Social media and democratic deliberation

Nick Lewis
London School of Economics and Political Science

A thesis submitted for the degree of
Doctor of Philosophy

January 2025

Statement of Originality

I certify that the thesis I have presented for examination for the PhD degree at the London School of Economics and Political Science is solely my own work, other than where I have clearly indicated that it is the work of others. In this case, the proportion of any collaborative work carried out by me is clearly identified.

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 52,936 words.

Statement of co-authored work

I confirm that the second paper presented in this thesis was co-authored with Professor Sara Hobolt and Professor James Tilley, and I contributed 80% of this work.

As the candidate's primary supervisor, I hereby confirm that the extent of the candidate's contribution to the joint-authored papers was as indicated above.

Professor Sara Hobolt

Acknowledgements

Completing a PhD is challenging for anyone. Beginning one during a pandemic is almost irresponsible. The isolation, the uncertainty, and the ever-changing landscape made this journey even more difficult, and at times, even overwhelming. Through all of it, though, I have been incredibly fortunate to have the most amazing support network, without whom I would not be writing this today. To my friends, family, colleagues, and mentors – this thesis would not exist without your unwavering encouragement, guidance, and love.

First, my heartfelt thanks go to my supervisors. Sara Hobolt, for encouraging me to follow my intellectual curiosity and take on this challenge. Your endless patience and insight - not to mention a willingness to indulge some of my more ambitious ideas - have shaped not only this thesis but also my growth as a researcher. From day one, you have made me feel like I belonged in the academic world. To Dan Berliner, thank you for being an invaluable mentor throughout this process. Your methodological rigour has been pivotal in the improvement of all three papers here, and your creative approach to problem-solving has bailed me out on more than one occasion. Your feedback was always insightful and constructive, and your passion for the subject matter was infectious. I feel incredibly privileged to have been supervised by you both.

The broader LSE community has also played a pivotal role in shaping this thesis. I'd like to thank Sarah Brierley and Lawrence Ezrow for acting as moderators at my MRes upgrade, and helping me set the trajectory of my thesis at an early stage. I would also like to thank Doctors Christopher Prosser and Nick Anstead for serving as my PhD examiners, for your thoughtful engagement with my work, and for helping me bring this project to its final form.

The LSE Political Behaviour seminar has been one of the highlights of my academic life; it is a rare privilege to be part of a community that attracts some of the world's leading scholars. Presenting my work in this forum was both nerve-wracking and exhilarating, but the insightful comments I received helped to shape and improve the third paper in this thesis significantly. A special thanks to Florian Foos and Mathias Poertner, for your detailed and expert comments on its design and analysis; your input has been invaluable. A big shout out goes to everyone on the 4th floor: Ceren, Tarsha, Stephanie, Martin, Max, Zach, Morris, and many others. Your humour and humanity kept me sane. Thanks for always reminding me there's life outside of research and for making sure we knew when it was time to head to the pub. Also, I didn't mean to start the wedding trend, but here we are.

Teaching at LSE has been a wonderful experience. I'd like to thank Ryan Jablonski for his constant guidance and for helping me to continually improve as a teacher. All the wonderful teaching fellows I've had the pleasure of working with: Victor, Thalia, Leo, Matthias, Christine, Federico, and Kiwi - thank you for your support, collaboration, and friendship. Finally, to all the talented students on GV101, who could have quite easily sat there in silence. Not only did you turn up at 9am every Monday, you were fun, engaged, perceptive, and receptive to new ideas.

The research presented in this thesis could not have been possible without the generous funding from both the Department of Government and the LSE Phelan US Centre. I'm also really grateful to Ade and Chris for having me on the Ballpark podcast to chat about my work. I owe a debt of gratitude to the incredible support staff in the Department of Government. In particular, I want to thank Emily Crook for keeping me on track, even when I could have quite easily found the answer to my question in the PhD handbook (!). Your patience, good humour, and willingness to help at every turn made such a difference.

This thesis is dedicated, in part, to those I lost along the way, including my two grand-dads, Drew and Alfie. Though you're not here to celebrate with me, your spirits have been with me throughout this journey. Grandad Lewis, with whom I shared many heated political debates, would have particularly enjoyed picking apart some of my assumptions. I'm sure this work would have made you both proud. To my friends in London, thank you for providing the balance I so desperately needed between academic life and personal life. To my friends at Merton Saints BMX Club: David, Isla, Dorchie, Mimi, Albert, Gonzalo, Craig, and everyone else I may have forgotten to name here; riding, racing, and running the club has been the perfect contrast to my research. Thank you for the early-morning race lifts, particularly those damp 6am starts. Your commitment to the local community is inspiring, and I feel so lucky to have been part of it.

To my long-term friends – Ed, Mike, Ross, Ali, Luke, Blair, Darren, Graeme, Malcs, and anyone else I may have missed - thank you for keeping an old man alive during his stag weekend and ensuring I returned to finish this thesis. I'm grateful for all of your constant support and friendship. I've also spent an inordinate amount of time with some of the musicians who have helped me focus throughout countless long writing sessions. Rival Consoles, Four Tet, Burial, Floating Points, Bicep, Kelly Lee Owens, Oneohtrix Point Never, Overmono, Nathan Micay, and others – thank you for creating the soundtrack to my life over the last four years.

To Mum and Dad, for always trusting me to find my own path, to learn from my own successes and failures, and for supporting me unconditionally every step of the way. To my sister Lisa, for being a constant source of encouragement and love. And to Sean and Vicki, thank you for book-ending this

PhD journey with the arrival of my niece Maya and nephew Luke, who have brought so much joy and perspective into my life. Your smiles and laughter have been a wonderful reminder of what truly matters.

Last, but by no means least, Alice. Your love, understanding, and constant pep-talks have anchored me throughout this entire process. Thank you for believing in me, for supporting me in the toughest moments, and for planning our incredible wedding day together. I couldn't have done any of this without you, and this achievement is as much yours as it is mine. I can't wait to see where our journey takes us next.

Abstract

The internet continues to have a profound influence upon political communication. Have social media changed the language we use to talk about politics, made particular voices more prominent in online discussions, or even led people to turn away from democratic deliberation entirely? In three distinct papers, this thesis explores each of these questions in turn, making the following theoretical, methodological, and practical contributions. First, using a novel dataset of 4 million tweets, paper 1 shows how political elites in the United Kingdom have used increasingly emotive rhetoric over time on Twitter. I argue this changing rhetorical behaviour has been influenced by, and rewarded with, greater engagement. Paper 2 also employs new data, linking a nationally-representative survey of the UK population with Twitter accounts, to show that people who discuss politics on Twitter are more ideologically and affectively extreme than those who do not. These findings have important implications for understanding who we are most likely to hear in online political discussion, and for increasingly-polarised online political debate. Finally, in a pre-registered lab-in-the-field experiment on Facebook, I argue that people disengage from discussing contentious political issues in groups of people with contrasting opinions and beliefs. In conducting a real-world experiment which attempts to causally identify why people turn away from political discussion, this study represents a methodological advance on existing survey-based research, and understanding the key drivers of online self-censorship may help mitigate some of its most negative consequences. Overall, these findings contribute to our understanding of how social media is influencing the scope and tone of democratic debate.

Contents

List of Tables

List of Figures

1 Introduction

“Statistically impossible to have lost the 2020 Election. Big protest in D.C. on January 6th. Be there, will be wild.”. This is how Donald Trump, the 45th President of the United States, ended his tweet sent at 1.42am on the 19th of December, 2020. The tweet is now considered by the Congressional Select Committee investigating the January 6th attack on the United States Capitol to have been instrumental in not only bringing protesters to Washington D.C., but in fomenting the violence which occurred that day (Dreisbach, 2022; Sheerin, 2022). In the words of Senator Jamie Raskin, a member of the House select committee investigating the Capitol riots, “Donald Trump would issue a tweet that would galvanize his followers, unleash a political firestorm, and change the course of our history as a country” (Stanley-Becker and Alemany, 2022). That a single post on a website could be considered so consequential for democracy speaks volumes about how social media have revolutionised the way we talk about politics.

Trump is widely considered to be an expert social media communicator, and regularly uses emotive rhetoric in his tweets[1]. Is he an outlier, or is this a ‘new normal’; in other words, have politicians become more emotional in the language they use on social media? This has consequences for democratic

[1] Indeed, one of his final two tweets before being banned in January 2021 would read “The 75,000,000 great American Patriots who voted for me, AMERICA FIRST, and MAKE AMERICA GREAT AGAIN, will have a GIANT VOICE long into the future. They will not be disrespected or treated unfairly in any way, shape or form!!!”.

deliberation. Emotive tweets like this are designed to influence behaviour and help elites achieve their political goals. We know that evoking emotion can motivate people to engage in a wide range of political activities (Marcus, 2000; Brader, 2005; Valentino et al., 2011). Indeed, Trump's tweets about the 'liberation' of Michigan, Minnesota, and Virginia during the Covid pandemic led to reduced compliance with stay-at-home orders and a surge in arrests of white Americans for offenses tied to civil disobedience and rebellion (Dickson and Hobolt, 2024). Emotionally-charged language also frequently taps into primal instincts, reinforcing pre-existing beliefs while simultaneously deepening ideological and emotional divides (Iyengar and Westwood, 2015), raising critical questions about its broader impact on the growing polarisation of societies worldwide.

In 'galvanising' his Twitter followers to travel to Washington D.C. on January 6th, was Trump simply preaching to the choir, and exhorting an already politically-extreme audience? After all, members of ultra-nationalist groups like the Proud Boys and Oath Keepers were found to have played a key role in the Capitol riots (Thompson and Ford, 2021). Of course, from Trump's point of view, Twitter's extremism actually stems from the other side: "We will not be SILENCED! Twitter is not about FREE SPEECH. They are all about promoting a Radical Left platform where some of the most vicious people in the world are allowed to speak freely..." (Trump, 2021). Are people who talk about politics on X representative of the population, or is it dominated by people closer to the edges of the ideological spectrum? Any answer to this question must not lose sight of the fact that the vast majority of people do not discuss politics online (McClain, 2021). If political discussion on social media is dominated by people with stronger opinions and beliefs, what happens to those we *don't* hear, and why do they turn away from discussing politics online?

This thesis presents evidence that social media have transformed the language used by political elites over time, are dominated by voices closer to the edges of the ideological spectrum and, in certain situations, represent environments which contribute to disengagement from democratic debate. Understanding each of these changes is crucial for assessing their impact on the health of democracy and, in addressing them, this thesis makes theoretical, methodological, empirical, and practical contributions. While each of the papers presented here address different debates in political science, all three build on a wide body of research exploring how emerging technologies have revolutionised political communication. From the influence of Gutenberg's printing press on the rapid spread of new ideas (Eisenstein, 1980), through Roosevelt's radio 'fireside chats' forging an intimate connection between politicians and the electorate (Briggs and Burke, 2009), to the advent of televised debates placing increased emphasis upon oratory and presentation as political skills (Druckman, 2003), new technologies have transformed the way we talk about politics. Two decades from the inception of the first global social media sites[2], we are still reckoning with how social media have changed - and are changing - the way we talk about politics.

1.1 Contributions to debates

Rhetorical adaptation

This thesis contributes to several strands of political communication literature. First, I build on research which examines how political elites employ emotive rhetoric in different environments, and ask whether a similar rhetorical adaptation has occurred with the advent of social media. As strategic rational actors (e.g. Downs 1957), politicians recognise the power of the

[2] Myspace launched in August 2003 and Facebook in February 2004

language they use and aim to maximise utility by changing their rhetorical behaviour depending on their goals in different situations. For example, after the extension of Britain's voting franchise in 1867, cabinet members used more simple language to appeal to the less educated citizens who were now part of the electorate (Spirling, 2016). To this day, dissenting parliamentarians in the House of Commons use simpler language and more first-person pronouns (Slapin and Kirkland, 2020), while more ideologically-extreme politicians alter their language depending on whether they are in government or opposition (Slapin et al., 2018). This rhetorical flexibility is particularly evident in shifting levels of emotion; politicians adjust their emotive rhetoric according to changing external events, their party's official status, or even their ideological position. For example, government parliamentarians use more positive language than opposition politicians (Rheault et al., 2016; Crabtree et al., 2020), populists use significantly more negative emotional appeals (Widmann, 2021), while politicians in general use emotion more frequently in higher-profile debates (Osnabrügge et al., 2021) to maximise the reach and impact of their speeches. Ultimately, according to Westen (2007) and Caplan (2007), voters are emotional animals who vote with their hearts rather than their heads, so strategically employing emotion in different situations can help politicians improve their electoral appeal. While this existing research has primarily focused on traditional political settings, we know far less about how the advent of new communication technology has influenced the emotional intensity of elite political rhetoric. In the first paper of this thesis, I contribute to the literature by providing empirical evidence that MPs have adapted their language over time on Twitter, and connect this work with a broader body of research that explores the relationship between communication media and emotional intensity.

Emotion and social media

A substantial body of research has explored how the advent of new communication media - particularly television - contributed to the growing emphasis on emotional appeals in elite rhetoric (e.g., Iyengar 1994; Druckman and Holmes 2004; Brader 2005). Looking specifically at social media, studies have performed qualitative content analysis of elite rhetoric on Twitter (Van Kessel and Castelein, 2016; Stier et al., 2018; Munger et al., 2019; van Vliet, 2021), the issues they discuss (Peeters et al., 2021) and their ability to use social media to set the political agenda (Shapiro and Hemphill, 2017). These, however, represent snapshots of elite rhetorical online behaviour, and our understanding of how digital media have shaped political language over time remains relatively underdeveloped.

There are good reasons to think that social media favour more emotionally-extreme rhetorical behaviour. First, emotionally- and morally-charged posts tend to receive more likes, shares, and comments than ‘rational’ content, leading to their amplification across networks (Berger and Milkman, 2012; Stieglitz and Dang-Xuan, 2013; Brady et al., 2017; Weismueller et al., 2022). Social media platforms are designed to maximise user engagement, and their algorithms tend to prioritise content which elicits strong emotional responses. Second, compared with ‘traditional’ channels, social media allow politicians to communicate directly with their audience without relying on journalistic media ‘gatekeepers’ who might moderate their message (Bennett and Manheim, 2006). Indeed, parties are more likely to use emotionally-charged populist rhetoric on social media than they are in more traditional channels (Ernst et al., 2019). This direct-to-public communication style, described by Papacharissi (2015) as ‘affective publics’, emphasises emotional expression over rational deliberation. An immediate feedback loop between politician and audience allows elites to see in real-time which rhetorical strategies resonate most, and further encourages emotive behaviour. Third, and finally, the rise of social media has been linked to the intensification of polarisation, where the formation of filter bubbles (Sunstein, 2018) limit exposure to opposing viewpoints and reinforce extreme perspectives. In particular, the literature on affective polarisation (Iyengar et al., 2012; Mason, 2015; Nordbrandt, 2021) suggests that individuals increasingly identify emotionally with their political tribe, heightening animosity toward the opposition. In such a polarised environment, have politicians adopted more emotionally extreme rhetoric to appeal to their base and amplify their reach on social media? The first paper in this thesis sheds new light on this question, using longitudinal Twitter data to explore whether MPs in the United Kingdom have become more emotional during their time on Twitter. In doing so, I build a bridge between the literature on rhetorical adaptation and research which examines the role of emotion in online political discussion.

Polarisation

This thesis also adds to the substantial literature on polarisation by exploring whether social media represent spaces which exacerbate or alleviate political division. Specifically, this thesis examines whether the prominent political communication medium X, formerly known as Twitter, is dominated by ideologically and affectively-extreme partisan voices. Some scholars argue that the seeds of polarisation can be traced back to a pre- social media era, when the advent of hyper-partisan broadcast media helped reinforce biases and negative perceptions of the opposing party (DellaVigna and Kaplan, 2007; Levendusky, 2013). These outlets, it is argued, provide content which aligns with the ideological preferences of their audience, leading to selective exposure where individuals consume media that reinforces their pre-existing beliefs (Stroud, 2010). Such selective exposure limits cross-ideological communication and promotes ideological isolation.

Have social media exacerbated or alleviated this political polarisation? Many argue the former. To an even greater extent than ‘traditional’ partisan media, social media platforms allow self-selection and the curation of homogeneous networks of like-minded individuals, providing an easy way for individuals to express their political identity and reinforce group norms. The development of these networks reinforce existing beliefs and create echo chambers (Sunstein, 2018), exacerbating political polarisation by limiting exposure to diverse viewpoints (Colleoni et al., 2014). Augmenting this process, self-selection is particularly appealing to those who are already highly polarised (Mason, 2018); individuals with strong political beliefs are more likely to seek out, engage with, and share content that confirms their views, leading to selective exposure (Garrett, 2009; Nyhan and Reifler, 2010). This points towards the existence of a ‘feedback loop’, where social media use and polarisation reinforce each other. Highly polarised individuals are drawn to social media to express and validate their views, while social media platforms, in turn, amplify polarising content (Tucker et al., 2018). This is often a consequence of design: most social media platforms are refined to maximise user engagement, often promoting sensational and divisive content, which tends to generate more clicks, shares, and comments (Bakshy et al., 2015). Negative and emotionally charged content is more likely to be shared and engaged with on social media, exacerbating and intensifying affective polarisation (Brady et al., 2017; Tucker et al., 2018). This negative emotional engagement frequently fosters hostility towards opposing groups (Weeks, 2015), deepening socio-cultural divisions, while platforms like YouTube and Facebook use recommendation systems that can lead users down paths of increasingly extreme content, contributing to radicalisation and polarisation (Ribeiro et al., 2020). Allied to the increased prominence of emotional content, we know that anger makes political partisans more likely to participate in motivated reasoning when it comes to ‘fake news’ (Weeks, 2015). Social media are fertile ground for the spread of misinformation, which spreads faster and more widely than factual news (Vosoughi et al., 2018). This contributes to fervent disagreement over the veracity of everyday events, distorting perceptions of political issues and opponents, and exacerbating polarisation (Allcott and Gentzkow, 2017).

Are people who discuss politics on social media more ideologically extreme, and more hostile to people on the other side? Anecdotally, at least, we might expect this to be the case. We already know that social media users are not representative of wider populations (Larsson and Moe, 2012; Vaccari et al., 2013). For example, in the UK, they overall tend to be more left-wing and more politically-engaged (Mellon and Prosser, 2017). This is also the case on Twitter; Barberá and Rivero (2015) show that users in the United States and Spain are majority male and disproportionately live in urban areas. Surprisingly, nobody has conclusively identified whether those people who discuss politics on social media are more ideologically and affectively extreme, and whether there are differences in this extremity between partisan groups. This represents another crucial piece of the puzzle of whether social media sites facilitate political polarisation. Paper 2 of this dissertation takes existing research a step further by finding that British political Twitter is both ideologically *and* affectively more extreme than the rest of the population, and that this extremity is more pronounced among particular partisan groups.

Self-censorship

If social media are increasingly emotional and politically-extreme spaces, what happens to the people we don't hear? The third debate explored in this thesis examines the reasons why individuals choose to avoid discussing politics online. The vast majority of social media research inevitably focuses on the behaviour of those we are able to observe - but most users never or rarely post about politics (McClain, 2021). We know far less about why people are so reluctant talk about politics online, and why they disengage. In paper 3, I provide fresh insights into the relevance of existing theories of self-censorship in the digital age. I show that, on Facebook, the presence of contentious issues in mixed partisan groups leads to disengagement from online political discussion.

Disengagement from online discussion should be of particular concern to political scientists. Healthy democratic debate is desirable for several reasons. First, it helps foster an informed citizenry, essential for making educated voting decisions (Delli Carpini and Keeter, 1996). Second, it promotes civic engagement, leading to higher participation in democratic processes (Verba et al., 1995). Third, it also allows for the exchange of diverse viewpoints, enhancing understanding and tolerance (Dryzek, 2002). Finally, open political discourse helps hold leaders accountable and ensures that a variety of voices are heard, contributing to more representative and responsive governance (Fishkin, 2009). Countries around the world are increasingly politically- polarised (Mason, 2018; Iyengar et al., 2019), while people have become less willing to discuss political issues - particularly with those from outside their partisan group (Settle and Carlson, 2019). When individuals do not engage in political discussion, especially with those holding differing viewpoints, it can lead to echo chambers, which often reinforce existing beliefs without challenge. This can increase political polarisation as people become more entrenched in their views and less open to compromise or understanding other perspectives (Mutz, 2001; Garrett, 2009). If people are turning away from discussing politics online, we should attempt to find out why, in the hope of mitigating some of the most potentially damaging consequences of this disengagement.

There is a substantial amount of work which has investigated the phenomenon of disengagement and self-censorship, with most existing studies tending to focus on the existence of 'spirals of silence', a popular but contested theory proposed by Elisabeth Noelle-Neumann (1974). Spirals of silence occur when, out of fear of social isolation, individuals refrain from expressing their opinions if they perceive themselves to be in the minority. People constantly observe the social environment to gauge prevailing opinion and assess the risk of voicing dissent. As a result, dominant views become even more amplified, while minority views are suppressed, creating a self-perpetuating cycle. Although it is acknowledged that these mechanisms are present to at least some degree in offline settings (Salmon and Neuwirth, 1990; Glynn et al., 1997; Matthes et al., 2018), our understanding of how they function online is much more limited. It is challenging to test; as Hayes and Matthes (2014) note, it is almost impossible to construct a falsifiable hypothesis which conclusively proves or disproves its existence. Taken literally, the interdependence of such a wide range of conditions and assumptions render it almost impossible to test in its entirety.

Further challenges are presented by trying to apply the theory to 21st-century digital communication. The spiral of silence theory was originally conceived in an era of the mass media as arbiters of a consonant 'opinion climate'. Difficulties in reliably measuring media exposure raise questions about the accuracy reliability of obtaining one's 'quasi-statistical sense' (Noelle-Neumann, 1974) of an opinion climate[3]. This also raises questions about how SoS mechanisms function in an online environment, which consist of many smaller and localised majority opinion climates depending on the content or network selected by the user (Schulz and Roessler, 2012). This would suggest that spiral of silence theory, at least in its entirety, might have limited applicability to social media.

Nevertheless, research has found that individual components of the spirals of

[3] The design of paper 3 addresses this particular challenge in the following section.

silence theory do occur on social media. First, alignment with a perceived online opinion climate does indeed affects one's likelihood of expression. Consistent with existing research on the fear of social isolation (Neuwirth et al., 2007), Kushin et al (2019) found support for an increased fear of social isolation for those who believed their opinion to be in the minority on Facebook. In other words, people who felt their opinions of Hillary Clinton and Donald Trump were similar to others in their friend network felt less fear of social isolation and therefore an increased likelihood of expressing support for either candidate. Further, they found partisan attachment to be positively related with fear of social isolation. This research helps demonstrate that, on social media, people do consider their opinion environment, fear of social isolation is real, and that partisan attachment matters.

I adapt and build on key findings from the literature to argue for the primacy of two key factors in online self-censorship. First, and consistent with Elisabeth Noelle-Neumann, I argue that in order for an individual to disengage, the issue being discussed must be controversial and invoke questions of ethics or morality (Noelle-Neumann, 1974). When the issue being discussed is controversial and divisive, an increased likelihood of social sanctions causes people to disengage, and evidence exists that spirals of silence do occur in relation to issues such as abortion (Salmon and Neuwirth, 1990) or affirmative action (Moy et al., 2001). Further, while some research has tested the role of issue type in self-censorship (Gearhart and Zhang, 2018), there has been no comparison made between self-censorship over controversial and uncontroversial issues in an online environment. My research aims to address this omission.

Second, my argument diverges from spiral of silence theory on the role of group identity. Social psychology research (e.g. Festinger 1950; Schachter 1951; Taber and Lodge 2006) shows that individuals are more likely to focus on and engage with content that aligns with their political beliefs. When it comes to online self-censorship, group identity matters. Fox and Warber's (2015) study highlights how different levels of openness about LGBT+ identity affect individuals' ability to express themselves politically on social media. Those who were 'out' and openly expressed an LGBT+ identity, used social media to assert their views, while those who did not often remained silent due to perceived social pressure. Political group identity, certainly in the United States, tends to be represented by partisan attachment (Green et al., 2004; Iyengar et al., 2019), and I argue this group identity acts as a heuristic for understanding the perceived opinions of opposing political groups on specific issues.

1.2 Methodological and data contributions

Making a meaningful contribution to these debates, and furthering our understanding of the issues discussed, presents a number of methodological challenges. In addition to my substantive findings, I adopt a range of techniques and examine novel datasets in each of the three papers presented.

Measuring large-scale shifts in emotive rhetoric over time

There is a robust and growing body of research which examines large-N data of emotion in elite parliamentary rhetoric, and how it changes in different environments or situations (Rheault et al., 2016; Slapin et al., 2018; Osnabrügge et al., 2021). Continually-evolving computational methods also allow analysis of the extremely large amounts of observational data generated online. While these studies have, for example, examined millions of tweets to detect changes in political hate speech (Siegel et al., 2019; Brown and Sanderson, 2020), they tend to be restricted to particular periods of time or to specific actors of interest. The availability of this data represented a unique opportunity to study the entire lifespan of political communication on a social media site. To measure the changing rhetorical behaviour of political elites, I gathered all tweets posted by elected MPs in the United Kingdom between May 2007 and December 2019. From an original dataset of more than 8 million tweets, I removed retweets and quote tweets to isolate over 3 million 'original' tweets by sitting MPs. Identifying a change in emotional rhetoric over time presented challenges, however. Certain political topics are discussed with greater levels of emotion than others; the issue of abortion, for example, would generally invoke greater strength of feeling than discussion of macro-economic policy. Overall levels of emotion, then, might fluctuate as certain issues increased and decreased in salience over the 12 years of interest. Therefore, in order to control for this heterogeneity, each tweet had to be broadly categorised according to the issue it discussed, and incorporated into a fixed effects OLS regression model. While the manual categorisation of even a representative sample of my overall dataset would be infeasible, computational methods again afford a solution to challenges presented by such a large dataset. Alongside the large-scale text analysis, I employed Latent Dirichlet Allocation (LDA), an unsupervised probabilistic model for uncovering topics from a collection of otherwise unstructured collection of text documents (Blei et al., 2003). Applying topic modelling to such a large dataset represents an innovative solution to the problem of issue-based emotional heterogeneity.

Linking survey data with social media behaviour

Paper 2 represents a methodological contribution to the study of polarisation on social media by bridging a gap between the type of observational 'digital trace' data (e.g. Barberá, 2015) gathered in paper 1 and survey methods traditionally used to study polarisation (e.g. Banda and Cluverius 2018; Butters and Hare 2022). While the former can offer rich observational data on large-scale human behaviour, it is difficult or even impossible to derive truly accurate representations of ideological or affective attitudes. By contrast, survey methods can provide relatively precise measures of values and affect, but are limited by their ability to measure (usually self-reported) social media use. Recognising this gap, scholars have linked survey data with online behaviour to explore, for example, elite campaign behaviour (Karlsen and Enjolras, 2016) and the extent to which people live in online ideological echo chambers. Together with my co-authors, we shed new light on the link between political extremity and social media use. We do this by measuring a range of ideological and affective attitudes through a nationally-representative survey of the UK population, and link these responses to actual observable Twitter behaviour, evaluating the relationship between their political extremity and online activity.

Identifying factors in online self-censorship with a lab-in-the-field experiment

Most social media research inevitably focuses on those people we are most likely to hear in online political discussion. By shifting focus to those who *disengage*, my third paper gains fresh perspective on how and why this happens. Existing studies, which mainly examine the existence of a 'spiral of silence' (Noelle-Neumann, 1974), face methodological challenges related to ecological validity. Primarily, a reliance on survey data (e.g. Gearhart and Zhang, 2014, 2015; Chan, 2021) leaves a gap in our understanding of how these mechanisms might occur in real-world scenarios. Using either surveys or lab experiments not only removes participants from their networks, but also removes their fear of social isolation: a key component of the spiral of silence. Therefore, to effectively identify the mechanisms which underpin self-censorship on social media, it is imperative to conduct tests directly on social media sites. To my knowledge, no studies have conducted empirical tests of self-censorship in the field.

To address this gap in the literature, I conduct an innovative pre-registered lab-in-the-field experiment which replicates a common everyday experience on social media. Using several of Facebook's native features allows me to control important aspects of my experiment and find innovative solutions to further experimental challenges. First, I used Facebook groups as discrete treatment conditions, varying the partisan composition of each group and exposing participants to either contentious or consensus political issues. Second, to encourage daily participation, I used Facebook polls on various topics, prompting participants to visit their assigned group. This approach enabled me to monitor compliance daily and, if needed, remind those who did not participate after the first day of their commitment. Third, to indicate the partisan composition, and by extension the opinion environment, of each treatment group, on the Tuesday of each study period I posted a poll asking "Do you think of yourself as being closer to the Republican or Democratic party?". This poll was 'pinned' to the top of each group as soon as it was posted, so that participants would have to scroll past it each day[4] to complete their task. Further, the 'seen by' metric for each treatment, allowed me to check that each participant had viewed the post in question.

A second major challenge lies in accurately measuring willingness to self-censor in hypothetical situations. Matthes and Hayes (2014), in examining research on spiral of silence mechanisms, point to difficulties in measuring subjects' willingness to express an opinion. Building on Noelle-Neumann's original conception of a 'train test', where respondents are asked if they would speak with a fellow train passenger expressing a strong opinion on a particular issue, researchers have variously used other hypothetical situations such as an airplane flight (Lasorsa, 1991), speaking with a reporter (Shamir,

[4] In each group, settings were amended so that content was presented to participants in reverse chronological order, i.e. with the most recent post appearing first.

1997), or a social gathering (Moy et al., 2001). The problem with these situations is precisely that they are hypothetical; by measuring engagement directly, I can accurately test respondents' willingness to make their opinion on a subject known. Using Facebook's engagement metrics allowed me to gather rich data on how participants engaged with content on both an individual and group level. Additionally, as Fox and Holt (2021) highlight, social media sites provide various methods for self-expression; by analysing willingness to both react and comment, this presents a more comprehensive and rounded understanding of what influences different types of interaction with political content.

The methodological innovations presented across these three papers address significant challenges in studying political communication in digital spaces. By combining large-scale data analysis, the linking of observational data with traditional survey methods, and novel experimental designs, this thesis offers a broad-based approach to understanding how people engage with politics online. The following sections provide a road map of the thesis structure, outlining the contributions and findings of each paper in detail.

1.3 Road map

This thesis is comprised of five chapters. In the next section, I briefly summarise its three individual research articles, before presenting each paper in turn. This is followed by a discussion of their limitations and future research avenues, and a conclusion which explores the broader normative and policy implications of my findings.

Paper 1: Once more with feeling: How political elites are changing their rhetorical behaviour on Twitter

Exploring the interaction between social media and emotive rhetoric, the first paper in this thesis makes a unique contribution to the political communication literature by building on three strands of research. First, rhetorical adaptation has been examined extensively in an offline context, primarily with regard to parliamentary speech (Rheault et al., 2016; Slapin et al., 2018; Crabtree et al., 2020), but we know much less about how the behaviour of politicians interacts with the advent of new communication technology. I connect this with a literature exploring elite behaviour on social media (e.g. Bessone et al., 2019; Ernst et al., 2019; Martella and Bracciale, 2022) and show that politicians adapt their speech over time and provide a detailed empirical analysis of how language has evolved over the lifespan of a new communication medium. By examining a large dataset of over 3 million tweets by UK Members of Parliament between 2007 and 2019, I offer evidence that politicians have used increasingly emotive rhetoric during their time on Twitter. This finding enriches our understanding of how digital platforms influence not just the content of political communication, but also its tone and emotional intensity. This is particularly valuable in the broader context of political science, where much of the focus has been on traditional media or face-to-face interactions.

Further, by demonstrating that tweets with higher emotional intensity receive more interactions, and that this likely incentivises further use of emotive language, I also contribute to the literature on the economics of attention in digital media (Berger, 2011; Berger and Milkman, 2012; Brady et al., 2017; Weismueller et al., 2022). This not only adds to the consensus that digital platforms amplify certain types of content over others, and in the process potentially skewing political discourse, but also advances our understanding of how politicians strategically adjust their rhetoric for maximum impact. Finally, by investigating whether social media platforms like Twitter exacerbate ideological and affective extremity, this paper also intersects with the literature on political polarisation. Emotionally-charged language often appeals to base instincts, reinforcing existing beliefs, and hardening both ideological and affective divides (Iyengar and Westwood, 2015). By linking the emotional tone of political communication to engagement metrics, and by exploring the broader implications for polarisation, my work provides evidence supporting the argument that social media may intensify partisan divides. In connecting micro-level rhetorical shifts with macro-level phenomena like political polarisation, I offer a comprehensive view of one of the consequences of digital political communication.

Paper 2: Extremely Online? Ideological and Affective Polarisation on British Political Twitter

The second paper of this thesis offers new insight into the extent of political polarisation among Twitter users compared to the general voting population. Analysing a novel dataset which links a nationally-representative survey to the Twitter accounts of respondents, we find that Twitter users are more ideologically and affectively-extreme than non-Twitter users. Further, of those people *on* Twitter, those who use it to discuss politics are more ideologically-extreme than those who do not, while 'Remain' supporters are more affectively polarised than their non-political counterparts. Finally, we show that politically-extreme users tweet about politics more and share more negatively-biased partisan content.

This paper represents an advance on existing research on social media polarisation in two ways. Our integration of survey responses with digital trace data acknowledges the limitations of obtaining measures of ideology and affect exclusively through either surveys (e.g., Banda and Cluverius, 2018; Butters and Hare, 2022) or digital trace data (e.g., Barberá, 2015; Yarchi et al., 2021). Although surveys can provide fairly accurate insights into ideology and affect, self-reported data on social media usage often lacks reliability (Henderson et al., 2021). Conversely, extracting accurate ideological or emotional measures solely from digital trace data is often very difficult. Linking validated measures of ideology and affect with actual observable usage data from Twitter gives us a more thorough understanding of the ideological and emotional extremes among respondents and their online activities. We also offer an additional advance on existing descriptive studies of social media polarisation by examining not only ideological, but affective polarisation.

In exploring the link between extreme Twitter users and the types of tweets they produce, we find that more ideologically-extreme users on the 'Remain' side of the Brexit debate are more likely to tweet about politics, and both Remainers and Leavers with strong negative feelings towards the 'other side' are more likely to tweet about politics. This highlights that, not only do we see an asymmetry in the most dominant and extreme voices on platforms like Twitter, but that social media can play an active role in the articulation of emergent political identities. Overall, these findings underscore the heightened potential for polarisation on social media, revealing that individuals with stronger ideological convictions are more likely to dominate political discourse. By demonstrating that these users are more inclined to tweet and share negatively-biased partisan content, we offer evidence that we disproportionately hear from the most ideologically and affectively-extreme users of platforms like Twitter. Social media not only promote homophily and reinforce selective exposure but also contribute to a more polarised and less deliberative public sphere.

Paper 3: Opting out of political discussion on Facebook

My third paper tests the effect of contentious political issues and group heterogeneity on disengagement from discussion. In a pre-registered lab-in-the-field experiment directly on Facebook, I do not find that either of these conditions alone lead to decreased likelihood of either commenting or reacting, but that their combination reduces discussion. The findings of this study have important implications for the study of online self-censorship by expanding on, and deepening, previous research. First, I offer a new theoretical perspective on Elisabeth Noelle-Neumann's concept of the 'spiral of silence' by refining and adapting its key tenets, while also acknowledging the challenges to its validity presented by new communication technology. Highlighting how the dynamics of mass political communication have changed since the theory was conceived in the 20th century, I contend that features of different social media networks can either encourage or inhibit discussion of contentious topics. Additionally, I draw on the literature in political psychology and social influence to argue that group identity plays a crucial role in determining whether individuals engage in or disengage from political debate.

Second, this paper also offers a methodological advance on previous studies of online political self-censorship. Designing a pre-registered lab-in-the-field experiment which closely mirrors everyday social media experiences, I conduct a more ecologically-valid investigation of the mechanisms which underpin online political disengagement. By examining the individual and combined effects of exposure to contentious political issues and partisan group composition, I isolate these factors and provide deeper insight into their interaction and its impact on disengagement. Additionally, this study expands on existing research by measuring a wider range of political engagement, including comments and reactions. Instead of focusing solely on the willingness to comment, I include reactions such as likes and emotive responses, offering a more comprehensive understanding of how people engage with political content on social media. Lastly, by including four distinct political issues and examining how opinions are distributed among participants, I address the possibility that self-censorship might be limited to specific issues, thereby enhancing the robustness of my findings. As such, compared with existing studies, this paper provides a more realistic and contextually-relevant assessment of self-censorship on social media. Third, in light of the influence of social media on public discourse, and the escalating concern over political polarisation, these findings have important implications for wider democratic debate. They suggest that private or semi-public social media platforms might be exacerbating division and discouraging debate instead of promoting inclusive dialogue; these insights are critical for grasping the potential impact of digital communication on the integrity and functioning of democratic processes.

2 Once more with feeling: How political elites are changing their rhetorical behaviour on Twitter

Abstract

The rise of social media as a communication tool has created an entirely new way for politicians to communicate with voters. However, little is known about how politicians have adapted their rhetorical behaviour over the course of its existence. Has the emergence of Twitter been linked to shifts in elite communication strategies? Analysing a dataset of over 3 million tweets by UK Members of Parliament between 2007 and 2019, I find a small but steady increase in their use of emotive rhetoric. I demonstrate that this was driven to varying extents both by an adaptive process, as MPs changed their behaviour over time, and by replacement, as new cohorts of MPs were elected in 2015 and 2017. Finally, to explore the mechanisms behind these changes, I uncover a relationship between emotional intensity and levels of engagement on Twitter, showing that negative tweets are rewarded with a greater number of interactions. This suggests a 'feedback loop' has reinforced an adaptive change in rhetorical behaviour. These results contribute to the study of online political rhetoric, highlighting the central role played by social media in the changing tone and scope of democratic debate.

2.1 Introduction

How has social media influenced the use of emotional rhetoric by politicians? Emotions such as anger (Valentino et al., 2011), positivity (Kosmidis et al., 2019) or fear (De Castella et al., 2009) are regularly leveraged by prominent political elites, carefully selecting emotional over 'rational' or 'logical' language in pursuit of their goals[1]. We know that appeals to emotion can result in a range of political outcomes (Marcus, 2000; Brader et al., 2011), yet we know less about how the use of these appeals has been shaped by new communication technologies.

For centuries, the advent of new media has influenced political persuasion; Gutenberg's press moved verbal appeals to the printed page, allowing increasingly literate populations to read the revolutionary scientific, philosophical, and political ideas which underpinned the Enlightenment in Western Europe (Eisenstein, 1980). In the 20th century, broadcast media carried the voices of politicians into the homes of millions, and placed renewed emphasis upon oratory as a political skill. Indeed, the televised U.S. presidential debate between Kennedy and Nixon in 1960 is often perceived as a defining moment in political communication (Druckman, 2003). More recently, social media have become an almost essential part of a politician's toolkit, with Twitter[2] used regularly by over 85% of the United Kingdom's members of parliament. Twitter afforded politicians something entirely new: the ability to communicate directly with large numbers of voters in real time. Free from journalistic gatekeepers, political elites gained unprecedented control over their messaging — choosing the content, timing, and tone entirely on

[1] For example, "Enough is enough, time to tell the arrogant, unelected EU bullies where to go. No British Prime Minister should be treated like this." (Farage, 2018), "The Red Wall voted to take back control of immigration, not to surrender the English Channel to criminal trafficking gangs." (Farage, 2022)

[2] Now known as X; as this research note focuses on a period between 2007 and 2019, I refer to it as Twitter throughout.

their own terms. Its popularity increased substantially from 2008 onwards and it quickly became a core communication tool for the UK's legislators; Figure 2.1 details the date at which MPs first created their Twitter account, showing a surge in joining the site between 2009 and 2011. Over 85% of MPs elected in the 2019 UK general election had an active Twitter account and used them on a daily basis to spread their messages directly to large audiences. Figure 2.2 shows that, as a group, members of parliament rapidly increased the number of tweets shared from their accounts in the first half of the last decade, with this number reaching over 500,000 in both 2016 and 2019. While Twitter is unrepresentative of the population, and used by just under 20% of UK adults (Mellon and Prosser, 2017), importantly it is used extensively by politically-engaged people, playing an agenda-setting role in news and public policy (Gilardi et al., 2022).

