *29*(3), 384–405. https://doi.org/10.1093/ijpor/edw010

Gervais, B. T. (2019). Rousing the Partisan Combatant: Elite Incivility, Anger, and Antideliberative Attitudes. *Political Psychology*, *40*(3), 637–655. https://doi.org/10.1111/pops.12532

Gidron, N. (2020). Many Ways to be Right: Cross-Pressured Voters in Western Europe. *British Journal of Political Science*, 1–16. https://doi.org/10.1017/S0007123420000228

Gidron, N., Adams, J., & Horne, W. (2019a). *How Ideology, Economics and Institutions Shape Affective Polarization in Democratic Polities*. Center for European Studies Harvard. https://ces.fas.harvard.edu/uploads/files/events/GAH-Affective-Polarization-in-Democratic-Polities.pdf

Gidron, N., Adams, J., & Horne, W. (2019b). Toward a Comparative Research Agenda on Affective Polarization in Mass Publics. *APSA-CP Newsletter*, *29*(1), 30–36. https://scholar.harvard.edu/files/gidron/files/gidron_et_al._2019_cp_newsletter.pdf

Gidron, N., Adams, J., & Horne, W. (2020). *American Affective Polarization in Comparative Perspective*. Cambridge University Press. https://doi.org/10.1017/9781108914123

Gift, K., & Gift, T. (2015). Does Politics Influence Hiring? Evidence from a Randomized Experiment. *Political Behavior*, *37*(3), 653–675. https://doi.org/10.1007/s11109-014-9286-0

Giger, N., & Lefkofridi, Z. (2014). Salience-Based Congruence Between Parties & Their Voters: The Swiss Case. *Swiss Political Science Review*, *20*(2), 287–304. https://doi.org/10.1111/spsr.12069

Gil de Zúñiga, H., Correa, T., & Valenzuela, S. (2012). Selective Exposure to Cable News and Immigration in the U.S.: The Relationship Between FOX News, CNN, and Attitudes Toward Mexican Immigrants. *Journal of Broadcasting & Electronic Media*, *56*(4), 597–615. https://doi.org/10.1080/08838151.2012.732138

Gimpel, J. G., & Hui, I. S. (2015). Seeking Politically Compatible Neighbors? The Role of Neighborhood Partisan Composition in Residential Sorting. *Political Geography*, *48*, 130–142. https://doi.org/10.1016/j.polgeo.2014.11.003

Goggin, S. N., Henderson, J. A., & Theodoridis, A. G. (2020). What Goes with Red and Blue? Mapping Partisan and Ideological Associations in the Minds of Voters. *Political Behavior*, *42*(4), 985–1013. https://doi.org/10.1007/s11109-018-09525-6

Golec de Zavala, A. (2011). Collective Narcissism and Intergroup Hostility: The Dark Side of 'In-Group Love'. *Social and Personality Psychology Compass*, *5*(6), 309–320. https://doi.org/10.1111/j.1751-9004.2011.00351.x

González, R., Manzi, J., Saiz, J. L., Brewer, M., De Tezanos-Pinto, P., Torres, D., Aravena, M. T., & Aldunate, N. (2008). Interparty Attitudes in Chile: Coalitions as Superordinate Social Identities. *Political Psychology*, *29*(1), 93–118. https://doi.org/10.1111/j.1467-9221.2007.00614.x

Goren, P. (2002). Character Weakness, Partisan Bias, and Presidential Evaluation. *American Journal of Political Science*, *46*(3), 627–641. https://doi.org/10.2307/3088404

Green, D., Palmquist, B., & Schickler, E. (2002). *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.

Greene, S. (2004). Social Identity Theory and Party Identification. *Social Science Quarterly*, *85*(1), 136–153. https://doi.org/10.1111/j.0038-4941.2004.08501010.x

Groenendyk, E. (2013). *Competing Motives in the Partisan Mind: How Loyalty and Responsiveness Shape Party Identification and Democracy*. Oxford University Press.

Garry, J. (2007) Making 'Party Identification' More Versatile: Operationalising the Concept for the Multiparty Setting. *Electoral Studies*, *26*(2), 346-358. https://doi.org/10.1016/j.electstud.2006.07.003

Haas, I. J., & Cunningham, W. A. (2014). The Uncertainty Paradox: Perceived Threat Moderates the Effect of Uncertainty on Political Tolerance. *Political Psychology*, *35*(2), 291–302. https://doi.org/10.1111/pops.12035

Haidt, J., Graham, J., & Joseph, C. (2009). Above and Below Left-Right: Ideological Narratives and Moral Foundations. *Psychological Inquiry*, *20*(2–3), 110–119. https://doi.org/10.1080/10478400903028573

Halberstam, Y., & Knight, B. (2016). Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence From Twitter. *Journal of Public Economics*, *143*(1), 73–88. https://doi.org/10.1016/j.jpubeco.2016.08.011

Halevy, N., Weisel, O., & Bornstein, G. (2012). "In-Group Love" and "Out-Group Hate" in Repeated Interaction Between Groups. *Journal of Behavioral Decision Making*, *25*(2), 188–195. https://doi.org/10.1002/bdm.726

Hamilton, D. (2015). *Cognitive Processes in Stereotyping and Intergroup Behavior*. Taylor & Francis Group.

Hamilton, D., & Trolier, T. (1986). Stereotypes and Stereotyping: An Overview of the Cognitive Approach. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, Discrimination, and Racism* (pp. 127–163). Academic Press Inc.

Han, S. H., & Brazeal, L. A. M. (2015). Playing Nice: Modeling Civility in Online Political Discussions. *Communication Research Reports*, *32*(1), 20–28. https://doi.org/10.1080/08824096.2014.989971

Hansard Society. (2019). *Audit of Political Engagement: Public Opinion on Brexit and Politics in 2019*. Hansard Society. https://www.hansardsociety.org.uk/publications/reports/audit-of-political-engagement-16

Hanson, K., O'dwyer, E., & Lyons, E. (2019). The Individual and the Nation: A Qualitative Analysis of US Liberal and Conservative Identity Content. *Journal of Social and Political Psychology*, *7*(1), 378–401. https://doi.org/10.5964/jspp.v7i1.1062

Hargittai, E., Gallo, J., & Kane, M. (2008). Cross-Ideological Discussions Among Conservative and Liberal Bloggers. *Public Choice*, *134*(1), 67–86. https://doi.org/10.1007/s11127-007-9201-x

Harris, R. B., & Paradice, D. (2007). An Investigation of the Computer-Mediated Communication of Emotions. *Journal of Applied Sciences Research*, *3*(12), 2081–2090.

Harwood, J., Giles, H., & Palomares, N. (2005). Intergroup Theory and Communication Processes. In H. Giles & J. Harwood (Eds.), *Intergroup Communication: Multiple Perspectives* (pp. 1–17). Peter Lang Publishing.

Harteveld, E. (2021). Fragmented Foes: Affective Polarization in the Multiparty Context of the Netherlands. *Electoral Studies,* 71. https://doi.org/10.1016/j.electstud.2021.102332

Hasell, A., & Weeks, B. E. (2016). Partisan Provocation: The Role of Partisan News Use and Emotional Responses in Political Information Sharing in Social Media. *Human Communication Research*, *42*(4), 641–661. https://doi.org/10.1111/hcre.12092

Hayward, L. E., Tropp, L. R., Hornsey, M. J., & Barlow, F. K. (2017). Toward a Comprehensive Understanding of Intergroup Contact. *Personality and Social Psychology Bulletin*, *43*(3), 347–364. https://doi.org/10.1177/0146167216685291

Heider, F. (1958). *The Psychology of Interpersonal Relations*. Lawrence Erlbaum Associates, Inc.

Helbling, M., & Jungkunz, S. (2020). Social Divides in the Age of Globalization. *West European Politics*, *43*(6), 1187–1210. https://doi.org/10.1080/01402382.2019.1674578

Herek, G. M., Norton, A. T., Allen, T. J., & Sims, C. L. (2010). Demographic, Psychological, and Social Characteristics of Self-Identified Lesbian, Gay, and Bisexual Adults in a US Probability Sample. *Sexuality Research and Social Policy*, *7*(3), 176–200. https://doi.org/10.1007/s13178-010-0017-y

Hern, A., & Sabbagh, D. (2018, March 27). EU Referendum Won Through Fraud, Whistleblower Tells MPs. *The Guardian*. https://www.theguardian.com/uk-news/2018/mar/27/brexit-groups-had-common-plan-to-avoid-election-spending-laws-says-wylie

Hersh, E. D., & Goldenberg, M. N. (2016). Democratic and Republican Physicians Provide Different Care on Politicized Health Issues. *Proceedings of the National Academy of Sciences*, *113*(42), 11811–11816. https://doi.org/10.1073/pnas.1606609113

Hetherington, K. (1996). Identity Formation, Space and Social Centrality. *Theory, Culture & Society*, *13*(4), 33–52. https://doi.org/10.1177/0263276496013004002

Hetherington, M. J., & Rudolph, T. J. (2015). *Why Washington Won't Work: Polarization, Political Trust, and the Governing Crisis*. University of Chicago Press.