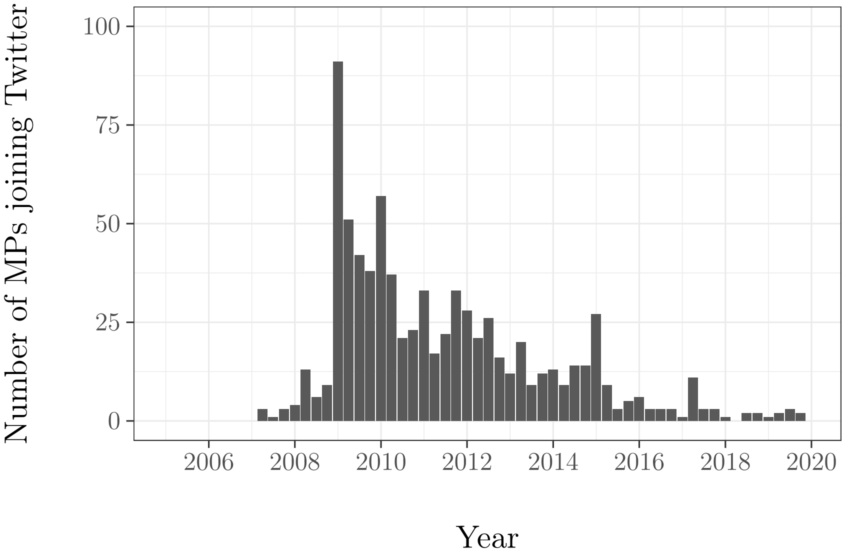


Figure 2.1: Number of MPs joining Twitter, 2005-2019

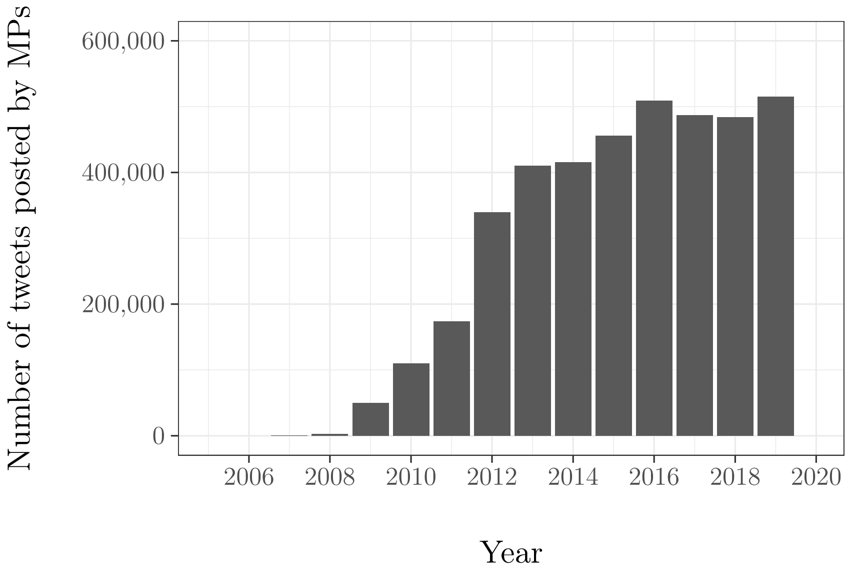


Figure 2.2: Number of tweets posted by MPs, 2005-2019

I argue that politicians have, over time, become more emotional in the rhetoric they use on Twitter. Further, I show it is a platform which rewards emotive language and suggest that, as strategic rational actors, MPs have learned to employ increasingly emotional rhetoric to maximise the reach of their messages and raise their electoral profiles. These arguments build on, and contribute to, three strands of literature. First, a large body of literature has examined the effect of new forms of mediated communication, particularly television, on the increasing prominence of emotion in elite rhetoric (e.g. Iyengar 1994; Druckman and Holmes 2004; Brader 2005). From this, we know that the introduction of these new media had - and continue to have - profound effects upon the behaviour of politicians. However, we still know relatively little about the effect of newer, online media upon the type of language they use. Second, research has shown that politicians adapt their rhetoric in different environments. For example, styles of parliamentary speech change in response to external events (Rheault et al., 2016), potential audience (Osnabrügge et al., 2021), or ideological position and status within a political party (Slapin et al., 2018). Third, research has shown that language can have a wide range of important effects upon online political behaviour. For example, we know that morally and emotionally-charged content spreads further and faster (Berger and Milkman, 2012; Brady et al., 2017) than non-emotional (or 'rational') content, but have yet to establish if elites have adapted their rhetoric to capitalise on this phenomenon and maximise the reach of their messages.

To test my arguments empirically, I use unsupervised dictionary methods on a large-N dataset of all tweets made by U.K. Members of Parliament between the 15th of May, 2007[3], and the 12th of December, 2019, the date of that year's General Election. I examine more than three million tweets sent by serving MPs, measuring emotive rhetoric with the VADER automated sentiment dictionary (Hutto and Gilbert, 2014). I find the language used by MPs on Twitter has become more emotional over time. To demonstrate that this rhetorical shift was driven both by an adaptive process, as politicians gradually change their language, and by replacement, as newer cohorts are elected, I model emotion in tweets made by MPs elected in the 2015 and 2017 General Elections against tweets made by MPs from the older cohorts of 2005 and 2010. Finally, while stopping short of making causal claims, I suggest this adaptation is encouraged by Twitter's relationship between emotion and levels of engagement, incorporating 'likes', 'retweets', 'replies', and 'quote tweets'. This feedback loop of positive reinforcement, I argue, incentivises politicians to become more emotive in their tweets.

[3] The first tweet sent by an MP: Liberal Democrat Duncan Hames tweeted "Testing my Twitter account over lunch".

2.2 New technologies and political debate

An extensive body of literature has examined the relationship between the introduction of new technologies and their effect on elite rhetorical behaviour. The introduction of mass radio communication in the early 20$^{\text{th}}$ century enabled political appeals to transcend borders and reach populations previously inaccessible by print media, quickly becoming a central part of any high-profile politician's communication strategy. Franklin D. Roosevelt's 'fireside chats' during the 1930s and 1940s allowed the U.S. president to speak directly to millions of Americans, heralding the start of a changing relationship between politicians and voters (Winfield, 1994). While this new medium offered unprecedented opportunities, it also imposed new demands on the capacity of politicians to engage and persuade the electorate, favouring those able to elicit emotions through their rhetoric. Marshall McLuhan (1964) famously described radio as a 'hot' medium, noting the ability of spoken addresses by populist demagogues to inflame passions in mass audiences. The eventual introduction of television, compared with the audio technology it superseded, further changed the way in which audiences evaluated the relative qualities of politicians. By priming people to rely more on perceptions of personality, televised debates emphasised image at the expense of issue agreement (Hellweg et al., 1992; Druckman, 2003), accelerating the 'personalisation' of candidates (Iyengar, 1996; Clarke and Stewart, 1998). In other words, television, compared with more traditional forms of communication, placed far greater emphasis upon the Aristotelian idea of emotion (*pathos*) and credibility (*ethos*) than on logic (*logos*) (Aristotle, 1909). If television moved political elites away from 'rationality' towards more emotive and personal appeals, have newer communication media reinforced or mitigated this effect?

The way that digital and social media platforms are designed shapes political behaviour in many ways; for example, changes to government websites can ‘nudge’ particular actions (Hale et al., 2018) and, more specifically, influence the performance of e-petitions (Margetts et al., 2015). On Twitter, a timeline algorithm ensures that content with higher levels of ‘engagement’ (e.g. likes, retweets, quote tweets and replies) is visible to a greater number people[4], increasing the probability of this content being shared, which in turn increases the likelihood of exponential or ‘viral’ information spread. This is demonstrated by the correlation between follower counts and engagement levels (Lu et al., 2014; Margetts et al., 2015; Hale et al., 2018), ensuring that Twitter remains an elite-led medium where the socially ‘rich’ tend to get richer. We also know that elite behaviour is influenced by social media; for example, research by Petrova et al. (2021) shows that opening a Twitter account helps candidates running for US Congress get higher campaign contributions, while Bessone et al. (2019) found Brazilian legislators became more active on Facebook as their constituencies were connected to 3G mobile technology. At the same time, the structure of these digital communication channels plays a pivotal role in the framing and content of their messages. For example, Brady et al. (2017) make a clear distinction between *message* factors (characteristics of the medium) and how they interact with *source* factors (characteristics of elites) to impact the diffusion of messages though online social networks. However, we know relatively little about how social media might be shaping their rhetorical behaviour.

[4] Since Elon Musk’s takeover of Twitter in 2022, the manipulation of this algorithm has become increasingly controversial, with research finding that misinformation and right-leaning content is artificially amplified (Corsi, 2024).

2.3 Emotion in online political rhetoric

Emotive rhetoric, as in Aristotle's conception of *pathos*, can generally be defined as a message intended to stimulate an emotional reaction in its recipient (Marcus, 2000; Havas and Chapp, 2016). Political appeals regularly leverage emotions such as fear or enthusiasm (Lazarsfeld et al., 1968; Brader, 2005), negativity (Baumeister et al., 2001; Soroka, 2006), or morality over rationality (Haidt, 2012; Brady et al., 2019). These emotions can shape political outcomes in a number of ways. First, high-arousal emotions motivate ordinary citizens to engage with, and participate in, political activity (Marcus, 2000); an imperative which is particularly pronounced when faced with negative emotions such as anger and enthusiasm (Brader, 2005; Valentino et al., 2011). Second, an interaction between emotional rhetoric and selective exposure can result in the rapid spread of information, particularly when framed upon partisan divisions (Halberstam and Knight, 2016), deepening the problem of political polarisation (e.g. Colleoni et al., 2014). Finally, and perhaps most importantly for elites, emotional appeals can influence cognitive processes and positively influence the persuasiveness of a message (Petty and Cacioppo, 1986; Brader, 2005; Petty and Briñol, 2015). Affective language frequently exploits latent cognitive bias in an attempt to shape political behaviour, being used as an important strategy to convince people of the veracity of a message (Bless et al., 1992; Charteris-Black, 2011; Arceneaux, 2012).

Adopting a conventional rational choice approach (Downs, 1957), which views elites as utility maximisers, politicians strategically employ everything at their disposal to build electoral appeal. In terms of communication, this means continually seeking the most effective methods of conveying their rhetorical appeals: to potential voters, other political actors, opinion leaders, and journalists. We might expect, then, that social media encourages politicians to become more emotional for three main reasons. First, emotional and moral words are given greater priority than non-moral or non-emotional words in early cognitive processes and therefore have greater capacity to capture attention (Brady et al., 2017, 2020). This type of content consequently tends to spread further and faster (Berger and Milkman, 2012; Stieglitz and Dang-Xuan, 2013; Brady et al., 2017; Weismueller et al., 2022). It follows that, if the aim of politicians is to communicate their messages to as many people as possible, greater emotion will maximise the potential reach and influence of their tweets. Second, elites have traditionally communicated largely through speeches, the press, and broadcast media. Twitter has given politicians a direct connection with their audiences without the filtering effect of journalistic intermediaries who might distort or moderate the affective intensity of their message (Bennett and Manheim, 2006). We know, for example, that political parties use emotive populist rhetoric more readily on social media than through traditional channels (Ernst et al., 2019). Third, and finally, given that Twitter audiences have increased substantially over the last 15 years, we can also build on work by Osnabrügge et al. (2021) showing that legislators use more emotional rhetoric when addressing larger audiences in higher-profile debates. Consequently, I argue that political elites have, over time, increasingly leveraged these emotional imperatives as rhetorical devices for strategic gain:

- H_1: *Politicians in the United Kingdom have adapted their rhetorical strategies on Twitter to become increasingly emotional over time.*

Is there a correlation between the emotional intensity of tweets shared by MPs and the engagement they receive, potentially shedding light on this adaptive process? Answering this question may shine a light on the existence of a potential incentive for political elites to adapt their rhetorical styles on Twitter. We know that negative political content is associated with greater engagement (Stieglitz and Dang-Xuan, 2013; Heiss et al., 2019; Antypas et al., 2023), and we know that Twitter's real-time 'feedback loop' allows elites to see which messaging strategies resonate with their followers, and modify their language accordingly. This process, I argue, occurs for both existing MPs and those with ambitions of office:

- H_2: *An increase in emotional rhetoric over time is driven both by an adaptive process, where elites have gradually refined the language of their tweets, and by a process of replacement, as new MPs are elected.*

Social media platforms are constantly optimized to tap into our underlying psychological instincts, profoundly influencing how we perceive political issues (Muchnik et al., 2013; Hale et al., 2018). Faced with an overwhelming volume of daily information, social media users are often influenced in heightened emotional states. In uncertain situations, individuals rely on heuristics — or mental shortcuts which simplify decision-making (Tversky and Kahneman, 1974; Chaiken, 1980; Petty and Cacioppo, 1986). Given the vast amount of content available and the rapid pace at which it is delivered, these psychological tendencies significantly affect how people filter, process, engage with, and share information on social media (Mosleh et al., 2020). On Twitter, users can react to tweets instantly with a single click or tap, often before they have fully evaluated, researched, or rationalised the content. This immediacy amplifies the influence of emotional impulses. At the same time, Twitter's algorithm[5] is designed to prioritize engagement — measured through retweets, likes, replies, and quote tweets — which increases a tweet's visibility. For instance, if user 1 follows user 2, and user 2 likes' a tweet from a politician not followed by user 1, user 1 is likely to see that

[5] At the time of writing, May 2023.

politician's tweet. This creates a strong link between levels of engagement and the spread of information. The concept of a feedback loop is central to this process: politicians can track platform metrics in real time, observing how well their tweets resonate with their audience and adjusting their rhetorical strategies accordingly. Existing research shows that emotional content spreads further and faster online (Stieglitz and Dang-Xuan, 2013; Heiss et al., 2019; Antypas et al., 2023). Building on this, and recognising that politicians often adapt their rhetorical behavior to suit different audiences, I argue that Twitter's feedback loop incentivises elites to use rhetorical styles which generate the greatest engagement. Emotional appeals, in particular, are more likely to resonate with Twitter users, and politicians strategically leverage them to achieve their objectives:

- H_3: *The level of emotive rhetoric used by MPs on Twitter is positively associated with levels of engagement.*

2.4 Data and methods

With over 330 million users globally and approximately 500 million tweets sent daily, Twitter is an ideal platform for large-scale observational studies of human behaviour. Its open-access data[6] and widespread use among politicians offer a unique opportunity to study digital discourse involving elected representatives. To investigate shifts in political rhetoric over time, I constructed a dataset comprising tweets from UK Members of Parliament (MPs) serving between the 15th of May, 2007 and the 12th of December, 2019. Analysing tweets across this period serves two primary goals. First, Twitter's popularity as a political communication tool has grown markedly,

[6] At least prior to Elon Musk's takeover in 2022.

with over 85% of MPs active on the platform by 2019. Second, this timeframe spans a broad spectrum of political issues, each potentially impacting the emotional tone of tweets. By employing topic modelling[7], I can control for issue-specific effects, revealing how rhetoric has shifted independently of the political agenda. Furthermore, this study period includes three UK General Elections and four distinct cohorts of MPs, allowing for comparisons between the rhetorical styles of seasoned MPs and newer members of parliament. From an initial set of 8,632,561 tweets, this study focuses on the rhetoric embedded within MPs' original tweets, excluding retweets, duplicates, single-word posts, non-alphanumeric characters (e.g., @ and #), and URLs. After processing, 4,011,139 tweets remained, and were parsed for natural language processing. Table 2.1 details tweet counts per party during this period[8].

Table 2.1: Total number of tweets by party

Party	Tweets
Labour	2, 044, 648
Conservative	1, 315, 049
Liberal Democrat	299, 489
SNP	279, 030
Sinn Féin	26, 841
Green Party	15,961
Plaid Cymru	10, 802
DUP	9, 429
SDLP	2,933
UUP	866
Respect	210
Total	4,011,139

[7] This is discussed further in the 'Robustness and alternative mechanisms' section and the appendix.

[8] N.B. UKIP's sole elected MP, Douglas Carswell, did not tweet during his term.

2.4.1 Measuring emotive rhetoric

To track shifts in emotional valence over the study period, I employ unsupervised dictionary methods. Quantitative text analysis has traditionally been applied to reveal latent political characteristics within party manifestos (Laver et al., 2003), press releases (Grimmer, 2010), and government bills (Martin and Vanberg, 2011). However, the digitisation of almost all political communication offers unprecedented opportunities to examine elite rhetoric. Here, I quantify emotive rhetoric in tweets using the VADER sentiment dictionary, which provides a compound sentiment score per tweet (Hutto and Gilbert, 2014). While several validated sentiment dictionaries, such as LIWC (Pennebaker et al., 2001), SentiStrength, and Lexicoder (Young and Soroka, 2012), are suited for social media, VADER was selected as the most appropriate tool for this analysis. Its advantages make it ideal for studying Twitter rhetoric: it is externally validated and widely used in social science, and it performs exceptionally well on brief, informal social media texts. Further, its lexicon captures sentiment-specific expressions prevalent on Twitter, including amplifications, booster words (e.g., “really really good”), sentiment-bearing acronyms (e.g., “LOL”, “OMG”), and slang terms such as “nah” and “meh”. Additionally, VADER’s linguistic rules account for negations—words like “but” and “however” that alter sentiment or polarity. It scores each word on a scale from -4 to +4, with -4 representing intense negative sentiment and +4 representing intense positive sentiment. A tweet’s overall sentiment is derived from its compound score — a normalised, weighted sum of individual word scores. To gauge emotional intensity, I use the absolute values of compound scores, standardising each tweet with a valence score. Hutto and Gilbert’s formula for calculating compound sentiment is shown in Equation 2.1:

$$\frac{x}{\sqrt{x^2 + a}} \tag{2.1}$$

where x is the sum of the sentiment scores of words included in the body of text and a is a normalisation parameter set to 15. The compound value returned is an overall score in the range [-1,1]. Further, positive (*pos*), neutral (*neu*), and negative (*neg*) scores represent ratios for the proportions of text in each document (in this case, each tweet) that fall in each category. Accordingly, for each document, the *pos*, *neu*, and *neg* scores will sum to 1. These values afford greater insight into the type of sentiment conveyed, and its context, in any given document.

Compound, *pos* and *neg* scores are demonstrated in Table 2.2. Firstly, Nick Boles' tweet (1) from 2019, which yields a compound score of 0.992, contains a number of high-valence words such as brilliant, kind, and thoughtful alongside the "*very very*" amplification of funny. The second half of the tweet carries a darker, but no less emotive, sentiment with the inclusion of the phrase "*dodged a bullet*" in reference to Michael Gove's elimination from the Conservative leadership race. Example (2), of a high positive score from Kate Green MP in 2012, contains an extremely high proportion of positive terms in the tweet: "*beautiful, relaxed, stunning, inspired...*". Conversely, of the 6 words in veteran Conservative Party MP Nicholas Soames' colourful tweet (3), all could be considered negative. In contrast to examples 1-3, Andrew Stephenson MP uses entirely functional, neutral language to tweet about ministerial business in the House of Commons. As such, the absence of any terms registered as either positive or negative in VADER's lexicon shows *pos*, compound, and *neg* scores of 0.000.

Table 2.2: Examples of emotive, positive, negative and neutral tweets

1) **High compound score** (20/6/2019)

compound	positive	negative	neutral	text	Twitter account
0.992	0.518	0.482	0.000	We may now be on opposite sides but I will always be proud to call @michaelgove my friend. He is brilliant, brave, mischievous, kind, thoughtful, affectionate and very very funny. In the fullness of time I hope he will realise that today he dodged a bullet.	Nick Boles MP

2) **High positive score** (27/7/2012)

compound	positive	negative	neutral	text	Twitter account
0.917	0.928	0.000	0.072	Beautiful, relaxed, stunning, inspired... #olympics2012 xx	Kate Green MP

3) **High negative score** (4/10/2016)

compound	positive	negative	neutral	text	Twitter account
-0.912	0.000	0.872	0.128	@GregoryTaylor86 liars shysters amoral cynical shits #brutes	Nicholas Soames MP

4) **High neutral score** (11/8/2011)

compound	positive	negative	neutral	text	Twitter account
0.000	0.000	0.000	1.000	the Prime Minister has just taken 160 questions in 165 minutes, now the Chancellor's statement on the Global Economy	Andrew Stephenson MP

Table 2.3 summarises the mean compound, *pos* and *neg* scores for tweets made by each party represented in the House of Commons during this period, showing tweets by Conservative MPs generally score higher in compound and positivity, with lower mean negativity than most other parties. In line with existing research (e.g. Crabtree et al., 2020), I suggest this imbalance can be attributed in part to the incumbency and opposition status of the respective parties, where the opposition is much more likely to use tweets to attack government. In this case, higher recorded mean *neg* scores attached to tweets by Labour MPs are partly a result of the number of years spent in opposition by their party during this period. Conversely, government MPs can be expected to spend more time highlighting policy proposals and achievements[9].

Table 2.3: Most emotive parties on Twitter, ordered by mean compound score

Party	Mean compound	Mean pos	Mean neg
Ulster Unionist	0.364	0.215	0.041
Conservative	0.316	0.175	0.045
Green	0.299	0.233	0.070
Liberal Democrat	0.269	0.202	0.046
Democratic Unionist	0.254	0.165	0.054
Sinn Féin	0.252	0.160	0.042
SDLP	0.231	0.150	0.056
SNP	0.206	0.155	0.055
Labour	0.196	0.162	0.062
Plaid Cymru	0.125	0.088	0.037
Alliance	0.104	0.128	0.073

[9] This is further supported when sentiment metrics are examined at an individual level. Tables can be found in the appendix.

2.5 Results

In this section, I first present a time series plot which uncovers overarching trends in the emotional intensity of MPs' tweets over time. Next, I present findings from Ordinary Least Squares (OLS) regression models, which shed light on the evolution of emotive rhetoric, positivity, and negativity in MPs' tweets. Finally, I employ OLS regression to examine the relationship between emotive rhetoric and Twitter engagement. In the following section, I detail the steps taken to address potential threats to validity. Figure 2.3 illustrates a simple moving average of compound scores over time, showing a clear upward trend in emotive rhetoric among MPs. Autocorrelation testing reveals strong persistence in mean compound scores between consecutive days (lag-1 autocorrelation = 0.24)[10]. To account for this short-run autocorrelation, I applied a 30-day moving average, which reveals a consistent increase in compound scores over time, warranting further examination.

To examine this in greater detail, I use linear regression to model the relationship between time and levels of emotive rhetoric. Table 2.4 summarises these results. I run four different models with different specifications, with standard errors clustered at MP levels in each. Model 1 is a simple baseline model including the explanatory variable *decades elapsed* (a continuous measure of time in decades), with compound, *pos* and *neg* scores. I observe a small but statistically-significant increase in both compound and pos scores, while the very small increase in negativity observed is not robust. In Model 2, I add controls for age and gender to my estimations, under the assumption that MPs with particular characteristics may be more likely to tweet emotionally. Compound score is still positively-associated with decades elapsed and remains statistically-significant, while the effect disappears for the mea-

[10] This is expected, as emotive political issues often extend beyond a single day. Additionally, a 7-day autocorrelation pattern may reflect recurring weekly events, such as Prime Minister's Questions, which drive more emotive tweets.

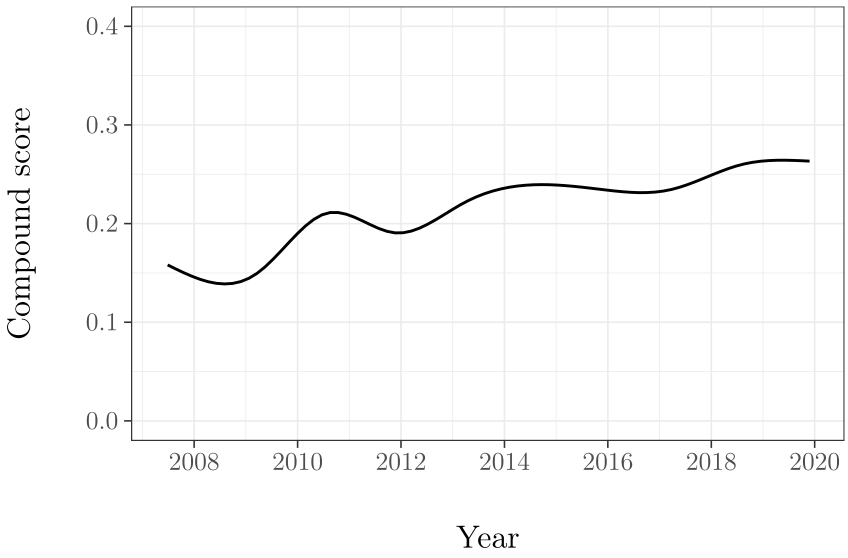


Figure 2.3: 30-day simple moving average of compound score

sures of *pos* and *neg* proportions. In Model 3, I retain the controls for age and gender and introduce party fixed effects. Party fixed effects account for any partisan heterogeneity in rhetoric due to a change in the composition of MPs[11]. For example, opposition MPs may be more likely to tweet in a more emotionally-extreme or negative fashion in criticism of the government, while ruling party MPs may be more likely to tweet positively in support of the government's achievements. Finally, in Model 4, *topic* fixed effects are incorporated to account for heterogeneity in emotion used by MPs when discussing more emotive political issues[12]. While the introduction of these controls in models 3 and 4 attenuates the effect of time on compound and *pos* score very slightly, these relationships remain positive and statistically significant. This indicates that the observed increase in emotive rhetoric over

[11] This process occurs primarily through elections, but can also be as a result of other factors such as death or suspensions.

[12] Brexit, in particular, has a strong positive correlation with compound score. The approach taken to topic modelling is discussed in detail in the *'Robustness and alternative mechanisms'* section, and in the appendix.

time is not driven by characteristics of individual MPs or the issues they discuss.

Table 2.4: OLS: Change in compound, *Pos* and *neg* score over time (decades elapsed)

(*Ind. var.=*	Model			
decades elapsed)	(1)	(2)	(3)	(4)
Dep. var.:				
Compound score	0.011***	0.011***	0.012***	0.0010***
	(0.001)	(0.001)	(0.001)	(0.008)
Pos score	0.002*	0.001	0.002**	0.002***
	(0.008)	(0.008)	(0.007)	(0.005)
Neg score	0.002	0.004	0.004*	0.000
	(0.108)	(0.002)	(0.002)	(0.001)
Controls		X	X	X
Party fixed effects			X	X
Topic fixed effects				X
Observations	3,629,439	3,629,439	3,629,439	3,629,439
Adjusted R^2	0.011	0.011	0.014	0.082
Note:	*p<0.1; **p<0.05; ***p<0.01			

The next stage of analysis examines the extent to which this increase is driven by an *adaptive* process, where elites gradually refine the language of their tweets, or by a process of *replacement*, as new MPs are elected. To identify the effect of the MP cohorts elected in the 2010, 2015, and 2017 General Elections, I use linear regression to model compound scores of tweets made

a) in tweets by all MPs over time in decades elapsed, b) in tweets made by members of parliament elected in the 2010 General Election, c) in tweets by members of parliament elected in the 2015 General Election, and d) in tweets by members of parliament elected in the 2017 General Election. This time, I run four different models with standard errors clustered at MP levels in each. Model 1 is a simple baseline model, again using the explanatory variable *decades elapsed*. However, this time, I also include the variables *2010 cohort* (a dummy variable indicating if an MP was elected in the 2010 General Election), *2015 cohort* (a dummy variable indicating if an MP was elected in the 2015 General Election) and *2017 cohort* (a dummy variable indicating if an MP was elected in the 2017 General Election). In Models 2, 3, and 4, for the reasons discussed above, I again introduce controls for age and gender, with party and topic fixed effects.

The results of these regression models can be seen in Table 2.5. The stable positive correlation between decades elapsed and compound score across all models reveals a small but statistically-significant within-MP increase in emotive rhetoric over time. This echoes the previous finding that a slight adaptive linguistic process was undertaken by MPs over time. At the same time, I observe varied effects across different cohorts. While I find no relationship between the 2010 cohort and the intensity of sentiment in their tweets, the 2015 and 2017 cohorts employ a greater degree of emotional language. To check this is not a result of the increased character limit introduced in 2017, I cross-check against both pre-2017 140-character limit and post-2017 280-character limits[13]. This suggests that, while MPs have exhibited linguistic adaptation over time on Twitter, the 2015 and 2017 cohorts also adopted substantially more emotional rhetoric than longer-serving MPs. This appears to support my claim that, while MPs are gradually adapting their rhetoric, newer generations of British politicians are employing greater levels of emo-

[13] See the 'Robustness and alternative mechanisms' section.

tion in their tweets, and that increasing levels of emotions are being driven by process of both adaptation and replacement.

Table 2.5: OLS regression: Compound score over time and by MP cohort

	Model			
	(1)	(2)	(3)	(4)
Decades elapsed	0.0102***	0.0103***	0.0101***	0.0083***
	(0.012)	(0.0012)	(0.0011)	(0.0009)
2010 cohort	-0.0026	-0.0009	-0.0005	-0.0049
	(0.0104)	(0.0098)	(0.0098)	(0.0078)
2015 cohort	0.0309***	0.0280**	0.0360***	0.0299***
	(0.0082)	(0.0088)	(0.0086)	(0.0071)
2017 cohort	0.0610***	0.0670***	0.0704***	0.0633***
	(0.0082)	(0.0088)	(0.0089)	(0.0075)
Controls		X	X	X
Party fixed effects			X	X
Topic fixed effects				X
Observations	3,629,439	3,629,439	3,629,439	3,629,439
Adjusted R^2	0.012	0.012	0.016	0.083
Note:	*p<0.1; **p<0.05; ***p<0.01			

Finally, to explore the relationship between emotive rhetoric and engagement, I create further OLS regression models, modelling the compound, positivity, and negativity scores of tweets against log engagement (Table 2.6). This time, controls include age and gender, but also introduce follower count as an additional control variable. This is based on the assumption that a greater number of followers significantly increases the initial visibility of a tweet, and is therefore inherently positively correlated with likely engagement levels. I retain party and topic fixed effects, along with weighting by tweet length. This time, I also introduce MP fixed effects to account for within-MP heterogeneity in rhetorical style. Surprisingly, these results show a negative relationship between compound score and log engagement, demonstrating that tweets with a higher level of emotional valence generate *lower* levels of engagement. Similarly, a high *pos* score is negatively correlated with levels of engagement, showing that more positive tweets are also generally met with fewer likes, retweets, quote tweets, and replies. However, perhaps most importantly for the tone of political discourse, incorporating negative language in tweets appears to generate far greater levels of engagement.

Table 2.6: OLS regression: Emotion and sentiment against Twitter engagement

	Dependent variable:		
	Log engagement + 1		
	(1)	(2)	(3)
Compound score	-0.172^{***}		
	(0.022)		
Pos score		-0.406^{***}	
		(0.083)	
Neg score			1.601^{***}
			(0.096)
Controls	X	X	X
MP fixed effect	X	X	X
Party fixed effects	X	X	X
Topic fixed effects	X	X	X
Observations	3,619,421	3,619,421	3,619,421
Adjusted R^2	0.427	0.427	0.427
Note:	*p<0.1; **p<0.05; ***p<0.01		

2.6 Robustness and alternative mechanisms

2.6.1 Topic-specific heterogeneity

A key challenge in making a valid claim about a genuine rise in emotive rhetoric is the problem of topic heterogeneity. As shown in Figure 2.4, certain political topics are discussed with greater levels of emotion than others. For example, the issues of Scottish independence and Brexit were, unsurprisingly, discussed with greater emotion as their respective referendum campaigns started to build momentum. Therefore, in order to control for this heterogeneity, each tweet should be broadly categorised according to the issue it discusses. Manual classification of such a large dataset would be prohibitively time-consuming, so I employ Latent Dirichlet Allocation (LDA), an unsupervised probabilistic model for uncovering topics from a collection of otherwise unstructured collection of text documents (Blei et al., 2003). LDA's demonstrable efficacy in classifying latent topics from large-scale Twitter text data (e.g. Barberá, 2015; Zhou et al., 2021) makes it a sensible choice for this application. I use `quanteda`'s *LDA* function to apply Gibbs sampling, iterating over the entirety of my text corpus from $k = 30$ to $k = 200$, in increments of 10 to minimise computation time. Second, I qualitatively evaluate each iteration of k by examining 20 words of highest probability of association with each topic to uncover broad themes. I ultimately select the ideal number of k as 135 topics; based on both human and computational validation[14], this offers an accurate representation of the topic in the tweet corpus. I extracted the most likely topic for each tweet, associating each one with a label, before appending these to the existing dataset. Finally, 500 tweets were selected at random and checked manually to ensure correlation of tweet text to topics. A full topic list is included in Appendix A2. As discussed, while the introduction of topic controls in Model 4 of tables 2.4 and 2.5 reduce the effect

[14] The methods used are discussed in greater detail in the appendix.

of time on compound and positivity score very slightly, these relationships remain positive and statistically significant. This indicates that the observed increase in emotive rhetoric over time is not driven by the changing issues discussed by MPs.

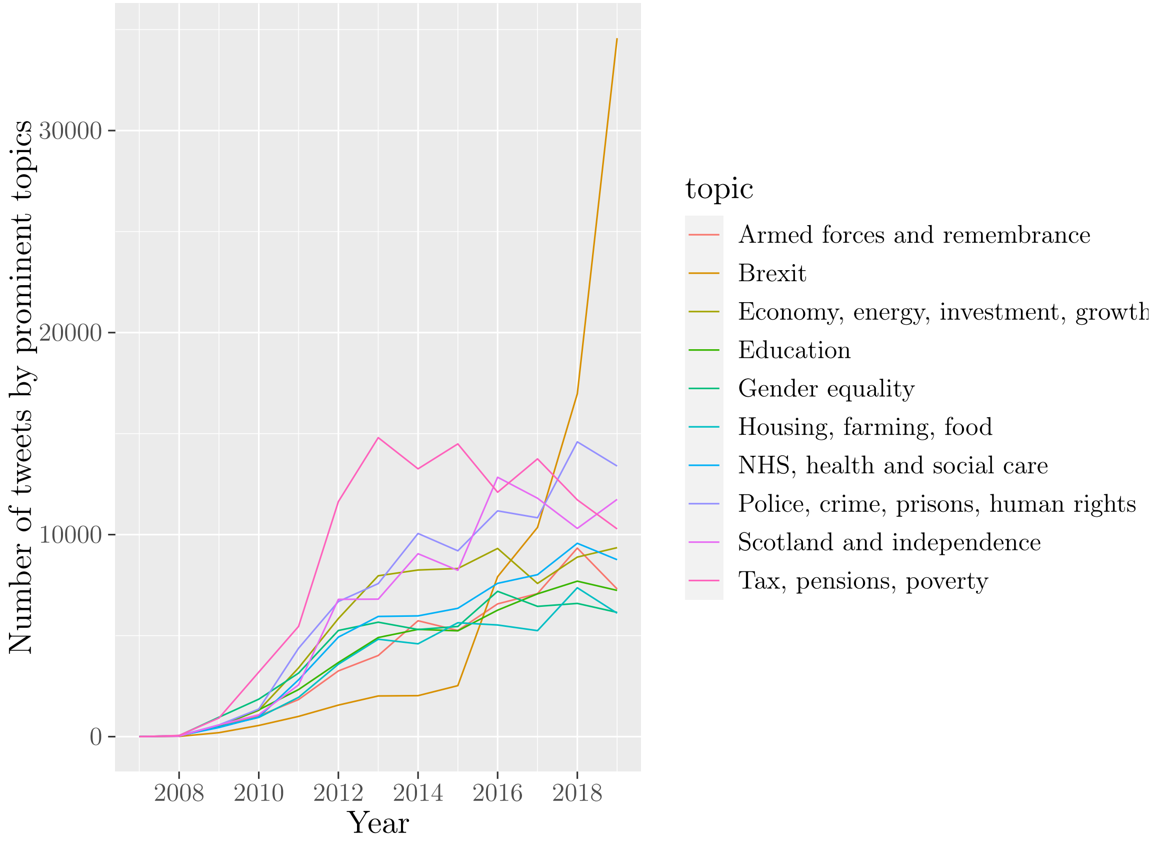


Figure 2.4: Prevalence of selected topics tweeted about by MPs, 2007-2019

2.6.2 Tweet length

In November 2017, Twitter announced a doubling of its character limit, from 140 characters to 280 characters. In the data, this resulted in an increase of the mean number of characters per tweet from 104.5 in the pre-platform change era to 160.9 post-change, while the mean number of words increased from 17.9 to 26.9. This has implications for measuring compound score, in that the method used to calculate it means that the returned value is often proportional to the length of the document. In other words, as the number of words in a speech or tweet grows larger, VADER's compound score tends towards -1 or 1. While this is much more pronounced in corpora with greater variance in length, for example speeches or books, it nonetheless presents a challenge for using compound scores in this specific situation. To examine trends before and after Twitter's platform change, I split the corpus of tweets into before and after the 8th of November, 2017. An OLS regression model with the same controls and fixed effects as applied to the entire corpus (Table 2.7) reveals that the trend within each dataset is still positive; compound scores generated by the new extended character limit still increase over time. All three dependent variables remain positively associated with the passing of time, have a similar effect size to previous models, and remain statistically significant. We can conclude that the increase in these metrics is not driven by the 2017 increase in tweet length, and the results support the hypothesis that politicians in the United Kingdom have adapted their rhetorical strategies between 2007 and 2019 to become increasingly emotional over time.

Table 2.7: Change in compound score over decades elapsed

	Dependent variable:	
	Compound score	
	(1)	(2)
140-character limit era	0.039***	
	(0.011)	
280-character limit era		0.073***
		(0.017)
Controls	X	X
MP fixed effect	X	X
Party fixed effects	X	X
Topic fixed effects	X	X
Observations	2,678,187	950,534
Adjusted R^2	0.074	0.091
Note:	*p<0.1; **p<0.05; ***p<0.01	

Alternative measures of emotion

To strengthen my core claim, I further test its validity by applying three other forms of natural language methods measuring similar outcomes. Finding a positive correlation will underline that elite political rhetoric has become more emotional over time. To test this argument, this time I use the `syuzhet` R package (Jockers, 2017) to apply three new dictionaries to my corpus of tweets. First, I consult the AFINN lexicon (Nielsen, 2011). AFINN assigns a score to each word in a range from -5 to +5, with positive scores indicating positive sentiment and negative scores indicating negative sentiment. I aggregate and normalise the valence scores for all words in a tweet to obtain an overall measure of emotional valence. Second, I consult the Bing lexicon (Hu and Liu, 2004), from which `syuzhet` sums the scores for each (positive or negative) sentiment and divides the result by the total number of matching words to obtain an average sentiment score. The overall sentiment score returned for each tweet is the difference between the average positive sentiment score and the average negative sentiment score. A positive score indicates a more positive sentiment, while a negative score indicates a more negative sentiment. Again, I normalise these values to measure emotional intensity. Finally, I turn to the National Research Council (NRC) lexicon (Mohammad and Turney, 2013). This provides scores for eight pre-defined categories of emotion: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, along with a measure of negative or positive sentiment. Like Bing, the overall sentiment score returned for each tweet is the difference between the average positive sentiment score and the average negative sentiment score. Again, I normalise these values to measure emotional intensity and allow easy interpretation in an OLS model, and apply all controls included in previous regression models. Table 2.8 shows these results; while the effect strength varies by the lexicon consulted, importantly these all show a statistically-significant positive correlation between the passing of time and emotional intensity of tweets by MPs.

Table 2.8: OLS regression: Alternative sentiment measures over time (decades elapsed)

Dependent variable	
AFINN	0.0817*** (0.0052)
Bing	0.0347*** (0.0023)
NRC	0.0251*** (0.0021)
Controls	X
Party fixed effects	X
Topic fixed effects	X
Observations	3,628,823
Note:	*p<0.1; **p<0.05; ***p<0.01

2.7 Conclusion

Emotions can have a profound impact upon a range of political outcomes, such as mobilisation (Marcus, 2000), the spread of information (Halberstam and Knight, 2016), and persuasion (Brader, 2005). Amid rising polarisation and escalating electoral hostility around the world (Layman et al., 2006; Mason, 2018; Iyengar et al., 2019), it is crucial that we understand the underlying drivers of our increasingly divisive politics. I focus on the shifting nature of rhetoric employed by political elites on Twitter, establishing an increase in their use of emotion in the years spanning 2007-2019, while also highlighting a relationship between these emotive tweets and the engagement they generate.

This paper makes three contributions to the existing literature on use of language by political elites. First, I argue that members of the House of Commons increased the strength of emotion in their Twitter rhetoric during the period 2007-2019. In accordance with my main expectations, these results show that mean levels of emotional valence were higher per tweet in 2019 than they were in 2007. Second, I establish that this change in emotion was driven both by a process of adaptation, as politicians modified their rhetorical styles over time, and by a process of replacement, as newer cohorts of MPs employed different rhetorical strategies. Finally, I find that rhetoric using more emotion is not necessarily associated with greater levels of Twitter engagement; a greater proportion of negative words within a tweet is associated with higher levels of engagement while a greater proportion of positive words is negatively associated with the same outcome variable. This demonstrates that, whether or not politicians choose to capitalise upon it, there exists an incentive to use more negative rhetoric. Given that greater levels of engagement result in increased visibility of an MP's messages, this has important implications for the scope and tone of political debate on Twitter.

This paper is not without its limitations. First, the unsupervised dictionary methods I use to show increasing emotional extremity are effective but lack contextual nuance; they are able to detect emotional words but struggle with varying intensity and context. The recent upsurge in powerful yet accessible deep learning models provide opportunities to improve precision and robustness. Incorporating a machine learning model like BERT or RoBERTa would, for example, enable context-aware sentiment analysis. Similarly, traditional topic modelling approaches, like Latent Dirichlet Allocation, fail to account for word context. Dynamic topic modelling or BERT embeddings would further improve the topic labelling I employ to hold issues constant. Second, a focus on emotive rhetoric may oversimplify the complexity of political discourse. While emotional language is significant, future research might seek to encompass more complex rhetorical dimensions such as the use of humour, irony, metaphor, or populist language. Further, expanding the research scope to include platforms like Facebook and TikTok, with their distinct user dynamics, might reveal varied rhetorical adaptations. Finally, my study cannot establish the *influence* of this increasingly emotive elite by addressing whether we see similar patterns in the mass public. Future research might address whether extreme rhetoric by politicians is reflected by voters.