Hewstone, M., Lolliot, S., Swart, H., Myers, E., Voci, A., Al Ramiah, A., & Cairns, E. (2014). Intergroup Contact and Intergroup Conflict. *Peace and Conflict: Journal of Peace Psychology*, *20*(1), 39–53. https://doi.org/10.1037/a0035582

Hill, M. L. (2018). "Thank You, Black Twitter": State Violence, Digital Counterpublics, and Pedagogies of Resistance. *Urban Education*, *53*(2), 286–302. https://doi.org/10.1177/0042085917747124

Hillmann, R., & Trier, M. (2012). Sentiment Polarization and Balance among Users in Online Social Networks. *AMCIS 2012 Proceedings*, *10*. https://core.ac.uk/download/pdf/301363564.pdf

Hillygus, S., & Shields, T. (2009). *The Persuadable Voter: Wedge Issues in Presidential Campaigns*. Princeton University Press.

Hinck, A. S., & Carr, C. T. (2020). Advancing a Dual-Process Model to Explain Interpersonal Versus Intergroup Communication in Social Media. *Communication Theory*. https://doi.org/10.1093/ct/qtaa012

Hmielowski, J. D., Hutchens, M. J., & Cicchirillo, V. J. (2014). Living in an Age of Online Incivility: Examining the Conditional Indirect Effects of Online Discussion on Political Flaming. *Information Communication and Society*, *17*(10), 1196–1211. https://doi.org/10.1080/1369118X.2014.899609

Hobolt, S. B., Leeper, T. J., & Tilley, J. (2020). Divided by the Vote: Affective Polarization in the Wake of the Brexit Referendum. *British Journal of Political Science*, 1–18. https://doi.org/10.1017/S0007123420000125

Hobolt, S. B., & Rodon, T. (2020). Cross-Cutting Issues and Electoral Choice. EU Issue Voting in the Aftermath of the Brexit Referendum. *Journal of European Public Policy*, *27*(2), 227–245. https://doi.org/10.1080/13501763.2019.1701535

Hobolt, S. B., & Tilley, J. (2016). Fleeing the Centre: The Rise of Challenger Parties in the Aftermath of the Euro Crisis. *West European Politics*, *39*(5), 971–991. https://doi.org/10.1080/01402382.2016.1181871

Hobolt, S. B., Tilley, J., Fieldhouse, E., Surridge, P., Jennings, W., Duffy, B., Henderson, A., Goodwin, M., Begum, N., Martin, N., McLaren, L., & Curtice, J. (2021). *Brexit and Beyond: Public Opinion*. UK in a Changing

Europe. https://ukandeu.ac.uk/wp-content/uploads/2021/02/87864-Brexit-and-Beyond-Public-Opinion.pdf

Hogg, M. A. (2003). Intergroup Relations. In J. Delamater (Ed.), *Handbook of Social Psychology* (pp. 479–501). Springer.

Hogg, M. A., Abrams, D., Otten, S., & Hinkle, S. (2004). The Social Identity Perspective: Intergroup Relations, Self-Conception, and Small Groups. *Small Group Research*, *35*(3), 246–276. https://doi.org/10.1177/1046496404263424

Hogg, M. A., Turner, P. J., & Smith, P. M. (1984). Failure and Defeat as Determinants of Group Cohesiveness. *British Journal of Social Psychology*, *23*(2), 97–111. https://doi.org/10.1111/j.2044-8309.1984.tb00619.x

Holbert, R. L., Garrett, K., Gleason, L. S., Garrett, R. K., & Gleason, L. S. (2010). A New Era of Minimal Effects? A Response to Bennett and Iyengar. *Journal of Communication*, *60*(1), 15–34. https://doi.org/10.1111/j.1460-2466.2009.01470.x

Hong, G. (2015). Explaining Vote Switching to Niche Parties in the 2009 European Parliament Elections. *European Union Politics*, *16*(4), 514–535. https://doi.org/10.1177/1465116515586428

Hornik, J., Shaanan Satchi, R., Cesareo, L., & Pastore, A. (2015). Information Dissemination via Electronic Word-of-Mouth: Good News Travels Fast, Bad News Travels Faster! *Computers in Human Behavior*, *45*, 273–280. https://doi.org/10.1016/j.chb.2014.11.008

Horwitz, S. N., & Nir, L. (2015). How Politics-News Parallelism Invigorates Partisanship Strength. *International Political Science Review*, *36*(2), 153–167. https://doi.org/10.1177/0192512113516900

Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). *The IRA, Social Media and Political Polarization in the United States, 2012-2018*. Computational Propaganda Project. https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/The-IRA-Social-Media-and-Political-Polarization.pdf

Howard, P. N., & Hussain, M. M. (2011). The Role of Digital Media. *Journal of Democracy*, *22*(3), 35–48. https://doi.org/10.1353/jod.2011.0041

Huber, G., & Malhotra, N. (2017). Political Homophily in Social Relationships: Evidence From Online Dating Behavior. *Journal of Politics*, *79*(1), 269–283. https://doi.org/10.1086/687533

Huddy, L., & Bankert, A. (2017). Political Partisanship as a Social Identity. In *Oxford Research Encyclopedia of Politics*. https://doi.org/10.1093/acrefore/9780190228637.013.250

Huddy, L., Bankert, A., & Davies, C. (2018). Expressive Versus Instrumental Partisanship in Multiparty European Systems. *Political Psychology*, *39*(1), 173–199. https://doi.org/10.1111/pops.12482

Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity. *American Political Science Review*, *109*(1), 1–17. https://doi.org/10.1017/S0003055414000604

Huddy, L., & Yair, O. (2021). Reducing Affective Polarization: Warm Group Relations or Policy Compromise? *Political Psychology*, *42*(2), 291–309. https://doi.org/10.1111/pops.12699

Hunt, E. (2017, April 3). Twitter Drops "Egg" Avatar in Attempt to Break Association With Internet Trolls. *The Guardian*.

https://www.theguardian.com/technology/2017/apr/03/twitter-drops-egg-avatar-in-attempt-to-break-association-with-internet-trolls

Hutchens, M. J., Hmielowski, J. D., & Beam, M. A. (2019). Reinforcing Spirals of Political Discussion and Affective Polarization. *Communication Monographs*, *86*(3), 357–376. https://doi.org/10.1080/03637751.2019.1575255

Igielnik, R., & Parker, K. (2019). *Most Americans Say the Current Economy Is Helping the Rich, Hurting the Poor and Middle Class*. Pew Research Center. https://www.pewresearch.org/social-trends/wp-content/uploads/sites/3/2019/12/SDT.12.11.19_economic.recovery.REPORT_forrelease.pdf

Isenberg, D. J. (1986). Group Polarization. A Critical Review and Meta-Analysis. *Journal of Personality and Social Psychology*, *50*(6), 1141–1151. https://doi.org/10.1037/0022-3514.50.6.1141

Iyengar, S., & Hahn, K. S. (2009). Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication*, *59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Iyengar, S., Konitzer, T., & Tedin, K. (2018). The Home as a Political Fortress: Family Agreement in an Era of Polarization. *The Journal of Politics*, *80*(4), 1326–1338. https://doi.org/10.1086/698929

Iyengar, S., & Krupenkin, M. (2018). The Strengthening of Partisan Affect. *Political Psychology*, *39*(S1), 201–218. https://doi.org/10.1111/pops.12487

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the

United States. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly*, *76*(3), 405–431. https://doi.org/10.1093/poq/nfs038

Iyengar, S., & Westwood, S. J. (2015). Fear and Loathing across Party lines: New Evidence on Group Polarization. *American Journal of Political Science*, *59*(3), 690–707. https://doi.org/10.1111/ajps.12152

Iyer, A., & Leach, C. W. (2008). Emotion in Inter-group Relations. *European Review of Social Psychology*, *19*(1), 86–125. https://doi.org/10.1080/10463280802079738

Jamieson, K. H., & Capella, J. (2010). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford University Press.

Jamieson, K. H., & Kenski, K. (2017). Political Communication: Then, Now and Beyond. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford Handbook of Political Communication* (pp. 4–9). Oxford University Press.