Being able to appeal directly to the electorate has given a generation of politicians far greater influence than their predecessors and, in certain situations, supplanted the traditional role of the mass media. As a result, the rhetorical actions of politicians will have profound implications for wider democratic discourse, representation, and accountability. If emotive rhetoric by political opinion leaders is correlated with higher levels of social media engagement than non-emotional (or 'rational') content, it will spread faster, further, and have implications for the overall tone of debate. In seeking to understand how the introduction of Twitter has shaped the rhetorical behaviour of elites, my results yield new insights into how digital communication is shaping our political discourse.

3 Extremely Online? Ideological and Affective polarisation on British Political Twitter

Nick Lewis, James Tilley, and Sara B. Hobolt

Abstract

Social media are now key arenas for political discussion. Rather than facilitating reasoned debate and compromise, however, such online discussions often foster disagreement and even hostility. Yet, we still know relatively little about who uses Twitter to engage in political discussion, and if these people are more likely to be ideologically and affectively polarised compared to the general population. Analysing a novel dataset representative of the British population, we link survey respondents to their Twitter accounts, finding that people who discuss politics on Twitter are more ideologically and affectively extreme than those who do not. We also demonstrate that these politically extreme users are more likely to tweet politically and share negatively-biased partisan content. These findings enhance our understanding of the type of voices most likely to be heard on social media platforms, with wider implications for the extremity of online political debate.

3.1 Introduction

Disagreement is a necessary part of a well-functioning democracy. However, political polarisation undermines the ability of people to deliberate in a healthy and productive way. The concern over polarisation is growing; scholars have highlighted its increase not only in the United States (Layman et al., 2006; Mason, 2018; Iyengar et al., 2019) but in many other parts of the world (Gidron et al., 2020; Harteveld, 2021; Reiljan and Ryan, 2021) at both an elite (Robison and Mullinix, 2016) and mass (Frimer et al., 2017) level. Polarisation also extends beyond partisan divisions; more recently, an emerging literature has identified similar trends in the United Kingdom centred around the Brexit debate (Sobolewska and Ford, 2020; Hobolt et al., 2021). Many argue that social media has expedited the process of political sorting (Levendusky, 2013; Sunstein, 2018) by enabling the wide and rapid spread of partisan cues (Levendusky, 2013; Bolsen and Druckman, 2015; Garrett et al., 2016) which, in turn, leads to homophily, or the creation of 'echo chambers' (Barberá, 2015; Sunstein, 2018). During the 2010s, the social media site *X*, formerly known as Twitter[1], increasingly became a focal point for the discussion of politics in many countries, including the United Kingdom. While unrepresentative of the population (Mellon and Prosser, 2017), it is used extensively by politicians[2] and now plays a crucial agenda-setting role in news and public policy (Barberá et al., 2019; Gilardi et al., 2022). Outside of elites, however, we still know surprisingly little about who uses Twitter to talk about politics and the nature of the content they share. Establishing whether certain groups are under- or over-represented in political discussion, the extremity of their attitudes, and how they behave, is important for our understanding of the nature of online democratic delib-

[1] As this research note focuses on a period between 2019 and 2021, we refer to it as Twitter throughout.

[2] For better or for worse, over 85% of MPs elected in the 2019 UK general election had an active Twitter account and used them on a daily basis.

eration. Analysing a novel dataset representative of the British population which we can directly link to respondents' Twitter accounts, we therefore examine how political Twitter users compare with others in terms of their ideological and affective polarisation. The negative consequences of this polarisation are well-documented, including increased out-group prejudice and discrimination (Iyengar et al., 2019), extreme activism and anger (Mason, 2015), decline in the quality of political discourse (Layman et al., 2006), political violence (Kalmoe and Mason, 2022), policy gridlock, and even growing mistrust of democratic institutions (Kingzette et al., 2021). Crucially, polarisation undermines the ability of politically engaged people to interact with each other in a way which promotes healthy and productive democratic deliberation (Mutz, 2006).

What has caused this rise in polarisation? A recurring debate is whether social media has expedited it through a process of partisan sorting. The ability to curate one's own social circle, it is argued, inevitably leads to the creation of homogeneous online networks, or 'echo chambers' (Conover et al., 2011; Colleoni et al., 2014; Terren and Borge-Bravo, 2021), which increases political polarisation (Sunstein, 2018; Settle, 2018). This argument is contested since, even in relatively homogeneous follower networks on Twitter, there remains a substantial amount of 'cross-ideological' communication (Barberá et al., 2015; Barberá, 2020)[3]. It is also plausible that levels of affective polarisation drive social media use rather than the other way round (Nordbrandt, 2021). And it is certainly the case that online behaviour is shaped to some degree by group identity and political attitudes; for example, Osmundsen et al. (2021) found partisanship to be a strong predictor of the likelihood of sharing pro-partisan fake news on Twitter. Equally, ideological alignment with a social media post is of greater importance than the presence of mis-

[3] Indeed, in certain situations, social media use has even been shown to have a *de*-polarising effect (Beam et al., 2018b).

information when sharing information (Bowyer and Kahne, 2019).

Whether social media is driving polarisation, or simply reflecting it, we might expect the people who choose to discuss politics online to be more affectively and ideologically extreme than the general public. Yet, to date, we have little evidence of this. What we do tends to focus on people's ideological positions and social characteristics. With regard to both, Twitter tends not to be representative of wider populations (Larsson and Moe, 2012; Vaccari et al., 2013). In the U.S., Twitter users are younger and more likely to be Democrats than the general public (Wojcik and Hughes, 2019). In Britain, we see a similar picture: older people and Conservative voters are less likely to use Twitter (Blank and Lutz, 2017; BES, 2021). Indeed, social media users *overall* in general are more liberal, pay more attention to politics and are more likely to support the Labour Party (Mellon and Prosser, 2017). What we do not know is whether people who talk politics on social media are more ideologically extreme and more hostile to people on the other side, and whether this extremity is associated with their online behaviour. Anecdotal evidence might suggest that this is the case, but we currently have little sense of the extent of this polarisation, especially with regard to affective polarisation, and no real answer to whether this is a symmetrical effect that impacts both sides.

This paper makes two contributions. First, we use representative survey data linked to Twitter accounts to examine the ideological and affective extremity of the minority of people who use Twitter to discuss politics. We find that Twitter users are more ideologically extreme than those who don't use Twitter, but only those who identify as Labour partisans and those on the 'Remain' side of the Brexit debate. Further, of those people *on* Twitter, those who use it to discuss politics are more ideologically-extreme than those who do not, while 'Remain' supporters are more affectively polarised than their non-political counterparts. Second, we explore the link between extreme Twitter users and the type of tweets they produce: more ideologically extreme Remain supporters are more likely to tweet politically, while affective extremity is linked to a greater number of political tweets by Leavers *and* Remainers. These results highlight the increased potential for polarisation on social media, where individuals with stronger convictions appear more likely to engage in online political discussion. They also highlight asymmetries in how people on different sides of the political spectrum engage in political discussion.

Political polarisation

Political polarisation primarily appears in two forms: ideological and affective polarisation. Ideological polarisation refers to the increasing divergence in political views or issue positions among individuals and groups, as highlighted by Dalton (1987) in his study of generational differences in political attitudes. Affective polarisation, on the other hand, is rooted in social identity theory, as proposed by Tajfel et al. (1979), and is characterised by a growing animosity toward members of the opposing partisan group. This phenomenon is not merely about disagreement over policies but extends to personal dislike and distrust, as explored by Iyengar and Westwood (2015) and Mason (2015). Affective polarisation often leads to profound divisions, where partisans view members of the out-group as fundamentally different and morally inferior.

The negative consequences of polarisation are extensive and multifaceted. Ideological polarisation often results in policy gridlock, where the deepening divide between political parties makes compromise increasingly difficult. This can hinder effective governance and the implementation of necessary reforms. Moreover, it can lead to a decline in the quality of political discourse, as debates become more about winning arguments than solving problems

(Layman et al., 2006). Affective polarisation exacerbates these issues by fostering out-group prejudice and discrimination; Iyengar and Krupenkin (2018) argue that such polarisation can lead to extreme activism and anger, further entrenching divisions within society. The emotional intensity associated with affective polarisation can also incite political violence, as individuals become more willing to endorse or engage in aggressive actions against perceived enemies (Kalmoe and Mason, 2022). Another significant consequence of polarisation is the erosion of trust in democratic institutions; when political discourse becomes increasingly hostile, and compromise appears impossible, public confidence in the effectiveness and fairness of democratic processes can wane, as noted by Kingzette et al. (2021). This mistrust can undermine the legitimacy of governmental institutions and threaten the stability of democratic systems. Ultimately, polarisation undermines the ability of politically-engaged individuals to interact in ways that promote healthy and productive democratic deliberation. For example, Mutz (2006) highlights that exposure to opposing viewpoints is crucial for democratic health, yet polarised environments discourage such interactions, leading to the formation of 'echo chambers' and a less informed electorate.

While the majority of research on polarisation has focused on the United States, where partisan polarisation often manifests in stark contrasts between Democrats and Republicans, similar patterns have emerged in other democracies. In the United Kingdom, for example, polarised Brexit identities have become prominent in the wake of the country's 2016 referendum on leaving the European Union. Attitudes toward Europe have created significant divisions, with distinct electoral realignments based on age, education, and geography (Hobolt et al., 2021). This realignment has shifted the traditional socio-economic predictors of voting behaviour, leading to new forms of political fragmentation, as noted by Evans and Tilley (2017).

3.1.1 Political polarisation on social media

Many scholars argue that social media encourage and expedite these political divisions, having become increasingly important spaces for political communication (Stier et al., 2018), mobilisation (Bond et al., 2012; Coppock et al., 2016; Yasseri et al., 2016) and persuasion (De Zúñiga et al., 2022). Notably, digital technology provides elites with powerful new ways to communicate their messages directly to the electorate, and enables voters to connect with people beyond their families and local communities to discuss political ideas. While social media sites have also become an important source of political news - around 25% of people in the United Kingdom get most of their news about current affairs from social media (YouGov, 2023) - people in many countries are more comfortable discussing politics face-to-face than they are online (Smith et al., 2019).

Social media represent fertile ground for polarisation due to several key factors. Firstly, it is argued that the ability to curate one's social circle leads to the creation of homogeneous online networks, or 'echo chambers' (Conover et al., 2011; Colleoni et al., 2014; Terren and Borge-Bravo, 2021), which subsequently heighten political polarisation (Sunstein, 2018; Settle, 2018). This process primarily occurs through three mechanisms. First, motivated reasoning, which leads people to overvalue information from in-group sources and undervalue information from out-group sources (Nyhan and Reifler, 2010). Second, attitude reinforcement, where exposure to similar views solidifies one's own attitudes (Visser and Mirabile, 2004) and third, social conformity, where opinions are adjusted to align with group norms (Festinger, 1950). Additionally, constant exposure to political content and the pressure to conform to group norms can lead to increased stress and anxiety, which in turn can exacerbate political polarisation. Research has shown that individuals who are more emotionally invested in political issues are more likely to engage in polarised behaviour online (Weeks, 2015). This emotional investment can create a hostile online environment where dissenting opinions are met with aggression and hostility, further entrenching polarised attitudes.

Second, the absence of traditional mainstream media ‘gatekeepers’ allows political elites to communicate more extreme positions directly to their followers (Rogowski and Sutherland, 2016; Banda and Cluverius, 2018). This increased elite polarisation prompts partisans to change their issue perceptions (Druckman et al., 2013) and express higher levels of affective polarisation (Banda and Cluverius, 2018). Twitter, for example, amplifies this effect, with negative tweets leading to increased polarisation beyond the effects of self-selection into partisan media communities (Banks et al., 2021). We also see this effect across other social media platforms, where recommendation algorithms on Facebook and YouTube promote polarising content (Bakshy et al., 2015; Bessi et al., 2016; Cho et al., 2020). *Deactivating* Facebook accounts around elections has been shown to reduce affective polarisation (Allcott et al., 2020), while heavy social media users are less likely to engage in face-to-face political discussions (Hampton et al., 2017). Third, social media’s role in shaping political polarisation has been witnessed during major political events. For example, during the 2016 U.S. Presidential election, social media platforms played a crucial role in spreading misinformation and polarising content, with fake news and hyper-partisan content more widely shared on social media than factual news (Allcott and Gentzkow, 2017; Vosoughi et al., 2018). In addition to the spread of misinformation, social media have also facilitated the rise of populist movements and extremist groups. These groups often use social media to organise, recruit, and disseminate ideas, further polarising the political landscape. The ability of these groups to reach a wide audience with minimal oversight has led to concerns about the role of social media in fostering political extremism and violence (Marwick and Lewis, 2017; Benkler et al., 2018).

Whether social media is driving polarisation or simply *reflecting* it remains debated. While Lelkes et al. (2017) demonstrate that broadband internet access increases partisan hostility, a cross-country study by Boxell et al. (2022) finds no correlation between online news consumption and affective polarisation trends. The existence of online echo chambers is also disputed; Flaxman et al. (2016) found that while social networks and search engines increase ideological distance, they also expose individuals to opposing viewpoints. Substantial 'cross-ideological' communication persists even in homogeneous Twitter networks (Barberá et al., 2015; Barberá, 2020), and in some cases, social media use has a de-polarising effect (Beam et al., 2018b). Additionally, evidence suggests that affective polarisation may drive social media use rather than vice versa (Nordbrandt, 2021). Indeed, political polarisation may be both a cause and an effect of social media use; it is argued that a 'feedback loop' exists between social media and polarisation, exacerbated by algorithms designed to maximise engagement by promoting sensational and divisive content (Tucker et al., 2018). As a result, individuals become trapped in echo chambers where pre-existing beliefs are constantly reinforced, leading to more extreme viewpoints and greater polarisation. While we broadly know who uses social media, we know surprisingly little about who uses it to discuss politics, and how ideologically and affectively extreme these people are. Gaining answers to these questions will give us insight into whether social media sites provide fertile ground for polarisation, and if political discussion is dominated by voices from a particular part of the ideological or affective spectrum.

3.1.2 Who discusses politics on Twitter?

We expect people who use Twitter to discuss politics to be more ideologically and affectively extreme than those who do not. First, we know that Twitter tends not to be representative of wider populations (Larsson and Moe, 2012;

Vaccari et al., 2013; Mellon and Prosser, 2017). For instance, Twitter users who write about politics in the United States and Spain are majority male, disproportionately live in urban areas, and tend to have more extreme ideological viewpoints (Barberá and Rivero, 2015). In the United States, Twitter users are younger and more likely to be Democrats than the general public (Wojcik and Hughes, 2019) and similarly, in the United Kingdom, social media users overall are also more liberal, pay more attention to politics, and are more likely to support the Labour Party (Mellon and Prosser, 2017).

Second, individuals who are actively engaged in politics, particularly members of political parties, are often more ideologically extreme than the general population (Poletti et al., 2019). Twitter provides a low-cost method of political participation, enabling users to easily connect with like-minded individuals. The platform's accessibility and opportunities for selective engagement are particularly appealing to highly-polarised individuals (Mason, 2018), who are more likely to interact with and disseminate content which reinforces their existing beliefs. The cultivation of these networks reinforce existing beliefs (Sunstein, 2018), exacerbating political polarisation by limiting exposure to diverse viewpoints (Colleoni et al., 2014; Barberá, 2015). Third, in addition to self-selection effects, social media platforms amplify polarising content (Tucker et al., 2018). Negative and emotionally charged content tends to be more widely shared and engaged with on Twitter, amplifying and deepening affective polarisation (Brady et al., 2017; Tucker et al., 2018). Such emotionally-driven engagement often fuels out-group animosity (Weeks, 2015), with posts about the political 'other side' generating far greater levels of engagement - and therefore information spread (Rathje et al., 2021). Partisanship also strongly predicts the likelihood of sharing pro-partisan fake news on Twitter (Osmundsen et al., 2021), and ideological alignment is more influential than the presence of misinformation when sharing information (Bowyer and Kahne, 2019). With all of these factors in mind, we anticipate that political Twitter users in the United Kingdom - where polarisation has intensified since the 2016 referendum on the European Union (Hobolt et al., 2021) — will exhibit greater ideological and affective extremity than those who do not use Twitter to discuss politics:

- H_1: *British Twitter users are more ideologically extreme than individuals who don't use Twitter.*
- H_2: *British political Twitter users are more ideologically extreme than individuals who don't use Twitter to talk about politics.*
- H_3: *British political Twitter users are more affectively polarised than individuals who don't use Twitter to talk about politics.*

Asymmetries also exist in online political behaviour between polarised group identities. For example, liberal Twitter users are more likely to engage in cross-ideological or cross-partisan communication (Barberá, 2015; Barberá et al., 2015) than their conservative counterparts. Conservative political elites who use moral-emotional language in their tweets experienced greater spread of their messages than liberal elites (Brady et al., 2019), suggesting that conservatives react differently to certain types of rhetoric. Different groups also curate their social networks in different ways: Colleoni et al. (2014) found that Democrats exhibit higher levels of political homophily than Republicans, except when the latter follow official Republican accounts. The extremity of polarisation on Twitter also varies from country to country; Urman (2020) found a strong relationship between a country's electoral system and the degree to which its political Twitter networks are homogeneous or heterogeneous. We also see unequal distributions of network segregation on Twitter by *party*; in a study of followers of political parties on Twitter, Rusche (2022) found the AfD's supporters to be the most homogeneous. With this in mind, we might also expect to see similar differences in attitudes and Twitter behaviour between certain partisan groups in the UK.

3.2 Design

To examine the ideological and affective extremity of the average Twitter user, and how they use the site, we need to gain accurate measures of their attitudes and online behaviour. Research on polarisation and social media tends to use either survey methods (e.g. Banda and Cluverius, 2018; Butters and Hare, 2022) or digital trace data (e.g. Barberá, 2015; Yarchi et al., 2021) exclusively. While we can gain relatively accurate estimates of ideology and affect from surveys, self-reported measures of social media usage are often inaccurate (Henderson et al., 2021). At the same time, obtaining accurate estimates of ideological or affective attitudes from digital trace data can be extremely challenging. Research linking survey and Twitter data has so far attempted to validate self-reported estimates of Twitter use (Henderson et al., 2021), understand the effect of self-reported happiness on tweet sentiment (Al Baghal et al., 2021), highlight differing communication styles of elites in online campaigns (Karlsen and Enjolras, 2016), and examine the extent to which people live in ideological online echo chambers (Eady et al., 2019). Further, Guess et al. (2019a) link survey data, including strength of partisan identity with rates of posting - and posting about politics - on Twitter. We take this a step further, examining positive and negative partisan identity, alongside ideological and affective attitudes of survey respondents. We then link these self-reported measures to actual observable online behaviour, including the likelihood of posting about politics and the *type* of tweets our respondents produce. The results raise important questions about who we are most likely to hear in online political discussions.

To systematically examine the political attitudes and online behaviours of Twitter users, we use data from an original survey ($N = 4149$) fielded in July 2021 by YouGov, a well-known research firm which uses quota sampling and re-weighting methods to generate nationally representative samples from a panel of over two million British adults. Uniquely, we linked this survey data to respondents' actual online activities on Twitter. To ascertain which respondents were active Twitter users, we relied on responses to the question 'How often do you look at Twitter?'. Those answering either 'Almost never' or 'Do not have an account' were classified as 'not on Twitter', with the rest classified as Twitter users. The latter were asked if they would voluntarily share their Twitter username, with responses entered into a free text box. 786 respondents supplied at least some text. A number of these could be immediately discounted, including responses such as 'Don't know', 'I do not have one', and 'None'. Those which looked like plausible account names were manually verified in the first instance through Twitter's search function. A small number of usernames were not immediately verifiable, potentially caused by spelling errors. In these instances, variations on the supplied name were searched and then cross-checked with survey data variables such as geographic location, age, and gender. Once found, these accounts were cross-checked and debated within the team to ensure robustness. 599 accounts were verified, and their account IDs retrieved via Twitter's API. 56 of these had protected status, which left us unable to access tweets from these accounts. Ultimately, tweets from 543 accounts were available. All tweets from these users between 2011 and 2021 were gathered via Twitter's API using the `academictwitteR` R package (Barrie and Ho, 2021).

Tables 3.1 and 3.2 show a breakdown of our survey respondents' Twitter usage. This is categorised by their partisan and Brexit identity, as these were the most salient political identities in the UK at the time (Hobolt et al., 2021). Interestingly, the results show that a significant proportion (42%) report that they spend some time on Twitter. Table 3.1 also shows a significant disparity between the proportion of Conservative and Labour partisans on Twitter (35% vs. 54%). Table 3.2 tells a similar story. Despite the relatively even balance of the 2016 referendum, Leavers appear to be significantly outnumbered by Remainers on Twitter. Two thirds of Leavers do not use Twitter compared to about half of Remainers[4].

Table 3.1: Twitter use by partisan identity

	All	**Con**	**Lab**	**Other**	**None/DK**
Not on Twitter	58%	66%	46%	52%	64%
On Twitter	42%	35%	54%	48%	36%

Table 3.2: Twitter use by Brexit identity

	All	**Leave**	**Remain**	**Neither**	**DK**
Not on Twitter	58%	65%	48%	63%	68%
On Twitter	42%	35%	52%	37%	32%

While a significant proportion of respondents were willing to share their Twitter accounts with us, it was essential that we identify any systematic differences between 1) non-Twitter users, 2) self-reported Twitter users, and 3) those who shared their account details with us for a study of their Twitter behaviour. The composition of these groups can be found in Appendix 1. To ensure generalisability, and the validity of our results, we had particular interest in identifying potential imbalance between groups 2 and 3. While age and education profiles of each look similar, on closer examination response

[4] Remainers are also more active on Twitter: 7% of Remainers are politically active on Twitter, compared with only 3% of Leavers. More detail can be found in Appendix B2.

bias was clear in both gender and political identity. Men, Labour partisans, and Remain supporters were more likely to share their account details and were therefore over-represented in the study selection. To account for this inherent bias and ensure robust results, we apply Heckman corrections (Heckman, 1979) to all OLS models. These two-stage models first estimate the probability of each respondent being selected into the sample based on party ID, Brexit ID, and gender. Then, the outcome OLS model incorporates these probabilities, via an inverse Mills ratio, to help correct for selection bias and obtain unbiased estimates of the coefficients in the outcome equation. A more detailed discussion of response bias, along with balance tables, can be found in Appendix B.7.

To analyse the online behaviour of users, we first divide their tweets into two six-month study periods: 1) 1st September 2019 - 29th February 2020, and 2) 1st May 2021 - 31st October 2021. Doing this around the same time of the relevant survey wave meant we could link Twitter activity to survey responses while also avoiding a few potential confounders. Period 1 avoided the initial onset of the COVID-19 pandemic, during which political tweets were inevitably dominated by the government's response to the pandemic, and period 2 was a relatively stable period in which the UK got back to 'normal' following three COVID lockdowns. We selected a random sample of maximum 25 tweets from each active user in each study period. Some users did not tweet at least 25 times, resulting in a 5,075-tweet random sample from period 1 and 5,835-tweet random sample from period 2, giving us a total of 10,910 tweets. These tweets were hand-coded by three research assistants using 19 standardised coding categories, including evaluations of whether a tweet or retweet was political, whether it represented support for (or opposition to) each of the main UK parties, and support of (or opposition to) the UK's decision to leave the European Union. We categorise 'political Twitter users' as those users who tweeted about politics at least once during both periods. As each research assistant coded a separate dataset, the lead researcher coded a 10% random sample of each dataset, and checked inter-coder reliability using Krippendorff's alpha[5].

To compare the attitudes of both Twitter users and non-users, we turn to the survey. For ideological attitudes, we average agreement with a series of statements along three dimensions: 'left-right' economic attitudes, liberal-conservative 'social' attitudes, and attitudes towards the European Union. Responses run from 'strongly disagree' to 'strongly agree', which are recorded 1-5 on a Likert scale, with 1 indicating strong agreement and 5 indicating strong disagreement[6]. These were averaged and ordered so that all high scores indicate more left-wing, more socially-liberal, and more pro-EU[7]. A measure of extremity was then calculated by taking ideological averages on each dimension, for each respondent, and subtracting the population mean: the average of all responses in our sample. To measure affective attitudes, we take a standard 'thermometer' score (as used by Gidron et al., 2020; Reiljan, 2020; Wagner, 2021); the difference between two 0-100 ratings, indicating feelings of favourability/unfavourability towards voters on either side of either the Conservative/Labour or Leave/Remain divide[8], and again calculate extremity as distance from the population mean. Higher scores equal greater 'warmth' or favourability, so a greater difference between the scores given to in-group and out-group equals greater affective polarisation.

[5] These scores were 0.937, 0.926, and 0.952 respectively, indicating strong inter-coder reliability.

[6] The full list of questions can be found in Appendix B.2.

[7] This choice was taken for ease of interpretation. Higher scores on these attitudes tend to be those shared by Labour and Remain partisans, with lower scores more typical of Conservative and Leave supporters.

[8] The full wording of this question can be found in Appendix B.4.

3.3 Results

3.3.1 Ideological and affective extremity

First, are people who use Twitter more ideologically extreme than those who don't? Figure 3.1 plots the results of a series of OLS models comparing the ideological extremity of individuals in our sample who report using Twitter with those who do not. Overall, across all three ideological dimensions (left-right, liberal-conservative, and EU attitudes) self-reported Twitter users are more extreme than non-Twitter users[9]. Sub-sample analyses by partisan identity, however, reveal this ideological extremity to be primarily driven by Labour and 'Remain' supporters. Online Labour supporters are particularly extreme on the Liberal-Conservative axis when compared to their non-Twitter using co-partisans. At the same time, Twitter-using Remain supporters are more extreme when it came to attitudes towards the European Union. Interestingly, Conservative supporters on Twitter are notably *less* extreme than non-Twitter-using Conservatives on the Liberal-Conservative dimension.

[9] Full results can be found in Appendix B.3

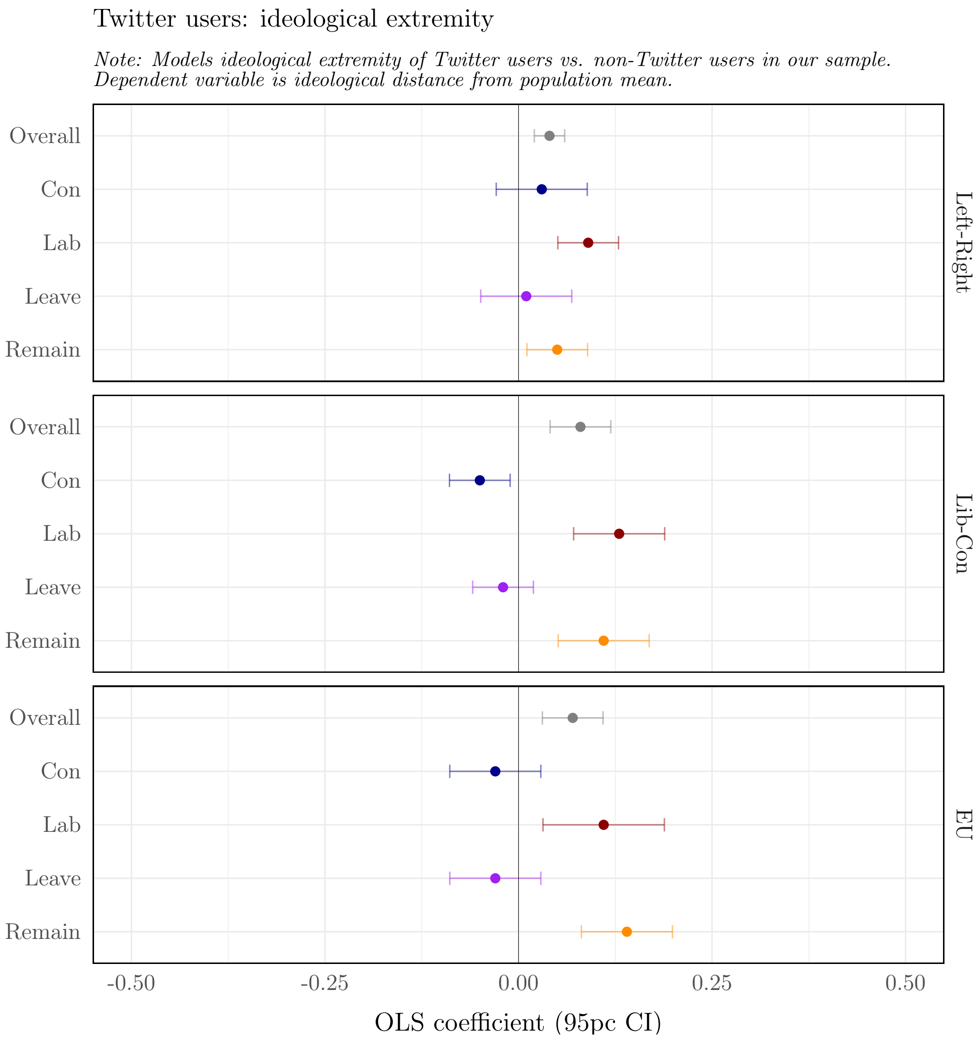


Figure 3.1: Ideological extremity: Twitter users vs. non-Twitter users.

Note: Plot shows ideological extremity of Twitter users vs. non-Twitter users in our sample. Dependent variable is ideological distance from population mean.

Narrowing our focus to those who use Twitter, we next turn our attention to the question of whether people who talk about politics on Twitter are more ideologically extreme than individuals who use Twitter, but do not talk about politics. Figure 3.2 plots the coefficients of a series of OLS regressions, which compare political Twitter users against non-political Twitter users, in both the whole sample ('Overall') and within each of our major partisan groups. This time, to account for potential selection bias in those people who shared their Twitter account details, we apply Heckman corrections[10]. Again, we see heterogeneity across partisan groups. Conservative and Leave-supporting political Twitter users seem ideologically similar to non-political Twitter users, suggesting their values are roughly in line with their offline peers.

[10] Full results are in Appendix B.3.

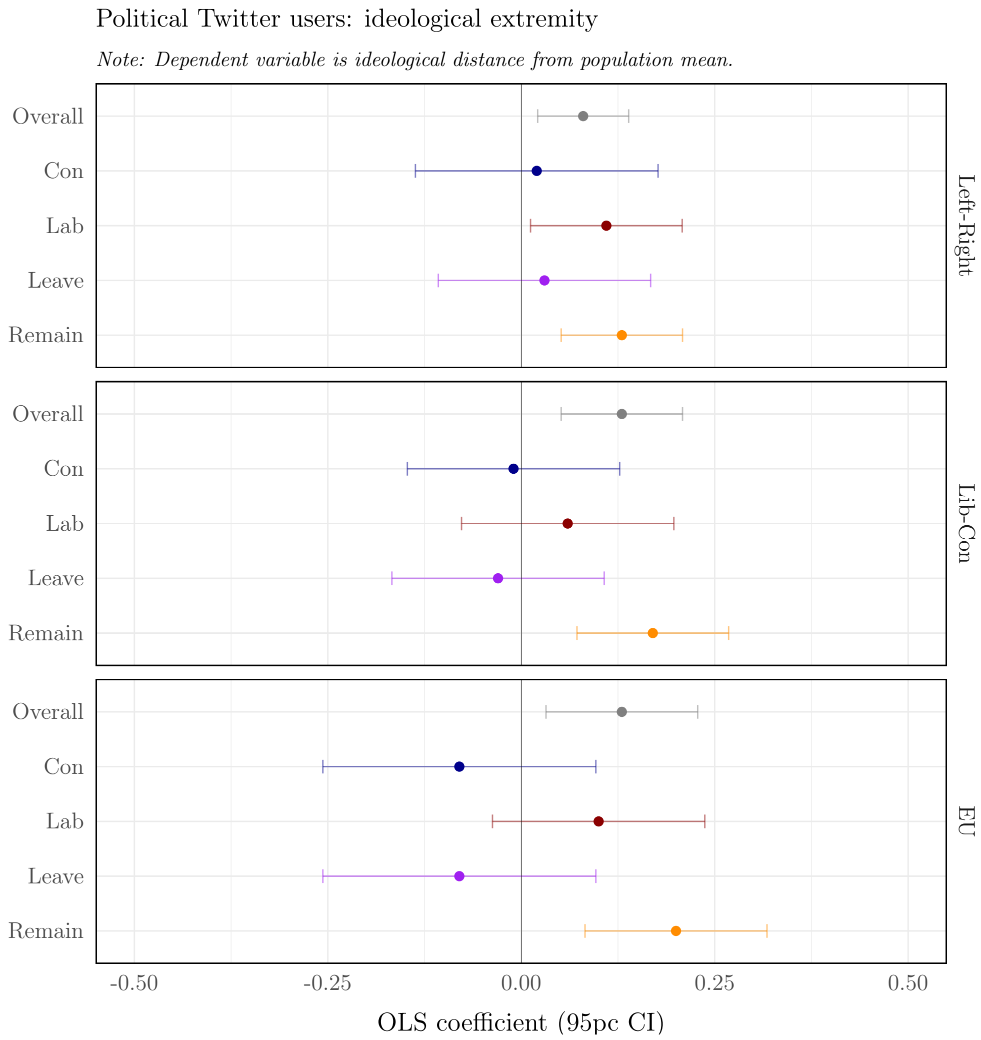


Figure 3.2: Ideological extremity: Political Twitter users vs. non-political Twitter users.

Note: Plot shows within-group comparison between political and non-political Twitter users, controlling for age and education, with Heckman corrections applied. Dependent variable is ideological distance from population mean.

By comparison, political Twitter's Labour partisans and 'Remainers' are more ideologically extreme. Labour partisans who discuss politics on Twitter are more left-wing than Labour partisans who do not use Twitter. Equally, politically active Remain supporters are the most ideologically extreme across all three measures. This may not mean we are more likely to hear their voices in political discussion[11], but those we do hear are likely to be more ideologically extreme than the rest of the population.

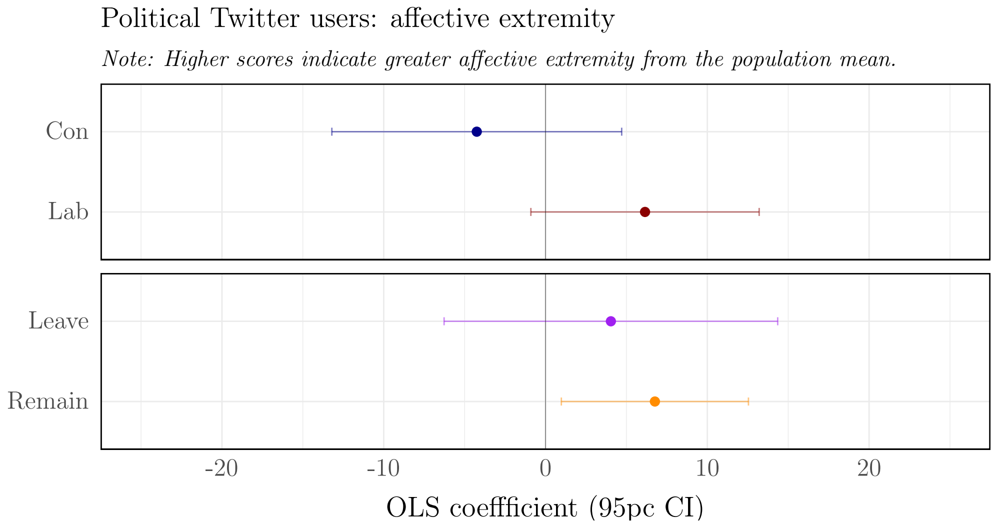
Moving on to affective polarisation, in Figure 3.3 we again plot coefficients of a series of OLS regressions, with Heckman corrections applied, this time modelling the affective extremity of political Twitter. We do this first by calculating the difference in feelings of favourability towards in-group and out-group voters via standard 'thermometer' measures, a widely-used indicator in political science (Lelkes and Westwood, 2017) which is effective in capturing emotional attitudes towards both out-group parties and voters (Druckman and Levendusky, 2019) and in multi-party contexts (Gidron et al., 2020; Reiljan, 2020; Wagner, 2021). First, we measure the difference between two 0-100 ratings, indicating feelings of favourability/unfavourability towards voters on either side of either the Conservative/Labour or Leave/Remain divide. Higher scores equal greater 'warmth' or favourability, so a greater difference between the scores given to in-group and out-group equals greater affective polarisation. To measure affective extremity, we then calculate the distance of this score from the population mean[12].

As with ideological attitudes, we find Remain-supporting political Twitter to be more affectively extreme than its non-political Twitter users, while this is not the case for Conservative, Labour, or Leave partisans. Interestingly, unlike any other group, Conservative political Twitter users actually report

[11] Twitter (now X)'s algorithms still generally result in tweets from higher-follower accounts being more visible.

[12] Full results and wording of questions are in Appendix B.5.

slightly lower degrees of affective polarisation; however this is not statistically significant. With both Leave and Remain partisans reporting relatively high levels of affect - particularly those engaged with politics online - the UK appears to be more divided along EU attitudes than traditional party lines on Twitter. Clearly, Remainers who discuss politics on Twitter are significantly more affectively extreme than those non-political users, and have more negative attitudes towards Leavers than vice-versa.



Political Twitter users: affective extremity

Note: Higher scores indicate greater affective extremity from the population mean.

Figure 3.3: Affective extremity: Political Twitter users vs. non-political Twitter users.

Note: Plot shows within-group comparison between political and non-political Twitter users, controlling for age and education and with Heckman corrections applied. Dependent variable is affective extremity, measured as distance from population mean in thermometer score difference between in and out-group partisans.

3.3.2 Tweet behaviour and extremity

While our results show that certain types of political Twitter users are more ideologically and affectively extreme, we still do not know how these attitudes might relate to their online behaviour. Does ideological or affective extremity predict the volume of political tweets? Figure 3.4 plots the relationship between ideological extremity and the number of political tweets posted by our respondents with identifiable Twitter accounts. This time, ideological extremity becomes an independent variable, normalised to represent a simple measure of extremity from the population mean[13]. While we find no statistically-significant relationship between ideological attitudes and Twitter behaviour in Leavers, Conservatives, or Labour partisans, ideologically-extreme Remainers on the liberal-conservative and EU dimensions are more likely to tweet politically[14]. Of course, these results speak to the high salience attitudes towards the European Union in the years following the Brexit referendum, particularly among those on the 'losing' side. However, given that we collected a maximum of 50 tweets from each individual, over two relatively short periods, this is significant. Continued over several years, this pattern might result in a far greater likelihood of seeing polarised political content on similar digital platforms.

[13] As opposed to Figure 1, where ideological extremity is either negative or positive depending on whether users are more or less left/right-wing, more or less socially-liberal/conservative, or more or less pro-EU than the population average.

[14] Full results can be found in Appendix B.6.

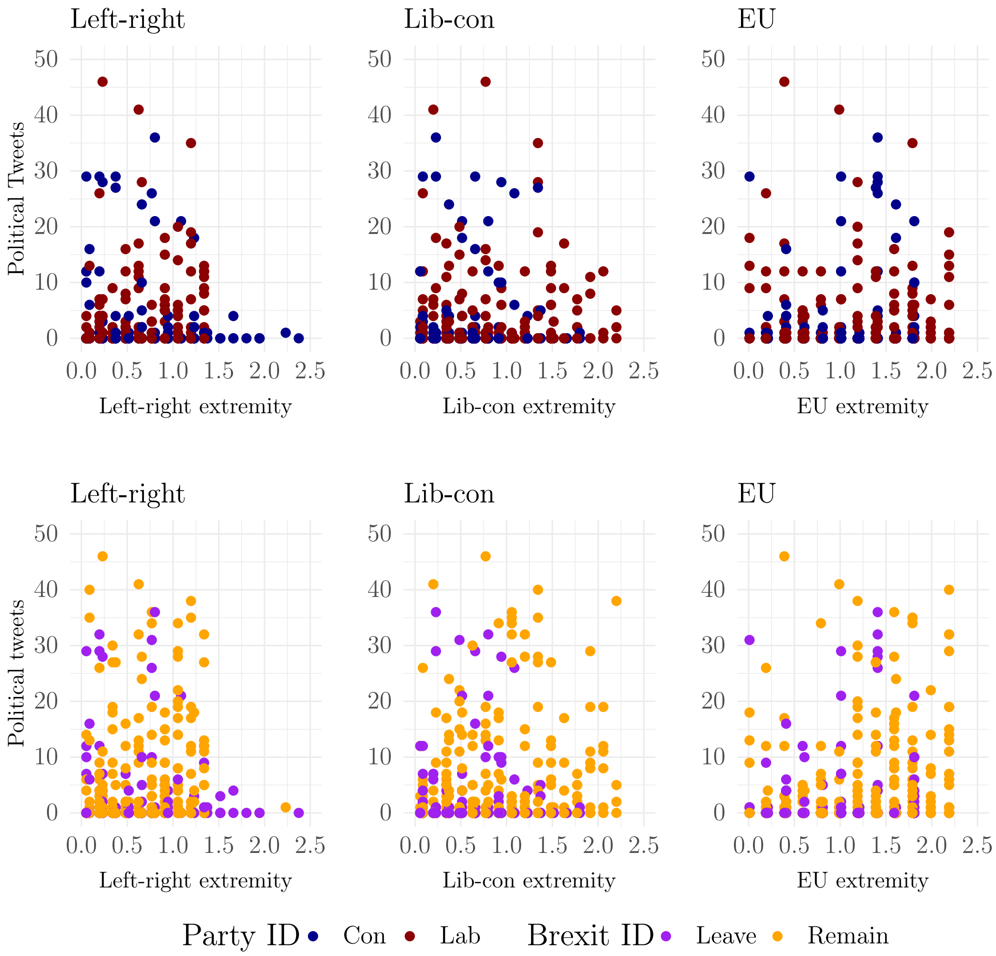


Figure 3.4: Ideological extremity vs. political tweets shared

Note: X-axis plots the ideological extremity of Twitter respondents, as distance from population mean across three separate dimensions, against the number of political tweets shared.