Jerit, J. (2008). Issue Framing and Engagement: Rhetorical Strategy in Public Policy Debates. *Political Behavior*, *30*(1), 1–24. https://doi.org/10.1007/s11109-007-9041-x

Jessee, S., & Malhotra, N. (2013). Public (Mis)Perceptions of Supreme Court Ideology. *Public Opinion Quarterly*, *77*(2), 619–634. https://doi.org/10.1093/poq/nft017

Jilke, S. (2018). Citizen Satisfaction Under Changing Political Leadership: The Role of Partisan Motivated Reasoning. *Governance*, *31*(3), 515–533. https://doi.org/10.1111/gove.12317

Jimmy Kimmel Live. (2016, July 28). *Democrats vs. Republicans [Video]*. YouTube. https://www.youtube.com/watch?v=4HyXiu927WU

John, N. A., & Dvir-Gvirsman, S. (2015). I Don't Like You Any More: Facebook Unfriending by Israelis During the Israel-Gaza Conflict of 2014. *Journal of Communication*, *65*(6), 953–974. https://doi.org/10.1111/jcom.12188

Jones, J. (2019, January 16). *Trump Job Approval Sets New Record for Polarization*. Gallup. https://news.gallup.com/poll/245996/trump-job-approval-sets-new-record-polarization.aspx

Jost, J. T., Stern, C., Rule, N. O., & Sterling, J. (2017). The Politics of Fear: Is There an Ideological Asymmetry in Existential Motivation? *Social Cognition*, *35*(4), 324–353. https://doi.org/10.1521/soco.2017.35.4.324

Joyce, E., & Kraut, R. E. (2006). Predicting Continued Participation in Newsgroups. *Journal of Computer-Mediated Communication*, *11*(3), 723–747. https://doi.org/10.1111/j.1083-6101.2006.00033.x

Jungherr, A., Rivero, G., & Gayo-Avello, D. (2020). *Retooling Politics: How Digital Media are Shaping Democracy*. Cambridge University Press.

Jungherr, A., Schoen, H., Posegga, O., & Jürgens, P. (2017). Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support. *Social Science Computer Review*, *35*(3), 336–356. https://doi.org/10.1177/0894439316631043

Kalmoe, N. P., & Mason, L. (2019). *Lethal Mass Partisanship: Prevalence, Correlates, & Electoral Contingencies*. NCAPSA American Politics Meeting. https://www.dannyhayes.org/uploads/6/9/8/5/69858539/kalmoe___mason_ncapsa_2019_-_lethal_partisanship_-_final_lmedit.pdf

Kerman, J. (2011). Neutral Noninformative and Informative Conjugate Beta and Gamma Prior Distributions. *Electronic Journal of Statistics*, *5*, 1450–1470. https://doi.org/10.1214/11-EJS648

Kessler, T., & Hollbach, S. (2005). Group-Based Emotions as Determinants of Ingroup Identification. *Journal of Experimental Social Psychology*, *41*(6), 677–685. https://doi.org/10.1016/j.jesp.2005.01.001

Kim, N., & Wojcieszak, M. (2018). Intergroup Contact Through Online Comments: Effects of Direct and Extended Contact on Outgroup Attitudes. *Computers in Human Behavior*, *81*, 63–72. https://doi.org/10.1016/j.chb.2017.11.013

Kim, Y., & Kim, Y. (2019). Incivility on Facebook and Political Polarization: The Mediating Role of Seeking Further Comments and Negative Emotion. *Computers in Human Behavior*, *99*, 219–227. https://doi.org/10.1016/j.chb.2019.05.022

Klar, S. (2018). When Common Identities Decrease Trust: An Experimental Study of Partisan Women. *American Journal of Political Science*, *62*(3), 610–622. https://doi.org/10.1111/ajps.12366

Klar, S., & Krupnikov, Y. (2016). *Independent Politics: How American Disdain for Parties Leads to Political Inaction*. Cambridge University Press.

Klar, S., Krupnikov, Y., & Ryan, J. B. (2018). Affective Polarization or Partisan Disdain? Untangling a Dislike for the Opposing Party From a Dislike of Partisanship. *Public Opinion Quarterly*, *82*(2), 379–390. https://doi.org/10.1093/poq/nfy014

Klein, E. (2020). *Why We're Polarized*. Simon and Schuster.

Knapp, M. L., & Daly, J. A. (John A. (2011). *The Sage Handbook of Interpersonal Communication*. SAGE Publications.

Knight, A. P., & Eisenkraft, N. (2015). Positive Is Usually Good, Negative Is Not Always Bad: The Effects of Group Affect on Social Integration and Task Performance. *Journal of Applied Psychology*, *100*(4), 1214–1227. https://doi.org/10.1037/apl0000006

Knobloch-Westerwick, S., & Kleinman, S. B. (2012). Preelection Selective Exposure: Confirmation Bias versus Informational Utility. *Communication Research*, *39*(2), 170–193. https://doi.org/10.1177/0093650211400597

Knobloch-Westerwick, S., & Meng, J. (2009). Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information. *Communication Research*, *36*(3), 426–448. https://doi.org/10.1177/0093650209333030

Knobloch-Westerwick, S., Westerwick, A., & Johnson, B. K. (2015). Selective Exposure in the Communication Technology Context. In *The Handbook of the Psychology of Communication Technology* (pp. 405–424). Wiley. https://doi.org/10.1002/9781118426456.ch18

Knudsen, E. (2021). Affective Polarization in Multiparty Systems? Comparing Affective Polarization Towards Voters and Parties in Norway and the United States. *Scandinavian Political Studies*, *44*(1), 34–44. https://doi.org/10.1111/1467-9477.12186

Kohut, A., Doherty, Dimock, M., & Keeter, S. (2012, April 11). *What the Public Knows about the Political Parties*. Pew Research Center. https://www.pewresearch.org/politics/2012/04/11/what-the-public-knows-about-the-political-parties/

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Krämer, N. C., Feurstein, M., Kluck, J. P., Meier, Y., Rother, M., & Winter, S. (2017). Beware of Selfies: The Impact of Photo Type on Impression Formation Based on Social Networking Profiles. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00188

Kriesi, H. (2010). Restructuration of Partisan Politics and the Emergence of a New Cleavage Based on Values. *West European Politics*, *33*(3), 673–685. https://doi.org/10.1080/01402381003654726

Krosnick, J. (1988). The Role of Attitude Importance in Social Evaluation: A Study of Policy Preferences, Presidential Candidate Evaluations, and Voting Behavior. *Journal of Personality and Social Psychology*, *55*(2), 196–210. https://doi.org/10.1037//0022-3514.55.2.196

Krosnick, J. (1990). Government Policy and Citizen Passion: A Study of Issue Publics in Contemporary America. *Political Behavior*, *12*(1), 59–92. https://doi.org/10.1007/BF00992332

Kugler, M., Jost, J. T., & Noorbaloochi, S. (2014). Another Look at Moral Foundations Theory: Do Authoritarianism and Social Dominance Orientation Explain Liberal-Conservative Differences in "Moral" Intuitions? *Social Justice Research*, *27*(4), 413–431. https://doi.org/10.1007/s11211-014-0223-5

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lakoff, G. (2002). *Moral Politics*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226471006.001.0001

Lau, R. R., Andersen, D. J., Ditonto, T. M., Kleinberg, M. S., & Redlawsk, D. P. (2017). Effect of Media Environment Diversity and Advertising Tone on Information Search, Selective Exposure, and Affective Polarization. *Political Behavior*, *39*(1), 231–255. https://doi.org/10.1007/s11109-016-9354-8

Lauka, A., McCoy, J., & Firat, R. B. (2018). Mass Partisan Polarization: Measuring a Relational Concept. *American Behavioral Scientist*, *62*(1), 107–126. https://doi.org/10.1177/0002764218759581

Lavine, H. G., Johnston, C. D., & Steenbergen, M. R. (2013). *The Ambivalent Partisan: How Critical Loyalty Promotes Democracy*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199772759.001.0001

Lavine, H. G., Johnston, C. D., Steenbergen, M. R., Lavine, H. G., Johnston, C. D., & Steenbergen, M. R. (2013). Partisan Ambivalence and the Contingent Nature of Political Judgment. In *The Ambivalent Partisan: How Critical Loyalty Promotes Democracy* (pp. 1–26). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199772759.003.0001

Layman, G. C., Carsey, T. M., & Horowitz, J. M. (2006). Party Polarization in American Politics: Characteristics, Causes, and Consequences. *Annual Review of Political Science*, *9*(1), 83–110. https://doi.org/10.1146/annurev.polisci.9.070204.105138

Lea, M., Spears, R., & De Groot, D. (2001). Knowing Me, Knowing You: Anonymity Effects on Social Identity Processes Within Groups. *Personality and Social Psychology Bulletin*, *27*(5), 526–537. https://doi.org/10.1177/0146167201275002

Lee, F. E. (2015). How Party Polarization Affects Governance. *Annual Review of Political Science*, *18*(1), 261–282. https://doi.org/10.1146/annurev-polisci-072012-113747

Lee, G. (2019, November 18). *Corbyn's Changing Brexit Stance*. Channel 4. https://www.channel4.com/news/factcheck/factcheck-corbyns-changing-brexit-stance

Lelkes, Y. (2016). Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly*, *80*(S1), 392–410. https://doi.org/10.1093/poq/nfw005

Lelkes, Y. (2021). Policy Over Party: Comparing the Effects of Candidate Ideology and Party on Affective Polarization. *Political Science Research and Methods*, *9*(1), 189–196. https://doi.org/10.1017/psrm.2019.18

Lelkes, Y., Sood, G., & Iyengar, S. (2017). The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect. *American Journal of Political Science*, *61*(1), 5–20. https://doi.org/10.1111/ajps.12237

Lelkes, Y., & Westwood, S. J. (2017). The Limits of Partisan Prejudice. *Journal of Politics*, *79*(2), 485–501. https://doi.org/10.1086/688223

Levendusky, M. (2009). *The Partisan Sort: How Liberals Became Democrats and Conservative Became Republicans*. University of Chicago Press.