Moving on to examine affective attitudes, again we look at thermometer scores. Table 3.3 shows the results of a series of OLS regressions, modelling the relationship between our measure of affective extremity and the number of a) political tweets and b) negatively-political tweets posted. Unlike ideological extremity, we see a relationship between affective attitudes and tweeting politically in three out of our four partisan groups. Leave and Remain supporters reporting higher affective extremity on our feelings thermometers share more political and negatively-partisan tweets, while more affectively-extreme Labour supporters are also more likely to share negative tweets. Interestingly, affectively-extreme Leavers are slightly less likely to tweet negatively than Labour and Remain supporters. This is perhaps unsurprising, not only given our previous results, but due to the political dynamics in play during the study period. The United Kingdom was undergoing the difficult and fractious process of leaving the European Union and, as a result, there was much for both sides to discuss. Strong Leavers were more likely to share tweets in support of the UK leaving the European Union. This is in contrast to affectively extreme Remainers, who were more likely to post negatively-political tweets, highlighting the perceived drawbacks of the Brexit process.

Table 3.3: OLS models: Affective extremity against tweet frequency

	Con	**Lab**	**Leave**	**Remain**
Political tweets	-0.01	0.02	0.04*	0.04*
	(0.02)	(0.02)	(0.02)	(0.02)
Negative political tweets	-0.01	0.03*	0.02*	0.04*
	(0.01)	(0.01)	(0.01)	(0.02)
Observations	153	215	159	339

*Note: * = p<0.05. Models the relationship between affective extremity (as distance from the population mean) and number of political tweets posted, controlling for age and education. Heckman corrections applied and standard errors in parentheses.*

3.3.3 Polarisation of language

Are political Twitter users polarised in the language they use? Instead of looking at self-reported measures, we can apply natural language processing methods to our rich dataset of over 1.7m tweets, finding answers through our respondents' behaviour. Running these tasks over such a large dataset would be prohibitively expensive, in terms of both time and computational power, so we take a 10% random sample (168,000 tweets) from the dataset. Then, zero-shot classification is used to label political and non-political tweets using the BART Large MNLI (Multi-Genre Natural Language Inference) transformer model (Lewis et al., 2019). Zero-shot learning differs from traditional machine learning, in that a model can recognise and classify objects or concepts without training data. We extract all tweets classified as 'political' with 70% confidence reported by BART-MNLI. The resulting dataset comprises 31,167 tweets. Additionally, we apply this process to classify the sentiment of each tweet as either positive, negative, or neutral, and label tweets as such if they are classified with 70% confidence. A 1,000 tweet random sample was then manually checked to ensure they had been correctly classified as political.

Figure 3.5 shows the monthly totals of positive and negative political tweets shared, grouped by partisan identity. Negative political tweets are shown with a solid line, while positive political tweets are shown with a dashed line. While the number of positive tweets shared is low, and stays relatively stable across time for both partisan and Brexit identity, interestingly the proportion of negative tweets increased for Leave, Remain, and Conservative supporters, but not for Labour partisans. Midway through 2016, the number of negative tweets begins to increase and Remain supporters generally shared more negative tweets than Leave supporters in the period 2015-2017, during the referendum campaigns and in its immediate aftermath. From 2018-2020, however, this trend was reversed, perhaps due to frustration at the lack of progress in the United Kingdom leaving the European Union. There is an obvious correlation in the number of negative tweets shared by Conservatives and Leavers, but not to the same extent between Labour and Remain partisans. Surprisingly, Labour supporters shared relatively fewer negative political tweets, despite their party being in opposition. Given that online negativity increased for those on both sides of the Brexit - but not partisan - divide, this helps to provide further evidence of two things which occurred over the last decade. One, the increased animosity of British political Twitter and, two, the increased salience of Brexit identity - and the strength of feeling associated with it.

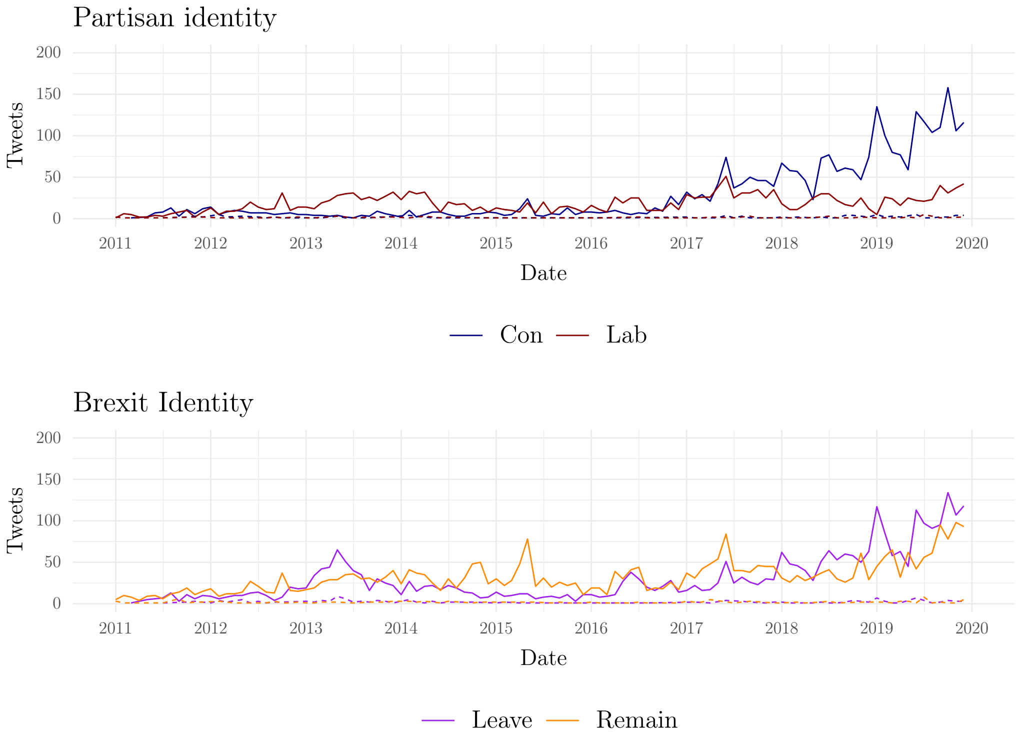


Figure 3.5: Negative and positive tweets shared, by month and partisan identity.

Note: Negative tweets shown by solid line, and positive tweets shown by dashed line. Number of tweets grouped and plotted by month and partisan identity.

While British political Twitter may have become increasingly negative, are they tweeting negatively about the same things? First, to try and understand whether our partisan groups have become increasingly polarised in their language on Twitter, I plot the cosine similarity of their tweets over time. Cosine similarity (Equation 3.1) compares how related two numerical vectors are to one another by calculating the cosine of the angle they make:

$$\cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{3.1}$$

The resulting cosine similarity metric ranges from 0 to 1, with higher values indicating greater similarity between the vectors. In this case, we take a broad measure of how similar the tweets of both Labour and Conservative, and Remain and Leave, partisans are year-by-year. Figure 3.6 shows a clear pattern: cosine similarity for both party and Brexit groupings actually increases over time. Despite a divergence in the ideological and affective extremity of certain groups within British political Twitter, there is a convergence in the language they use to discuss it. While outside the scope of this paper, future research might examine whether ideologically and affectively-extreme British political Twitter is using increasingly similar rhetoric, or even if its discussion has converged on a limited number of political issues.

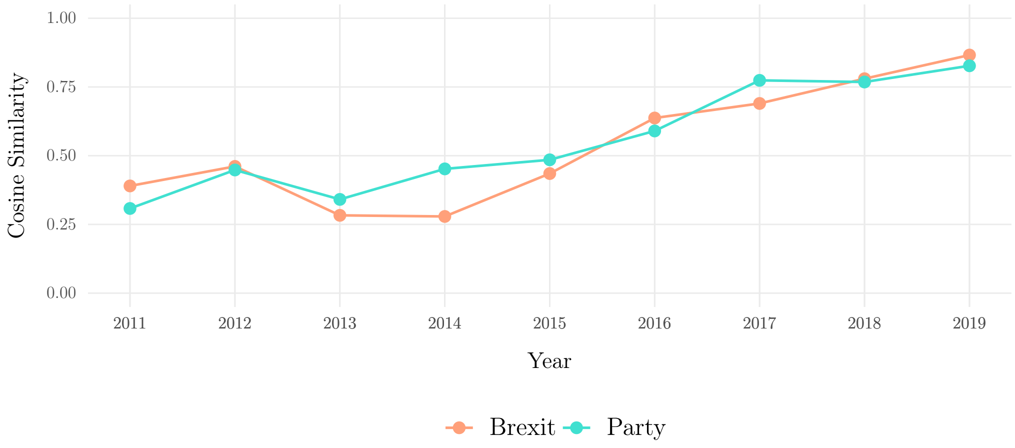


Figure 3.6: Cosine similarity of tweets over time, by political identity

3.4 Conclusion

Political polarisation remains a growing problem for democracies around the world and it is crucial to understand its underlying causes (Mason, 2018; Iyengar et al., 2019; Reiljan, 2020; Wagner, 2021). While we make no causal claims, our findings make the following contributions to identifying the ideological and affective extremity of online debate. First, if social media *is* driving political polarisation, we should expect to see that users of these social media sites are more ideologically and affectively extreme. We demonstrate that people who discuss politics on Twitter are generally more ideologically and affectively extreme than the rest of the population. Second, we identify asymmetry in the extremity of different partisan groups. We find that, on different ideological and affective dimensions, political Twitter's Labour and Remain supporters are attitudinally different to their co-partisans who are not so online. Third, and finally, we find a positive relationship between different types of extremity and the likelihood of producing political tweets, therefore amplifying the voices of those further towards the ideological and affective periphery.

Why are more extreme partisans from these groups disproportionately represented? It could be partially attributed to self-selection effects; perceptions of Twitter being more left-leaning or more pro-EU may have made it a more appealing space for Labour and Remain partisans to discuss politics[15]. If 'echo chambers' do exist, we might expect attitudes to become reinforced over time, as well as attracting new partisans to become politically active on the platform. Future research into how the partisan composition of social media sites, and the attitudes of their users, change over time could shed light on this subject. Given the timing of the study, our findings also

[15] This perception, however, has almost certainly shifted towards the 'right' since Elon Musk's takeover of Twitter in October 2022.

raise questions about some of the underlying political dynamics at play. On most of our measures, Brexit appears to be a more salient and divisive fault-line than traditional party politics. The most ideological and emotionally attached Remainers also tweet more about politics, while the more prolific tweeters in the Leave camp of British political Twitter are more affectively, but not *ideologically* extreme. There certainly seems to be a clear divide between political 'winners' and 'losers', particularly in relation to the EU referendum, where many Remainers felt a degree of anger, not only about the result, but about the process itself (Tilley and Hobolt, 2023). We know that high-arousal emotions such as anger are related to partisan sorting (Webster and Abramowitz, 2017) and can motivate people to participate in political activity (Marcus, 2000; Brader, 2005). Twitter represents a low-cost method of political participation, and a convenient outlet for these emotions. This appears to be supported by our finding that greater affective extremity correlates with a greater number of political, and negatively-partisan tweets. It follows that the asymmetry we witness on Twitter could also be a function of a government-opposition dynamic. For the vast majority of Twitter's existence, the Conservative Party was in government. Knowing that opposition parties are more likely to use negative rhetoric than incumbent parties (Crabtree et al., 2020), it is perhaps unsurprising that Labour supporters are likely to be more motivated to take to Twitter to attack the government. Again, this appears to be supported by our finding that more ideologically-extreme Remain and Labour Twitter users are more likely to share political, and negatively-partisan, tweets. Overall, we highlight the worrying potential for social media platforms such as Twitter to exacerbate polarisation, since it is dominated by individuals who are more ideologically and affectively polarised.

4 | Opting out of political discussion on Facebook

Abstract

On our most sensitive political topics, social media sites often represent centres of conflict rather than consensus. Do online platforms facilitate engagement with these issues - and why might people opt out of discussing them? In a pre-registered lab-in-the-field experiment on Facebook, I assign political partisans in the United States to groups containing news coverage of either divisive or unifying political issues, comprised of either all like-minded partisans or a mixture of Democrats and Republicans. While I do not find that either of these treatments lead to lower engagement alone, I show that the combination of exposure to contentious political topics and mixed partisan groups causes people to disengage. These findings have important implications for our understanding of online democratic debate, and how social media may be contributing to political polarisation.

4.1 Introduction

Around three-quarters of the population of the United States use social media in some form (Pew, 2021). They connect with friends and family, share news, find entertainment and, sometimes, discuss politics. Open political dialogue is a crucial part of a healthy democracy (Habermas, 1989) and, in its early days, the internet promised a 'purer' version of the Habermasian public sphere. With easy access to diverse perspectives, low barriers to accessing information, and freedom from traditional institutional gatekeepers, it offered the potential for exposure to increasingly heterogeneous networks and an unprecedented ability to organise collective action (Shirky, 2009). In theory, this would promote an informed citizenry, capable of making better policy decisions through rational debate, in turn radically improving democratic deliberation and political outcomes (Scott, 1999; Delli Carpini, 2000; Kent Jennings and Zeitner, 2003). Despite this, 70% of social media users in the United States never or rarely post about political issues (McClain, 2021), with many people increasingly weary and frustrated by the political content they encounter online (Sveningsson, 2014; Duggan and Smith, 2016). Rather than representing a space where everyone feels able to articulate their views, people shy away from discussing politics online, and social media appears to replicate face-to-face political avoidance (Eliasoph, 1998) in a digital environment (Mor et al., 2015; Mascheroni and Murru, 2017). Moreover, frequent social media use has been shown to reduce the likelihood of discussing politics *offline* as well (Hampton et al., 2017).

Why are people reluctant to discuss politics on social media? I argue this is primarily caused by two factors. First, when the issue being discussed is controversial and divisive, the increased likelihood of social sanctions causes people to disengage. This dynamic closely aligns with Elisabeth Noelle-Neumann's spiral of silence theory (1974), which suggests that individuals hesitate to express minority opinions due to fear of social isolation. It creates a self-reinforcing cycle where dominant opinions become more prominent, and consequently perceived as the majority, while dissenting views are suppressed. Second, I argue that group identity plays a critical role; when contentious topics arise, individuals often engage or disengage in alignment with perceived group norms. While it is generally accepted that self-censorship mechanisms like the spiral of silence exist, to at least some extent, in offline situations (Salmon and Neuwirth, 1990; Glynn et al., 1997), we know far less about how they function online.

Most research in online political communication has focused upon who we hear and how they express their views. However, examining who withdraws from political discussion, and the reasons behind their disengagement, can offer us deeper insights into how social media shapes political discourse. To date, studies on self-censorship in social media contexts have relied almost exclusively on surveys or lab experiments, leaving open questions about how these dynamics unfold in real-world settings. To address this gap, I conduct a pre-registered lab-in-the-field experiment on Facebook, recruiting political partisans from the United States and assigning them to groups discussing either contentious or consensus political topics, composed of either all co-partisans or a mixture of Democrats and Republicans. While I do not find that, in isolation, contentious issues or mixed partisan groups result in disengagement, I find evidence this occurs when the treatments are combined. These findings have important implications for our understanding of online democratic debate, and how social media may be contributing to political polarisation.

4.2 Social media and political discussion

One of the most common, enduring, and important ways that people engage with politics is through discussion. Social media have introduced entirely new spaces for deliberation, albeit with significant representational imbalances where certain voices dominate and others are excluded. On a basic level, we know that platforms like Facebook and Twitter do not represent the general population (Mellon and Prosser, 2017). In the U.S., for instance, those discussing politics on Twitter tend to be disproportionately male, urban, and ideologically extreme (Barberá and Rivero, 2015). Meanwhile, political discourse on Facebook also skews toward particular demographics. For example, older and more conservative users were more likely to share misinformation during the 2016 U.S. elections than their younger, liberal counterparts (Guess et al., 2019b). Further, social media tends to amplify voices that employ particular communication styles; populist politicians often use negative, high-arousal rhetoric to evoke emotions like enthusiasm or fear, which are particularly persuasive and drive engagement (Brader, 2005; Arceneaux, 2012). Recommendation algorithms reward this kind of content by making it more visible, leading to a cycle where sensationalist rhetoric spreads faster and reaches wider audiences than more measured discourse (Brady et al., 2017; Berger and Milkman, 2012). A bias towards particular demographics means it is likely that much online discourse is unrepresentative of wider populations, while the prominence of emotive rhetoric signals a shift towards more sensationalist and less factual political communication. Neither of these things are desirable for a healthy democracy. We can, and have, learnt a lot from the wealth of research dedicated to who we hear online and how they communicate. However, we know far less about who decides to switch off from online political discussion, and why.

4.3 Online ‘spirals of silence’

What might be causing people to disengage from political discussion? One explanation of disengagement and self-censorship is Elisabeth Noelle-Neumann’s ‘spiral of silence’ theory (Noelle-Neumann, 1974), which has become one of the most studied and debated works in political communication. Based on the premise that individuals fear social isolation, one of its key hypotheses is that people are motivated by this fear to constantly monitor their environment and, based on whether they perceive themselves to be in the minority or majority, choose whether to publicly express their opinion on contentious issues. Evidence supporting the theory, however, is mixed. Meta-analyses have found limited support for the existence of spiral of silence mechanisms in society as a whole (Glynn et al., 1997), pointing towards methodological challenges in testing the theory on a holistic level (Matthes and Hayes, 2014). There is some evidence to suggest they might occur online; Hayes and Matthes’ (2017) meta-analysis of 66 studies, incorporating more than 27,000 participants, found no discernible difference in willingness to express opinions between online and offline environments. Specifically examining social media, Gearhart and Zhang (2014; 2015) also found support for the existence of spirals of silence on Facebook.

It is plausible that social media fosters self-censorship. First, these platforms provide myriad ways to monitor one’s online opinion environment — from engagement metrics supporting posts (Von Sikorski and Hänelt, 2016) to simple polls. This heightened visibility can increase fears of social isolation, affecting the likelihood of posting, especially when one’s online network is large or represents a conflicting opinion climate (Chen, 2018). Second, in most instances, social media are either public or semi-public environments. On public sites, even where anonymity is possible like X or Reddit, the risk attached to sharing unpopular opinions is too great for most people. A lack of trust or potential risk of a range of sanctions - social or otherwise - play a role in changing peoples behaviours. For example, Stoycheff (2016) found that people were less willing to discuss the Edward Snowden-NSA story on social media than they were in person, knowing that one's online activities could be subject to monitoring. Third, and finally, content posted online can remain online indefinitely and the lifetime of a message can be unpredictable (Bäck et al., 2019). Even if the consequences of dissonant opinion don't exist in the 'real-world', online sanctions can still be highly unpleasant, most often in the form of online abuse (Eckert, 2018; Jones et al., 2020), humiliation, threats, or even the possibility of being 'doxxed' (Douglas, 2016).

On the other hand, there are several reasons social media might inhibit spirals of silence. First, while Noelle-Neumann initially grounded her theory in a fear of 'real-world' social isolation, the advent of online communication has introduced a distinctive socio-psychological space which is significantly different from traditional face-to-face interactions. Studies by Ho and McLeod (2008) and Woong Yun and Park (2011) indicate that computer-mediated communication can circumvent instincts toward conflict avoidance or social isolation; people may be just as willing to express their opinions online, regardless of whether they perceive themselves as holding a minority or majority view. The absence of physical presence on social media, it is argued, shields individuals from fear of judgement, making the digital realm a less intimidating space for expressing political opinions. This builds on Meyrowitz (1986)'s seminal work on situationism and the media, which explored how electronic media eroded traditional communication barriers and accelerated socio-political change. We know that a similar type of 'context collapse' (i.e. dilution of how 'personal' a user's network is) can occur on social media, where the lines between personal networks and 'audiences' become blurred (Marwick and Boyd, 2011). The greater the context collapse, the more likely someone is to share content (Beam et al., 2018a). Further, we know that social media generally lowers the threshold for political activity (Boulianne, 2015; Vaccari and Valeriani, 2018); they offer many more ways to engage with content, such as simple 'one-click' reactions and shares (Pang et al., 2016). These actions do not necessarily require the articulation of a viewpoint, requiring neither the same cognitive effort or risk of isolation present in a face-to-face interaction. Therefore, the decision to share one's views may be less dependent on other factors.

Second, spirals of silence are dependent on the existence of a consonant opinion climate. Noelle-Neumann conceived the theory when traditional broadcast media was the dominant source of news; controversial and morally-laden issues would often receive extensive media coverage and public discussion, making it easier for people to gauge prevailing views and determine whether they align with the majority or minority. This dynamic evolved to some extent in the second half of the 20th century: Moy and Hussain (2014) highlight the potential impact of hyper-partisan media in the United States, where increasingly polarised media on both 'left' and 'right' meant that people became less likely to feel politically-isolated and therefore censor their opinion. Unlike Noelle-Neumann's conception of a consonant and dominant opinion climate reinforced by mass media, online environments offer a huge range of opinions. Social media makes it even easier to find like-minded partisans and have one's opinion, no matter how 'fringe' it might be, validated. Instead, the internet might offer not one universal majority opinion climate, but many smaller and localised majority opinion climates depending on the content or network selected by the user (Schulz and Roessler, 2012). It follows that individuals are more likely to find consonant opinions and self-select into like-minded political networks, making majority pressure weak and the perceived cost of social isolation low. The anonymity offered by certain social media sites reduces this cost even further, making it easier to voice unpopular opinions (Wu and Atkin, 2018). Third, and finally, spirals of silence may only apply to specific types of people. We know that heavy social media use has a direct, negative relationship with deliberation in many *offline* settings, with only those holding the strongest attitudes immune (Hampton et al., 2017). Do these challenges mean that, in the words of Katz and Fialkoff (2017), the spiral of silence is a 'concept about to retire'? I argue that two of its central tenets can nevertheless offer us insight into why people might disengage from online political discussion.

4.4 Issue type

In line with the spiral of silence theory, the first key factor in disengagement from online political discussion is, I argue, the issue under discussion. Noelle-Neumann asserted that a spiral of silence can only happen when the issue being discussed contains a strong moral component or is, in other words, controversial and emotionally-laden (Noelle-Neumann, 1974; Noelle-Neumann and Petersen, 2004). There is modest support for the notion that self-censorship may indeed occur in relation to these types of issues - for example with abortion (Salmon and Neuwirth, 1990) or affirmative action (Moy et al., 2001). Moreover, the significance of the issue extends specifically to social media contexts. Research by Gearhart and Zhang (2018), utilising Facebook surveys, reveals differences in individuals' willingness to express opinions based on enduring, emerging, and transitory issues. However, their study focuses exclusively on the contentious topics of abortion and gay marriage, and, to my knowledge, no research has yet differentiated between contentious and non-contentious issues. This contrast between issue types is central to our understanding of political disengagement for two reasons. First, controversial issues tend to evoke strong emotional responses. Some authors argue that opinions are purely evaluative or 'rational' appraisals of issues but, when imbued with emotion, they become *attitudes*, which are much more deep-rooted and harder to change (Aronson, 1972). Second, issues that are morally-charged also often come with strong societal norms about what is considered acceptable or unacceptable. When discussing such issues, individuals may face normative pressure to conform to these societal expectations, even if it means suppressing their true opinions to avoid conflict or moral condemnation. Combined, these factors increase the likelihood of social sanctions for those who hold minority views, resulting in their disengagement from discussion. In an online environment, such sanctions might manifest as negative comments that harm an individual's social standing, personal abuse, or even the aforementioned threat of doxxing. Given Noelle-Neumann's premise that controversial, morally laden issues exert greater social pressure for conformity than uncontroversial topics, I hypothesise that Facebook users will engage more with the latter than the former:

- H_1: *Participants exposed to contentious political issues will engage less than those exposed to consensus political issues.*

4.5 Group identity

Diverging from the spiral of silence theory, I contend that group identity serves as a second key factor influencing disengagement from online political discussions. The phenomenon of social conformity is well-documented in social science: people either share or censor their opinions to fit in with a group norm (Festinger, 1950), being either socially rewarded for expressing views that reinforce a majority opinion, or socially penalised for views at odds with the majority opinion (Schachter, 1951). In public settings, people typically present a social identity that conforms to prevailing norms, even when their private beliefs diverge from the majority (Goffman, 1959). Solomon Asch's classic experiments on group conformity showed that individuals often yield to group pressure, even when an opinion is clearly incorrect or is counter to their own perceptions or beliefs (Asch, 1955).

Social identity theory (Tajfel and Turner, 2004) has helped us understand that group membership shapes social perceptions and behaviours, influencing how members relate to others in and outside of their social groups. For example, we know that homogeneity can reaffirm the certainty of group members attitudes (Druckman and Nelson, 2003). We also know that partisans in *mixed* groups are less likely to indulge in motivated reasoning during political discussion (Klar, 2014), perhaps through the 'perspective-taking' (Mutz, 2002) that occurs through exposure to cross-cutting beliefs. Such exposure, however, has become less and less common; in today's particularly-polarised United States (Layman et al., 2006; Mason, 2018), people are less willing to discuss political issues than non-political issues, particularly with out-group partisans (Settle and Carlson, 2019). This increasing polarisation has made political identity more salient when it comes to political discussion, where group identity tends to manifest in partisan attachment (Green et al., 2004; Iyengar et al., 2019), and often acts as a heuristic for the perceived opinions of out-group partisans on selected political issues.

Identity cues also matter in online environments. Users are more likely to interact positively with posts created by those with known and positive reputations, when compared with anonymous information-sharers (Taylor et al., 2022). These users also react quicker to 'identified' content, leveraging prior reputation as a heuristic for their opinion of a piece of content. In a study of university students, Miller et al. (2015) found that centrality in a network predicted the likelihood of engaging in political discussion - but at the same time these pivotal individuals tended to minimise social risk by confining their political interactions to like-minded peers.

These motivations and risks are not uniform across all social media platforms. Willingness to self-censor, I argue, can vary according to the structure of social network sites. On private (or at least semi-public) platforms like Facebook, online and offline networks more frequently overlap (Pang et al., 2016), heightening the risk of social isolation. Fox and Holt (2021) argue that, on Facebook, any fear of social isolation is reduced due to being surrounded by our closest ties who either share our opinions - or at least love us in spite of them. While Facebook networks tend to be quite homogeneous (Lönnqvist and Itkonen, 2016), their more personal real-world nature - compared with platforms like X or Reddit - limits users' ability to curate their networks based on ideological conviction or partisan identity. Research has shown that perceived alignment of opinion with friends and family, rather than society at large, predicts a willingness to speak out (Moy et al., 2001). Consequently, network homogeneity can make it more difficult to express a dissenting opinion. A fear of social isolation exists on a site like Facebook, precisely *because* one is surrounded by the strong social ties of family, friends, and co-workers. To recall Marwick and Boyd (2011), a user's context 'collapses' to a far lesser extent on personal social networks, and real-life relationships can be imperilled by disagreement. Finally, in a more general sense, we know people are less willing to discuss political issues than non-political issues, particularly with out-group partisans (Settle and Carlson, 2019). Therefore, I argue that group identity is important for political discussion; people in groups dominated by their partisan in-group are more likely to put forward an opinion, with those in mixed groups more likely to self-censor:

- H_2: *Participants in mixed partisan groups will engage less than those in homogeneous groups.*
- H_3: *Participants exposed to contentious political issues in mixed partisan groups will engage less than those in other treatment conditions.*

In summary, I argue that group composition and issue type are paramount in determining the likelihood of individuals engaging in political discussions.

4.6 Experiment design

Existing studies of self-censorship on social media have been conducted exclusively through either surveys or lab experiments (e.g. Gearhart and Zhang, 2014, 2015; Chan, 2021). As such, we have yet to understand how these dynamics might play out in a real-world environment. Gaining insight into these mechanisms will further our understanding of how democratic deliberation is articulated online. To address this gap, I conduct a pre-registered[1] lab-in-the-field online experiment to test my hypotheses[2]. Using the unique structure of Facebook groups, I assign participants to different study conditions and isolate the effect of different variables. Importantly, the experiment is designed to maximise ecological validity by replicating participants' daily experiences on Facebook, including scenarios where they may engage in political discussion.

4.6.1 Participant recruitment

I studied active social media users from both sides of the partisan political divide in the United States, focusing on supporters of the Democratic and Republican parties. Recruitment commenced in August 2023 using targeted advertisements on Facebook which invited participants to complete a short survey for a chance to earn a $20 gift card. The survey had three primary objectives. First, it collected standard demographic information, including age,

[1] The pre-registration plan can be found here: `https://osf.io/p3vdk`. Hypotheses reference differences in 'average engagement' between treatment groups; this was subsequently considered to be a less precise measure of testing the outcome of interest - disengagement - than a simple binary indicator. Count models, as specified in the pre-registration plan, find broadly similar results to a binary variable. These are detailed through robustness checks in Appendix C. H_1, H_4, and H_5 were ultimately not tested in this paper.

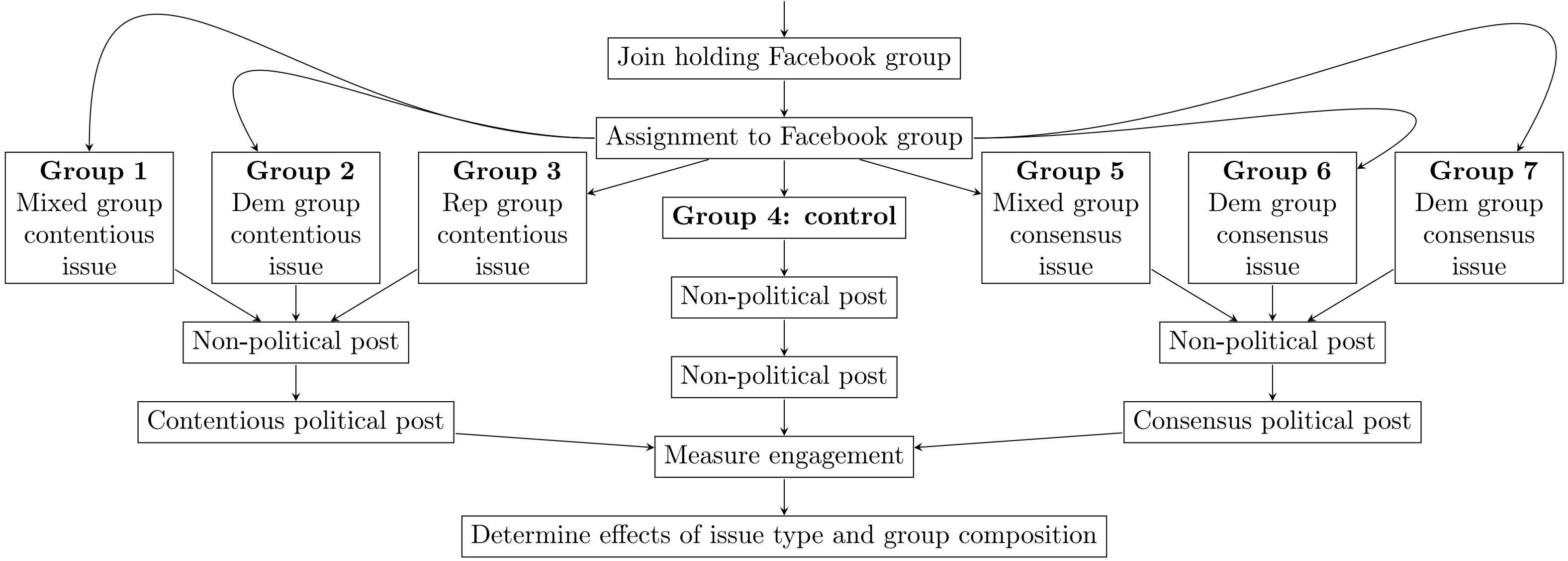
[2] Ethics approval was granted by the Research Ethics Committee of the London School of Economics and Political Science

education, and gender. Second, while contentious and consensus issues were selected in part based on public opinion polls, the survey aimed to validate these choices. Specifically, I assessed the partisan distribution of opinions regarding the ‘contentious’ topics of immigration and abortion, as well as the ‘consensus’ issues of the economy and education. As anticipated, opinions were significantly more polarised among Republicans and Democrats on the contentious issues compared to the consensus ones[3]. Third, and finally, I measured partisan identity and its strength using a battery of five statements that focused on parties and voters as objects of affect. Responses ranged from ‘strongly disagree’ to ‘strongly agree’, recorded on a Likert scale from 1 to 5, where 1 indicated strong agreement and 5 signified strong disagreement.

At the end of the survey, respondents were asked to provide their Facebook profiles for potential inclusion in the main study. Verified accounts were then invited to join a Facebook holding group, which served as a centralised platform for study-related communication. To ensure data integrity, a three-step verification process was implemented. First, Qualtrics settings were configured to flag suspicious or automated responses. Next, responses were manually reviewed to exclude non-U.S. participants where Qualtrics’ geo-filter might have failed. Finally, Facebook profiles were scrutinised for authenticity: new accounts, limited activity, few friends, or lack of a profile picture were potential indicators of fake profiles. Obviously fake accounts, or suspicious accounts with a combination of these ‘red flags’ were filtered out to help preserve data quality. From the holding group, participants were assigned to treatment or control groups based on their partisan identity and gender, ensuring balanced representation. An overview of the study design is shown in Figure 4.1.

[3] The selection of these topics is discussed in the next section, while graphs showing the partisan balance of opinion on each of these four issues can be found in appendix C.1.

Figure 4.1: Experiment design



4.6.2 Issue choice

To assess the effect of different issue types on participants' willingness to share their opinions, it was essential to identify and substantiate contentious issues. In line with Noelle-Neumann (1974), these topics needed to meet key criteria for spiral of silence mechanisms to emerge. First, there had to be a lack of consensus. Second, the topic required a moral dimension. Third, the issue needed to be salient, representing an ongoing, significant debate.

1. Abortion

The first contentious issue selected was the legal right to terminate a pregnancy. Abortion has long been a polarising issue in the United States; by 2023, 76% of Republicans identified as 'pro-life', while 84% of Democrats consider themselves 'pro-choice' (Gallup, 2023). This divide grew even wider, and the issue gained salience, following the Supreme Court's decision to overturn Roe vs. Wade in 2022. Abortion is also clearly a contested moral issue; according to Pew, 31% of U.S. adults consider abortion morally acceptable in either all or most cases, while 46% consider it morally wrong in either all or most cases. Only one in five adults (21%) believe it is not a moral issue at all (Pew, 2022b).

2. Immigration

The question of whether the United States should maintain open or closed borders is a long-standing ethical and moral debate (Zolberg, 2012). Public opinion is similarly split on the priority of addressing immigration, with 70% of Republicans but only 37% of Democrats considering it a top concern for the president and Congress (Pew, 2022a). Additionally, Republicans and Democrats diverge significantly on the goals of U.S. immigration policy (Oliphant and Cerda, 2023). For instance, 91% of Republicans believe that enhancing security along the U.S. - Mexico border is essential, compared to just 59% of Democrats.

3. The economy

The first consensus issue selected was strengthening the economy. In 2023, this was the public's most important policy priority (Pew, 2023), making it a highly-salient political issue. However, unlike the contentious issues above, partisans are in relative agreement: 84% of Republicans and 68% of Democrats believe that strengthening the economy should be a priority for the president and Congress.

4. Education

Similarly, improving education represents an uncontroversial issue among Americans. Like the economy, it remains a salient political issue, with almost 60% of voters believing it should be a top priority for the president and Congress (Pew, 2023). The partisan gap is also narrow, with 62% of Republicans and 51% of Democrats holding this view. It should also be noted that, particularly on education[4], the way that issues are framed can turn a relatively uncontroversial topic into a polarised debate. Steps taken to avoid priming participants when administering treatments are discussed in the methodology section. These choices were corroborated through answers to the pre-study survey, with figures showing the partisan balance of opinion on each of these four issues in appendices C.1-C.4.

[4] For example, the debate over school curricula in certain states.

4.7 Data collection

Participants were assigned randomly to either homogeneous or mixed partisan groups through a stratified approach. Table 4.1 shows an overview of these groups.

Table 4.1: Group overview

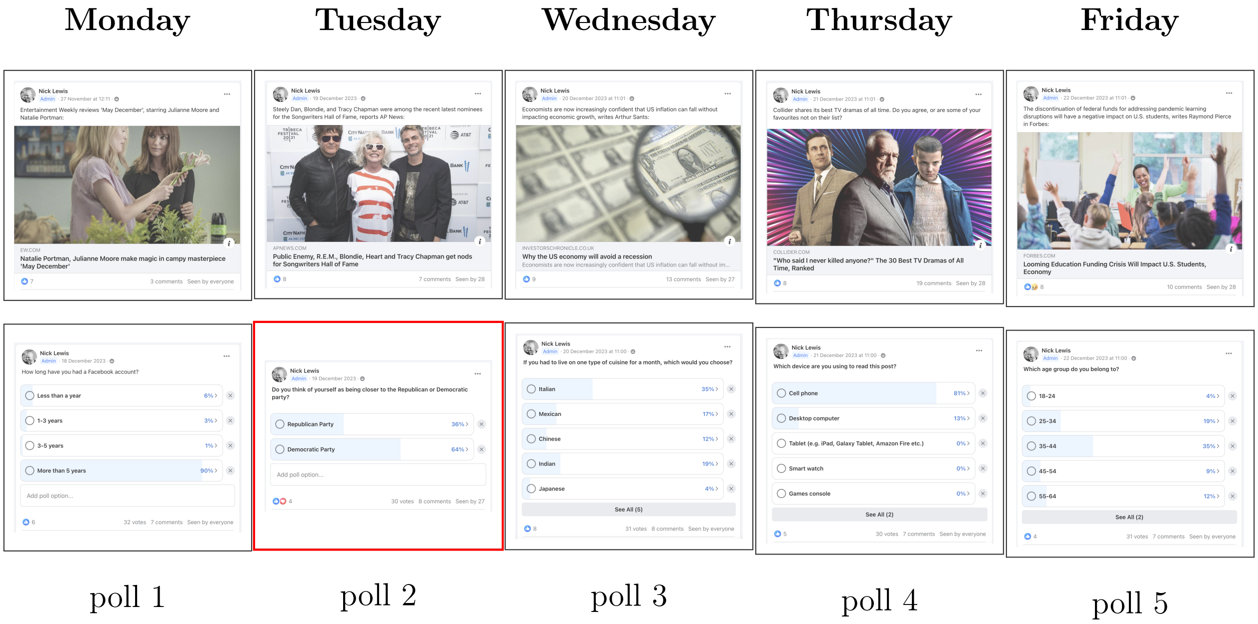
	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**	**Group 7**
group:	Mixed	All-Dem	All-Rep	Mixed	Mixed	All-Dem	All-Dem
content:	contentious	contentious	contentious	control	consensus	consensus	consensus
N	39	33	29	38	33	29	36
Mean age	38.01	41.7	40.4	35.9	46.6	40.4	35.1
Gender %							
Female	38.5	63.6	58.6	47.8	56.2	58.6	61.1
Male	61.5	36.4	41.4	52.3	41.4	41.4	39.9
Partisan ID %							
Democrat	43.6	96.1	0.0	54.5	53.1	100.0	100.0
Republican	56.4	3.9	100.0	45.5	46.9	0.0	0.0

Figure 4.2 gives an overview of required daily tasks and content shared within groups. Each study period took place over five days, from Monday to Friday. Participants were required to visit their assigned Facebook group on each of these five days and vote in a poll, posted at 6am Eastern Time. These polls asked participants to choose from a range of options in response to a range of set non-political questions: "*How long have you had a Facebook account?*", "*If you had to live on one type of cuisine for a month, which would you choose?*", "*Which device are you using to read this post?*" and "*Which age group do you belong to?*". These questions were asked on the same day in each study group. On day two of each study, participants were asked "*Do you think of yourself as being closer to the Republican or Democratic*

party?". As a central tenet of the spiral of silence, it was important to allow participants to easily monitor their opinion environment, through more than just the comments and 'likes' of their peers. In reality, this "quasi-statistical sense" (Noelle-Neumann, 1974) of an opinion environment is developed over a much longer period and uses less 'blunt' instruments; for example just using the likes or comments of those in one's network to paint a picture of the opinions and beliefs of others. Time and resource constraints meant this had to be expedited. Consequently, this poll was pinned to the top of each group as soon as it was posted, so that participants would have to scroll past it each day to complete their task. This served as an additional indicator of the distribution of partisan identity within the group, and as a heuristic for its likely balance of opinion. In each group, settings were amended so that content was presented to participants in reverse chronological order, i.e. with the most recent post appearing first. Task completion was monitored each day, with non-compliers sent a polite reminder about their involvement in the study at the end of the first day. Participants who did not complete their task on two consecutive days were removed from the study.

In each study condition, a daily media article was shared by the researcher. In the control group, this content provided a baseline of Facebook engagement with non-political topics; both the polls and articles were non-political in nature and sourced accordingly. Articles included movie reviews, an obituary, human interest stories, recipe suggestions, and a 'listicle' of TV recommendations. The treatment groups followed the same pattern but, on days three and five, articles on either consensus or contentious political issues were posted. To avoid any partisan priming, and isolate the effect of the issue discussed, these articles were selected from apolitical or centrist media outlets (e.g. Reuters, APNews, BBC News etc.) which presented a neutral or unbiased 'take'. The text used to share these articles to Facebook simply stated the focus of the piece and the author. These posts were scheduled to

Figure 4.2: Example daily content



Monday

Nick Lewis
Admin · 27 November at 12:11 ·

Entertainment Weekly reviews 'May December', starring Julianne Moore and Natalie Portman:

EW.COM
Natalie Portman, Julianne Moore make magic in campy masterpiece 'May December'

7 · 3 comments · Seen by everyone

Nick Lewis
Admin · 18 December 2023 ·

How long have you had a Facebook account?

- ○ Less than a year 6%
- ○ 1-3 years 3%
- ○ 3-5 years 1%
- ○ More than 5 years 90%

Add poll option...

6 · 32 votes · 7 comments · Seen by everyone

poll 1

Tuesday

Nick Lewis
Admin · 19 December 2023 ·

Steely Dan, Blondie, and Tracy Chapman were among the recent latest nominees for the Songwriters Hall of Fame, reports AP News:

APNEWS.COM
Public Enemy, R.E.M., Blondie, Heart and Tracy Chapman get nods for Songwriters Hall of Fame

8 · 7 comments · Seen by 28

Nick Lewis
Admin · 19 December 2023 ·

Do you think of yourself as being closer to the Republican or Democratic party?

- ○ Republican Party 36%
- ○ Democratic Party 64%

Add poll option...

4 · 30 votes · 8 comments · Seen by 27

poll 2

Wednesday

Nick Lewis
Admin · 20 December 2023 at 11:01 ·

Economists are now increasingly confident that US inflation can fall without impacting economic growth, writes Arthur Sants:

INVESTORSCHRONICLE.CO.UK
Why the US economy will avoid a recession
Economists are now increasingly confident that US inflation can fall without im...

9 · 13 comments · Seen by 27

Nick Lewis
Admin · 20 December 2023 at 11:00 ·

If you had to live on one type of cuisine for a month, which would you choose?

- ○ Italian 35%
- ○ Mexican 17%
- ○ Chinese 12%
- ○ Indian 19%
- ○ Japanese 4%

See All (5)

8 · 31 votes · 8 comments · Seen by everyone

poll 3

Thursday

Nick Lewis
Admin · 21 December 2023 at 11:01 ·

Collider shares its best TV dramas of all time. Do you agree, or are some of your favourites not on their list?

COLLIDER.COM
"Who said I never killed anyone?" The 30 Best TV Dramas of All Time, Ranked

8 · 19 comments · Seen by 28

Nick Lewis
Admin · 21 December 2023 at 11:00 ·

Which device are you using to read this post?

- ○ Cell phone 81%
- ○ Desktop computer 13%
- ○ Tablet (e.g. iPad, Galaxy Tablet, Amazon Fire etc.) 0%
- ○ Smart watch 0%
- ○ Games console 0%

See All (2)

5 · 30 votes · 7 comments · Seen by everyone

poll 4

Friday

Nick Lewis
Admin · 22 December 2023 at 11:01 ·

The discontinuation of federal funds for addressing pandemic learning disruptions will have a negative impact on U.S. students, writes Raymond Pierce in Forbes:

FORBES.COM
Looming Education Funding Crisis Will Impact U.S. Students, Economy

8 · 10 comments · Seen by 28

Nick Lewis
Admin · 22 December 2023 at 11:00 ·

Which age group do you belong to?

- ○ 18-24 4%
- ○ 25-34 19%
- ○ 35-44 35%
- ○ 45-54 9%
- ○ 55-64 12%

See All (2)

4 · 31 votes · 7 comments · Seen by everyone

poll 5

appear in each group immediately after the daily poll, meaning participants had to scroll past the treatment articles to reach their task. Treatment was gauged by Facebook's 'seen by' metric, showing which group members have viewed a post. At the end of the week, participants were compensated for their time with a $20 Amazon digital gift card.

4.7.1 Measures and estimation

As a test of the willingness of participants to express their views, I use two types of engagement as key dependent variables. First, *reactions*, which are measured at an individual level through reactions (e.g. 'likes', 'loves', 'sad', 'angry', 'wow' etc.), both directly to the post and to comments by other participants. Second, I measure *comments*, where a participant has posted a comment either in reply to a post or to comments made by other participants. I then break down the likelihood of reacting and commenting both 1)*overall*, which encompasses engagement with political *and* non-political content, and 2) purely *political* engagement, in relation to political posts or comments only. Participant engagement was higher than expected: Figure 4.3 shows the total number of engagements with posted content across all groups in the experiments, broken down by engagement type. This focuses on engagement with the treatment only: comments and reactions related to the daily task, which often prompted discussion[5], were excluded. Perhaps unsurprisingly, reactions, which require the least time, cognitive effort, or risk of social isolation, were by far the most common method of interaction, and I observed a substantial difference between the number of reactions to non-political content (354) and political content (169).

Comments, of which there were 234 in total, were far less common than simple reactions. Again, this is perhaps unsurprising given the relative time,

[5] Most notably in relation to the question 'if you could only eat one type of cuisine for a month, what would that be?'

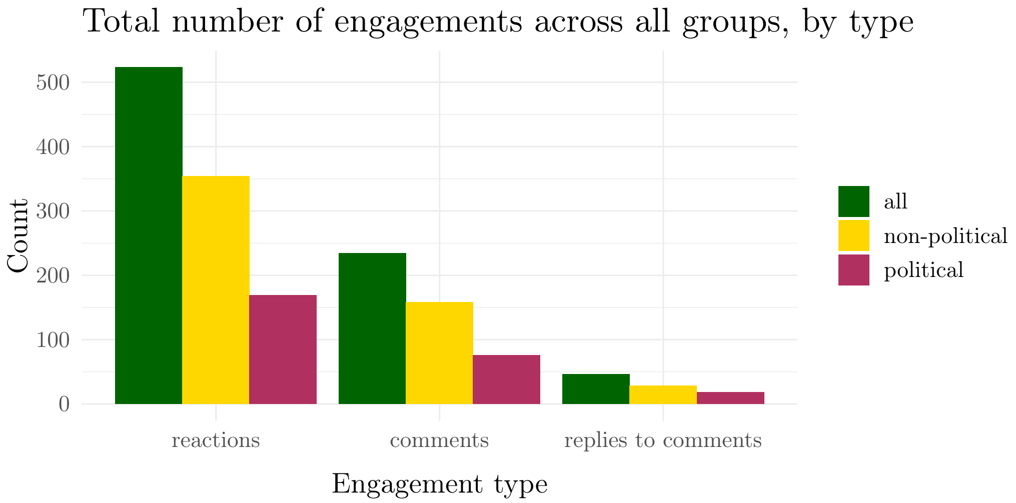


Figure 4.3: Engagement totals

cognitive effort, and potential cost of isolation involved in this type of post. Both political comments and political reactions comprise roughly a third of the total of all comments and reactions respectively. While replies to other comments were the least common type of engagement (46), it is reassuring that participants were engaged in at least some discussion with each other - as well as in direct response to each post. In terms of content, people expressed opinions in the group and were generally willing to engage in political discussion. This discussion was overwhelmingly civil and generally conducted within group rules, while no moderation action was required to either warn participants or remove content. In relation to contentious issues, comments ranged from nuanced (“My personal views are that abortion is murder of an unborn child ... but my views cannot be forced upon others, nor should they be.”) to the more strident (“Abortion IS murder”). Interestingly, on a few occasions, participants indicated that they were actually self-censoring their views. For example, in relation to abortion, comments included “*I try to stay out of this particular debate. So, I can’t really comment on it.*" and “if I speak I’ll be in trouble”. More succinctly, one person simply posted a ‘zipped mouth' emoji. Occasionally, discussion extended for a day or two after the official end of the study period each Friday. To account for this, data was typically gathered on the following Monday.

4.8 Analysis and Results

My first two hypotheses seek to test individual pillars of the spiral of silence theory: whether 1) issue type and 2) group composition influence the likelihood of engaging in discussion. Figure 4.4 shows the average treatment effect of exposure to contentious political issues and assignment to a mixed partisan group[6]. Looking at the average effect of these treatments in isolation, I find no evidence that 1) participants exposed to contentious political issues engage less than those exposed to consensus political issues and 2) participants in mixed partisan groups will engage less than those in homogeneous groups.

To further explore the nature of these relationships, Table 4.2 shows results from a series of OLS regression models which examine two-way interactions between issue and group type as treatments. As my intention is to measure whether specific scenarios lead to either engagement or disengagement, reactions and comments are coded as binary variables, indicating 1 where the number of reactions or comments is greater than 0. All reactions are visible to other participants, so come with a risk of social sanction, require time and cognitive effort. As such, I consider this a more useful measure of willingness to engage in both political and non-political discussion than a raw count of engagements. The latter, while interesting[7], could be subject to influence by a small number of highly-active users. With these binary variables, I use OLS

[6] This is calculated through a simple difference-in-means between collapsed treatment groups. The proportion of respondents engaged in each treatment condition, and calculation of ATEs, can be found in tables C.1 and C.2 in the appendix.

[7] OLS models examining engagement as a count variable can be found in appendix C.3.

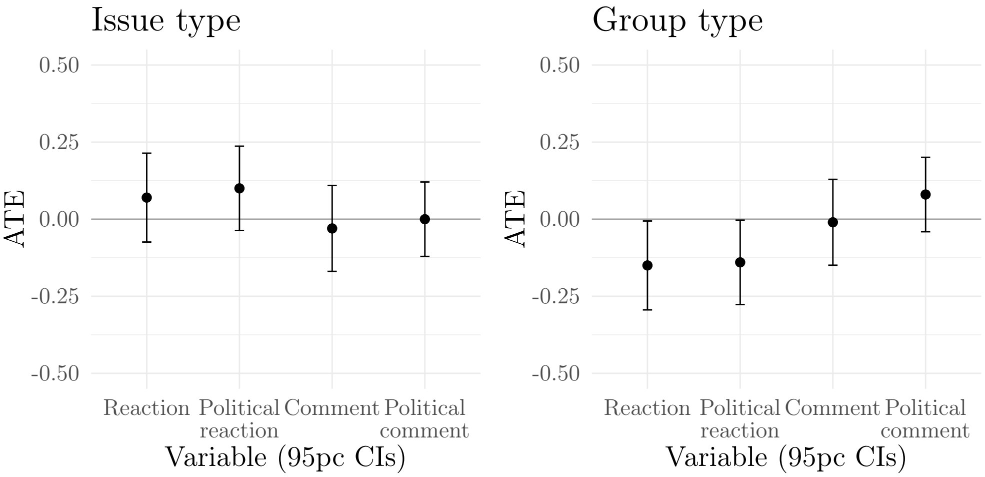


Figure 4.4: ATEs

Note: Average treatment effects are calculated from the proportion of participants who engaged in each study condition. Dependent variables are binary, indicating whether participants engaged or commented.

to estimate a linear probability model. Treatment is assigned at the group level and measured at the individual level, so standard errors are clustered by group. Due to the small number of groups, I rely on wild cluster bootstrapping (Cameron et al., 2008; Roodman et al., 2019; MacKinnon et al., 2023), implemented via the `fwildclusterboot` package (Fischer, 2021). All models control for relevant pre-treatment individual-level covariates, including age, gender, and party identification.

Across all types of engagement there is a small positive association between exposure to a contentious issue and the likelihood of reacting or commenting. While none of these are statistically significant, the direction of these relationships runs counter to my initial expectations. In retrospect, it might not be surprising that politically-engaged individuals, particularly in an election year, would be motivated to make their feelings known on highly-salient and divisive issues. Perhaps a larger, more representative sample of the U.S. population - encompassing people who are less engaged with politics - would see these coefficients tend even closer to zero. The negative relationship between group heterogeneity and engagement, for three out of four outcomes at least, is more in line with my initial hypothesis. Again, however, these are some distance from being significant.

Focusing on the 'reaction' outcome variables as a group, there appears to be no significant association between either treatment type and the likelihood of reacting to content. There also appears to be no discernible effect of an interaction between contentious issues and mixed groups upon the likelihood of reacting. Reactions (or at the very least 'likes') are an imperfect measure of participants' propensity to meaningfully engage in political discussion for at least three reasons. First, 'likes' are very low-effort, requiring only one click: it was notable that a relatively high proportion of participants would simply 'like' all posts without contributing to discussion. Second, 'likes' are a

Table 4.2: Treatment effects on types of participant engagement.

	Reaction	**Political reaction**	**Comment**	**Political comment**
Contentious issue	0.08 (0.12)	0.10 (0.13)	0.07 (0.13)	0.06 (0.07)
Mixed group	-0.18 (0.10)	-0.16 (0.14)	-0.02 (0.21)	0.06 (0.12)
*Contentious*mixed*	-0.16 (0.28)	-0.14 (0.35)	-0.42* (0.21)	-0.21* (0.13)
N	200	200	200	200

*Note: *p<0.05. This table shows estimates and group-clustered standard errors (in parentheses) from OLS regressions of four types of engagement on the two treatment dummies of interest, and their interactions. These models control for age, gender, and party identification. Bootstrap standard errors are clustered by group and reported in parentheses, with asterisks denoting statistical significance at a 95% confidence level. These bootstrap standard errors and p-values are based on Rademacher weights and 1000 repetitions.*

relatively banal and low-risk way of engaging: participants understand that others are unlikely to monitor and sanction each other for their reactions. Third, even if 'likes' *are* sanctioned, the likelihood of social isolation in a non-personal and time-limited network like this is very low. Comments, on the other hand, require much greater cognitive effort and risk of social isolation, and therefore represent a more accurate measure of willingness to engage in political discussion. This is where a more interesting picture emerges. Examining the interaction between contentious issues and group heterogeneity, I see the expected negative relationship: those in mixed groups discussing contentious politics are considerably *less* likely to comment. In other words, while contentious issues generally motivate people to discuss politics, when individuals are aware they are in a group containing people likely to hold different views, they hold back from expressing their opinion. Alternatively, group heterogeneity does not attenuate discussion until it is combined with a polarising issue. Group identity therefore becomes highly salient and influences behaviour. This appears to show that a spiral of silence only emerges when more than one of the theory's constituent components is present.

This provides some evidence in support of my third hypothesis. However, with reference to Brambor et al. (2006), I am cautious about interpreting these coefficients as unconditional marginal effects. Instead, based on the previous OLS regression models, Figure 4.5 shows the marginal predicted probabilities of commenting based on assignment to each treatment condition, with marginal effects at representative values of both 'contentious' and 'heterogeneous' (0 or 1) calculated through the `margins` R package (Leeper, 2017). This shows a clear interaction which reflects the results of my OLS models. People appear more likely to comment across both measures in groups exposed to contentious political issues. When these divisive issues are discussed in mixed political groups, the probability of commenting drops. This treatment condition, in accordance with hypothesis 3, is where I ex-

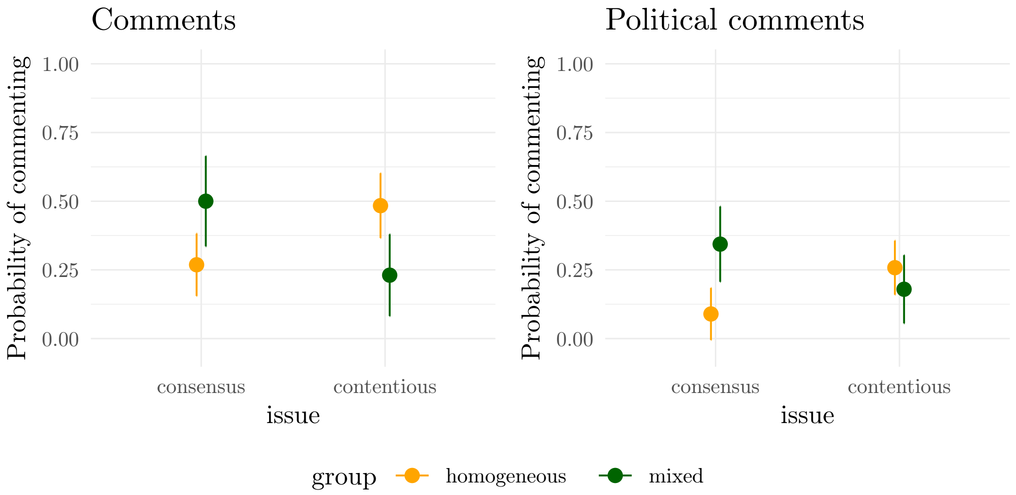


Figure 4.5: Marginal predicted probabilities of commenting (95% CIs)

pected a spiral of silence most likely to occur. While the pattern is clear, the statistical significance is not; non-political comments represent the only statistically-significant marginal effect. It is interesting that I find an interaction effect for both types of comments, and it could be that a spiral of silence 'chills' discussion outside of politics. In other words, those who disengage stay quiet about everything - not just politics - rather than risk social isolation. While the confidence intervals for political comments in homogeneous and mixed groups overlap, an obvious issue is that of power. In a larger study with a greater number of participants, one might expect the pattern to persist, and for statistical significance to emerge. Further, given that these people are politically-engaged individuals, I suggest this represents a conservative estimate of their likelihood of disengaging from online political discussion.

4.9 Conclusion

People in countries around the world are becoming increasingly polarised, undermining potential for healthy and productive democratic debate. At the same time, they are increasingly unwilling to discuss politics online with those from across the partisan divide. Given that so much of our lives are spent online, this is a toxic combination, and understanding the conditions that lead to self-censorship on social media might help us mitigate some of the most negative consequences of polarisation. My pre-registered lab-in-the-field experiment shows that, while exposure to contentious issues and membership of mixed groups alone can lead to greater engagement with political discussion, their combination leads people to *disengage*. This experiment builds on and furthers our understanding of online self-censorship in several ways. First, consistent with previous studies (Gearhart and Zhang, 2015; Kushin et al., 2019; Chan, 2021), this study presents evidence that self-censorship does indeed take place on Facebook. I take this a step further, however, and conduct tests in a setting which replicates common 'real-world' experiences on social media, ensuring ecological validity and generalisability. Second, my key treatments – exposure to contentious issues and partisan group composition – enable me to establish their individual and combined effects on political discussion. Third, as highlighted by Fox and Holt (2021), these sites offer a range of self-expression methods. By examining willingness to react, as well as comment, we now have a more rounded picture of what shapes interaction with political content. Fourth, and finally, most spiral of silence studies examine only one issue in one context. Incorporating four issues and showing the distribution of opinion among study participants helps eliminate any concern that self-censorship may be related to a single specific issue.

This study has some limitations. The most obvious shortcoming is statistical power; the costs associated with advertising, recruitment and compensation dictated that a maximum of 250 subjects could take part in the experiment. The execution of this study should at least provide proof of concept, while demonstrating the feasibility of future iterations which refine and adapt its experimental approach with a greater number of participants. Other limitations relate to the challenges of studying group discussion in an experimental setting. While the experiment was designed to replicate an everyday social media experience, some important caveats remain which might curb its generalisability. First, the groups created here do not include the closest social ties represented by real-world 'offline' friends and family. Therefore, it could be argued, that any fear of real-world social isolation is actually quite low in this specific setting. Second, due to resource constraints, the timeline of the study was necessarily quite short. As such, attempts were made to expedite the process of identifying one's opinion climate using the relatively blunt instrument of partisan balance within medium-sized groups. While this is a reliable approximation of values and beliefs in polarised two-party systems like the United States or United Kingdom, its applicability may be limited in more fragmented party systems. In terms of time constraints, one's mental image of their opinion climate will be painted over a much longer period than five days, and use many more data points. Whether this paints a more nuanced and detailed picture, or simply exacerbates divisions, is open to debate and presents an opportunity for future study. Finally, the participants recruited have higher levels of political interest than the average person - and even that of the average political partisan. While this means that the study is not perfectly representative of the wider population, it may in fact mean I found a conservative estimate of willingness to self-censor. The average person's likelihood of engaging in political discussion may be even lower. Nevertheless, this study still represents a methodological and theoretical advance on our previous knowledge of self-censorship in online settings.

What are the wider implications of my study? An apparent unwillingness to discuss contentious topics in mixed groups may well elicit concern. Participation and inclusion are generally good things for democracy and, at present, we are more likely to hear extreme voices online (Barberá and Rivero, 2015). We know that face-to-face discussion of political issues in mixed groups can help mitigate polarisation (Hobolt et al., 2024). However these results suggest that, in an online environment, people are more likely to switch off than make their voice heard. Whether this reticence stems from fear of social isolation, intransigence, or even apathy, the consequences for political debate are worrying. It is clear that social media represent a unique way of engaging with politics, and that technology companies are still reckoning with their responsibility. While this study tested specific treatments under quasi-realistic conditions, i.e. passive contact with political content, future studies might test whether specific and direct prompts might encourage a civic, conciliatory environment in which controversial issues can be explored in depth. Of course, more discussion in an online environment does not necessarily equal *better* discussion. Related research, for example, might examine algorithmic or architectural changes which highlight and accentuate similarity with other users, rather than conflict.

5 Conclusion

In 2023, Elon Musk proclaimed that "The reason I acquired Twitter is because it is important to the future of civilization to have a common digital town square, where a wide range of beliefs can be debated in a healthy manner" (Musk, 2023a). Whether a virtual town square is achievable, or even desirable, is of course much-debated (De Zuniga, 2015; Kruse et al., 2018). Habermas' concept of a public sphere refers to a space where individuals can come together to freely discuss and debate societal issues. In his eyes, it can only function effectively if citizens have unrestricted access to information and, without this, the ability of the public to form reasoned opinion and hold the powerful to account are compromised (Habermas, 1989). Access to factually-correct information is crucial for the health of deliberative democracy (Hochschild and Einstein, 2015) and, several decades ago, technology optimists believed that a global internet would facilitate such access. This, they argued, would lead to the democratisation of information, and even a form of 'cognitive surplus' (Shirky, 2010) formed out of the collective wisdom of newly-connected individuals. The prospect of a global town square is particularly appealing to technology entrepreneurs like Musk and Mark Zuckerberg, who often espouse hyper-libertarian ideology[1], with their products representing a 'marketplace of ideas' through which the strongest ideas naturally prevail.

[1] Musk, for example, calls himself a 'free-speech absolutist'

There are good reasons to believe that social media companies are not fully committed to these principles. First and foremost, any physical town square needs a set of rules, laws, and governing principles to facilitate the open discussion of ideas; as ex-Vice President of Twitter Bruce Daisley argues, "as much as X/Twitter loves framing itself as the 'global town square', such common spaces only thrive when everyone knows anti-social behaviour isn't going to be tolerated" (Daisley, 2024). Shortly after taking over Twitter, Musk fired roughly 50% of his workforce, which included large content moderation teams, reflecting his belief in minimal interference. Meta is set to follow suit, with Mark Zuckerberg recently announcing plans to eliminate third-party fact-checking within the company. On X, we know this move directly resulted in the widespread proliferation of misinformation (O'Carroll, 2023) and, despite advocating for free interaction between users, Musk has been selectively removing journalists from his platform who challenge his views or values (Timm, 2023). At best, the perception is that X has become a digital version of the 'Wild West', flooded with hate speech and misinformation, and does not represent a space for productive democratic debate (Faverio, 2023; McClain et al., 2024). At worst, X has become a 'pro-Trump echo chamber' (Ingram, 2024) reflecting the views of its owner who, in the summer of 2024, declared his full-throated endorsement of the Republican candidate. With its move to minimise interference and recommend more political content, it seems plausible that Facebook could follow a similar trajectory.

Misinformation is just the tip of the iceberg, however, and one of many challenges social media present for the health of democratic debate. At the outset of this thesis, I asked three questions: is elite political rhetoric on social media becoming more emotive, are extreme voices more prominent in online discussions, and why do people turn away from talking about contentious issues? The answers to all three have important implications for the effective functioning of an idealised ‘digital town square’. Sadly, the evidence presented here helps validate a perception that social media are currently not environments which facilitate productive democratic dialogue. First of all, excessively-emotive rhetoric is generally regarded as less desirable than rationality for political discourse (Arkes, 1993; Marcus, 2000); in paper 1, analysing a novel dataset of over 4 million tweets spanning more than 12 years, I find that MPs in the United Kingdom have used increasingly emotive language in their tweets. This is in line with the idea that politicians change their rhetorical styles in different situations (Rheault et al., 2016; Slapin et al., 2018; Crabtree et al., 2020), and that they become more emotional to maximise their appeal to larger audiences (Osnabrügge et al., 2021). Twitter grew exponentially during this time, becoming an increasingly influential political forum, and the easiest way to find a larger audience on Twitter during this time was through increased likes, shares, retweets, or quote tweets. I demonstrate a link between heightened emotion and levels of engagement; showing that increased emotion occurs through adaptation, where existing politicians change their rhetorical style, and through replacement, as newer MPs enter parliament, I argue that MPs have learned to become more effective Twitter communicators over time, maximising their appeal with the widest possible pool of potential voters.

Elites may be becoming more emotionally extreme in their language, but is political discussion on social media increasingly dominated by those towards the edges of the political spectrum? If these platforms are polarised - or even hostile - political environments, it stands to reason they leave little space for moderate or nuanced voices. To explore this question in greater depth, in paper 2 we asked whether people who talk about politics on Twitter between 2019 and 2021 were more ideologically and affectively-extreme than those who do not use Twitter to talk about politics. As with paper 1, we use new data, this time linking a nationally-representative survey of the UK population with behavioural data. We show that certain types of political Twitter users - primarily those on the ‘Remain’ side of the Brexit debate - are more politically extreme than their offline counterparts. We also show that those closer to the political periphery simply post more tweets, further amplifying the voices of more extreme people. These findings have important implications for understanding who we are most likely to hear in online political discussion, and for increasingly-polarised online political debate.

What about those people we don’t hear in political discussion, and why might they turn away? In contrast to the first two papers, and much political communication research, in my third paper I shift focus towards the people least likely to engage. In a pre-registered lab-in-the-field experiment on Facebook, I show that, in mixed partisan groups, people disengage from discussing contentious political issues. Building on existing research which examines self-censorship, I offer a new theoretical perspective which argues that group identity and issue type are particularly important when it comes to turning away from online political debate. Further, by conducting a real-world experiment which attempts to causally identify why people turn away from political discussion, this study represents a methodological advance on existing survey-based research on the topic. Understanding the key drivers of self-censorship may help mitigate some of the most negative consequences of political polarisation. In this final chapter, I highlight the main contributions and limitations of each paper, suggesting potential avenues for future research. I then broaden the discussion to encompass the wider implications of this research, including some of the real-world consequences of my findings.

Contributions

The first paper in this thesis makes a unique contribution to the literature on political communication by exploring the interaction between social media use and emotive elite rhetoric. While rhetorical adaptation has been examined extensively in an offline context, primarily with regard to parliamentary speech (Rheault et al., 2016; Slapin et al., 2018; Crabtree et al., 2020), we know much less about how new communication media have changed political rhetoric overall. My study provides a detailed empirical analysis of how language has evolved over the lifespan of politicians on Twitter. By examining a large dataset of over 3 million tweets, between 2007 and 2019, by UK Members of Parliament, I offer robust evidence of a trend toward increasingly emotional rhetoric over time. This finding enriches our understanding of how digital platforms, as a new communication medium, influence not just rhetorical content, but its tone, intensity, and extremity. Further, by demonstrating that tweets with higher emotional intensity receive more interactions, and that this likely incentivises further use of emotive language, I contribute to the literature on the economics of attention in digital media (Berger, 2011; Berger and Milkman, 2012; Brady et al., 2017; Weismueller et al., 2022). This not only adds to the consensus that digital platforms amplify certain types of content over others, and in the process potentially skew politicial discourse, but also advances our understanding of how politicians strategically adjust their rhetoric for maximum impact. Finally, by investigating whether social media platforms like Twitter might exacerbate ideological and affective extremity, this paper also intersects with the literature on political polarisation. Emotionally-charged language often appeals to base instincts, reinforcing existing beliefs, and hardening both ideological and affective divides (Iyengar and Westwood, 2015). By linking the emotional tone of political communication to engagement metrics, and by exploring its broader implications for polarisation, my work provides evidence supporting the argument that social media may intensify partisan divides. In connecting micro-level rhetorical shifts with macro-level phenomena like political polarisation, I offer a comprehensive view of one of the consequences of digital political communication.

The second paper in this thesis offers a slightly different perspective on the role of social media in political polarisation. First, acknowledging the limitations of exclusively using either surveys (e.g. Banda and Cluverius, 2018; Butters and Hare, 2022) or digital trace data (e.g. Barberá, 2015; Yarchi et al., 2021) to research the topic, we use nationally-representative survey data linked to Twitter accounts to examine the ideological and affective extremity of those who use Twitter to discuss politics. This methodological advance allows us to compare robust measures of polarisation with the real-world online behaviour of respondents. In doing so, we find that political Twitter users identifying on the 'Remain' side of the Brexit debate are more ideologically and affectively extreme than those who don't talk about politics on Twitter. Second, we also explore the link between extreme Twitter users and the type of tweets they produce. Remain supporters with stronger ideological convictions are more likely to tweet about politics, while Remainers and Leavers with strong affectively-negative feelings towards the 'other side' are more prolific political tweeters. Further, in finding the Brexit stance of respondents to be a stronger predictor of extremity than traditional party identity, we highlight asymmetries in how people on different sides of the political argument engage in political discussion. Our focus on Brexit as a political divide shows how social media platforms help articulate and reinforce emergent political identities, offering new insights into the unique dynamics of political discussion on Twitter.

By showing that individuals with stronger convictions are more likely to dominate the political conversation, these results underscore the increased potential for polarisation on social media. It represents further evidence that social media promote disproportionately amplify more extreme voices, reinforce homophily and selective exposure, and contribute towards a more polarised and less deliberative public sphere. Overall, this paper extends our understanding of polarisation in online political communication, and provides a foundation for future research into the question of whether social media platforms exacerbate, mitigate, or simply reflect political divides.

If social media sites are dominated by increasingly extreme discourse, this invites questions about what happens to more moderate people who choose not to enter the political conversation. The vast majority of political communication research focuses upon the type of observable online behaviour examined in the first two papers of this thesis; in contrast, paper 3 shifts this gaze towards those who *disengage*. In the process, it makes an innovative theoretical and methodological contribution to the political science literature by exploring the dynamics of online political engagement. In concordance with much of the research into the existence of spirals of silence on Facebook (Gearhart and Zhang, 2015; Kushin et al., 2019; Chan, 2021), I find that self-censorship does indeed occur on social media sites. This paper, however, takes such studies a step further in both theoretical and methodological terms.

Building on Elisabeth Noelle-Neumann's Spiral of Silence theory (1974), I refine and extend the concept to provide fresh insights into its relevance in the digital age. First, the structure of the social media site matters. On private (or at least semi-public) platforms like Facebook, online and offline networks more frequently overlap (Pang et al., 2016), resulting in a greater risk of social isolation. On public sites like X, where anonymous posting is possible, users are likely to encounter fewer potential social sanctions. Second, I argue that when the issue being discussed is controversial and divisive, an increased likelihood of social sanctions causes people to disengage. To my knowledge, no study examines the difference between contentious and consensus issues in examining online self-censorship. Third, I argue that group identity matters. An extensive literature shows that individuals tend to conform to group norms, sharing or censoring opinions to 'fit in' and avoid social penalties. Group membership influences perceptions and behaviour, and particularly political behaviour. In polarised environments, such as the contemporary United States, political identity becomes more salient, and becomes a heuristic for the opinion of others on a range of contentious topics, leading people to avoid discussing politics with those outside their partisan group. Noelle-Neumann theorised that people evaluate their immediate environment before offering a potentially unpopular view. Social media sites, through comments, reactions, and follower metrics, actually facilitate this 'quasi-statistical' sense (Noelle-Neumann, 1974) of the balance of opinion.

This paper also represents an empirical and methodological advance on existing studies, by conducting a pre-registered lab-in-the-field experiment in a setting which closely resembles an everyday real-world social media experience. In doing so, I offer a more comprehensive and ecologically-valid examination of the factors which influence online political disengagement, strengthening the argument that such behaviour is not merely a result of laboratory conditions or survey response bias, but a prevalent aspect of online political communication. Examining the individual and combined effects of exposure to contentious political issues and partisan group composition allows me to isolate these factors and provide valuable insights into how their interaction governs disengagement. The study also contributes to the literature by expanding the measurement of online political engagement beyond mere willingness to comment. By including reactions, such as likes or emotive responses, I offer a more rounded picture of how people interact with political content on social media. This broader approach aligns with the findings of Fox (2021), who emphasises the diverse range of self-expression methods available on these platforms. By capturing a wider range of user interactions, this approach provides a more comprehensive understanding of the factors which shape online political engagement. Finally, by incorporating four different political issues and analysing the distribution of opinions among participants, I help mitigate concerns that self-censorship could be an issue-specific phenomenon, and add robustness to my findings. This contribution is particularly valuable given the increasing importance of social media in shaping public discourse and the growing concern over political polarisation. Rather than fostering inclusive political discussions, social media platforms may be contributing to increasing polarisation by discouraging cross-cutting interactions and reinforcing echo chambers. This insight is crucial for understanding the potential consequences of digital communication on the health of democratic processes.

Limitations and directions for future research

Nevertheless, the papers presented here are not without their limitations. The methods used to find that politicians have become more emotional over time - 'bag-of-words' models such as VADER - are effective but have a number of constraints. Primarily, these models do not capture the context in which emotional words are used; they may detect the presence of emotional words in isolation but are less effective than, say, transformer models at distinguishing between varying levels of intensity. Incorporating a more sophisticated supervised machine learning approach would strengthen the paper by improving its accuracy and robustness. A deep learning model, such as BERT, GPT, or RoBERTa, would provide more nuance and context-aware sentiment analysis. Unlike traditional natural language processing, which analyses the words in a document sequentially, these models use a bi-directional approach to consider the entire context of a sentence and allow them to capture complex relationships between words. Similarly, employing Latent Dirichlet Allocation as a probabilistic topic model does not account for the context in which words appear. More sophisticated methods, such as dynamic topic modelling or topic modelling using BERT embedding, may offer richer and more accurate insight.

Expanding the scope of Paper 1 highlights potential opportunities for improvement. First of all, the study's reliance on the concept of emotive rhetoric as a primary lens for analysing political communication may oversimplify the complexity of political discourse. While emotional language is undoubtedly significant, it is far from the only factor influencing political communication. This paper's focus on emotional intensity might overlook other critical aspects of rhetoric, such as the use of humour and irony. Further, in framing my theoretical argument that emotional *pathos* contrasts directly with rational *logos*, I focus solely on conceptualising and measuring the former. Future research could aim to clarify, define, and measure the latter. These rhetorical strategies differ, of course, depending on the medium used; expanding scope to include traditional and other social media platforms would provide a more comprehensive understanding of how different online environments influence political communication. Platforms like Facebook and TikTok, with their distinct user bases and content algorithms, may reveal different patterns of rhetorical adaptation among politicians. A comparison of the language used by politicians in press releases, for example, may shed light on the effect of unmediated communication on rhetorical extremity. Finally, understanding how politicians in different countries adapt their communication strategies to their specific political and media environments could offer valuable cross-national insights.

Given the transnational nature of social media, an exclusive focus on one country is a limitation which extends to paper 2, which presents a snapshot of one particular country at one particular time. Our results show that, if anything, perceptions of Twitter around this time as a platform which leaned more towards left-wing, 'liberal', or progressive viewpoints (e.g. Wojcik and Hughes, 2019; Bacon, 2021) were justified. Since Elon Musk's takeover of Twitter (now *X*) in 2022, however, perceptions of its ideological balance have altered radically. Musk has reinstated previously banned right-wing accounts such as Tommy Robinson, Donald Trump, and Marjorie Taylor-Greene, and drastically scaled back what he perceived as left-leaning content moderation. In contrast to previous Twitter CEOs, Musk uses his account to not only tweet politically, but to frequently amplify right-leaning politicians and talking points to his 195 million followers. For example, in response to recent far-right riots on the streets of the United Kingdom, Musk echoed anti-immigrant rhetoric in proclaiming that "Civil war is inevitable" (Musk, 2024) as a consequence of mass migration and open borders. Musk has conducted high-profile (albeit technologically-challenged) interviews with Republican politicians such as Ron DeSantis and Donald Trump, and seemingly made algorithmic changes which disproportionately benefit engagement levels of far-right accounts (Barrie, 2023). This has certainly reassured those on the right: the proportion of Republicans believing that X is 'mostly good' for democracy has increased sharply since Musk's takeover (McClain et al., 2024). At the same time, these changes seem to have prompted many left-leaning accounts, including those of the United Kingdom's Labour Party, to either leave X or substantially scale back their activity on the site (Courea, 2024).

Altogether, Musk's behaviour since his takeover have significantly altered X's user base, the nature of its political discourse, and the dynamics of polarisation on the platform. Future studies will need to revisit these questions in the context of the new X landscape to assess whether the patterns of polarisation have persisted, shifted, or intensified under the platform's evolving political orientation. Additionally, the study's emphasis on the Brexit debate and UK political context may not translate directly to other countries with different political landscapes and issues. The nature of political discourse on Twitter in the UK, particularly surrounding Brexit, is unique and may not reflect how political polarisation manifests in other regions or on other issues. Similarly, our focus on Twitter may give us insight into political polarisation on one platform only, and these findings may not tell us very much about polarisation on social media in general. User demographics and behaviour on other platforms like Facebook, Instagram, or TikTok, vary significantly.

Paper 3, while offering a novel and ecologically-valid approach to studying the phenomenon of disengagement on social media, has three main limitations. First, due to resource constraints, the study is still significantly underpowered. Second, while the experiment comes very close to replicating an everyday Facebook experience, there are some important caveats concerning the ability of the study to accurately replicate the fear of real-world social isolation. The groups in this experiment were, to the best of my knowledge, completely comprised of strangers. This means that participants were not interacting in networks including some of their closest social ties. For participants, while the possibility of online consequences[2] remain, a reduced fear of potential real-world social isolation would theoretically make participants more likely to interact with political content. This does have an upside for my findings, however. A diminished fear of social isolation, coupled with study participants' high average level of political interest, suggests that any measure of disengagement is likely conservative, particularly when compared to the general population.

A central argument of this paper is that social media are far from uniform; the psycho-social mechanisms shaping behaviour vary significantly across dif-

[2] E.g. the aforementioned discussion of cyber-bullying or 'doxxing'.

ferent platforms. One way to tease out whether this fear of social isolation truly exists on a close-social-tie network like Facebook, even in a group full of strangers, might be a comparison with another social network. We already know that perceived anonymity affects the likelihood of engagement with political discussion on social media platforms (Wu and Atkin, 2018). Future studies might employ an almost identical research design on Reddit, for example. Discrete treatment groups could again be used to administer the twin treatments of issue type and group heterogeneity, with participants now afforded full anonymity and, by extension, a lower fear of social isolation. This would then allow a comparison between a public (or semi-public) site like Facebook with a fully-anonymised social media site.

Third, and finally, generalisability of the findings from paper 3 may be limited by the length of the study. Due to limited resources, each study period took place over five days only. To speed up the process of opinion climate monitoring, I used a relatively crude measure of opinion: each group's partisan composition of Democrats and Republicans. While the predicted balance of opinion on each issue was corroborated by my pre-study survey, this heuristic of the opinions held by others is robust only in ideologically-polarised two party systems like the United States or (at least to some extent) the United Kingdom. This may not be generalisable to less-polarised multiparty systems, where there is likely to be a much more nuanced overlap between political partisans and their ideological beliefs. Additionally, people's perceptions of their opinion climate usually form over a much longer time than the five days I had, and are formed from a far greater number and diversity of sources. This longer process could either foster more nuanced understanding or deepen divisions, and is something that future research could explore. With more resources to compensate participants for a longer commitment, perhaps mid-study surveys could be introduced to gauge perceptions of the balance of opinion within their assigned group. While this study tested specific treatments under quasi-realistic conditions, i.e. passive contact with political content, future studies might test whether specific and direct prompts might encourage a civic, conciliatory environment in which controversial issues can be explored in depth. Of course, more discussion in an online environment does not necessarily equal *better* discussion. Related research, for example, might examine algorithmic or architectural changes which highlight and accentuate similarity with other users, rather than conflict.

Wider implications

The findings presented here have provided insight into how social media are reshaping political communication, with profound implications for voters, policymakers, the media, technology companies, and society at large. Here, I discuss some normative implications for the health of democracy, and offer a evidence-based suggestions for navigating some of the challenges and opportunities posed by this rapidly-evolving digital landscape. It seems clear that social media do not represent Elon Musk's idealised 'common digital town square'. What, then, might this mean for the 'future of civilization'? The evidence presented in this thesis suggests that, left unchecked, some of their more toxic features will have direct implications for democratic deliberation.