Levendusky, M. (2010). Clearer Cues, More Consistent Voters: A Benefit of Elite Polarization. *Political Behavior*, *32*(1), 111–131. https://doi.org/10.1007/s11109-009-9094-0

Levendusky, M. (2013). Why Do Partisan Media Polarize Viewers? *American Journal of Political Science*, *57*(3), 611–623. https://doi.org/10.1111/ajps.12008

Levendusky, M. (2017). Americans, Not Partisans: Can Priming American National Identity Reduce Affective Polarization? *The Journal of Politics*, *80*(1), 59–70. https://doi.org/10.1086/693987

Levendusky, M. (2018). When Efforts to Depolarize the Electorate Fail. *Public Opinion Quarterly*, *82*(3), 583–592. https://doi.org/10.1093/poq/nfy036

Levendusky, M., & Malhotra, N. (2016a). Does Media Coverage of Partisan Polarization Affect Political Attitudes? *Political Communication*, *33*(2), 283–301. https://doi.org/10.1080/10584609.2015.1038455

Levendusky, M., & Malhotra, N. (2016b). (Mis)perceptions of Partisan Polarization in the American Public. *Public Opinion Quarterly*, *80*(S1), 378–391. https://doi.org/10.1093/poq/nfv045

Li, S., & Zhang, G. (2018). Intergroup Communication in Online Forums: The Effect of Group Identification on Online Support Provision. *Communication Research*. https://doi.org/10.1177/0093650218807041

Lickel, B. (2012). Retribution and Revenge. In L. Troop (Ed.), *The Oxford Handbook of Intergroup Conflict*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199747672.013.0006

Lippman, W. (1922). *Public Opinion*. Harcourt.

Little, R. (1992). Regression With Missing X's: A Review. *Journal of the American Statistical Association*, *87*(420), 1227–1237. https://doi.org/10.2307/2290664

Lodge, M., & Taber, C. S. (2013). *The Rationalizing Voter*. Cambridge University Press. https://doi.org/10.1017/CBO9781139032490

Lu, J. H., & Steele, C. K. (2019). 'Joy Is Resistance': Cross-Platform Resilience and (Re)Invention of Black Oral Culture Online. *Information, Communication & Society*, *22*(6), 823–837. https://doi.org/10.1080/1369118X.2019.1575449

Lupu, N. (2015). Party Polarization and Mass Partisanship: A Comparative Perspective. *Political Behavior*, *37*(2), 331–356.

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup Emotions: Explaining Offensive Action Tendencies in an Intergroup Context. *Journal of Personality and Social Psychology*, *79*(4), 602–616. https://doi.org/10.1037/0022-3514.79.4.602

Mackie, D. M., & Hamilton, D. (1993). *Affect, Cognition, and Stereotyping*. Academic Press Inc.

Mackie, D. M., & Silver, L. (2004). Intergroup Emotions. In L. Z. Tiedens & C. W. Leach (Eds.), *The Social Life of Emotions* (pp. 227–245). Cambridge University Press. https://doi.org/10.1017/CBO9780511819568

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

MacKinnon, J. G., & White, H. (1985). Some Heteroskedasticity-Consistent Covariance Matrix Estimators With Improved Finite Sample Properties. *Journal of Econometrics*, *29*(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

MacKuen, M., Marcus, G. E., Neuman, W. R., & Keele, L. (2007). The Third Way: The Theory of Affective Intelligence and American Democracy. In R. Neuman, G. Marcus, A. Crigler, & M. MacKuen (Eds.), *The Affect Effect: Dynamics of Emotion in Political Thinking and Behaviour* (pp. 124–152). University of Chicago Press. https://doi.org/10.7208/chicago/9780226574431.001.0001

Malka, A., & Lelkes, Y. (2010). More Than Ideology: Conservative-Liberal Identity and Receptivity to Political Cues. *Social Justice Research*, *23*(2), 156–188. https://doi.org/10.1007/s11211-010-0114-3

Maloney, S. (2018). Political Tribes and Tribal Scholarship: A Review of Amy Chua's Political Tribes and Lilliana Mason's Uncivil Agreement. *Journal of Political Power*, *11*(3), 442–447. https://doi.org/10.1080/2158379X.2018.1525182

Marcus, G., Neuman, R., & MacKuen, M. (2000). *Affective Intelligence and Political Judgment*. University of Chicago Press.

Margetts, H., John, Peter., Hale, Scott., & Yasseri, Taha. (2015). *Political Turbulence: How Social Media Shape Collective Action*. Princeton University Press.

Martherus, J. L., Martinez, A. G., Piff, P. K., & Theodoridis, A. G. (2019). Party Animals? Extreme Partisan Polarization and Dehumanization. *Political Behavior*. https://doi.org/10.1007/s11109-019-09559-4

Martini, S., & Torcal, M. (2013). The Politics of Interpersonal Trust and Reciprocity: An Experimental Approach. *Political Behavior*, *35*(1), 43–63. https://doi.org/10.1007/s11109-011-9181-x

Martini, S., & Torcal, M. (2019). Trust Across Political Conflicts: Evidence From a Survey Experiment in Divided Societies. *Party Politics*, *25*(2), 126–139. https://doi.org/10.1177/1354068816685933

Mason, L. (2015). "I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science*, *59*(1), 128–145. https://doi.org/10.1111/ajps.12089

Mason, L. (2018a). *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.

Mason, L. (2018b). Ideologues Without Issues: The Polarizing Consequences of Ideological Identities. *Public Opinion Quarterly*, *82*(S1), 280–301. https://doi.org/10.1093/poq/nfy005

Mason, L. (2018c). Losing Common Ground: Social Sorting and Polarization. *Forum*, *16*(1), 47–66. https://doi.org/10.1515/for-2018-0004

Mason, L., & Wronski, J. (2018). One Tribe to Bind Them All: How Our Social Group Attachments Strengthen Partisanship. *Political Psychology*, *39*(S1), 257–277. https://doi.org/10.1111/pops.12485

Mattila, M., & Raunio, T. (2006). Cautious Voters-Supportive Parties. *European Union Politics*, *7*(4), 427–449. https://doi.org/10.1177/1465116506069434ï

Mattila, M., & Raunio, T. (2012). Drifting Further Apart: National Parties and their Electorates on the EU Dimension. *West European Politics*, *35*(3), 589–606. https://doi.org/10.1080/01402382.2012.665742

Mauss, I. B., & Robinson, M. D. (2009). Measures of Emotion: A Review. *Cognition & Emotion*, *23*(2), 209–237. https://doi.org/10.1080/02699930802204677

Mayer, S. J. (2017). How Negative Partisanship Affects Voting Behavior in Europe: Evidence From an Analysis of 17 European Multi-Party Systems With Proportional Voting. *Research & Politics*, *7*(1). https://doi.org/10.1177/2053168016686636

McCarty, N., Poole, K., & Rosenthal, H. (2006). *Polarized America: The Dance of Unequal Riches*. MIT Press.