First and foremost, all three papers presented here show that social media sites, in different ways, can exacerbate political polarisation. In paper 1, I demonstrated the increased prevalence of emotive rhetoric by political elites on Twitter. The prevailing perspective in politics is that a healthy democracy should strive for more reason and less emotion in the pursuit of better policy-making (Arkes, 1993; Marcus, 2000). Reason, it is argued, promotes objectivity and impartiality in fair and just decision-making (Rawls, 1971), allowing allows policy-makers to evaluate evidence, consider diverse perspectives, and use logic and facts rather than personal biases or transient feelings. While the changing rhetorical strategies I identify might be effective in helping politicians reach a broader audience, their increasingly emotive language can also contribute to the intensification of political divides. We already know that emotionally-charged traditional media can frame issues in ways which evoke strong, polarised reactions, simplifying complex topics into binary ‘us-versus-them’ narratives (Iyengar and Hahn, 2009), and it seems as if we are witnessing the emergence of similarly sensationalist and divisive political discourse online. We might worry that such a shift leads voters to become more susceptible to emotional manipulation, leading to less-informed and more reactionary voting behaviour, or even greater political apathy. Certainly, emotionally-charged language often appeals to base instincts, reinforcing existing beliefs, and hardening both ideological and affective divides (Mutz, 2007). A greater focus on emotion signals a shift away from reasoned debate and a focus upon rallying core supporters, which can alienate moderate voices, discourage nuanced discussions, and normalise extreme rhetoric. As opinion leaders, politicians are influential, and we might expect their behaviour to heavily influence the tone of democratic debate, pushing people towards increasingly extreme positions. Farrer (2024) argues persuasively that the public sphere is, like other common pool resources, subject to a ‘tragedy of the commons’, where competition for public attention results in a huge volume of increasingly-extreme political messages. Those overwhelmed by this morass simply switch off, while those that remain become increasingly polarised in their beliefs. Over time, polarisation presents real dangers for democracy by eroding social cohesion and trust (Kingzette et al., 2021), increasing out-group prejudice (Iyengar et al., 2019), and even leading to anger and violence (Kalmoe and Mason, 2022).

The findings presented here also present significant challenges for legislators in forming effective public policy. Politicians face the challenge of engaging and persuading constituents who are increasingly entrenched in their views and less open to compromise. At the same time, the growing emotional intensity and polarisation of online environments may significantly shape the topics and issues that dominate public discourse. For example, we know that elite perceptions of public opinion are frequently distorted (Broockman and Skovron, 2018; Pereira, 2021) and that social media platforms are often used to set the policy agenda (Gilardi et al., 2022). As I highlight in paper 1, increased emotion on social media is associated with greater engagement, and greater engagement results in greater visibility. In paper 2, my co-authors and I show that political discussion on Twitter is dominated by those closest to the edges of the political spectrum. With more extreme positions particularly loud and visible on Twitter, elite perceptions are likely to become skewed towards extreme positions on controversial issues, potentially sidelining critical but less sensational topics. Combined with increasingly extreme echo chambers, which limit exposure to diverse perspectives and reinforce existing beliefs, it becomes difficult or even impossible to collectively reach consensus on important issues.

This distortion of our political discourse might make it seem as though there is less room for moderate or nuanced voices, particularly on the type of contentious issues examined in paper 3. On Facebook, I find that many voters choose to avoid political debates altogether, especially when dealing with contentious issues and faced with opposing views. Based on the findings of my first two papers, this might seem to be an increasingly common situation. This has real-world consequences; a less-informed electorate can lower the quality of democratic participation and, if voters self-censor to avoid conflict, these contentious and important issues may not receive the attention and debate they require for effective policy-making. Ultimately, this results in a lack of true substantive representation and a disconnect between the government and its citizens. Apathy can lead to a lack of trust in government institutions, and even the rise of anti-democratic values or authoritarianism as people lose faith in democracy's ability to serve their needs.

In polarised societies, political identity has become increasingly salient, and an ever-growing number of political issues become contentious. Does this mean that people will turn away from political discussion entirely? The findings of these three papers highlight a feedback loop in which social media amplifies extreme voices, leading to more polarised political discourse. This, in turn, discourages nuanced or balanced engagement; as discourse becomes more emotionally charged and polarised, moderates may feel disillusioned with, or unrepresented by, the political process. Consequently, active participants in political discussions may increasingly come from the extremes of the political spectrum. This tends to create profound problems for democratic institutions, for example in the achievement of consensus on important issues, legislative gridlock, and ineffective governance. Silencing of moderate voices and amplification of extremes may lead to feeling of being unrepresented by the system, and consequent lack of public trust in democratic institutions, apathy, erosion of democratic legitimacy, social fragmentation, and even violence (Kalmoe and Mason, 2022).

Industry responses

How might we mitigate some of these harmful effects? The social media companies themselves represent an obvious place to start. As the 21$^{\text{st}}$ century progresses, social media *could* play the role of the idealised 'town square', where people gather for the free exchange of ideas and information. As discussed earlier in this thesis, social media platforms maximise user engagement by promoting emotional and sensational content. This design choice has significant implications for the quality of political discourse, and social media companies have a responsibility to reconsider platform algorithms which prioritise engagement over accuracy and quality. By promoting more balanced content, they can help reduce polarisation and improve the overall quality of online discussions.

While algorithmic changes designed to reduce the extent of online 'echo chambers' (by increasing exposure to cross-cutting sources) don't actually appear to reduce polarisation (Nyhan et al., 2023), Meta have nevertheless in recent years deprioritised their focus on increasing users' time spent on Facebook in favour of 'time well spent'. In other words, this means time spent interacting with friends, rather than the media or businesses, in the hope users will have more 'meaningful' and healthy interactions, and has been presented as a shift away from the 'digital town square' to a more private form of social networking in a 'digital living room' (Zuckerberg, 2021). However, recent events have led Meta to signal a reverse to this approach and return to the recommendation of politics on its platforms, including Facebook, Instagram, and Threads (Booth, 2025). Combined with the removal of most of its fact-checking teams, the future of constructive dialogue seems uncertain. This is widely seen as a politically-expedient rather than a well-intentioned decision; despite its undeniable size and power, the suspicion is that Meta still requires Donald Trump's political help to stop the European Union from sanctioning American 'big tech' over antitrust laws (Hernández-Morales, 2025).

As the gatekeepers of online discourse, technology companies have an ethical responsibility to ensure that their platforms do not contribute to societal harm, and the findings of this thesis emphasise a need for greater corporate accountability in addressing the social impacts of their services. The leading social media firms seem either unwilling or incapable of addressing some of the most profoundly harmful issues inherent in their platforms, despite the significant real-world social consequences. This intransigence perhaps stems from a few places; first, they are beholden to shareholders, rather than to the general public or to government. Perhaps the free market will help propel them towards a solution. The perception is that social media remain hostile spaces for the discussion of politics; certainly, this thesis has found that extreme voices are likely to dominate the online conversation. The conditions which are poor for political debate are also bad for business; in the case of *X*, advertising revenue has plummeted since Elon Musk's arrival. A major source of X's current challenges appears to be Elon Musk himself and, rather than facilitating the free exchange of ideas, his decisions often have the opposite effect. Ultimately, a 'global town square' must be governed by a plurality of people, not by decree of an autocratic billionaire.

Policy responses

An idealised version of the public sphere should be free from both governmental and market influences, allowing for the formation of public opinion through rational-critical debate (Habermas, 1989). If social media companies are either unable or unwilling to implement meaningful changes, though, government legislation may become necessary to protect the integrity of public discourse. In recent years, many have questioned whether governments are actually able to enact meaningful laws which constrain the actions of technology companies, given the extent of their economic, social, and political influence. Meta has a market capitalisation larger than the GDP of the Netherlands, Sweden, and Switzerland (Wallach, 2021), making the prospect of fines largely trivial. Tech companies operate globally, often surpassing the reach of any single government, making it difficult for individual nations to regulate them effectively. In 2018, for example, Mark Zuckerberg refused to testify before a House of Commons select committee and, in 2023, X dropped out of the European Union's agreement to combat online misinformation. There is, however, hope for those advocating for state-led solutions. The alignment of the "tech industrial complex" (Holland and Singh, 2025) behind Donald Trump reflects their reliance on his support to "push back on governments around the world" (Zuckerberg, 2025). The recent US Congressional ban on TikTok demonstrates that governments can take decisive action to limit the influence of technology firms. As David Allen Green observes, despite the immense size and influence of social media firms, "In any ultimate battle, the state will prevail over a corporation ... those who control the law can, if they choose, regulate and tame any corporation within their jurisdiction" (Green, 2025). Ultimately, while technology companies wield considerable power, the authority of governments to legislate and enforce remains a potent force, provided there is the political will to act.

For policy-makers, the regulation of social media represents a tricky political and legal challenge. While there is an increasing appetite for some kind of governmental oversight (Fung, 2023; Anderson, 2024), a majority of people in most countries still believe that social media are good for democracy (Gubala and Austin, 2024). Social media, after all, helped foster democratic transition during the Arab Spring protests, with subsequent autocratic regimes jailing dissidents for expressing anti-government sentiment online (Gouvy, 2021). Further, media professionals have raised concerns that clamping down on social media could be damaging for their ability to hold the powerful to account. For over a decade, Twitter was the go-to platform for breaking news, but Elon Musk has actually made things progressively harder for journalists, from throttling traffic to specific outlets (Merrill and Harwell, 2023), removing headlines to 'improve the esthetics' (Musk, 2023b), or adjusting X's 'For You' algorithm to promote the content of users who pay $8 a month to be verified - rather than content which organically achieved high engagement (Kirshner, 2023). Notably, access to Twitter has historically only been blocked in authoritarian regimes such as Iran, North Korea, Russia, and Venezuela. In a test of the options available to democracies, and their potential consequences, in 2024 the Brazilian government ordered internet service providers to suspend access to X, after the company repeatedly refused requests to ban pro-Bolsonaro 'digital militias' spreading hate speech and disinformation. In such a polarised environment, the move inflamed significant socio-political tensions between those broadly on the 'right', in defence of free speech, and those on the 'left', determined to legislate against perceived societal harm. In considering an ethical and regulatory framework which addresses these competing concerns, policy-makers must walk this delicate tightrope. At the same time, any new legislative strategy must preserve the positive aspects of social media, such as encouraging political engagement and bridging political divides.

Australia's recent legislation banning teenagers from using social media seems like a similarly-blunt instrument as the Brazilian *X* ban. Could policy-makers instead legislate to enforce tech companies to make their products less addictive and less extreme? The implementation of the Stop Addictive Feeds Exploitation (SAFE) for Kids Act by the New York State Senate may give us some clues. The new law "prohibits the provision of addictive feeds to minors" and, from 2025, social media platforms will have to seek parental consent before children under 18 use apps with 'addictive feeds', in an attempt to regulate algorithmic recommendations. Addictive feeds, in this context, are defined as "platforms and services that recommend content based on information from the user's activity or device" (New York State Senate, 2024). While the law in its current iteration would be extremely difficult to extend to adults, if successful, it might provide a pathway towards a law which requires that social media companies design their apps to be less addictive. In a political sense, this might mean the de-prioritisation of emotionally or ideologically-extreme content. A balanced regulatory approach to social media must acknowledge its dual role as both a tool for democratic empowerment and a source of societal harm. In collaboration with governments, technology companies, and civil society, policy-makers can craft solutions that mitigate the risks of polarisation, and disinformation - while preserving the potential to encourage political engagement and bridge divides. Ultimately, the challenge lies not in silencing the digital public sphere but in ensuring it serves democracy.

A | Appendix: Paper 1

A.1 Data-gathering process

Details of the accounts (including username and Twitter ID) of current serving MPs were exported from *politics-social.com*, with those of historical MPs researched and checked manually in November 2021. A total of 957 members of parliament served from the date of the first tweet sent by an MP until the date of the 2019 General Election; of those, 780 (or 81.5%) used a verifiable Twitter account at some point in the period of interest. Data from these accounts was obtained using `academictwitteR` (Barrie and Ho 2021), an R package which queries Twitter's Academic Research Product Track, allowing access to Twitter's full-archive search and v2 API endpoints. `AcademictwitteR` gives the option of returning timelines based either on Twitter username or ID number. The latter is unique and immutable, whereas usernames can change frequently, particularly in the world of politics as elected representatives lose their seats or change roles. Therefore, ID number was the preferred option in obtaining an accurate historical representation of an individual's tweets; ID numbers of 780 MPs were checked with Twitter's API. Finally, the timelines of these accounts from 15th May 2007 (the date of the first tweet sent by an MP) to 12th December 2019 (the date of the General Elections in that year) were gathered.

Data on current serving MPs was exported from *They Work For You*, along with historical records of MPs elected at the 2005, 2010, 2015, and 2017 General Elections respectively. Tweet data was returned as series of json files, which were subsequently bound and converted to an R dataframe using `academictwitter`'s *bind_tweets* function. This resulted in a raw dataset of 8,632,561 tweets with 31 distinct variables, including name, text of the tweet, date of the tweet, and various engagement metrics such as 'like' count, retweet count, and 'quote' count. These metadata variables were appended with 11 new columns including full name, party, constituency and engagement: a sum of all engagement metrics for each individual tweet. A small number of empty rows were removed. Finally, using Twitter IDs as a common identifier, data was then merged with a number of MP characteristics, coded either manually or via They Work For You. Gender was coded as an indicator variable, indicating 1 if female and 0 if male, alongside age at the time of research (January 2022). With regard to deceased members of parliament, their age at death was entered. Start and end dates (date of first election and date of losing seat, if applicable) were converted to new binary variables cohort 2010, cohort 2015 and cohort 2017 depending on date of election. Data was trimmed to include tweets from accounts only during the period in which they served as MPs (between start and end dates). As this analysis focuses on the rhetoric of MPs, and therefore on the text of tweets created by their accounts only, it was decided that retweets, duplicates, and tweets containing only one word, should be removed. Further, non alphanumeric characters (such as @ and #), and URLs were removed. After this operation, 4,011,139 tweets remained, which were then parsed into appropriate units of analysis for natural language processing.

A.2 Topic modelling

In common with most applications of unsupervised topic modelling approaches like LDA, a prior number of topics, k, must be specified. As highlighted by many researchers (e.g. Roberts et al., 2016), there can be no 'ground truth' or definitive number of k; much depends on research objectives, context, and researcher expertise. Therefore I employ an iterative process, combining human evaluation and computational goodness-of-fit models, to arrive at an appropriate number of topics. First, I use `quanteda`'s *LDA* function to apply Gibbs sampling, iterating over the entirety of my text corpus from $k = 30$ to $k = 200$, in increments of 10 to minimise computation time. Second, I qualitatively evaluate each iteration of k by examining 20 words of highest probability of association with each topic to uncover broad themes. At $k = 50$ and above, discrete topics such as Brexit ('deal', 'remain', 'EU', 'referendum'), the NHS, health and social care ('NHS', 'care', 'hospital*', 'mental'), education ('children', 'student', 'colleg*', 'univers*') and Scottish independence ('independen*', 'Westminster', '#indyref') begin to emerge. Third, I apply the `ldatuning` package (Murzintcev and Chaney 2020) to LDA models between $k = 50$ and $k = 250$ in iterations of 25 to obtain two separate metrics; Griffiths2004 (Griffiths and Steyvers, 2004) and Caojuan2009 (Cao et al., 2009). The Griffiths metric maximises likelihood, while CaoJuan minimises divergence between topics, similar to a measure of perplexity. Applying these scoring algorithms over a wide range of k on such a large dataset is extremely (and prohibitively) computationally expensive. Instead, I run these metrics on a 10% random sample (500,000 tweets), the results of which are plotted in figure A.1. This places the ideal number of k somewhere between 125 and 150, providing a range which invites closer manual validation. Examining the distribution of likely key words across each value of k, I settled on the ideal number of topics as 135. Based on both human and computational validation, this offers an accurate representation of the topic in the tweet corpus. I extracted the most likely topic for each tweet, associating each one with a label, before appending these to the existing dataset. Finally, 500 tweets were selected at random and checked manually to ensure correlation of tweet text to topics. A full topic list is included below.

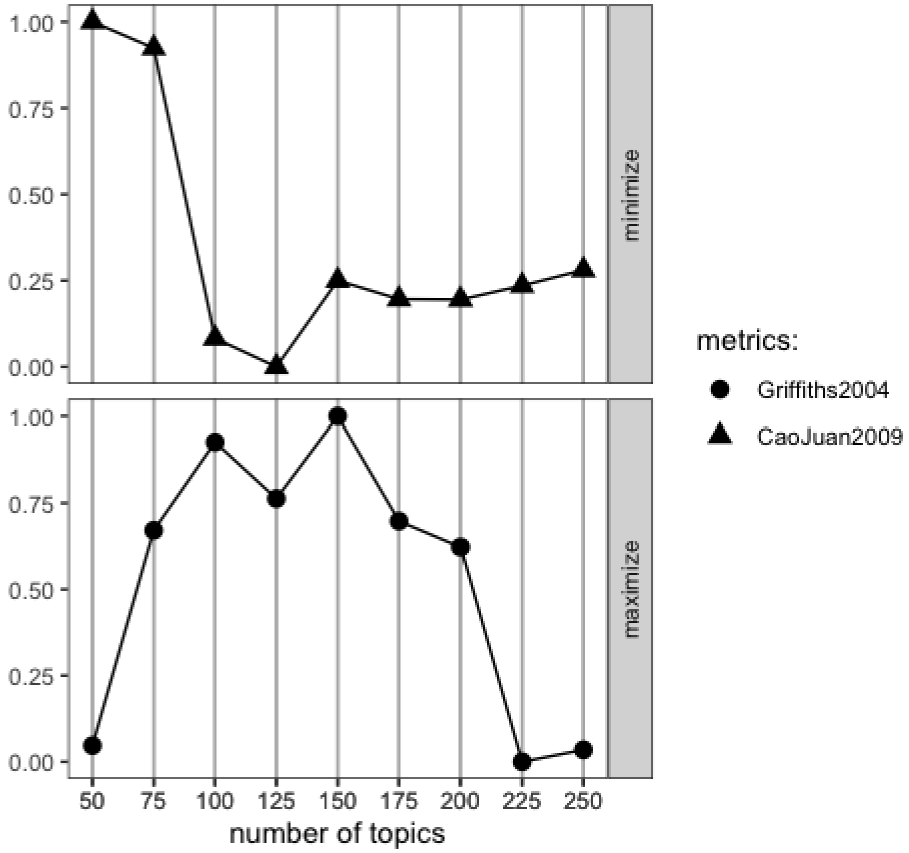


Figure A.1: A comparison of quantitative metrics for ideal number of k in LDA

Topic list: LDA 135			
No.	Label	No.	Label
1	Economic statistics	31	Calls to action 1
2	Weather	32	Extent (‘far’, ’enough’ etc.)
3	Pride and thanks	33	Verbs
4	Human rights and racism	34	Road transport
5	Public speaking	35	URLs
6	Help	36	Action (‘role’, ’take’, ’part’)
7	Small business	37	Media comment pieces
8	Meetings and discussion 1	38	Positive replies
9	War, foreign affairs, Syria	39	Thanks
10	Money and banking	40	Social media generic
11	Events	41	Replies to/from Lib Dems
12	Titles	42	Prime Minister’s Questions
13	Time 1	43	Thanks 2
14	Interesting ideas	44	The NHS and Health 2
15	Local communities	45	Replies and quote tweets
16	Police and crime	46	Animal rights
17	Polite debate	47	Time 2
18	Disagreement	48	Well wishes
19	Visits	49	Calls to action 2
20	Meetings and discussion 2	50	Abbreviations 1
21	Twitter	51	Replies and quote tweets 2
22	Labour Twitter mentions	52	Questions 1
23	The NHS and Health 1	53	Constituency issues
24	Contact with constituents	54	Abbreviations 2
25	Meetings and discussion 3	55	Constituency surgeries
26	Gender equality	56	Parliamentary business
27	Environmental protection	57	Croydon
28	Scotland	58	Time 3
29	Congratulations 1	59	Theresa May
30	Conjunctives	60	Feelings (‘like’, ’think’)

Topic list: LDA 135 (cont.)

No.	Label	No.	Label
61	Lincolnshire and the NE	91	Brexit - The single market
62	Brexit and London	92	The media 1
63	Election campaigns	93	Congratulations
64	Local government	94	House of Lords
65	Superlatives	95	Jeremy Corbyn
66	People and unity	96	Sport
67	Isolated letters / numbers	97	Campaign launches
68	Law and justice	98	Telecoms and tech
69	Agreement	99	Brexit - the referendum
70	Contact with constituents	100	Grimsby and Brighton
71	Non-alphanumeric chars	101	Colours and patterns
72	Time 4 (‘tonight’, ’last’)	102	Christmas
73	Time 5 (‘long’, ‘wait’)	103	Public sector pay
74	Calls for govt. action	104	Visits and meetings
75	Time 6 (‘look’, ‘forward’)	105	Wales
76	Grief and condolences	106	The media 2
77	Disability	107	Informal language
78	Agreement	108	Criticism of govt. policy
79	Calls to action 3	109	Election campaigning
80	Con election campaign ‘19	110	Division (‘torn’, ‘apart’)
81	Gambling and betting	111	Compass points/geography
82	Calls for govt. action 2	112	Defence and armed forces
83	Music	113	Listing words (‘plus’, ‘inc*’)
84	Requests for information	114	Truth and misinformation
85	Adjectives	115	Future national security
86	Ministerial questions	116	Investment and funding
87	Environment and energy	117	Social security
88	Numbers	118	Welcoming good news
89	Emojis and emoticons	119	Good luck wishes
90	Children and families	120	Coffee, tea, and cake

Topic list: LDA 135 (cont.)	
No.	Label
121	Housing
122	Reading (‘articl*’, ‘book’)
123	Brexit - deal with the EU
124	Lists
125	Labour election campaigns
126	Ashfield, Angela Eagle
127	Transport - rail and bus
128	Select committees
129	Questions 2
130	Calls to action 4
131	Feelings (‘seem’, ‘think’)
132	Definitive statements
133	The media 3
134	Employment and jobs
135	Education

A.3 Additional descriptive statistics

Table A.1: Top 10 most prolific tweeters

	Name	Party	Total no. of tweets
1	Tim Farron	Liberal Democrat	68,451
2	Denis MacShane	Labour	64,631
3	Stella Creasy	Labour	59,991
4	Robert Halfon	Conservative	54,424
5	Jamie Reed	Labour	53,783
6	Jess Phillips	Labour	44,458
7	Naomi Long	Alliance	43,279
8	Tom Watson	Labour	40,868
9	George Freeman	Conservative	40,298
10	Angus MacNeil	SNP	39,880

Table A.2: Most positive MPs on Twitter 2007-2019, by mean *pos* score

Name	Party	Mean positivity score
Norman Lamb	Liberal Democrat	0.327
Suella Braverman	Conservative	0.320
Gavin Shuker	Labour	0.318
Chris White	Conservative	0.305
Jonathan Lord	Conservative	0.294
Penny Mordaunt	Conservative	0.283
Wendy Morton	Conservative	0.281
Alberto Costa	Conservative	0.277
Tim Farron	Liberal Democrat	0.274
David Rutley	Conservative	0.264

Table A.3: Most negative MPs on Twitter 2007-2019, by mean *neg* score

Name	Party	Mean negativity score
Gavin Shuker	Labour	0.154
John Spellar	Labour	0.114
Yvette Cooper	Labour	0.105
Graham Jones	Labour	0.103
Roger Godsiff	Labour	0.101
Dame Joan Ruddock	Labour	0.100
Jack Dromey	Labour	0.100
Kate Osamor	Labour	0.098
Michael McCann	Labour	0.098
Malcolm Wicks	Labour	0.097

Table A.4: Most emotive parties on Twitter 2007-2019, by mean compound score

Party	Mean compound	Mean pos	Mean neg
Ulster Unionist	0.364	0.215	0.041
Green	0.299	0.233	0.070
Conservative	0.316	0.175	0.045
Liberal Democrat	0.269	0.202	0.046
Democratic Unionist	0.254	0.165	0.054
Sinn Féin	0.252	0.160	0.042
SDLP	0.231	0.150	0.056
SNP	0.206	0.155	0.055
Labour	0.196	0.162	0.062
Plaid Cymru	0.125	0.088	0.037
Alliance	0.104	0.128	0.073

B | Appendix: Paper 2

B.1 Who uses Twitter to discuss politics?

Table B.1: How many people use Twitter? Weighted percentages

Not on Twitter	71.9%	*Never use*	62.2%
		less than weekly	9.7%
On Twitter	28.1%	*Refused to share*	15.5%
		Identifiable account	8.7%
		Active	5.3%
		Not active	3.4%
		Active but not political	3.9%
		Active and political	1.5%

Note: these figures are percentages of a representative sample of the British adult population.

Table B.2: Twitter use by partisan identity - Weighted percentages

	All	**Con**	**Lab**	**LD**	**SNP**	**Other**	**None/DK**
Not on Twitter	71.9%	26.9%	14.3%	5.8%	2.1%	5.9%	17.0%
On Twitter	28.1%	7.4%	8.9%	2.7%	1.1%	2.8%	5.1%
Identifiable accounts:	9.0%	2.3%	3.2%	0.8%	0.4%	1.0%	1.2%
1. *Active*	5.5%	1.3%	2.0%	0.5%	0.3%	0.7%	0.7%
2. *Not active*	3.5%	1.0%	1.2%	0.4%	0.1%	0.4%	0.4%
3. *Active but not political*	3.0%	0.8%	1.0%	0.3%	0.2%	0.2%	0.5%
4. *Active and political*	2.5%	0.5%	1.0%	0.2%	0.1%	0.4%	0.2%

Note: these figures are percentages of a representative sample of the British adult population.

Table B.3: Twitter use by partisan identity: Whole sample, unweighted

Party	**Total**	**Not on Twitter**	**On Twitter**
Con	1358	897 *(66.1%)*	461 *(33.9%)*
Lab	1032	476 *(46.1%)*	556 *(53.9%)*
LD	360	193 *(53.6%)*	167 *(46.4%)*
SNP	138	69 *(50%)*	69 *(50%)*
Green	241	125 *(51.9%)*	116 *(48.1%)*
PC	38	10 *(26.3%)*	28 *(73.7%)*
Other	100	61 *(61%)*	39 *(39%)*
None	656	430 *(65.5%)*	226 *(34.5%)*
DK	226	136 *(60.2%)*	90 *(39.8%)*
Total	4149	2397 *(57.8%)*	1752 *(42.2%)*

B.2 Ideological polarisation measures

Respondents were asked to rate agreement on a scale of 1 to 5 (1 indicating 'strongly agree' and 5 'strongly disagree') with a randomised selection of the following statements in each of our three ideological dimensions. Note that scales were averaged and ordered so that all high scores indicate more left-wing, more socially-liberal, and more pro-EU values. This choice was taken for ease of interpretation. Higher scores on these attitudes tend to be those shared by Labour and Remain partisans, with lower scores more typical of Conservative and Leave supporters.

1. Left-right attitudes

- *Government should redistribute income from the better off to those who are less well off*
- *Government should redistribute income from the better off to those who are less well off*
- *Big business benefits owners at the expense of workers*
- *Ordinary working people do not get their fair share of the nation's wealth*
- *There is one law for the rich and one for the poor*
- *Management will always try to get the better of employees if it gets the chance*
- *Strong trade unions protect employees' working conditions and wages*
- *Major public services and industries ought to be in state ownership*

2. **Liberal-conservative attitudes**

- *Young people today don't have enough respect for traditional British values*
- *For some crimes, the death penalty is the most appropriate sentence*
- *Schools should teach children to obey authority*
- *Censorship of films and magazines is necessary to uphold moral standards*
- *People who break the law should be given stiffer sentences*
- *The amount of immigration to Britain should be decreased*
- *Gay couples should not be allowed to get married*

3. **European Union attitudes**

- *European courts should be able to make decisions about human rights cases in Britain*
- *Some laws are better made at the European level*
- *The British Parliament should not be able to override all EU laws*
- *Britain should hold another referendum on re-joining the EU*
- *Britain loses out by not being a member of the EU*

B.3 Are political Twitter users more ideologically extreme?

Figure B.1 plots the average ideological attitudes of our respondents, with 95% confidence intervals, across three dimensions (left-right, liberal-conservative, and EU attitudes) grouped by partisan identity (Conservative, Labour, Leave, and Remain) and segmented by 1) those who don't use Twitter ('Non- Twitter users') and 2) those who tweeted about politics at least once during the study period ('Political Twitter users') in each partisan grouping. The average of responses in the wider population is shown as a vertical dashed red line. As expected, we see a pronounced divergence in attitudes between Conservative/Leave supporters and Labour/Remain supporters on the liberal-conservative and EU dimensions, with a slightly greater similarity between these partisan groups on the left-right dimension. Perhaps unsurprisingly, the greatest disparity in viewpoints is found on the EU dimension, and between Leavers and Remainers.

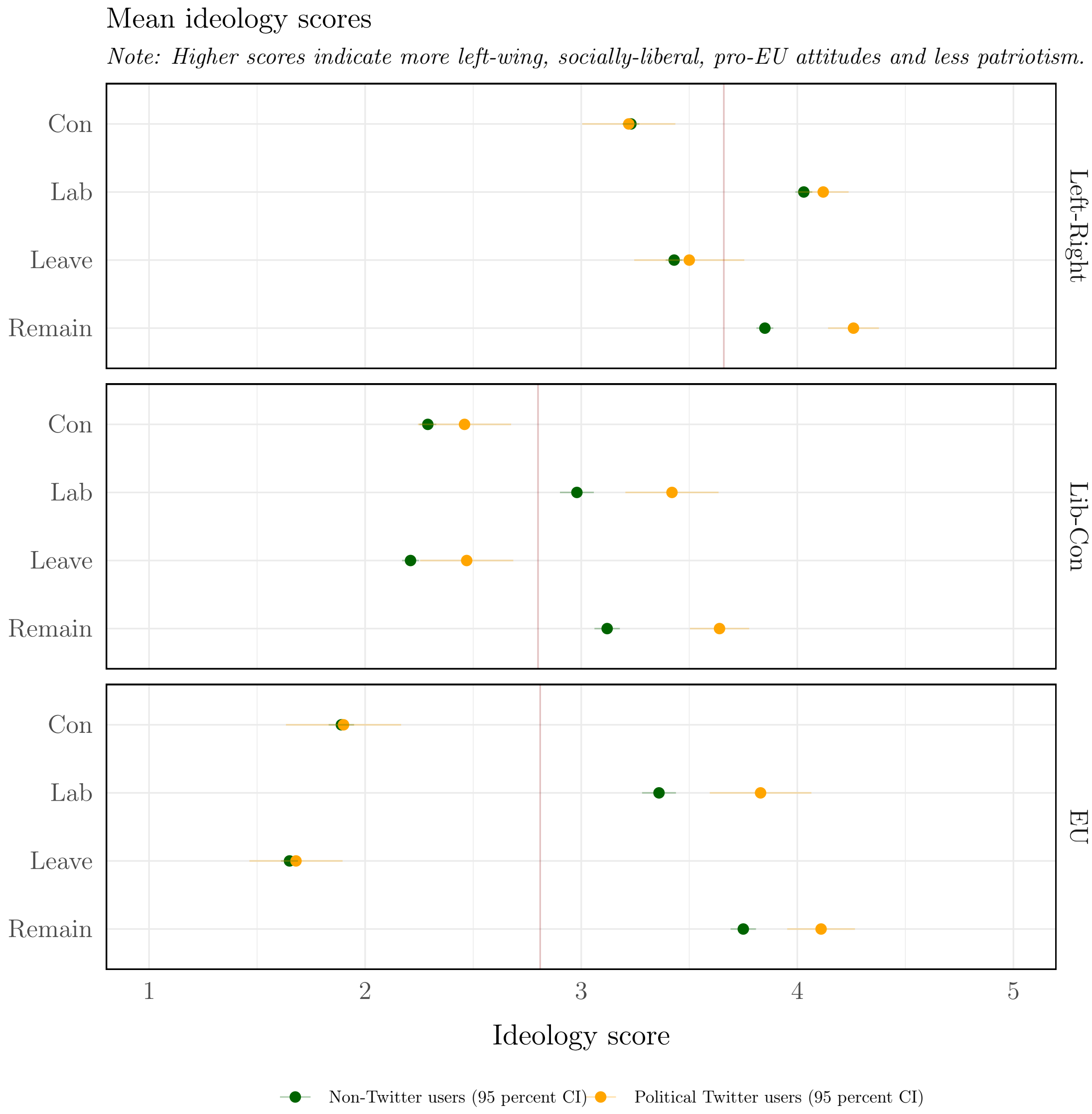


Figure B.1: Twitter use and ideological attitudes.

Table B.4: Are Twitter users more ideologically extreme than non-Twitter users?

	Conservatives			**Labour**		
	L-R	Lib-Con	EU	L-R	Lib-Con	EU
Twitter users	0.03	-0.05	-0.03	0.09*	0.13*	0.11*
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)
Observations	1358	1358	1358	1032	1032	1032
	Leavers			**Remainers**		
	L-R	Lib-Con	EU	L-R	Lib-Con	EU
Twitter users	0.01	-0.02	-0.03	0.05*	0.11*	0.14*
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Observations	1386	1386	1386	1733	1733	1733

*Note: * = p<0.05. Models ideological extremity of Twitter users vs. non-Twitter users. Dependent variable is ideological distance from population mean: higher scores = greater distance.*

Table B.5: Are political Twitter users more ideologically extreme than non-political Twitter users?

	Conservatives			**Labour**		
	L-R	Lib-Con	EU	L-R	Lib-Con	EU
Political Twitter users	0.02	-0.01	-0.08	0.11*	0.06	0.10
	(0.08)	(0.07)	(0.09)	(0.05)	(0.07)	(0.07)
Observations	461	461	461	556	556	556

	Leavers			**Remainers**		
	L-R	Lib-Con	EU	L-R	Lib-Con	EU
Political Twitter users	0.03	-0.03	-0.08	0.13*	0.17*	0.20*
	(0.07)	(0.07)	(0.09)	(0.04)	(0.05)	(0.06)
Observations	484	484	484	903	903	903

*Note: * = $p<0.05$. Models ideological extremity of users who talk about politics on Twitter vs. all Twitter users, with Heckman corrections applied. Dependent variable is ideological distance from population mean: higher scores = greater distance.*

B.4 Affective polarisation measures

As a measure of affective attitudes, we take two different versions of a standard 'thermometer' score (as used by Gidron et al., 2020; Reiljan, 2020; Wagner, 2021). First, the difference between two 0-100 ratings, indicating feelings of favourability/unfavourability towards voters on either side of either the Conservative/Labour or Leave/Remain divide. Higher scores equal greater 'warmth' or favourability, so a greater difference between the scores given to in-group and out-group equals greater affective polarisation. The question wording is as follows for parties:

- "*We'd like you to rate how you feel towards the Conservative party and the Labour party on a scale from 1-100, which we call a 'feeling thermometer'. Ratings between 0 and 49 mean that you feel unfavourable and cold. Ratings between 51 and 100 mean that you feel favourable and warm. A rating of 50 means that you have no feelings one way or the other. How would you rate your feelings towards the X?*"

We also ask thermometer scores for the two partisan groups, rather than parties, and the two Brexit groups. In principle, the difference between the two scores for these measures thus runs from -100 to +100, although, in practice, since extremely few people rate the out-group as better than their in-group, it runs from 0-100 with 100 as the maximum level of affective polarisation.

B.5 Are political Twitter users more affectively extreme?

Table B.6: OLS models: Twitter use against affective extremity

	Con	**Lab**	**Leave**	**Remain**
Political Twitter users	-4.25	6.15	4.04	6.76*
	(4.57)	(3.60)	(5.26)	(2.95)
Observations	1358	1032	1386	1733

*Note: * = p<0.05. Dependent variable is difference in thermometer score given to in-group and out-group: higher scores = greater affective extremity. Coefficients compare thermometer difference scores of political Twitter users to rest of population, broken down by partisan identification, with Heckman corrections applied. Standard errors in parentheses.*

B.6 Is ideological extremity related with Twitter activity?

Table B.7: OLS models: Ideological extremity against tweet frequency

	Con			Lab		
	L-R	Lib-Con	EU	L-R	Lib-Con	EU
Political tweets	-1.35	0.22	1.57	1.96	-0.10	-0.08
	(1.08)	(1.37)	(1.06)	(1.19)	(0.79)	(0.72)
Observations	153	153	153	215	215	215
	Leavers			**Remainers**		
	L-R	Lib-Con	EU	L-R	Lib-Con	EU
Political tweets	-0.99	-0.69	0.76	2.16	2.08*	2.11*
	(1.11)	(1.23)	(0.99)	(1.15)	(0.83)	(0.74)
Observations	159	159	159	339	339	339

*Note: * = p<0.05. Models the relationship between ideological extremity and the number of political tweets posted within each partisan group, controlling for age and education. Heckman corrections applied and standard errors in parentheses.*

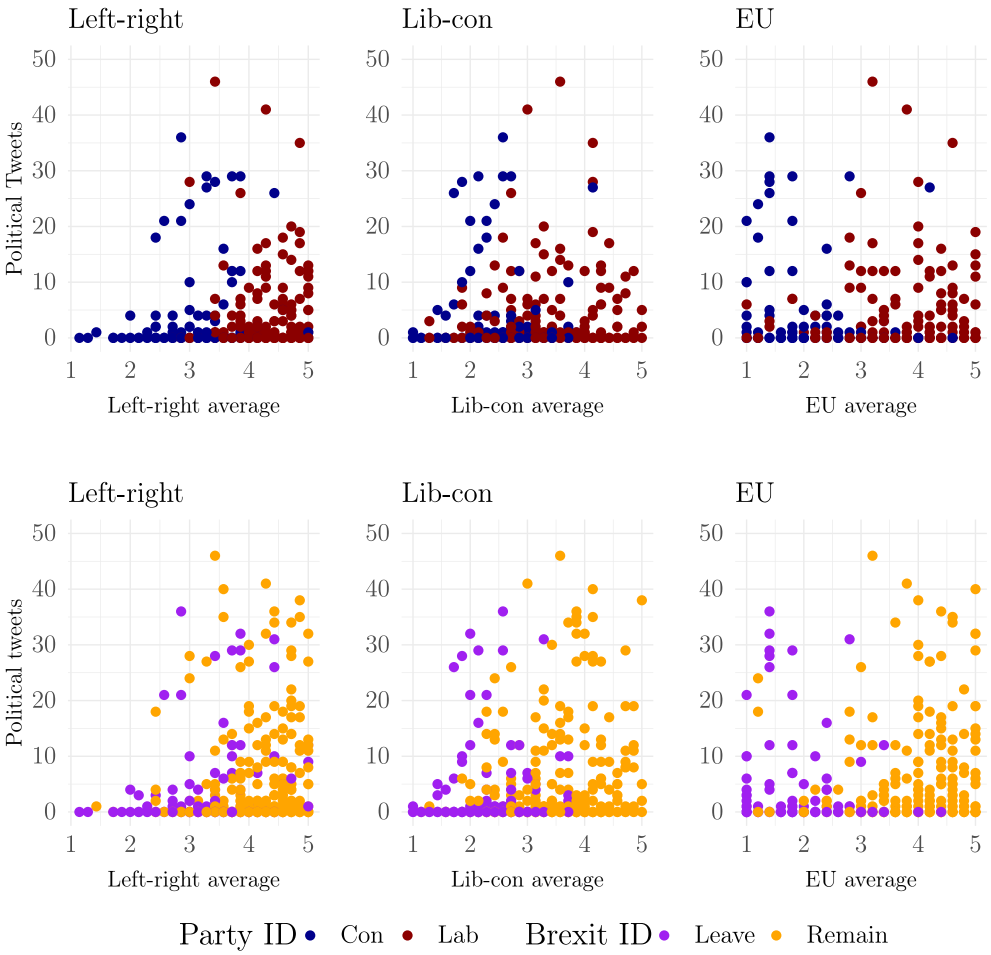


Figure B.2: Ideological values vs. political tweets shared

Note: Plots the ideological positions of Twitter users, across three separate dimensions, against the number of political tweets shared.

B.7 Response bias

As discussed, a significant proportion of respondents reported that they used Twitter, but were either unwilling or unable to share their account details with us. Consequently, it was essential that we identify any systematic differences between 1) non-Twitter users, 2) self-reported Twitter users, and 3) those who shared their account details with us for a study of their Twitter behaviour. The composition of these groups is shown in Table B.8; for party ID, Brexit ID, and Education, the count of respondents in each is given with the proportion of respondents in reach in brackets. To ensure generalisability, and the validity of our results, we had particular interest in identifying potential imbalance between groups 2 and 3. The age and education profiles of each look similar, however, on closer examination, response bias was clear in both gender and political identity. Men, Labour partisans, and Remain supporters were more likely to share their account details and were therefore over-represented in the study selection. To account for this inherent bias and ensure robust results, we employed Heckman corrections (Heckman, 1979) to all OLS models. Heckman, or two-stage models, first estimate the probability of each respondent being selected into the sample based on a specific set of variables. Party ID, Brexit ID, and gender were selected due to the above response bias and, to aid robustness, we also use gross personal income. This is employed as an instrument which helps predict stage 1 (propensity to share information - in this case self-reported Twitter use) but is unrelated to stage 2 (having a verified Twitter account). 876 users in the sample chose not to share their income; a simple binary variable was created indicating 1 if respondents shared their income and 0 otherwise. Subsequently, we calculate Inverse Mills Ratios and incorporate these probabilities into outcome OLS models to help correct for selection bias, obtaining unbiased estimates of the coefficients in the outcome equation. All OLS models specified in the Results section have Heckman corrections applied.