McConnell, C., Margalit, Y., Malhotra, N., & Levendusky, M. (2018). The Economic Consequences of Partisanship in a Polarized Era. *American Journal of Political Science*, *62*(1), 5–18. https://doi.org/10.1111/ajps.12330

McCoy, J., & Somer, M. (2019). Toward a Theory of Pernicious Polarization and How It Harms Democracies: Comparative Evidence and Possible Remedies. *The ANNALS of the American Academy of Political and Social Science*, *681*(1), 234–271. https://doi.org/10.1177/0002716218818782

McGarty, C., Bliuc, A. M., Thomas, E. F., & Bongiorno, R. (2009). Collective Action as the Material Expression of Opinion-Based Group Membership. *Journal of Social Issues*, *65*(4), 839–857. https://doi.org/10.1111/j.1540-4560.2009.01627.x

McLaughlin, B., Holland, D., Thompson, B. A., & Koenig, A. (2020). Emotions and Affective Polarization: How Enthusiasm and Anxiety About Presidential Candidates Affect Interparty Attitudes. *American Politics Research*, *48*(2), 308–316. https://doi.org/10.1177/1532673X19891423

Meleady, R., & Forder, L. (2019). When Contact Goes Wrong: Negative Intergroup Contact Promotes Generalized Outgroup Avoidance. *Group Processes & Intergroup Relations*, *22*(5), 688–707. https://doi.org/10.1177/1368430218761568

Merrill, S., Grofman, B., & Adams, J. (2001). Assimilation and Contrast Effects In Voter Projections of Party Locations: Evidence From Norway, France, and the USA. *European Journal of Political Research*, *40*(2), 199–223. https://doi.org/10.1023/A:1012975221087

Messing, S., & Westwood, S. J. (2014). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research*, *41*(8), 1042–1063. https://doi.org/10.1177/0093650212466406

Michelitch, K. (2015). Does Electoral Competition Exacerbate Interethnic or Interpartisan Economic Discrimination? Evidence from a Field Experiment in Market Price Bargaining. *American Political Science Review*, *109*(1), 43–61. https://doi.org/10.1017/S0003055414000628

Miller, P. R., & Conover, P. J. (2015). Red and Blue States of Mind: Partisan Hostility and Voting in the United States. *Political Research Quarterly*, *68*(2), 225–239. https://doi.org/10.1177/1065912915577208

Mosquera, R., Odunowo, M., McNamara, T., Guo, X., & Petrie, R. (2020). The Economic Effects of Facebook. *Experimental Economics*, *23*(2), 575–602. https://doi.org/10.1007/s10683-019-09625-y

Muddiman, A. (2017). Personal and Public Levels of Political Incivility. *International Journal of Communication*, *11*, 3182–3202. https://ijoc.org/index.php/ijoc/article/view/6137

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup Bias as a Function of Salience, Relevance, and Status: An Integration. *European Journal of Social Psychology*, *22*(2), 103–122. https://doi.org/10.1002/ejsp.2420220202

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The Generalizability of Survey Experiments. *Journal of Experimental Political Science*, *2*(2), 109–138. https://doi.org/10.1017/XPS.2015.19

Mummendey, A., Otten, S., Berger, U., & Kessler, T. (2000). Positive-Negative Asymmetry in Social Discrimination: Valence of Evaluation and Salience of Categorization. *Personality and Social Psychology Bulletin*, *26*(10), 1258–1270. https://doi.org/10.1177/0146167200262007

Munger, K. (2020). Don't @ Me: Experimentally Reducing Partisan Incivility on Twitter. *Journal of Experimental Political Science*, 1–15. https://doi.org/10.1017/XPS.2020.14

Muresan, S., Gonzalez-Ibanez, R., Ghosh, D., & Wacholder, N. (2016). Identification of Nonliteral Language in Social Media: A Case Study on Sarcasm. *Journal of the Association for Information Science and Technology*, *67*(11), 2725–2737. https://doi.org/10.1002/asi.23624

Murray, I., Plagnol, A., & Corr, P. (2017). *'When Things Go Wrong and People are Afraid': An Evaluation of Group Polarisation in the UK Post Brexit*. SSRN. https://doi.org/10.2139/ssrn.3041846

Mutz, D. C. (2002). The Consequences of Cross-Cutting Networks for Political Participation. *American Journal of Political Science*, *46*(4), 838. https://doi.org/10.2307/3088437

Mutz, D. C. (2015). *In-Your-Face Politics: The Consequences of Uncivil Media*. Princeton University Press. 10.1177/1077699016644196p

Myers, D. G., & Lamm, H. (1976). The Group Polarization Phenomenon. *Psychological Bulletin*, *83*(4), 602–627. https://doi.org/10.1037/0033-2909.83.4.602

Naim, M. (2019, January 23). The Globalization of Polarization. *El País*. https://english.elpais.com/elpais/2019/01/23/inenglish/1548242900_128216.html

Nakagawa, S., & Schielzeth, H. (2013). A General and Simple Method for Obtaining R 2 From Generalized Linear Mixed-Effects Models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nave, N. N., Shifman, L., & Tenenboim-Weinblatt, K. (2018). Talking It Personally: Features of Successful Political Posts on Facebook. *Social Media + Society*, *4*(3), 205630511878477. https://doi.org/10.1177/2056305118784771

Negroponte, N. (1995). *Being Digital*. Vintage Books.

Neundorf, A., & Adams, J. (2018). The Micro-Foundations of Party Competition and Issue Ownership: The Reciprocal Effects of Citizens' Issue Salience and Party Attachments. *British Journal of Political Science*, *48*(2), 385–406. https://doi.org/10.1017/S0007123415000642

Newman, N., Fletcher, R., Schulz, A., Andi, S., & Kleis Nielsen, R. (2020). *Digital News Report 2020*. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Nisbet, E. C., Mortenson, C., & Li, Q. (2021). The Presumed Influence of Election Misinformation on Others Reduces Our Own Satisfaction With Democracy. *Harvard Kennedy School (HKS) Misinformation Review*. https://doi.org/10.37016/mr-2020-59

Nithyanand, R., Schaffner, B., & Gill, P. (2017). *Measuring Offensive Speech in Online Political Discourse*. arXiv. https://arxiv.org/abs/1706.01875

Nordlinger, E. A. (1972). *Conflict Regulation in Divided Societies*. Harvard University Press. https://doi.org/10.2307/1960388

Ofcom. (2019, July 24). *Half of People Now Get Their News From Social Media*. Ofcom. https://www.ofcom.org.uk/about-ofcom/latest/features-and-news/half-of-people-get-news-from-social-media

Onishi, N. (2021, February 9). Will American Ideas Tear France Apart? Some of Its Leaders Think So. *The New York Times*. https://www.nytimes.com/2021/02/09/world/europe/france-threat-american-universities.html

Orr, L. V., & Huber, G. A. (2020). The Policy Basis of Measured Partisan Animosity in the United States. *American Journal of Political Science*, *64*(3). https://doi.org/10.1111/ajps.12498

Ott, B. L. (2017). The Age of Twitter: Donald J. Trump and the Politics of Debasement. *Critical Studies in Media Communication*, *34*(1), 59–68. https://doi.org/10.1080/15295036.2016.1266686

Ott, B. L., & Dickinson, G. (2019). *The Twitter Presidency: Donald J. Trump and the Politics of White Rage*. Routledge.

Ottati, V., Steenbergen, M., & Riggle, E. (1992). The Cognitive and Affective Components of Political Attitudes: Measuring the Determinants of Candidate Evaluations. *Political Behavior*, *14*(4), 423–442. https://doi.org/10.1007/BF00992043

Ottati, V., Wyer, R., Deiger, M., & Houston, D. (2002). The Psychological Determinants of Candidate Evaluation and Voting Preference. In V.

Ottati, S. Tindale, J. Edwards, F. Bryant, L. Heath, Y. Suarez-Balcazar, & E. Posavac (Eds.), *The Social Psychology of Politics* (pp. 3–29). Springer.

Panagopoulos, C., Green, D. P., Krasno, J., Schwam-Baird, M., Moore, E., & Endres, K. (2016). *Risky Business: Does Corporate Political Giving Affect Consumer Behavior?* Annual Meeting of the American Political Science Association. https://rubenson.org/wp-content/uploads/2016/10/Panagopoulos-etal.pdf

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419. https://doi.org/10.2139/ssrn.1626226

Papacharissi, Z. (2004). Democracy Online: Civility, Politeness, and the Democratic Potential of Online Political Discussion Groups. *New Media and Society*, *6*(2), 259–283. https://doi.org/10.1177/1461444804041444

Papacharissi, Z. (2014). *Affective Publics: Sentiment, Technology, and Publics*. Oxford University Press.

Papacharissi, Z., & De Fatima Oliveira, M. (2012). Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt. *Journal of Communication*, *62*(2), 266–282. https://doi.org/10.1111/j.1460-2466.2012.01630.x

Papacharissi, Z., & Mendelson, A. (2010). Look at Us: Collective Narcissism in College Student Facebook Photo Galleries. In Z. Papacharissi (Ed.), *The Networked Self: Identity, Community and Culture on Social Network Sites* (London, pp. 251–274). Routledge.