Table B.8: Balance table

	Not on Twitter	**On Twitter: no account shared**	**On Twitter: account shared**
Mean age (SD)	53.6 (17.0)	45.7 (16.5)	45.7 (15.5)
Gender (%)			
Male	1024 (42.7%)	422 (43.6%)	309 (51.6%)
Female	1373 (57.3%)	545 (56.4%)	290 (48.4%)
Party ID (%)			
Conservative	835 (34.8%)	232 (24.0%)	138 (23.0%)
Labour	432 (18.0%)	242 (25.0%)	199 (33.2%)
Brexit ID (%)			
Leave	717 (29.9%)	213 (22.0%)	123 (20.5%)
Remain	633 (26.4%)	366 (37.8%)	275 (45.9%)
Education (%)			
None	175 (7.3%)	37 (3.8%)	17 (2.8%)
Other	103 (4.3%)	26 (2.7%)	16 (2.7%)
GCSE or equiv.	442 (18.4%)	126 (13.0%)	79 (13.2%)
A Level or equiv.	500 (20.9%)	222 (23.0%)	132 (22.0%)
Higher below degree	195 (8.1%)	54 (5.6%)	30 (5.0%)
Degree	638 (26.6%)	380 (39.3%)	251 (41.9%)
Don't know	108 (4.5%)	45 (4.7%)	15 (2.5%)
Other technical	236 (9.8%)	77 (8.0%)	59 (9.8%)

C | Appendix: Paper 3

C.1 Average issue positions of participants

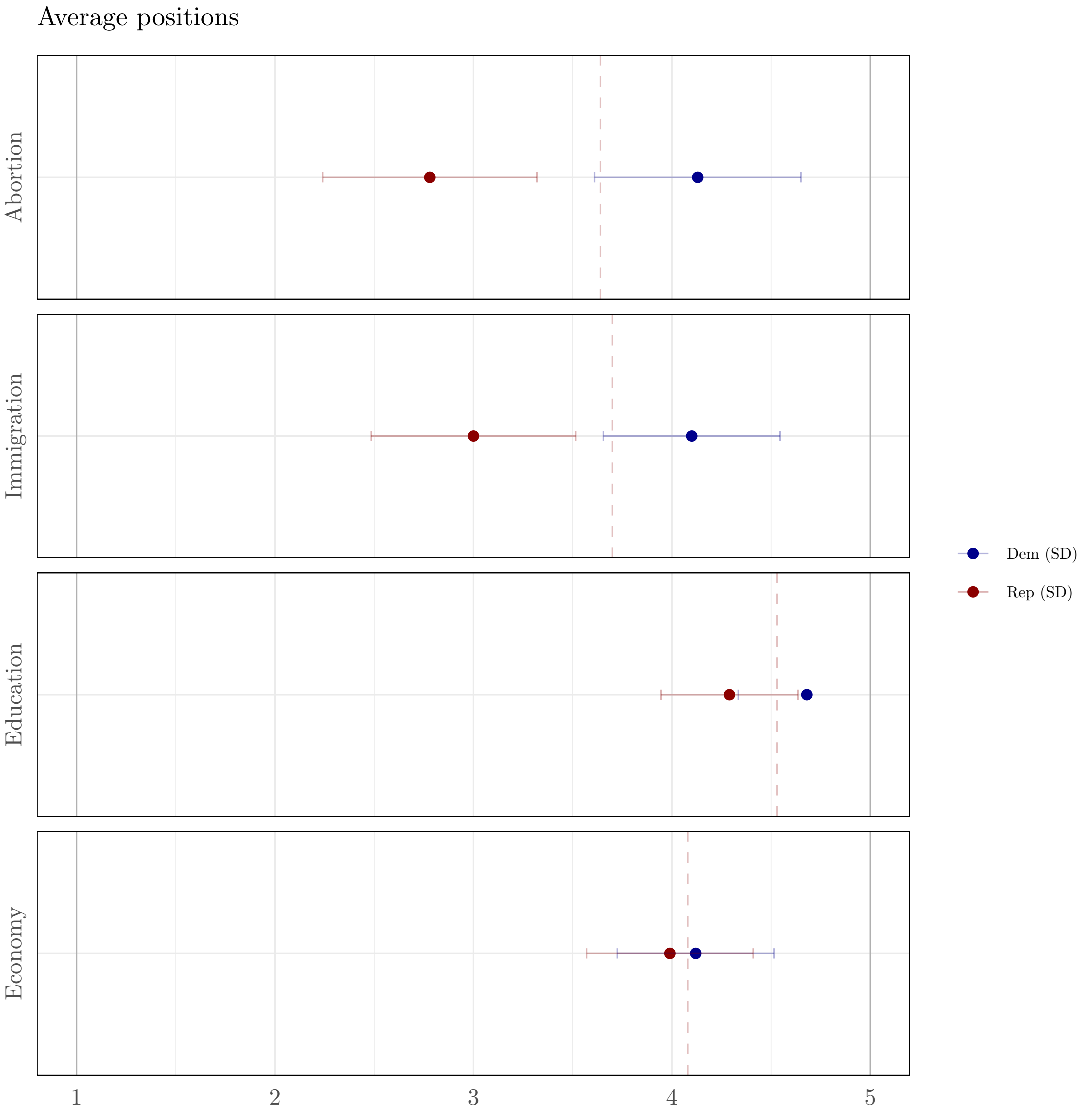


Figure C.1: Average issue positions from survey responses.

Note 5 = pro-abortion, pro-immigration, pro-educational opportunities for all, and pro-strong economy.

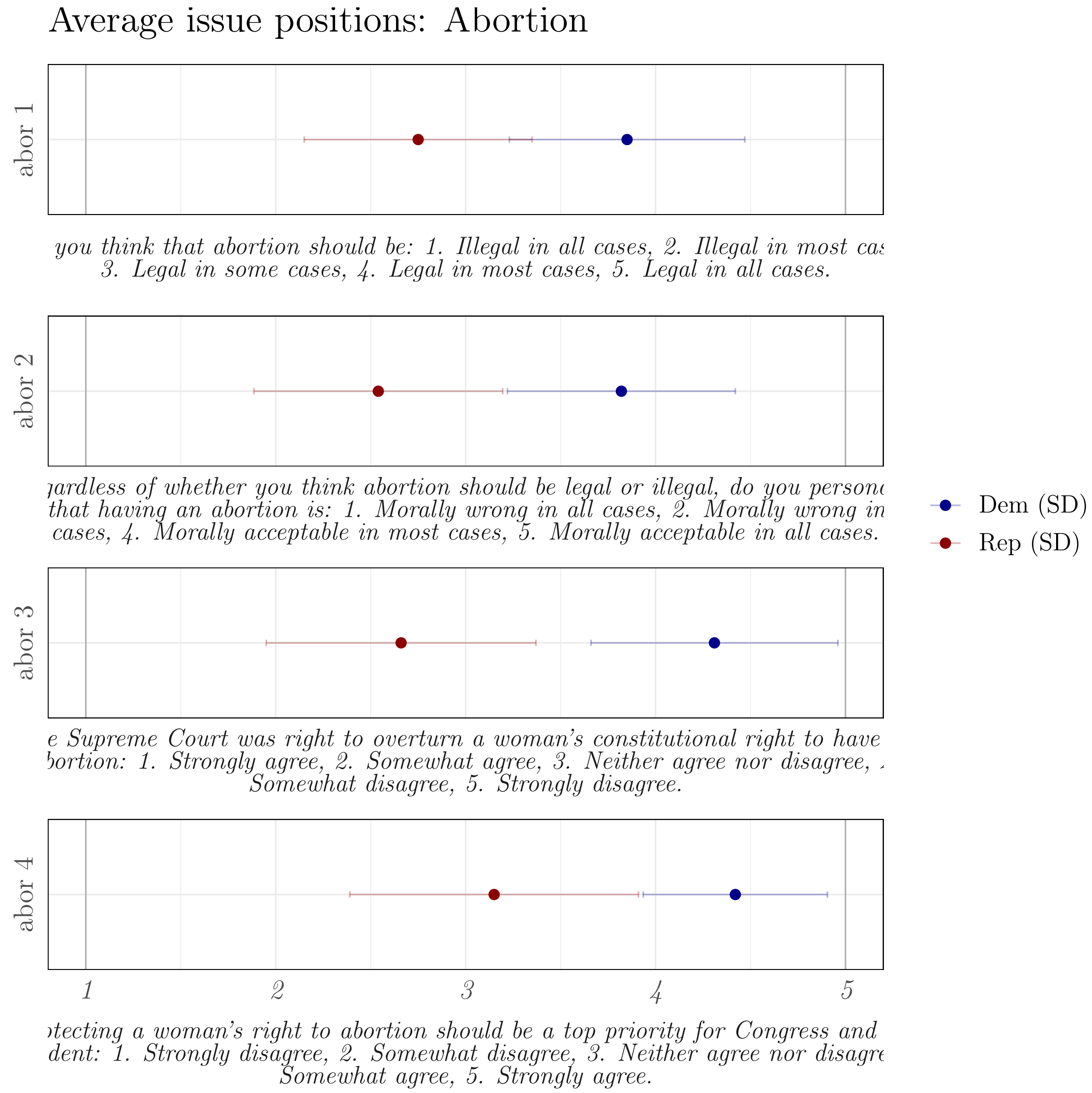


Figure C.2: Average issue positions on each abortion question, by partisan identity.

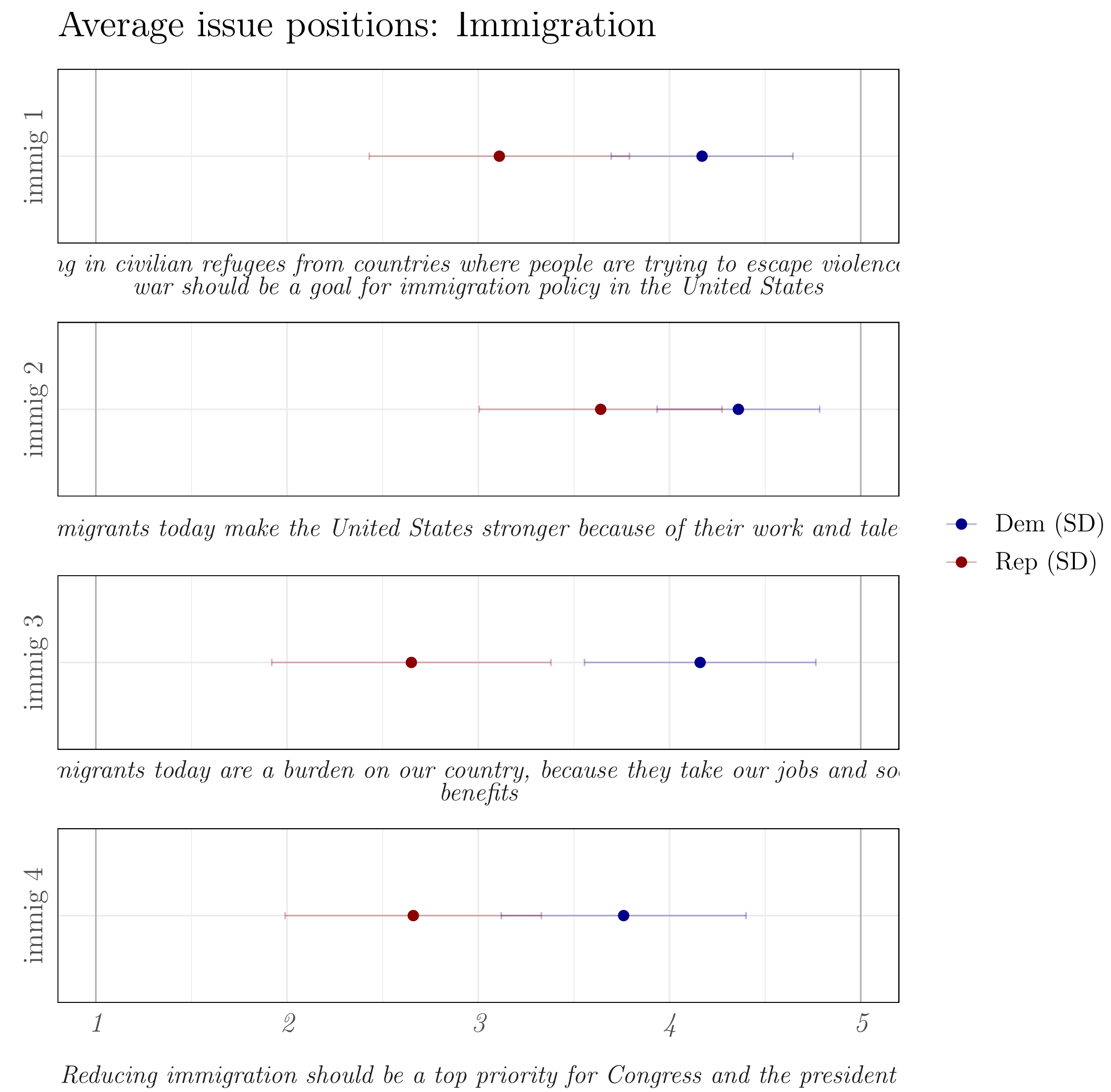


Figure C.3: Average issue positions on each immigration question, by partisan identity.

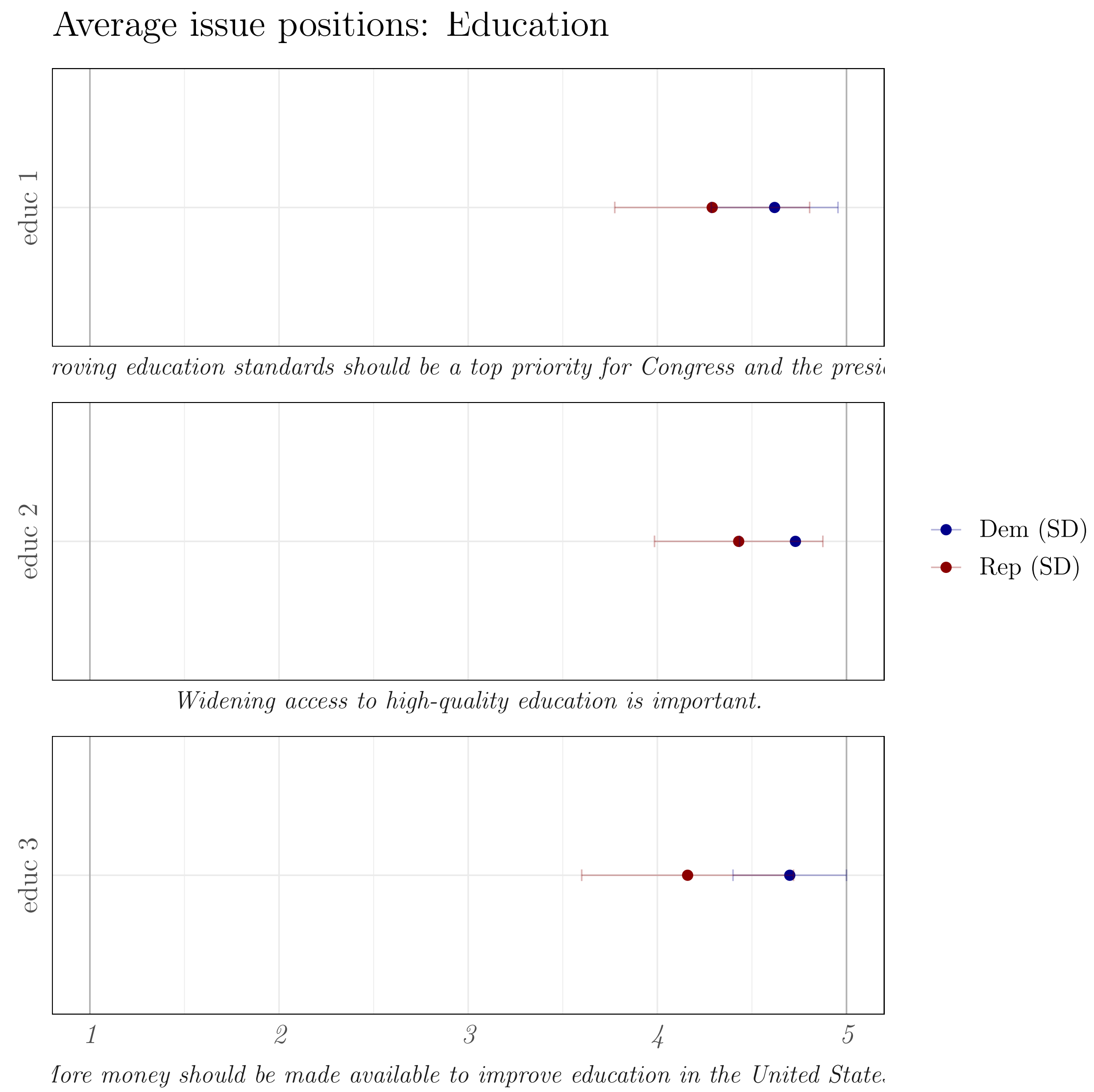


Figure C.4: Average issue positions on each education question, by partisan identity.

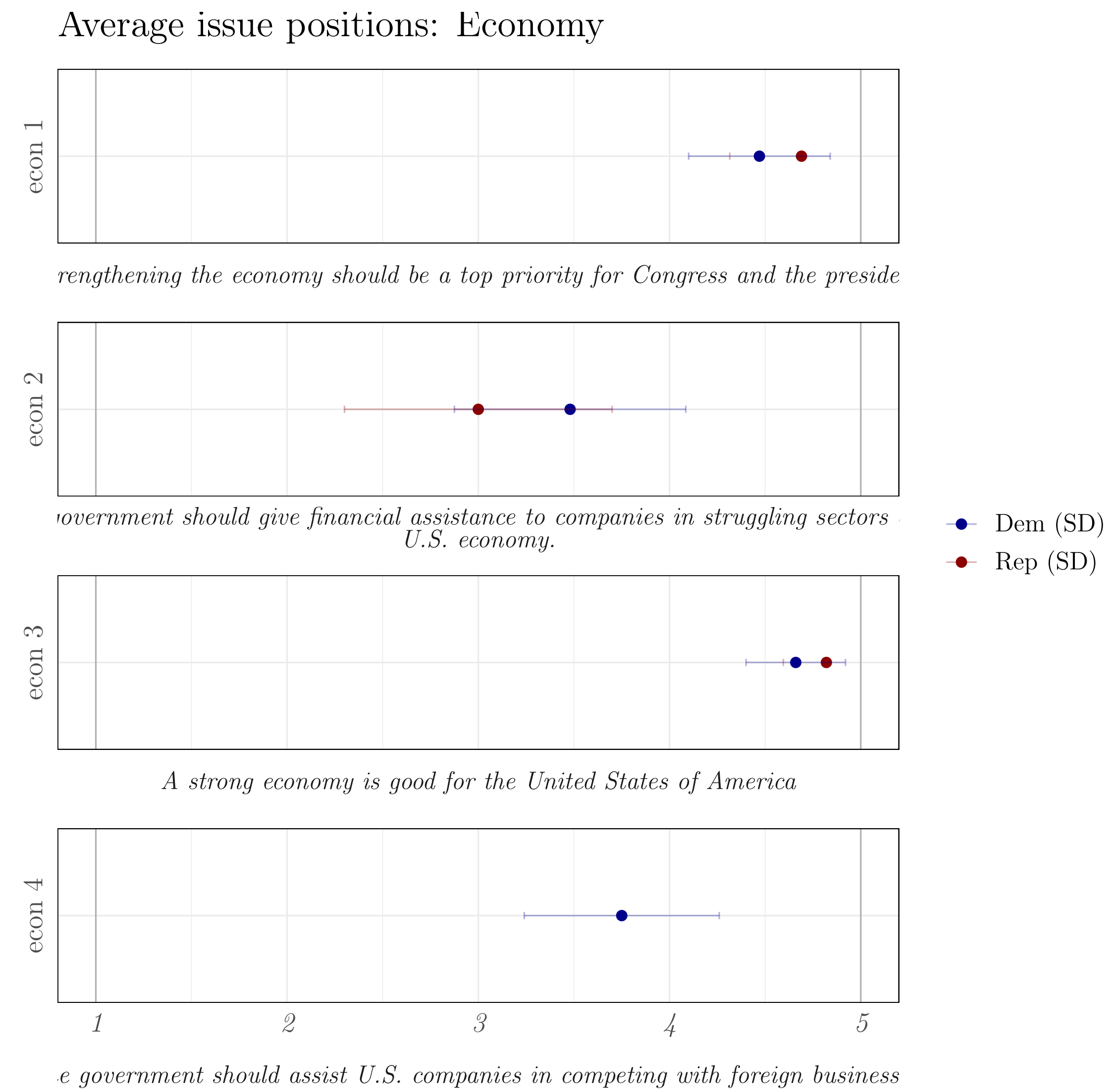


Figure C.5: Average issue positions on each economy question, by partisan identity.

C.2 Main effects

Variable	**Consensus issue**		**Contentious issue**		**Issue means**		**Issue effect**
	homog	mixed	homog	mixed	consensus	contentious	difference
Reaction	0.51	0.44	0.66	0.44	0.48	0.55	**0.07**
Comment	0.27	0.50	0.48	0.23	0.39	0.36	**-0.03**
Pol. reaction	0.40	0.31	0.55	0.36	0.36	0.46	**0.10**
Pol. comment	0.09	0.34	0.26	0.18	0.22	0.22	**0.00**

Table C.1: Main effects: Issue type

Note: Table shows proportion of participants engaged in each study condition and main effects of issue type. Dependent variables are binary, indicating whether participants engaged or commented.

Variable	**Consensus issue**		**Contentious issue**		**Group means**		**Group effect**
	homog	mixed	homog	mixed	homog	mixed	difference
Reaction	0.51	0.44	0.66	0.44	0.59	0.44	**-0.15**
Comment	0.27	0.50	0.48	0.23	0.38	0.37	**-0.01**
Pol. reaction	0.40	0.31	0.55	0.36	0.48	0.34	**-0.14**
Pol. comment	0.09	0.34	0.26	0.18	0.18	0.26	**0.08**

Table C.2: Main effects: Group composition

Note: Table shows proportion of participants engaged in each study condition and main effects of group composition. Dependent variables are binary, indicating whether participants engaged or commented.

C.3 Robustness checks

To ensure the main findings of this study are consistent across different model specifications, I conduct two further robustness checks alongside the application of cluster bootstrapping. First, in Table C.3, I re-run my OLS models using a raw count of engagement and comments as a dependent variable. This is intended to validate the direction of the relationships between exposure to contentious issues, mixed groups, and the likelihood of engagement - and the treatment interaction. As with the results of my main (binary) models shown in Table 2, these count models show a positive association between each treatment in turn, and a negative association when treatments are combined.

Second, I use a hurdle model to account for the large number of zeros in my dependent variable. Hurdle models are particularly suitable for data with an excess number of zero observations, as in my data, which might not be adequately captured by the OLS model. The results shown in C.3 confirm that my key findings are broadly robust to different model specifications. The direction of the relationships observed in the main OLS models persist, and the coefficient for 'comments' remains negative and statistically-signifiance. This indicates that the relationship observed in the OLS model is not driven by the presence of excess zeros in the data.

	Reactions	**Political reactions**	**Comments**	**Political comments**
Contentious issue	0.21* (0.06)	0.44 (0.29)	0.16 (0.39)	-0.38 (0.18)
Mixed group	0.79* (0.06)	-0.08 (0.04)	0.71* (0.04)	0.21* (0.07)
*Contentious*mixed*	-0.35 (0.28)	-1.13* (0.06)	-1.44* (0.24)	-0.69* (0.05)
N	200	200	200	200

Table C.3: Treatment effects on types of participant engagement: count models

*Note: *p<0.05. This table shows estimates and group-clustered standard errors (in parentheses) from OLS regressions of the raw counts of four different types of engagement on the two treatment dummies of interest, and their interactions. These models control for age, gender, and party identification. Bootstrap standard errors are clustered by group and reported in parentheses, with asterisks denoting statistical significance at a 95% confidence level. These bootstrap standard errors and p-values are based on Rademacher weights and 1000 repetitions.*

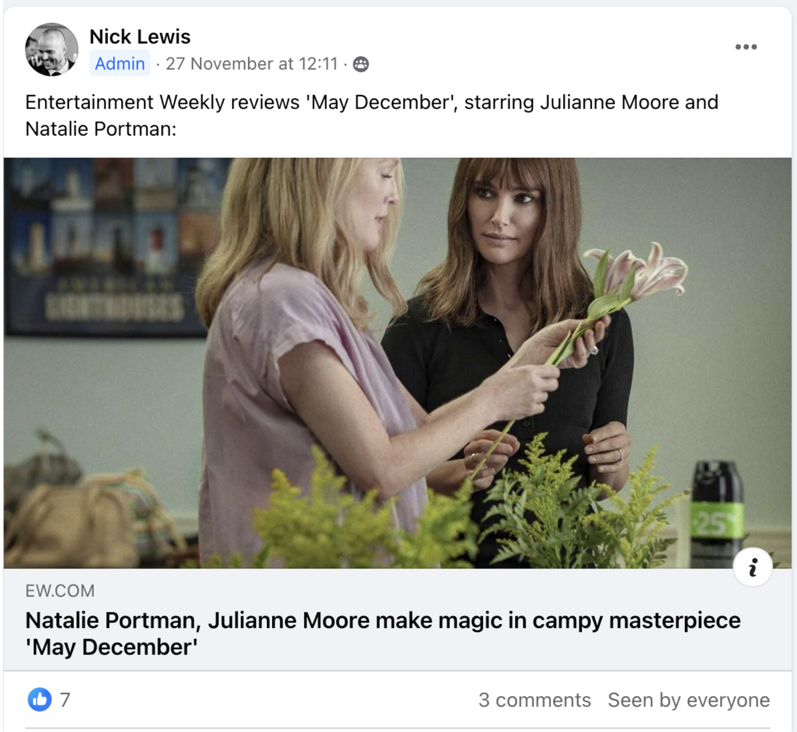
	Reactions	**Political reactions**	**Comments**	**Political comments**
Contentious issue	0.36	0.41	0.26	0.56
	(0.29)	(0.29)	(0.31)	(0.39)
Mixed group	-0.56	-0.55	-0.12	0.43
	(0.30)	(0.31)	(0.32)	(0.40)
*Contentious*mixed*	-0.68	-0.50	-1.80*	-1.49
	(0.62)	(0.64)	(0.66)	(0.82)
N	200	200	200	200

Table C.4: Treatment effects on types of participant engagement: hurdle models

*Note: *p<0.05. This table shows estimates and group-clustered standard errors (in parentheses) from hurdle models of four different types of (dummy) engagement variables on the two treatment dummies of interest, and their interactions. These models control for age, gender, and party identification.*

C.4 Treatment examples

C.4.1 Non-political content



Nick Lewis

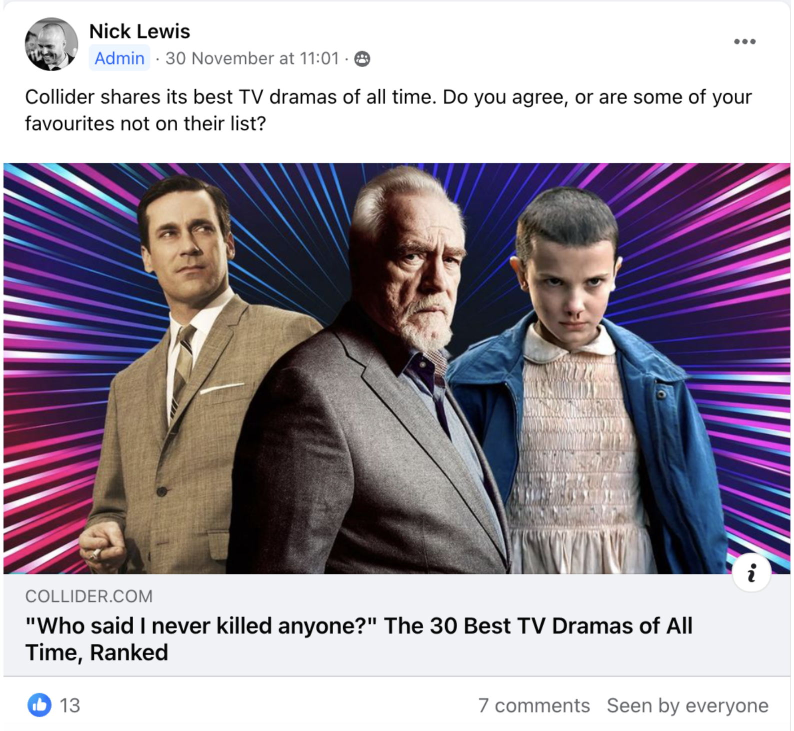
Admin · 27 November at 12:11 ·

Entertainment Weekly reviews 'May December', starring Julianne Moore and Natalie Portman:

EW.COM

Natalie Portman, Julianne Moore make magic in campy masterpiece 'May December'

👍 7 3 comments Seen by everyone



Nick Lewis

Admin · 30 November at 11:01 ·

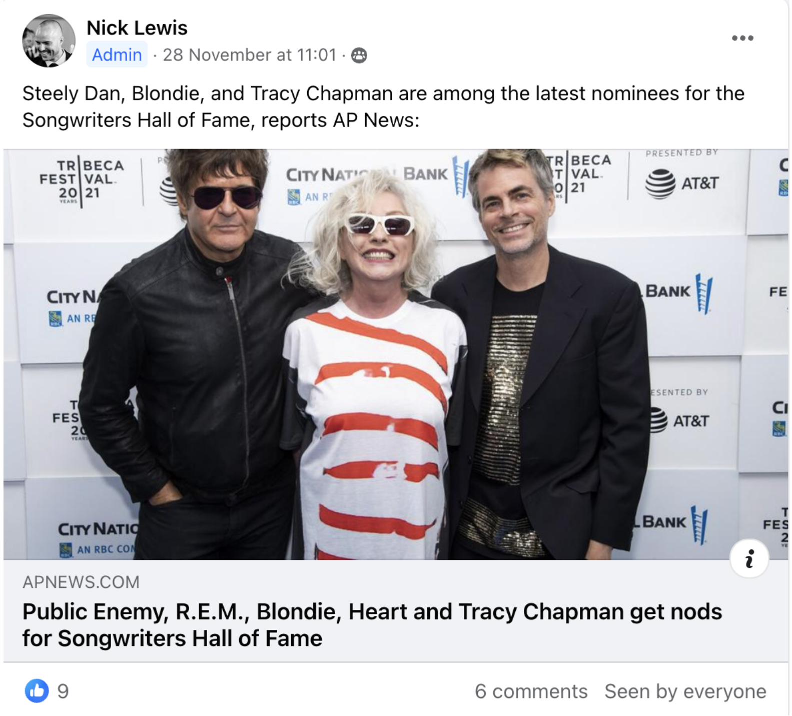
Collider shares its best TV dramas of all time. Do you agree, or are some of your favourites not on their list?

COLLIDER.COM

"Who said I never killed anyone?" The 30 Best TV Dramas of All Time, Ranked

13

7 comments Seen by everyone



Nick Lewis

Admin · 28 November at 11:01 ·

Steely Dan, Blondie, and Tracy Chapman are among the latest nominees for the Songwriters Hall of Fame, reports AP News:

APNEWS.COM

Public Enemy, R.E.M., Blondie, Heart and Tracy Chapman get nods for Songwriters Hall of Fame

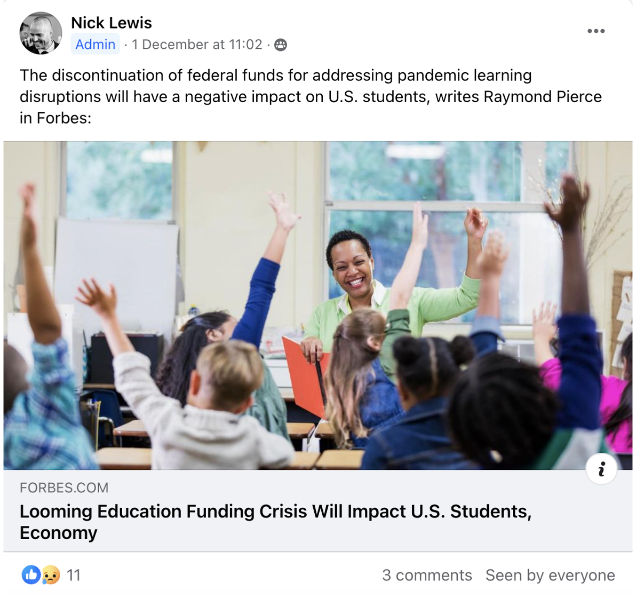
9

6 comments Seen by everyone

C.4.2 Consensus political content

These posts were scheduled at 6am Eastern Time each day, one minute before the required daily task (a poll - see below). This means that participants had to scroll past it, mimicking the appearance of a Facebook news feed/group. Articles were selected so that participants were not primed on:

- **Issue**: the contentious issue was selected based on polling and pre-study survey questions identifying less-divisive political issues.
- **Source**: Non-partisan or apolitical sources.
- **Author**: Avoid authors who are identifiably on either side of a debate or partisan divide.
- **Language**: Headline and author only, quoting directly from the piece where applicable. The language aimed to avoid priming participants.
- **Images** used in the posts avoided priming participants; i.e. with partisan figures.



Nick Lewis
Admin · 1 December at 11:02 ·

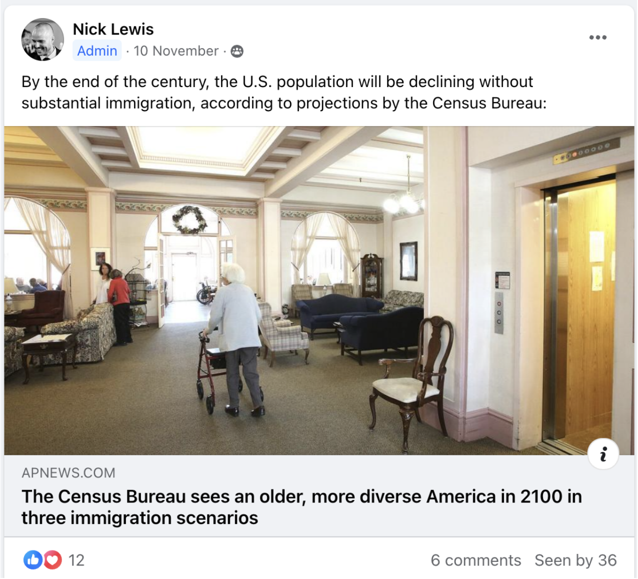
The discontinuation of federal funds for addressing pandemic learning disruptions will have a negative impact on U.S. students, writes Raymond Pierce in Forbes:

FORBES.COM
Looming Education Funding Crisis Will Impact U.S. Students, Economy

C.4.3 Contentious political content

These posts were scheduled to post before the required daily task (a poll - see below) so that participants scrolled past it, mimicking the appearance of a Facebook news feed/group. Articles were selected so that participants were not primed on:

- **Issue**: the contentious issue was selected based on polling and pre-study survey questions identifying divisive political issues.
- **Source**: Non-partisan or apolitical sources.
- **Author**: Avoid authors who are identifiably on either side of a debate or partisan divide.
- **Language**: Headline and author only, quoting directly from the piece where applicable. The language aimed to avoid priming participants.
- **Images** used in the posts avoided priming participants; i.e. with partisan figures.



Nick Lewis

Admin · 10 November ·

By the end of the century, the U.S. population will be declining without substantial immigration, according to projections by the Census Bureau:

APNEWS.COM

The Census Bureau sees an older, more diverse America in 2100 in three immigration scenarios

12

6 comments Seen by 36

C.5 Ethical considerations

Following ethical principles when conducting research which involves human subjects is of paramount importance. While due care was given to ethical considerations in all three of the papers presented in this thesis, the field experiment in paper 3 raised specific ethical concerns. Consequently, this study was designed carefully in order to address these; following the guidelines set out in the Helsinki Declaration (World Medical Association, 2022) I sought and was granted approval by the Research Ethics Committee (REC) of the London School of Economics and Political Science. Designing this experiment was a careful, iterative process, and the REC's advice and support was indispensable throughout. Here, I discuss the relevant ethical principles, and how they were incorporated into the experimental design.

Existing guidelines which govern biomedical research, such as the Belmont Report (1979) and the Declaration of Helsinki World Medical Association (2022), highlight a broad range of considerations. These include the minimisation of potential harm to participants, autonomy and fully-informed consent, and transparency and accountability of research. I closely adhere to these general principles, which are incorporated into a framework created by the American Political Science Association, and is more specifically designed for social science research (APSA, 2020). Based on these principles, a primary potential concern was the minimisation of potential harm to participants. Notably, the need to use real Facebook accounts to closely mimic a real-world social media environment where 'spiral of silence' mechanisms might occur, and therefore ensure generalisability, created a situation where anonymity for participants became impossible. This raised two main challenges. First and foremost, this raised concerns over the confidentiality and privacy for all participants. Second, and linked to the preceding concern, this design carried an attendant elevated risk of negative social repercussions for participants. For example, if a participant attempted to harass, bully or 'dox' another participant based on their comments within a closed group. Third, and finally, thought was given to potential psychological harm related to the discussion of contentious political issues. These issues, abortion and immigration, were chosen partly due to their ability to generate strong emotions in both Republicans and Democrats, but not to invoke distress. I summarise how my final design addressed each of these concerns in turn below.

Regarding the first two concerns, a clear justification is required for using real Facebook accounts. Running this experiment with anonymous profiles would fundamentally undermine the validity of one the core mechanisms tested in the paper: that participants fear social isolation and therefore disengage from political discussion if they believe their views to be in the minority. Anonymity, it is argued, removes some of that fear. Indeed, existing studies (Wu and Atkin, 2018) have shown that perceived anonymity affects the likelihood of engagement with political discussion on social media platforms. Regarding the risk of both negative social and psychological consequences, any risk or exposure to participants in this situation was concluded to be equal or less than subjects' day-to-day activities on Facebook. Anyone that chooses to comment on a public Facebook page can be messaged or receive friend requests from another use. Examples were provided from publicly-available political discussion on the U.K. Labour Party's page. By contrast, risk to participants in my experimental setting was judged to be lower for two main reasons. First, and unlike overtly political pages, where posts are specifically designed to invoke a reaction, my intention was not to provoke a reaction, instead simply measuring variation in engagement across issues. Participants were not encouraged or forced to engage with these issues: the only mandatory engagement was with a daily Facebook poll where other participants could not see who had voted or for which option. Second, the size of the groups were relatively small, drastically reducing the risk of exposure for participants when compared with their normal daily activities on Facebook. Further, existing published studies used real accounts in Facebook groups to issue similar issue-based treatments (Feezell, 2018). Again, the attendant risk of exposure in this paper was very similar to my final design, in that members of the same group were able to view group membership and any comments made beneath treatment news articles. In addition, I gave no encouragement or incentive for participants to engage - either with the treatment news articles or with other members of the group.

Finally, full informed consent was obtained before participants engaged in the study. Only those who gave their explicit written consent, to a comprehensive information document fully outlining the terms of their involvement, participated in the study. Details of the study's aims, its funding, participant compensation, data protection, and the presentation of data at an anonymised, aggregated level were all included. To further minimise risk and ensure full informed consent, the following four key pieces of information were emphasised. First, that the only required involvement was with daily short tasks, and that involvement in political discussion was not required. Second, that consent could be withdrawn at any point for any reason, with participants removed from the group immediately. Third, and to avoid distress or potential psychological harm related to the issues, participants were informed that online discussion would be political, so that anyone uncomfortable with political discussion could choose not to take part. Fourth, and finally, subjects were informed that, similar to all public activity on Facebook, other members of the group would be able to view their name and request a connection with them. Guidance was given as to how privacy settings could be amended; for example, setting messages to 'don't receive requests', so that other participants could not contact them privately.

As a further step to minimise potential harm, when joining their assigned

Facebook group, participants were required to read and agree to each group's four community rules. These required 1) respect for others in debates, 2) no self-promotion or spam, 3) no hate speech or discriminatory comments about race, religion, culture, sexual orientation, gender, or identity and 4) respect for every participant's privacy. Finally, suspected harmful or abusive posts were flagged by Facebook and moderated before appearing in each treatment group, and participants also had the ability to report content to group moderators. If any post was judged to be harmful or abusive, it would not appear. However, at no point did any participant share content of this kind.

Bibliography

Al Baghal, T., Wenz, A., Sloan, L., and Jessop, C. (2021). Linking Twitter and survey data: asymmetry in quantity and its impact. *EPJ Data Science*, 10(1):32. Publisher: Springer Berlin Heidelberg.

Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110(3):629–676. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Allcott, H. and Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2):211–236.

Anderson, M. (2024). Americans' Views of Technology Companies. *PewResearchCenter Washington, DC, USA*. Accessed 29th April, 2024. Available at: `https://www.pewresearch.org/internet/2024/04/29/americans-views-of-technology-companies-2/`.

Antypas, D., Preece, A., and Camacho-Collados, J. (2023). Negativity spreads faster: A large-scale multilingual Twitter analysis on the role of sentiment in political communication. *Online Social Networks and Media*, 33:100242.

APSA (2020). Principles and Guidance for Human Subjects Research. *APSA*. Accessed 11th July, 2024. Available at: `https://connect.apsanet.org/hsr/principles-and-guidance`.

Arceneaux, K. (2012). Cognitive biases and the strength of political arguments. *American Journal of Political Science*, 56(2):271–285. Publisher: Wiley Online Library.

Aristotle (1909). *Aristotle: Rhetoric.* Cambridge: University Press.

Arkes, H. (1993). Can Emotion Supply the Place of Reason? In Marcus, G.E. and Hanson, R.L., editor, *Reconsidering the Democratic Public*, pages 287–305. Penn State University Press.

Aronson, E. (1972). *The Social Animal.* Viking Press.

Asch, S. E. (1955). Opinions and Social Pressure. *Scientific American*, 193(5):31–35. Publisher: JSTOR.

Bacon, P. (2021). How Twitter became the media of America's left. *The Washington Post.* October 26th, 2021. Available at: `https://www.washingtonpost.com/opinions/2021/10/26/how-twitter-became-media-americas-left/`,.

Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132. Publisher: American Association for the Advancement of Science.