Park, Barum. (2018). How Are We Apart? Continuity and Change in the Structure of Ideological Disagreement in the American Public, 1980–2012. *Social Forces*, *96*(4), 1757–1784. https://doi.org/10.1093/sf/sox093

Park, Bernadette, & Judd, C. M. (2005). Rethinking the Link Between Categorization and Prejudice Within the Social Cognition Perspectiv

Perrin, A., & Atske, S. (2021, March 26). *About Three-in-Ten U.S. Adults Say They Are 'Almost Constantly' Online*. Pew Research Center. https://www.pewresearch.org/fact-tank/2021/03/26/about-three-in-ten-u-s-adults-say-they-are-almost-constantly-online/

Petrocik, J. R. (2009). Measuring Party Support: Leaners Are Not Independents. *Electoral Studies*, *28*(4), 562–572. https://doi.org/10.1016/j.electstud.2009.05.022

Phillips, D., Curtice, J., Phillips, M., & Perry, J. (2018). *British Social Attitudes 35*. NATCEN. https://www.bsa.natcen.ac.uk/media/39284/bsa35_full-report.pdf

Pitkin, H. F. (2004). Representation and Democracy: Uneasy Alliance. *Scandinavian Political Studies*, *27*(3), 335–342.

Platow, M. J., O'Connell, A., Shave, R., & Hanning, P. (1995). Social Evaluations of Fair and Unfair Allocators in Interpersonal and Intergroup Situations. *British Journal of Social Psychology*, *34*(4), 363–381. https://doi.org/10.1111/j.2044-8309.1995.tb01071.x

Plescia, C., & Staniek, M. (2017). In the Eye of the Beholder: Voters' Perceptions of Party Policy Shifts. *West European Politics*, *40*(6), 1288–1309. https://doi.org/10.1080/01402382.2017.1309623

Poole, K. T., & Rosenthal, H. (2001). D-Nominate after 10 Years: A Comparative Update to Congress: A Political-Economic History of Roll-Call Voting. *Legislative Studies Quarterly*, *26*(1), 5. https://doi.org/10.2307/440401

Postmes, T., Haslam, S. A., & Swaab, R. I. (2005). Social Influence in Small Groups: An Interactive Model of Social Identity Formation. *European Review of Social Psychology*, *16*(1), 1–42. https://doi.org/10.1080/10463280440000062

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or Building Social Boundaries? Side-Effects of Computer-Mediated Communication. *Communication Research*, *25*(6), 689–715. https://doi.org/10.1177/009365098025006006

Postmes, T., Spears, R., & Lea, M. (2000). The Formation of Group Norms in Computer-Mediated Communication. *Human Communication Research*, *26*(3), 341–371. https://doi.org/10.1111/j.1468-2958.2000.tb00761.x

Prosser, C. (2020). *Expert Survey Cumulative Codebook 2014-2019*. British Election Study. https://www.britishelectionstudy.com/wp-content/uploads/2020/02/BES-Expert-Survey-Codebook.pdf

Puglisi, R., & Snyder, J. M. (2011). Newspaper Coverage of Political Scandals. *Journal of Politics*, *73*(3), 931–950. https://doi.org/10.1017/S0022381611000569

Rabinowitz, G., Prothro, J. W., & Jacoby, W. (1982). Salience as A Factor in The Impact of Issues on Candidate Evaluation. *The Journal of Politics*, *44*(1), 41–63. https://doi.org/10.2307/2130283

Rains, S. A., Kenski, K., Coe, K., & Harwood, J. (2017). Incivility and Political Identity on the Internet: Intergroup Factors as Predictors of Incivility in Discussions of News Online. *Journal of Computer-Mediated Communication*, *22*(4), 163–178. https://doi.org/10.1111/jcc4.12191

Reddit. (May 11, 2019) *GunsAreCool Subreddit* [Screenshot]. Retrieved from: https://www.reddit.com/r/GunsAreCool/

Reddit. (May 12, 2019) *HillaryForPrison Subreddit* [Screenshot]. Retrieved from: https://www.reddit.com/r/HillaryForPrison

Reeve, E. (2015, April 12). A History of the Hot Take. *New Republic*. https://newrepublic.com/article/121501/history-hot-take

Reiljan, A. (2019). 'Fear and Loathing Across Party Lines' (Also) in Europe: Affective Polarisation in European Party Systems. *European Journal of Political Research*, *59*(2), 376–396. https://doi.org/10.1111/1475-6765.12351

Remnick, D. (2016). Obama Reckons with a Trump Presidency. *The New Yorker*. https://www.newyorker.com/magazine/2016/11/28/obama-reckons-with-a-trump-presidency?mbid=social_twitter

Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). SentiBench: A Benchmark Comparison of State-of-the-Practice Sentiment Analysis Methods. *EPJ Data Science*, *5*(23). https://doi.org/10.1140/epjds/s13688-016-0085-1

Ridge-Newman, A. (2019, October 30). A Single-Issue Democracy: Party Politics Ahead of the UK's General Election. *LSE British Politics and Policy*. https://blogs.lse.ac.uk/politicsandpolicy/single-issue-democracy/

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup Threat and Outgroup Attitudes: A Meta-Analytic Review. *Personality and Social Psychology Review*, *10*(4), 336–353. https://doi.org/10.1207/s15327957pspr1004_4

Roccas, S., & Brewer, M. B. (2002). Social Identity Complexity. *Personality and Social Psychology Review*, *6*(2), 88–106. https://doi.org/10.1207/S15327957PSPR0602_01

Rogowski, J. C., & Sutherland, J. L. (2016). How Ideology Fuels Affective Polarization. *Political Behavior*, *38*(2), 485–508. https://doi.org/10.1007/s11109-015-9323-7

Rohrschneider, R., & Whitefield, S. (2016). Responding to Growing European Union-Skepticism? The Stances of Political Parties Toward European Integration in Western and Eastern Europe Following the Financial Crisis. *European Union Politics*, *17*(1), 138–161. https://doi.org/10.1177/1465116515610641

Rokeach, M., & Rothman, G. (1965). The Principle of Belief Congruence and the Congruity Principle as Models of Cognitive Interaction. *Psychological Review*, *72*(2), 128–142. https://doi.org/10.1037/h0021702

Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018, March 17). How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

Rösner, L., Winter, S., & Krämer, N. C. (2016). Dangerous Minds? Effects of Uncivil Online Comments on Aggressive Cognitions, Emotions, and

Behavior. *Computers in Human Behavior*, *58*, 461–470. https://doi.org/10.1016/j.chb.2016.01.022

Rossini, P. (2020). Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research*, 1–27. https://doi.org/10.1177/0093650220921314

Rothschild, J. E., Howat, A. J., Shafranek, R. M., & Busby, E. C. (2019). Pigeonholing Partisans: Stereotypes of Party Supporters and Partisan Polarization. *Political Behavior*, *41*(2), 423–443. https://doi.org/10.1007/s11109-018-9457-5

Rothschild, Z. K., Keefer, L. A., & Hauri, J. (2021). Defensive Partisanship? Evidence that In-Party Scandals Increase Out-Party Hostility. *Political Psychology*, *42*(1), 3–21. https://doi.org/10.1111/pops.12680

Samii, C., & Aronow, P. M. (2012). On Equivalencies Between Design-Based and Regression-Based Variance Estimators for Randomized Experiments. *Statistics & Probability Letters*, *82*(2), 365–370. https://doi.org/10.1016/j.spl.2011.10.024

Sanders, D. (2017). The UK's Changing Party System: The Prospects for a Party Realignment at Westminster. *Journal of the British Academy*, *5*, 91–124. https://doi.org/10.5871/jba/005.091

Schmitt, C. (1932). *The Concept of the Political*. Rutgers University Press.

Schmitt, H., & Loughran, T. (2017). Understanding Ideological Change in Britain Corbyn, Brexit, and the BES Expert Surveys. *EPOP Conference*. https://www.nottingham.ac.uk/conference/fac-socsci/epop-2017/documents/conference-papers/understanding-ideological-change-in-britain-final-7.3.pdf

Schumacher, S. (2019, October 28). *Brexit Divides the UK, but Partisanship and Ideology are Still Key Factors*. Pew Research Center. https://www.pewresearch.org/fact-tank/2019/10/28/brexit-divides-the-uk-but-partisanship-and-ideology-are-still-key-factors/

Sclafani, J. (2018). *Talking Donald Trump: A Sociolinguistic Study of Style, Metadiscourse, and Political Identity*. Routledge. https://doi.org/9781138244504

Scott, C. R. (2007). Communication and Social Identity Theory: Existing and Potential Connections in Organizational Identification Research. *Communication Studies*, *58*(2), 123–138. https://doi.org/10.1080/10510970701341063

Sears, D. O., & Funk, C. L. (1999). Evidence of the Long-Term Persistence of Adults' Political Predispositions. *The Journal of Politics*, *61*(1), 1–28. https://doi.org/10.2307/2647773

Settle, J. E. (2018a). *Frenemies: How Social Media Polarizes America*. Cambridge University Press. https://doi.org/10.1017/9781108560573

Settle, J. E. (2018b). Moving Beyond Sentiment Analysis. In S. González-Bailón & B. F. Welles (Eds.), *The Oxford Handbook of Networked Communication*. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190460518.013.20

Settle, J. E., & Carlson, T. N. (2019). Opting Out of Political Discussions. *Political Communication*, *36*(3), 476–496. https://doi.org/10.1080/10584609.2018.1561563

Shafranek, R. M. (2020). Political Consequences of Partisan Prejudice. *Political Psychology*, *41*(1), 35–51. https://doi.org/10.1111/pops.12596

Shah, D. V. (2016). Conversation is the Soul of Democracy: Expression Effects, Communication Mediation, and Digital Media. *Communication and the Public*, *1*(1), 12–18. https://doi.org/10.1177/2057047316628310

Shearer, E., & Gottfried, J. (2017). *News Use Across Social Media Platforms*. Pew Research Center. https://www.journalism.org/wp-content/uploads/sites/8/2017/09/PJ_17.08.23_socialMediaUpdate_FINAL.pdf

Shearer, E., & Grieco, E. (2019). *Americans Are Wary of the Role Social Media Sites Play in Delivering the News*. Pew Research Center. https://www.journalism.org/wp-content/uploads/sites/8/2019/09/PJ_2019.09.25_Social-Media-and-News_FINAL.pdf

Shi, F., Teplitskiy, M., Duede, E., & Evans, J. A. (2019). The Wisdom of Polarized Crowds. *Nature Human Behaviour*, *3*(4), 329–336. https://doi.org/10.1038/s41562-019-0541-6

Sides, J., Tesler, M., & Vavreck, L. (2018). *Identity Crisis*. Princeton University Press. https://doi.org/10.1515/9781400888306

Silverman, C., Lytvynenko, J., & Kung, W. (2020). Disinformation For Hire: How A New Breed Of PR Firms Is Selling Lies Online. *Buzzfeed*.

Skoric, M. M., Zhu, Q., & Lin, J.-H. T. (2018). What Predicts Selective Avoidance on Social Media? A Study of Political Unfriending in Hong Kong and Taiwan. *American Behavioral Scientist*, *62*(8), 1097–1115. https://doi.org/10.1177/0002764218764251

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and Extremity Biases in Impression Formation: A Review of Explanations. *Psychological Bulletin*, *105*(1), 131–142. https://doi.org/10.1037/0033-2909.105.1.131

Skytte, R. (2020). Dimensions of Elite Partisan Polarization: Disentangling the Effects of Incivility and Issue Polarization. *British Journal of Political Science*, 1–19. https://doi.org/10.1017/S0007123419000760

Smith, E. R., & Mackie, D. M. (2015). Dynamics of Group-Based Emotions: Insights From Intergroup Emotions Theory. *Emotion Review*, *7*(4), 349–354. https://doi.org/10.1177/1754073915590614

Smith, E. R., Seger, C. R., & Mackie, D. M. (2007). Can Emotions be Truly Group Level? Evidence Regarding Four Conceptual Criteria. *Journal of Personality and Social Psychology*, *93*(3), 431–446. https://doi.org/10.1037/0022-3514.93.3.431

Smith, M. (2019). *Left-Wing vs Right-Wing: It's Complicated*. YouGov. https://yougov.co.uk/topics/politics/articles-reports/2019/08/13/left-wing-vs-right-wing-its-complicated

Sobieraj, S., & Berry, J. M. (2011). From Incivility to Outrage: Political Discourse in Blogs, Talk Radio, and Cable News. *Political Communication*, *28*(1), 19–41. https://doi.org/10.1080/10584609.2010.542360

Sood, G., & Iyengar, S. (2017). *Coming to Dislike Your Opponents: The Polarizing Impact of Political Campaigns*. SSRN. https://doi.org/10.2139/ssrn.2840225

Soroka, S. N. (2014). *Negativity in Democratic Politics: Causes and Consequences*. Cambridge University Press. https://doi.org/10.1017/CBO9781107477971

Soroka, S. N., Fournier, P., & Nir, L. (2019). Cross-National Evidence of a Negativity Bias in Psychophysiological Reactions to News. *Proceedings of the National Academy of Sciences*, *116*(38), 18888–18892. https://doi.org/10.1073/pnas.1908369116

Spears, R., Lea, M., Corneliussen, R. A., Postmes, T., & Ter Haar, W. (2002). Computer-Mediated Communication as a Channel for Social Resistance: The Strategic Side of Side. *Small Group Research*, *33*(5), 555–574. https://doi.org/10.1177/104649602237170

Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, *65*(3), 557–586. https://doi.org/10.2307/2171753

Stark, L. (2020). Empires of Feeling. In M. Boler & E. Davis (Eds.), *Affective Politics of Digital Media* (pp. 298–313). Routledge. https://doi.org/10.4324/9781003052272-15

Statista. (2021, February). *Leading Social Media Websites in the United States*. Statista. https://www.statista.com/statistics/265773/market-share-of-the-most-popular-social-media-websites-in-the-us/

Stecker, C., & Tausendpfund, M. (2016). Multidimensional Government-Citizen Congruence and Satisfaction With Democracy. *European Journal of Political Research*, *55*(3), 492–511. https://doi.org/10.1111/1475-6765.12147

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending With Group Image: The Psychology of Stereotype and Social Identity Threat. *Advances in Experimental Social Psychology*, *34*, 379–440. https://doi.org/10.1016/S0065-2601(02)80009-0

Steenbergen, M., Killburn, W., & Wolak, J. (2001). Affective Moderators of On-Line and Memory-Based Processing in Candidate Evaluation. *Annual Meeting of the Midwest Political Science Association*.

Stephan, W. G., & Stephan, C. W. (2000). An Integrated Threat Theory of Prejudice. In S. Oskamp (Ed.), *Reducing Prejudice and Discrimination* (pp. 23–45). Lawrence Erlbaum Associates Publishers.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, *29*(4), 217–248. https://doi.org/10.2753/MIS0742-1222290408

Stieglitz, S., & Dang-Xuan, L. (2012). Political Communication and Influence through Microblogging: An Empirical Analysis of Sentiment in Twitter Messages and Retweet Behavior. *2012 45th Hawaii International Conference on System Sciences*, 3500–3509. https://doi.org/10.1109/HICSS.2012.476

Stokes, D. E. (1963). Spatial Models of Party Competition. *American Political Science Review*, *57*(2), 368–377. https://doi.org/10.2307/1952828

Storbeck, J., & Clore, G. L. (2007). On the Interdependence of Cognition and Emotion. *Cognition & Emotion*, *21*(6), 1212–1237. https://doi.org/10.1080/02699930701438020

Stryker, S., & Serpe, R. T. (1994). Identity Salience and Psychological Centrality: Equivalent, Overlapping, or Complementary Concepts? *Social Psychology Quarterly*, *57*(1), 16–35. https://doi.org/10.2307/2786972

Suhay, E., Bello-Pardo, E., & Maurer, B. (2018). The Polarizing Effects of Online Partisan Criticism: Evidence from Two Experiments. *International Journal*

of Press/Politics, *23*(1), 95–115. https://doi.org/10.1177/1940161217740697

Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.

Sunstein, C. R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.

Tajfel, H. (1972). Social Categorization. In S. Moscovici (Ed.), *Introduction à la Psychologie Sociale* (pp. 272–302). Larousse.

Tajfel, H. (1974). Social Identity and Intergroup Behaviour. *Social Science Information*, *13*(2), 65–93. https://doi.org/10.1177/053901847401300204

Tajfel, H. (1981). *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge University Press.

Tajfel, H., & Turner, J. (1979). An Integrative Theory of Intergroup Conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–37). Brooks/Cole.

Tamburrini, N., Cinnirella, M., Jansen, V. A. A., & Bryden, J. (2015). Twitter Users Change Word Usage According to Conversation-Partner Social Identity. *Social Networks*, *40*, 84–89. https://doi.org/10.1016/j.socnet.2014.07.004

TED. (2018, July 25). *How Partisan Tribalism is Killing Democracy | Brian Klaas | TEDxThessaloniki [Video]*. YouTube. https://www.youtube.com/watch?v=gknM14P8XQ8

Tettegah, S. (2016). *Emotions, Technology, and Social Media*. Academic Press Inc.

The Economist. (2021, February 11). *Censorious Governments are Abusing "Fake News" Laws*.

https://www.economist.com/international/2021/02/11/censorious-governments-are-abusing-fake-news-laws

Theodoridis, A. G. (2017). Me, Myself, and (I), (D), or (R)? Partisanship and Political Cognition through the Lens of Implicit Identity. *The Journal of Politics*, *79*(4), 1253–1267. https://doi.org/10.1086/692738

Theriault, S. M. (2008). *Party Polarization in Congress*. Cambridge University Press.

Thomassen, J. (2012). The Blind Corner of Political Representation. *Representation*, *48*(1), 13–27. https://doi.org/10.1080/00344893.2012.653229

Thomassen, J., & Schmitt, H. (1997). Policy Representation. *European Journal of Political Research*, *32*(2), 165–184. https://doi.org/10.1111/1475-6765.00337

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. Hewlett Foundation. https://www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf

Tucker, J. A., Theocharis, Y., Roberts, M. E., & Barberá, P. (2017). From Liberation to Turmoil: Social Media and Democracy. *Journal of Democracy*, *28*(4). https://doi.org/10.1353/jod.2017.0064

Tyagi, A., Uyheng, J., & Carley, K. M. (2020). Affective Polarization in Online Climate Change Discourse on Twitter. *Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,*

ASONAM 2020, 443–447. https://doi.org/10.1109/ASONAM49781.2020.9381419

United States of America v. Internet Research Agency, (United States District Court for the District of Columbia 2018). https://www.justice.gov/file/1035477/download

Urbinati, N., & Warren, M. E. (2008). The Concept of Representation in Contemporary Democratic Theory. *Annual Review of Political Science*, *11*(1), 387–412. https://doi.org/10.1146/annurev.polisci.11.053006.190533

Usherwood, S. (2008). The Dilemmas of a Single-Issue Party – the UK Independence Party. *Representation*, *44*(3), 255–264. https://doi.org/10.1080/00344890802237023

Vaccari, C. (2017). Online Mobilization in Comparative Perspective: Digital Appeals and Political Engagement in Germany, Italy, and the United Kingdom. *Political Communication*, *34*(1), 69–88. https://doi.org/10.1080/10584609.2016.1201558

Vaccari, C., & Valeriani, A. (2018). Digital Political Talk and Political Participation: Comparing Established and Third Wave Democracies. *SAGE Open*, *8*(2), 1–14. https://doi.org/10.1177/2158244018784986

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., & Tucker, J. A. (2016). Of Echo Chambers and Contrarian Clubs: Exposure to Political Disagreement Among German and Italian Users of Twitter. *Social Media and Society*, *2*(3), 1–24. https://doi.org/10.1177/2056305116664221

Valkenburg, P. M., Peter, J., & Walther, J. B. (2016). Media Effects: Theory and Research. *Annual Review of Psychology*, *67*(1), 315–338. https://doi.org/10.1146/annurev-psych-122414-033608

van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stepińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheafer, T., & Stanyer, J. (2017). Political Communication in a High-Choice Media Environment: A Challenge for Democracy? *Annals of the International Communication Association*, *41*(1), 3–27. https://doi.org/10.1080/23808985.2017.1288551

van Der Brug, W., & van Spanje, J. (2009). Immigration, Europe and the "New" Cultural Dimension. *European Journal of Political Research*, *48*(3), 309–334. https://doi.org/10.1111/j.1475-6765.2009.00841.x

van Leeuwen, F., & Park, J. H. (2009). Perceptions of Social Dangers, Moral Foundations, and Political Orientation. *Personality and Individual Differences*, *47*(3), 169–173. https://doi.org/10.1016/j.paid.2009.02.017

van Zomeren, M., Fischer, A. H., & Spears, R. (2007). Testing the Limits of Tolerance: How Intergroup Anxiety Amplifies Negative and Offensive Responses to Out-Group-Initiated Contact. *Personality and Social Psychology Bulletin*, *33*(12), 1686–1699. https://doi.org/10.1177/0146167207307485

Vanbeselaere, N., Boen, F., Van Avermaet, E., & Buelens, H. (2006). The Janus Face of Power in Intergroup Contexts: A Further Exploration of the Noblesse Oblige Effect. *The Journal of Social Psychology*, *146*(6), 685–699. https://doi.org/10.3200/SOCP.146.6.685-699

Veitch, R., & Griffitt, W. (1976). Good News-Bad News: Affective and Interpersonal Effects. *Journal of Applied Social Psychology*, *6*(1), 69–75. https://doi.org/10.1111/j.1559-1816.1976.tb01313.x

Wagner, M. (2021). Affective Polarization in Multiparty Systems. *Electoral Studies*, *69*, 102199. https://doi.org/10.1016/j.electstud.2020.102199

Walgrave, S., & Lefevere, J. (2013). Ideology, Salience, and Complexity: Determinants of Policy Issue Incongruence between Voters and Parties. *Journal of Elections, Public Opinion and Parties*, *23*(4), 456–483. https://doi.org/10.1080/17457289.2013.810630

Webster, S., & Abramowitz, A. (2017). The Ideological Foundations of Affective Polarization in the U.S. Electorate. *American Politics Research*, *45*(4), 621–647. https://doi.org/10.1177/1532673X17703132

Weeks, B. E., & Garrett, K. (2019). Emotional Characteristics of Social Media and Political Misperceptions. In J. E. Katz & K. K. Mays (Eds.), *Journalism and Truth in an Age of Social Media* (pp. 236–251). Oxford University Press.

West, E. A., & Iyengar, S. (2020). Partisanship as a Social Identity: Implications for Polarization. *Political Behavior*. https://doi.org/10.1007/s11109-020-09637-y

Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving Political Polarization in the United States: Party Identity Strength and Attitude Extremity Exacerbate the Perceived Partisan Divide. *Perspectives on Psychological Science*, *10*(2), 145–158. https://doi.org/10.1177/1745691615569849

Westwood, S. J., Iyengar, S., Walgrave, S., Leonisio, R., Miller, L., & Strijbis, O. (2018). The Tie That Divides: Cross-National Evidence of the Primacy of Partyism. *European Journal of Political Research*, *57*(2), 333–354. https://doi.org/10.1111/1475-6765.12228

Westwood, S. J., & Peterson, E. (2020). The Inseparability of Race and Partisanship in the United States. *Political Behavior*. https://doi.org/10.1007/s11109-020-09648-9

Whitt, S., Yanus, A. B., McDonald, B., Graeber, J., Setzler, M., Ballingrud, G., & Kifer, M. (2020). Tribalism in America: Behavioral Experiments on Affective Polarization in the Trump Era. *Journal of Experimental Political Science*, 1–13. https://doi.org/10.1017/XPS.2020.29

Williams, C., & Spoon, J. J. (2015). Differentiated Party Response: The Effect of Euroskeptic Public Opinion on Party Positions. *European Union Politics*, *16*(2), 176–193. https://doi.org/10.1177/1465116514564702

Wilson, S. (2020, February). The Era of Antisocial Social Media. *Harvard Business Review*. https://hbr.org/2020/02/the-era-of-antisocial-social-media

Wojcieszak, M. (2008). False Consensus Goes Online Impact of Ideologically Homogeneous Groups on False Consensus. *Public Opinion Quarterly*, *72*(4), 781–791. https://doi.org/10.1093/poq/nfn056

Wojcieszak, M. (2010). "Don't Talk to Me": Effects of Ideologically Homogeneous Online Groups and Politically Dissimilar Offline Ties on Extremism. *New Media and Society*, *12*(4), 637–655. https://doi.org/10.1177/1461444809342775

Wojcieszak, M., & Warner, B. R. (2020). Can Interparty Contact Reduce Affective Polarization? A Systematic Test of Different Forms of Intergroup Contact. *Political Communication*, *37*(6), 789–811. https://doi.org/10.1080/10584609.2020.1760406

Wolak, J., MacKuen, M., Keele, L., Marcus, G., & Neuman, R. (2003). *How the Emotions of Public Policy Affect Citizen Engagement, Public Deliberation, and the*

Quality of Electoral Choice. Annual Meeting of the American Political Science Association, Philadelphia, PA.

Wozniak, K. H. (2017). Public Opinion About Gun Control Post–Sandy Hook. *Criminal Justice Policy Review*, *28*(3), 255–278. https://doi.org/10.1177/0887403415577192

Wu, J., & Hamada, M. (2009). *Experiments: Planning, Analysis, and Optimization*. Wiley-Blackwell Publishing Ltd.

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020). Political Polarization on the Digital Sphere: A Cross-Platform, Over-Time Analysis of Interactional, Positional, and Affective Polarization on Social Media. *Political Communication*, 1–42. https://doi.org/10.1080/10584609.2020.1785067

Yardi, S., & boyd, danah. (2010). Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. *Bulletin of Science, Technology & Society*, *30*(5), 316–327. https://doi.org/10.1177/0270467610380011

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics*, *12*(1), 415–438. https://doi.org/10.1146/annurev-economics-081919-050239

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should We Worry about Filter Bubbles? *Internet Policy Review*, *5*(1). https://doi.org/10.14763/2016.1.401