Banda, K. K. and Cluverius, J. (2018). Elite polarization, party extremity, and affective polarization. *Electoral Studies*, 56:90–101. Publisher: Elsevier.

Banks, A., Calvo, E., Karol, D., and Telhami, S. (2021). #polarizedfeeds: Three experiments on polarization, framing, and social media. *The International Journal of Press/Politics*, 26(3):609–634. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, 23(1):76–91. Publisher: Cambridge University Press.

Barberá, P. (2020). Social media, echo chambers, and political polarization. *Social media and democracy: The state of the field, prospects for reform*, 34. Publisher: Cambridge University Press Cambridge.

Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., and Tucker, J. A. (2019). Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, 113(4):883–901. Publisher: Cambridge University Press.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10):1531–1542. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Barberá, P. and Rivero, G. (2015). Understanding the political representativeness of Twitter users. *Social Science Computer Review*, 33(6):712–729. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Barrie, C. (2023). Did the Musk takeover boost contentious actors on Twitter? *Harvard Kennedy School Misinformation Review*. Accessed August 29th, 2023. Available at: `https://misinforeview.hks.harvard.edu/article/did-the-musk-takeover-boost-contentious-actors-on-twitter/`.

Barrie, C. and Ho, J. C.-t. (2021). academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software*, 6(62):3272.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4):323–370. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Beam, M. A., Child, J. T., Hutchens, M. J., and Hmielowski, J. D. (2018a). Context collapse and privacy management: Diversity in Facebook friends increases online news reading and sharing. *New Media & Society*, 20(7):2296–2314. Publisher: SAGE Publications Sage UK: London, England.

Beam, M. A., Hutchens, M. J., and Hmielowski, J. D. (2018b). Facebook news and (de)polarization: Reinforcing spirals in the 2016 US election. *Information, Communication & Society*, 21(7):940–958. Publisher: Taylor & Francis.

Belmont Report (1979). The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. *The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.* Accessed 11th July, 2024. Available at: `https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html`.

Benkler, Y., Faris, R., and Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

Bennett, W. L. and Manheim, J. B. (2006). The one-step flow of communication. *The Annals of the American Academy of Political and Social Science*, 608(1):213–232.

Berger, J. (2011). Arousal increases social transmission of information. *Psychological Science*, 22(7):891–893.

Berger, J. and Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2):192–205. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

BES (2021). Age and voting behaviour at the 2019 General Election. *British Election Study*. Publisher: British Election Study.

Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., and Quattrociocchi, W. (2016). Users polarization on Facebook and YouTube. *PloS one*, 11(8):e0159641. Publisher: Public Library of Science San Francisco, CA USA.

Bessone, P., Campante, F., Ferraz, C., and Souza, P. (2019). Internet access, social media, and the behavior of politicians: Evidence from Brazil. *Working Paper, Massachusetts Institute of Technology.*

Blank, G. and Lutz, C. (2017). Representativeness of social media in Great Britain: investigating Facebook, Linkedin, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, 61(7):741–756. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Bless, H., Mackie, D. M., and Schwarz, N. (1992). Mood effects on attitude judgments: Independent effects of mood before and after message elaboration. *Journal of Personality and Social Psychology*, 63(4):585. Publisher: American Psychological Association.

Bolsen, T. and Druckman, J. N. (2015). Counteracting the politicization of science. *Journal of Communication*, 65(5):745–769. Publisher: Oxford University Press.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298. Publisher: Nature Publishing Group UK London.

Booth, R. (2025). Meta to get rid of factcheckers and recommend more political content. *The Guardian.* 8th January, 2025. Available at: `https://www.theguardian.com/technology/2025/jan/07/meta-facebook-instagram-threads-mark-zuckerberg-remove-fact-checkers-recommend-`

Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information, Communication & Society*, 18(5):524–538. Publisher: Taylor & Francis.

Bowyer, B. and Kahne, J. (2019). Motivated circulation: How misinformation and ideological alignment influence the circulation of political content. *International Journal of Communication*, 13:25.

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2022). Cross-country trends in affective polarization. *Review of Economics and Statistics*, pages 1–60. Publisher: MIT Press, Cambridge, MA.

Brader, T. (2005). Striking a responsive chord: How political ads motivate and persuade voters by appealing to emotions. *American Journal of Political Science*, 49(2):388–405. Publisher: Wiley Online Library.

Brader, T., Marcus, G. E., and Miller, K. L. (2011). 'Emotion and Public Opinion'. In *The Oxford handbook of American public opinion and the media.*

Brady, W. J., Gantman, A. P., and Van Bavel, J. J. (2020). Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology: General*, 149(4):746. Publisher: American Psychological Association.

Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., and Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, 148(10):1802. Publisher: American Psychological Association.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., and Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318. Publisher: National Acad Sciences.

Brambor, T., Clark, W. R., and Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1):63–82. Publisher: Cambridge University Press.

Briggs, A. and Burke, P. (2009). *A social history of the media: From Gutenberg to the Internet*. Polity.

Broockman, D. E. and Skovron, C. (2018). Bias in perceptions of public opinion among political elites. *American Political Science Review*, 112(3):542–563.

Brown, M. and Sanderson, Z. (2020). How Trump impacts harmful Twitter speech: A case study in three tweets. Accessed October 22nd, 2020. Available at: `https://www.brookings.edu/articles/how-trump-impacts-harmful-twitter-speech-a-case-study-in-three-tweets/`.

Butters, R. and Hare, C. (2022). Polarized networks? New evidence on American voters' political discussion networks. *Political Behavior*, 44(3):1079–1103. Publisher: Springer.

Bäck, E. A., Bäck, H., Fredén, A., and Gustafsson, N. (2019). A Social Safety Net? Rejection sensitivity and political opinion sharing among young people in social media. *New Media & Society*, 21(2):298–316. Publisher: Sage Publications Sage UK: London, England.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.

Cao, J., Xia, T., Li, J., Zhang, Y., and Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7-9):1775–1781. Publisher: Elsevier.

Caplan, B. (2007). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton: Princeton University Press.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5):752. Publisher: American Psychological Association.

Chan, M. (2021). Reluctance to talk about politics in face-to-face and facebook settings: Examining the impact of fear of isolation, willingness to self-censor, and peer network characteristics. In *Social Media News and Its Impact*, pages 169–191. Routledge.

Charteris-Black, J. (2011). *Politicians and rhetoric: The persuasive power of metaphor*. Springer.

Chen, H.-T. (2018). Spiral of silence on social media and the moderating role of disagreement and publicness in the network: Analyzing expressive and withdrawal behaviors. *New Media & Society*, 20(10):3917–3936. Publisher: SAGE Publications Sage UK: London, England.

Cho, J., Ahmed, S., Hilbert, M., Liu, B., and Luu, J. (2020). Do search algorithms endanger democracy? An experimental investigation of algorithm effects on political polarization. *Journal of Broadcasting & Electronic Media*, 64(2):150–172. Publisher: Taylor & Francis.

Clarke, H. D. and Stewart, M. C. (1998). The decline of parties in the minds of citizens. *Annual Review of Political Science*, 1(1):357–378. Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

Colleoni, E., Rozza, A., and 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. Publisher: Oxford University Press.

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, pages 89–96. Issue: 1.

Coppock, A., Guess, A., and Ternovski, J. (2016). When treatments are tweets: A network mobilization experiment over Twitter. *Political Behavior*, 38:105–128. Publisher: Springer.

Corsi, G. (2024). Evaluating Twitter's algorithmic amplification of low-credibility content: an observational study. *EPJ Data Science*, 13(1):18.

Courea, E. (2024). Labour MPs begin quitting X over 'hate and disinformation'. *The Guardian.* 12th August, 2024. Available at: `https://www.theguardian.com/technology/article/2024/aug/12/labour-mps-begin-quitting-x-over-hate-and-disinformation`.

Crabtree, C., Golder, M., Gschwend, T., and Indridason, I. (2020). It is not only what you say, it is also how you say it: The strategic use of campaign sentiment. *The Journal of Politics*, 82(3):1044–1060. Publisher: The University of Chicago Press Chicago, IL.

Daisley, B. (2024). As an ex-Twitter boss, I have a way to grab Elon Musk's attention. If he keeps stirring unrest, get an arrest warrant. *The Guardian.* 12th August, 2024. Available at: https://www.theguardian.com/commentisfree/article/2024/aug/12/elon-musk-x-twitter-uk-riot-tweets-arrest-warrant.

Dalton, R. J. (1987). Generational change in elite political beliefs: The growth of ideological polarization. *The Journal of Politics*, 49(4):976–997. Publisher: Southern Political Science Association.

De Castella, K., McGarty, C., and Musgrove, L. (2009). Fear appeals in political rhetoric about terrorism: An analysis of speeches by Australian Prime Minister Howard. *Political Psychology*, 30(1):1–26. Publisher: Wiley Online Library.

De Zuniga, H. G. (2015). Toward a European public sphere? The promise and perils of modern democracy in the age of digital and social media. *International Journal of Communication*, 9:9.

De Zúñiga, H. G., Marné, H. M., and Carty, E. (2022). Abating Dissonant Public Spheres: Exploring the Effects of Affective, Ideological and Perceived Societal Political Polarization on Social Media Political Persuasion. *Political Communication*, pages 1–19. Publisher: Taylor & Francis.

DellaVigna, S. and Kaplan, E. (2007). The Fox News effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Delli Carpini, M. X. (2000). Gen. com: Youth, civic engagement, and the new information environment. *Political Communication*, 17(4):341–349. Publisher: Taylor & Francis.

Delli Carpini, M. X. and Keeter, S. (1996). *What Americans know about politics and why it matters.* Yale University Press.

Dickson, Z. P. and Hobolt, S. B. (2024). Elite cues and noncompliance. *American Political Science Review*, pages 1–17.

Douglas, D. M. (2016). Doxing: a conceptual analysis. *Ethics and Information Technology*, 18(3):199–210. Publisher: Springer.

Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, 65(2):135–150. Publisher: The University of Chicago Press.

Dreisbach, T. (2022). How Trump's 'will be wild!' tweet drew rioters to the Capitol on Jan. 6. *NPR*. 13th July, 2022. https://www.npr.org/2022/07/13/1111341161/how-trumps-will-be-wild-tweet-drew-rioters-to-the-capitol-on-jan-6.

Druckman, J. N. (2003). The Power of Television Images: The first Kennedy-Nixon debate revisited. *Journal of Politics*, 65(2):559–571. Publisher: Wiley Online Library.

Druckman, J. N. and Holmes, J. W. (2004). Does presidential rhetoric matter? Priming and presidential approval. *Presidential Studies Quarterly*, 34(4):755–778. Publisher: Wiley Online Library.

Druckman, J. N. and Levendusky, M. S. (2019). What do we measure when we measure affective polarization? *Public Opinion Quarterly*, 83(1):114–122. Publisher: Oxford University Press UK.

Druckman, J. N. and Nelson, K. R. (2003). Framing and deliberation: How citizens' conversations limit elite influence. *American Journal of Political Science*, 47(4):729–745. Publisher: Wiley Online Library.

Druckman, J. N., Peterson, E., and Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review*, 107(1):57–79. Publisher: Cambridge University Press.

Dryzek, J. S. (2002). *Deliberative democracy and beyond: Liberals, critics, contestations*. Oxford University Press, USA.

Duggan, M. and Smith, A. (2016). The Political Environment on Social Media. October 25th, 2016. Available at: https://www.pewresearch.org/internet/2016/10/25/the-political-environment-on-social-media/.

Eady, G., Nagler, J., Guess, A., Zilinsky, J., and Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *Sage Open*, 9(1):2158244019832705. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Eckert, S. (2018). Fighting for recognition: Online abuse of women bloggers in Germany, Switzerland, the United Kingdom, and the United States. *New Media & Society*, 20(4):1282–1302. Publisher: SAGE Publications Sage UK: London, England.

Eisenstein, E. L. (1980). *The Printing Press as an Agent of Change*, volume 1. Cambridge University Press.

Eliasoph, N. (1998). *Avoiding politics: How Americans produce apathy in everyday life*. Cambridge University Press.

Ernst, N., Blassnig, S., Engesser, S., Büchel, F., and Esser, F. (2019). Populists prefer social media over talk shows: An analysis of populist messages and stylistic elements across six countries. *Social Media + Society*, 5(1):2056305118823358.

Evans, G. and Tilley, J. (2017). *The new politics of class: The political exclusion of the British working class*. Oxford University Press.

Farage, N. (2018). “Enough is enough, time to tell the arrogant, unelected EU bullies where to go. No British prime minister should be treated like this.". *Twitter*. September 21st, 2018. Available at: https://twitter.com/Nigel_Farage/status/1043084994074357760.

Farage, N. (2022). “The Red Wall voted to take back control of immigration, not to surrender the English Channel to criminal trafficking gangs.". *Twitter*. January 6th, 2022. Available at: `https://twitter.com/nigel_farage/status/1479027078280404996`.

Farrer, B. (2024). Political communication as a tragedy of the commons. *Political Studies*, 72(2):701–718.

Faverio, M. (2023). Majority of U.S. Twitter users say they've taken a break from the platform in the past year. May 17th, 2023. Available at: https://www.pewresearch.org/short-reads/2023/05/17/majority-of-us-twitter-users-say-theyve-taken-a-break-from-the-platform-in-the-past-year/.

Feezell, J. T. (2018). Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political Research Quarterly*, 71(2):482–494.

Festinger, L. (1950). Informal social communication. *Psychological Review*, 57(5):271. Publisher: American Psychological Association.

Fischer, A. (2021). Fast & wild bootstrap inference. *PhD dissertation, Aarhus University, August 2021.*

Fishkin, J. (2009). *When the people speak: Deliberative democracy and public consultation*. Oxford University Press.

Flaxman, S., Goel, S., and Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1):298–320. Publisher: Oxford University Press.

Fox, J. and Holt, L. F. (2021). Fear of isolation and perceived affordances: The spiral of silence on social networking sites regarding police discrimination. In *Social Media News and Its Impact*, pages 147–168. Routledge.

Fox, J. and Warber, K. M. (2015). Queer identity management and political self-expression on social networking sites: A co-cultural approach to the spiral of silence. *Journal of Communication*, 65(1):79–100.

Frimer, J. A., Skitka, L. J., and Motyl, M. (2017). Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *Journal of Experimental Social Psychology*, 72:1–12. Publisher: Elsevier.

Fung, B. (2023). Elizabeth Warren and Lindsey Graham want a new agency to regulate tech. *CNN*. 27th July, 2023. Available at: https://edition.cnn.com/2023/07/27/tech/big-tech-regulation-new-federal-agency.

Gallup (2023). Abortion Trends by Party Identification. Available at: https://news.gallup.com/poll/246278/abortion-trends-party.aspx.

Garrett, R. K. (2009). Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication*, 59(4):676–699.

Garrett, R. K., Weeks, B. E., and Neo, R. L. (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. Publisher: Oxford University Press Oxford, UK.

Gearhart, S. and Zhang, W. (2014). Gay bullying and online opinion expression: Testing spiral of silence in the social media environment. *Social Science Computer Review*, 32(1):18–36. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Gearhart, S. and Zhang, W. (2015). "Was it something I said?" "No, it was something you posted!" A study of the spiral of silence theory in social media contexts. *Cyberpsychology, Behavior, and Social Networking*, 18(4):208–213. Mary Ann Liebert, New Rochelle, NY.

Gearhart, S. and Zhang, W. (2018). 'Same Spiral, Different Day? Testing the spiral of silence across issue types'. *Communication Research*, 45(1):34–54. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Gidron, N., Adams, J., and Horne, W. (2020). *American Affective Polarization in Comparative Perspective.* Cambridge University Press.

Gilardi, F., Gessler, T., Kubli, M., and Müller, S. (2022). Social media and political agenda setting. *Political Communication*, 39(1):39–60. Publisher: Taylor & Francis.

Glynn, C. J., Hayes, A. F., and Shanahan, J. (1997). Perceived Support for One's Opinions and Willingness to Speak Out: A Meta-Analysis of Survey Studies on the" Spiral of Silence". *Public Opinion Quarterly*, pages 452–463. Publisher: JSTOR.

Goffman, E. (1959). *The Presentation of Self in Everyday Life.* New York, NY: Bantam Doubleday Dell Publishing.

Gouvy, C. (2021). Social media curbs threaten 'last relic' of Tunisia's revolution. *Al Jazeera.* 16th January, 2021. Available at: `https://www.aljazeera.com/news/2021/1/16/social-media-ban-threatens-the-last-relic-of-tunisia-revolution`.

Green, D. A. (2025). The coming battle between social media and the state. *Financial Times.* 11th January, 2025. Available at: `https://www.ft.com/content/917c9535-1cdb-4f6a-9a15-1a0c83663bfd`.

Green, D. P., Palmquist, B., and Schickler, E. (2004). *Partisan hearts and minds: Political parties and the social identities of voters.* Yale University Press.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235. Publisher: National Acad Sciences.

Grimmer, J. (2010). A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. *Political Analysis*, 18(1):1–35. Publisher: Cambridge University Press.

Gubala, S. and Austin, S. (2024). Majorities in most countries surveyed say social media is good for democracy. February 23rd, 2024. Available at: `https://www.pewresearch.org/short-reads/2024/02/23/majorities-in-most-countries-surveyed-say-social-media-is-good-for-democracy/`.

Guess, A., Munger, K., Nagler, J., and Tucker, J. (2019a). How accurate are survey responses on social media and politics? *Political Communication*, 36(2):241–258. Publisher: Taylor & Francis.

Guess, A., Nagler, J., and Tucker, J. (2019b). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1).

Habermas, J. (1989). *The Structural Transformation of the Public Sphere*, volume 1. MIT Press.

Haidt, J. (2012). *The Righteous Mind: Why good people are divided by politics and religion.* Vintage.

Halberstam, Y. and Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143:73–88. Publisher: Elsevier.

Hale, S. A., John, P., Margetts, H., and Yasseri, T. (2018). How digital design shapes political participation: A natural experiment with social information. *PloS one*, 13(4):e0196068. Publisher: Public Library of Science San Francisco, CA USA.

Hampton, K. N., Shin, I., and Lu, W. (2017). Social media and political discussion: when online presence silences offline conversation. *Information, Communication & Society*, 20(7):1090–1107. Publisher: Taylor & Francis.

Harteveld, E. (2021). Fragmented foes: Affective polarization in the multiparty context of the Netherlands. *Electoral Studies*, 71:102332. Publisher: Elsevier.

Havas, D. A. and Chapp, C. B. (2016). Language for winning hearts and minds: Verb aspect in US presidential campaign speeches for engaging emotion. *Frontiers in Psychology*, 7:899. Publisher: Frontiers.

Hayes, A. and Matthes, J. (2017). Self-censorship, the Spiral of Silence, and Contemporary Political Communication. In *The Oxford Handbook of Political Communication*.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, pages 153–161. Publisher: JSTOR.

Heiss, R., Schmuck, D., and Matthes, J. (2019). What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions. *Information, Communication & Society*, 22(10):1497–1513.

Hellweg, S. A., Pfau, M., and Brydon, S. R. (1992). *Televised presidential debates: Advocacy in contemporary America*. Praeger Pub Text.

Henderson, M., Jiang, K., Johnson, M., and Porter, L. (2021). Measuring Twitter use: validating survey-based measures. *Social Science Computer Review*, 39(6):1121–1141. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Hernández-Morales, A. (2025). Zuckerberg urges Trump to stop the EU from fining US tech companies. *Politico*. 11th January, 2025. Available at: `https://www.politico.eu/article/zuckerberg-urges-trump-to-stop-eu-from-screwing-with-fining-us-tech-companies/`.

Ho, S. S. and McLeod, D. M. (2008). Social-psychological influences on opinion expression in face-to-face and computer-mediated communication. *Communication Research*, 35(2):190–207. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Hobolt, S. B., Lawall, K., and Tilley, J. (2024). The polarizing effect of partisan echo chambers. *American Political Science Review*, 118(3):1464–1479.

Hobolt, S. B., Leeper, T. J., and Tilley, J. (2021). Divided by the vote: Affective polarization in the wake of the Brexit referendum. *British Journal of Political Science*, 51(4):1476–1493. Publisher: Cambridge University Press.

Hochschild, J. L. and Einstein, K. L. (2015). Do facts matter? Information and misinformation in American politics. *Political Science Quarterly*, 130(4):585–624.

Holland, S. and Singh, K. (2025). Biden takes aim at 'tech industrial complex', echoing Eisenhower. *Reuters*. 16th January, 2025. Available at: `https://www.reuters.com/world/us/biden-raises-alarm-about-dangerous-concentration-power-among-few-wealthy-people`

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168–177.

Hutto, C. and Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, pages 216–225. Issue: 1.

Ingram, D. (2024). How Elon Musk turned X into a pro-Trump echo chamber. *NBC News*. 31st October, 2024.

Available at: `https://www.nbcnews.com/tech/social-media/elon-musk-turned-x-trump-echo-chamber-rcna174321`.

Iyengar, S. (1994). *Is Anyone Responsible?: How Television Frames Political Issues*. University of Chicago Press.

Iyengar, S. (1996). Framing responsibility for political issues. *The Annals of the American Academy of Political and Social Science*, 546(1):59–70. Publisher: SAGE Periodicals Press.

Iyengar, S. and Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1):19–39. Publisher: Oxford University Press.

Iyengar, S. and Krupenkin, M. (2018). The strengthening of partisan affect. *Political Psychology*, 39:201–218.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22:129–146. Publisher: Annual Reviews.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly*, 76(3):405–431.

Iyengar, S. and Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3):690–707. Publisher: Wiley Online Library.

Jockers, M. (2017). Package 'syuzhet'. `https://cran.r-project.org/web/packages/syuzhet`.

Jones, C., Trott, V., and Wright, S. (2020). Sluts and soyboys: MGTOW and the production of misogynistic online harassment. *New Media & Society*, 22(10):1903–1921. Publisher: Sage Publications Sage UK: London, England.

Kalmoe, N. P. and Mason, L. (2022). Radical American partisanship: Mapping violent hostility, its causes, and the consequences for democracy. In *Radical American Partisanship*. University of Chicago Press.

Karlsen, R. and Enjolras, B. (2016). Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with Twitter data. *The International Journal of Press/Politics*, 21(3):338–357. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Katz, E. and Fialkoff, Y. (2017). Six concepts in search of retirement. *Annals of the International Communication Association*, 41(1):86–91. Publisher: Taylor & Francis.

Kent Jennings, M. and Zeitner, V. (2003). Internet use and civic engagement: A longitudinal analysis. *Public Opinion Quarterly*, 67(3):311–334. Publisher: Oxford University Press.

Kingzette, J., Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., and Ryan, J. B. (2021). How affective polarization undermines support for democratic norms. *Public Opinion Quarterly*, 85(2):663–677. Publisher: Oxford University Press.

Kirshner, M. (2023). Twitter Was for News. *Slate*. October 5th, 2023. Available at: `https://slate.com/technology/2023/10/elon-musk-x-twitter-news-links-headlines-why.html`.

Klar, S. (2014). Partisanship in a social setting. *American Journal of Political Science*, 58(3):687–704. Publisher: Wiley Online Library.

Kosmidis, S., Hobolt, S. B., Molloy, E., and Whitefield, S. (2019). Party competition and emotive rhetoric. *Comparative Political Studies*, 52(6):811–837. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Kruse, L. M., Norris, D. R., and Flinchum, J. R. (2018). Social media as a public sphere? Politics on social media. *The Sociological Quarterly*, 59(1):62–84.

Kushin, M. J., Yamamoto, M., and Dalisay, F. (2019). Societal majority, Facebook, and the spiral of silence in the 2016 US presidential election. *Social Media + Society*, 5(2):2056305119855139.

Larsson, A. O. and Moe, H. (2012). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society*, 14(5):729–747. Publisher: Sage Publications Sage UK: London, England.

Lasorsa, D. L. (1991). Political outspokenness: Factors working against the spiral of silence. *Journalism Quarterly*, 68(1-2):131–140.

Laver, M., Benoit, K., and Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2):311–331. Publisher: Cambridge University Press.

Layman, G. C., Carsey, T. M., and Horowitz, J. M. (2006). Party polarization in American politics: Characteristics, causes, and consequences. *Annual Review of Political Science*, 9:83–110. Publisher: Annual Reviews.

Lazarsfeld, P. F., Berelson, B., and Gaudet, H. (1968). The People's Choice. In *The People's Choice*. Columbia University Press.

Leeper, T. J. (2017). Interpreting regression results using average marginal effects with R's margins. *Available at the comprehensive R Archive Network (CRAN)*, 32:1–32.

Lelkes, Y., Sood, G., and 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. Publisher: Wiley Online Library.

Lelkes, Y. and Westwood, S. J. (2017). The limits of partisan prejudice. *The Journal of Politics*, 79(2):485–501.

Levendusky, M. (2013). Partisan media exposure and attitudes toward the opposition. *Political Communication*, 30(4):565–581. Publisher: Taylor & Francis.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. (2019). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Lu, Y., Zhang, P., Cao, Y., Hu, Y., and Guo, L. (2014). On the frequency distribution of retweets. *Procedia Computer Science*, 31:747–753. Publisher: Elsevier.

Lönnqvist, J.-E. and Itkonen, J. V. (2016). Homogeneity of personal values and personality traits in Facebook social networks. *Journal of Research in Personality*, 60:24–35. Publisher: Elsevier.

MacKinnon, J. G., Nielsen, M. Ø., and Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, 232(2):272–299.

Marcus, G. E. (2000). Emotions in politics. *Annual Review of Political Science*, 3(1):221–250. Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

Margetts, H., John, P., Hale, S., and Yasseri, T. (2015). *Political Turbulence: How Social Media Shape Collective Action*. Princeton University Press.

Martella, A. and Bracciale, R. (2022). Populism and emotions: Italian political leaders' communicative strategies to engage Facebook users. *Innovation: The European Journal of Social Science Research*, 35(1):65–85.

Martin, L. W. and Vanberg, G. (2011). *Parliaments and coalitions: The role of legislative institutions in multi-party governance*. Oxford University Press.

Marwick, A. and Lewis, R. (2017). Media manipulation and disinformation online. *New York: Data & Society Research Institute*, pages 7–19.

Marwick, A. E. and Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1):114–133. Publisher: Sage Publications Sage UK: London, England.

Mascheroni, G. and Murru, M. F. (2017). "I can share politics but I don't discuss it": everyday practices of political talk on Facebook. *Social Media+ Society*, 3(4):2056305117747849. Publisher: SAGE Publications Sage UK: London, England.

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. Publisher: Wiley Online Library.

Mason, L. (2018). Ideologues without issues: The polarizing consequences of ideological identities. *Public Opinion Quarterly*, 82(S1):866–887. Publisher: Oxford University Press US.

Matthes, J. and Hayes, A. F. (2014). Methodological conundrums in spiral of silence research. *The spiral of silence: New perspectives on communication and public opinion*, pages 54–64. Publisher: Routledge New York, NY.

Matthes, J., Knoll, J., and von Sikorski, C. (2018). The "spiral of silence" revisited: A meta-analysis on the relationship between perceptions of opinion support and political opinion expression. *Communication Research*, 45(1):3–33. Publisher: Sage Publications Sage CA: Los Angeles, CA.

McClain, C. (2021). 70% of U.S. social media users never or rarely post or share about political, social issues. May 4th, 2021. Available at: https://www.pewresearch.org/short-reads/2021/05/04/70-of-u-s-social-media-users-never-or-rarely-post-or-share-about-political-soci

McClain, C., Anderson, M., and Gelles-Watnick, R. (2024). How Americans Navigate Politics on TikTok, X, Facebook and Instagram. June 12th, 2024. Available at: https://www.pewresearch.org/internet/2024/06/12/how-americans-navigate-politics-on-tiktok-x-facebook-and-instagram/.

McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. MIT press.

Mellon, J. and Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, 4(3):2053168017720008. Publisher: SAGE Publications Sage UK: London, England.

Merrill, J. B. and Harwell, D. (2023). Elon Musk's X is throttling traffic to websites he dislikes. *The Washington Post*. August 16th, 2023. Available at: https://www.washingtonpost.com/technology/2023/08/15/twitter-x-links-delayed/.

Meyrowitz, J. (1986). *No sense of place: The impact of electronic media on social behavior*. Oxford University Press.

Miller, P. R., Bobkowski, P. S., Maliniak, D., and Rapoport, R. B. (2015). Talking politics on Facebook: Network centrality and political discussion practices in social media. *Political Research Quarterly*, 68(2):377–391.

Mohammad, S. M. and Turney, P. D. (2013). NRC emotion lexicon. *National Research Council, Canada*, 2:234.

Mor, Y., Kligler-Vilenchik, N., and Maoz, I. (2015). Political expression on Facebook in a context of conflict: Dilemmas and coping strategies of Jewish-Israeli youth. *Social Media + Society*, 1(2):2056305115606750. Publisher: SAGE Publications Sage UK: London, England.

Mosleh, M., Pennycook, G., and Rand, D. G. (2020). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLOS one*, 15(2):e0228882. Publisher: Public Library of Science San Francisco, CA USA.

Moy, P., Domke, D., and Stamm, K. (2001). The spiral of silence and public opinion on affirmative action. *Journalism & Mass Communication Quarterly*, 78(1):7–25. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Moy, P. and Hussain, M. M. (2014). Media and public opinion in a fragmented society. *The spiral of silence: New perspectives on communication and public opinion*, pages 92–100. Publisher: Routledge New York, NY.

Muchnik, L., Aral, S., and Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, 341(6146):647–651. Publisher: American Association for the Advancement of Science.

Munger, K., Bonneau, R., Nagler, J., and Tucker, J. A. (2019). Elites tweet to get feet off the streets: Measuring regime social media strategies during protest. *Political Science Research and Methods*, 7(4):815–834.

Musk, E. (2023a). “The reason I acquired Twitter is because it is important to the future of civilization to have a common digital town square, where a wide range of beliefs can be debated in a healthy manner.”.

X. May 29th, 2023. Available at: `https://x.com/muskQu0tes/status/1663073525933182977`.

Musk, E. (2023b). “this is coming from me directly. will greatly improve the esthetics.". *X*. August 22nd, 2023. Available at: `https://x.com/elonmusk/status/1693843680904216619`.

Musk, E. (2024). “civil war is inevitable". *X*. August 4th, 2024. Available at: `https://x.com/elonmusk/status/1819933223536742771`.

Mutz, D. (2007). Effects of “In-Your-Face” Television Discourse on Perceptions of a Legitimate Opposition. *American Political Science Review*, 101(4):621 – 635.

Mutz, D. C. (2001). Facilitating communication across lines of political difference: The role of mass media. *American Political Science Review*, 95(1):97–114.

Mutz, D. C. (2002). The consequences of cross-cutting networks for political participation. *American Journal of Political Science*, pages 838–855. Publisher: JSTOR.

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

Neuwirth, K., Frederick, E., and Mayo, C. (2007). The spiral of silence and fear of isolation. *Journal of Communication*, 57(3):450–468.

New York State Senate (2024). Senate Bill S7695A. *New York State Senate*. 20th June, 2024. Available at: https://www.nysenate.gov/node/12033410.

Nielsen, F. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Noelle-Neumann, E. (1974). The Spiral of Silence: A Theory of Public Opinion. *Journal of Communication*, 24(2):43–51. Publisher: JoC.

Noelle-Neumann, E. and Petersen, T. (2004). *'The Spiral of Silence and the Social Nature of Man' in L. L. Kaid (Ed.), Handbook of Political Communication Research*. Mahweh, NJ: Lawrence Erlbaum Associates.

Nordbrandt, M. (2021). Affective polarization in the digital age: Testing the direction of the relationship between social media and users' feelings for out-group parties. *New Media & Society*, page 14614448211044393. Publisher: SAGE Publications Sage UK: London, England.

Nyhan, B. and Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2):303–330. Publisher: Springer.

Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., and others (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972):137–144. Publisher: Nature Publishing Group UK London.

O'Carroll, L. (2023). EU warns Elon Musk after Twitter found to have highest rate of disinformation. *The Guardian*. 26th September, 2023. Available at: `https://www.theguardian.com/technology/2023/sep/26/eu-warns-elon-musk-that-twitter-x-must-comply-with-fake-news-laws`.

Oliphant, J. B. and Cerda, A. (2023). Republicans and Democrats have different top priorities for U.S. immigration policy. September 8th, 2022. Available at: `https://www.pewresearch.org/short-reads/2022/09/08/republicans-and-democrats-have-different-top-priorities-for-u-s-immigration-pol`

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review*, 115(3):999–1015. Publisher: Cambridge University Press.

Osnabrügge, M., Hobolt, S. B., and Rodon, T. (2021). Playing to the Gallery: Emotive Rhetoric in Parliaments. *American Political Science Review*, 115(3):885–899. Publisher: Cambridge University Press.

Pang, N., Ho, S. S., Zhang, A. M., Ko, J. S., Low, W., and Tan, K. S. (2016). Can spiral of silence and civility predict click speech on Facebook? *Computers in Human Behavior*, 64:898–905. Publisher: Elsevier.

Papacharissi, Z. (2015). *Affective publics: Sentiment, technology, and politics*. Oxford University Press.

Peeters, J., Van Aelst, P., and Praet, S. (2021). Party ownership or individual specialization? a comparison of politicians' individual issue attention across three different agendas. *Party Politics*, 27(4):692–703.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic Inquiry and Word Count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Pereira, M. M. (2021). Understanding and reducing biases in elite beliefs about the electorate. *American Political Science Review*, 115(4):1308–1324.

Petrova, M., Sen, A., and Yildirim, P. (2021). Social media and political contributions: The impact of new technology on political competition. *Management Science*, 67(5):2997–3021. Publisher: INFORMS.

Petty, R. E. and Briñol, P. (2015). Emotion and persuasion: Cognitive and meta-cognitive processes impact attitudes. *Cognition and Emotion*, 29(1):1–26. Publisher: Taylor & Francis.

Petty, R. E. and Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In *Communication and Persuasion*, pages 1–24. Springer.

Pew (2021). Social Media Fact Sheet. April 7th, 2021. Available at: https://www.pewresearch.org/internet/fact-sheet/social-media/.

Pew (2022a). The economy remains the top issue for voters in the midterms. November 2nd, 2022. Available at: https://www.pewresearch.org/short-reads/2022/11/03/key-facts-about-u-s-voter-priorities-ahead-of-the-2022-midterm-elections.

Pew (2022b). Social and moral considerations on abortion. May 6th, 2022. Available at: https://www.pewresearch.org/religion/2022/05/06/social-and-moral-considerations-on-abortion/.

Pew (2023). Economy Remains the Public's Top Policy Priority; COVID-19 Concerns Decline Again. February 6th, 2023. Available at: https://www.pewresearch.org/politics/2023/02/06/economy-remains-the-publics-top-policy-priority-covid-19-concerns-decline-again

Poletti, M., Webb, P., and Bale, T. (2019). Why do only some people who support parties actually join them? Evidence from Britain. *West European Politics*, 42(1):156–172. Publisher: Taylor & Francis.

Rathje, S., Van Bavel, J. J., and Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26):e2024292118.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.

Reiljan, A. (2020). 'Fear and loathing across party lines'(also) in Europe: Affective polarisation in European party systems. *European Journal of Political Research*, 59(2):376–396. Publisher: Wiley Online Library.

Reiljan, A. and Ryan, A. (2021). Ideological tripolarization, partisan tribalism and institutional trust: The foundations of affective polarization in the Swedish multiparty system. *Scandinavian Political Studies*, 44(2):195–219. Publisher: Wiley Online Library.

Rheault, L., Beelen, K., Cochrane, C., and Hirst, G. (2016). Measuring emotion in parliamentary debates with automated textual analysis. *PLOS one*, 11(12):e0168843. Publisher: Public Library of Science San Francisco, CA USA.

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., and Meira Jr, W. (2020). Auditing Radicalization Pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 131–141.

Roberts, M. E., Stewart, B. M., and Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515):988–1003. Publisher: Taylor & Francis.

Robison, J. and Mullinix, K. J. (2016). Elite polarization and public opinion: How polarization is communicated and its effects. *Political Communication*, 33(2):261–282. Publisher: Taylor & Francis.

Rogowski, J. C. and Sutherland, J. L. (2016). How Ideology Fuels Affective Polarization. *Political Behavior*, 38:485–508. Publisher: Springer.

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., and Webb, M. D. (2019). Fast and wild: Bootstrap inference in stata using boottest. *The Stata Journal*, 19(1):4–60.

Rusche, F. (2022). Few voices, strong echo: Measuring follower homogeneity of politicians' Twitter accounts. *New Media & Society*. Publisher: SAGE Publications Sage UK: London, England.

Salmon, C. T. and Neuwirth, K. (1990). Perceptions of opinion "climates" and willingness to discuss the issue of abortion. *Journalism Quarterly*, 67(3):567–577. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Schachter, S. (1951). Deviation, rejection, and communication. *The Journal of Abnormal and Social Psychology*, 46(2):190. Publisher: American Psychological Association.

Schulz, A. and Roessler, P. (2012). The spiral of silence and the Internet: Selection of online content and the perception of the public opinion climate in computer-mediated communication environments. *International Journal of Public Opinion Research*, 24(3):346–367. Publisher: Oxford University Press.

Scott, C. R. (1999). Communication technology and group communication. In Frey, L. R., Gouran, D., and Poole, M. S., editors, *The Handbook of Group Communication Theory and Research*, pages 432–472. Sage.

Settle, J. E. (2018). *Frenemies: How social media polarizes America*. Cambridge University Press.

Settle, J. E. and Carlson, T. N. (2019). Opting out of political discussions. *Political Communication*, 36(3):476–496. Publisher: Taylor & Francis.

Shamir, J. (1997). Speaking up and silencing out in face of a changing climate of opinion. *Journalism & Mass Communication Quarterly*, 74(3):602–614.

Shapiro, M. A. and Hemphill, L. (2017). Politicians and the policy agenda: Does use of Twitter by the US Congress direct New York Times content? *Policy & Internet*, 9(1):109–132.

Sheerin, J. (2022). Capitol riots: ‘Wild’ Trump tweet incited attack, says inquiry. *BBC News*. 12th July, 2022. Available at: `https://www.bbc.co.uk/news/world-us-canada-62140410`.

Shirky, C. (2009). *Here comes everybody: How change happens when people come together*. Penguin UK.

Shirky, C. (2010). *Cognitive Surplus: Creativity and Generosity in a Connected Age*. Penguin.

Siegel, A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J., and Tucker, J. A. (2019). Trumping hate on Twitter? Online hate in the 2016 US election and its aftermath. *Social Media and Political Participation, New York University*.

Slapin, J. B. and Kirkland, J. H. (2020). The Sound of Rebellion: Voting Dissent and Legislative Speech in the UK House of Commons. *Legislative Studies Quarterly*, 45(2):153–176.

Slapin, J. B., Kirkland, J. H., Lazzaro, J. A., Leslie, P. A., and O'Grady, T. (2018). Ideology, grandstanding, and strategic party disloyalty in the British Parliament. *American Political Science Review*, 112(1):15–30. Publisher: Cambridge University Press.

Smith, A., Silver, L., Johnson, C., and Jiang, J. (2019). More people are comfortable discussing politics in person than on their phones or via social media. *Pew Research Center*.

Sobolewska, M. and Ford, R. (2020). *Brexitland: Identity, diversity and the reshaping of British politics*. Cambridge University Press.

Soroka, S. N. (2006). Good news and bad news: Asymmetric responses to economic information. *The Journal of Politics*, 68(2):372–385. Publisher: Cambridge University Press New York, USA.

Spirling, A. (2016). Democratization and linguistic complexity: The effect of franchise extension on parliamentary discourse, 1832–1915. *The Journal of Politics*, 78(1):120–136.

Stanley-Becker, I. and Alemany, J. (2022). Trump hid plan for Capitol march on day he marked as 'wild', panel says. *The Wash-*

ington Post. 12th July, 2022. https://www.washingtonpost.com/national-security/2022/07/12/january-6-hearing-trump/.

Stieglitz, S. and 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.

Stier, S., Bleier, A., Lietz, H., and Strohmaier, M. (2018). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Political Communication*, 35(1):50–74. Publisher: Taylor & Francis.

Stoycheff, E. (2016). Under surveillance: Examining Facebook's spiral of silence effects in the wake of NSA internet monitoring. *Journalism & Mass Communication Quarterly*, 93(2):296–311. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Stroud, N. J. (2010). Polarization and Partisan Selective Exposure. *Journal of Communication*, 60(3):556–576.

Sunstein, C. R. (2018). *#Republic: Divided democracy in the age of social media*. Princeton University Press.

Sveningsson, M. (2014). "i don't like it and I think it's useless, people discussing politics on Facebook": Young Swedes' understandings of social media use for political discussion. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 8(3):106–120.

Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3):755–769.

Tajfel, H. and Turner, J. C. (2004). The social identity theory of intergroup behavior. In *Political Psychology*, pages 276–293. Psychology Press.

Tajfel, H., Turner, J. C., Austin, W. G., and Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, 56(65):9780203505984–16.

Taylor, S., Muchnik, L., Kumar, M., and Aral, S. (2022). 'Identity Effects in Social Media'. *Nature Human Behaviour*. Publisher: Nature.

Terren, L. and Borge-Bravo, R. (2021). Echo chambers on social media: A systematic review of the literature. *Review of Communication Research*, 9:99–118.

Thompson, A. and Ford, F. (2021). Members of Several Well-Known Hate Groups Identified at Capitol Riot. *PBS Frontline*. 9th January, 2021. Available at: `https://www.pbs.org/wgbh/frontline/article/several-well-known-hate-groups-identified-at-capitol-riot/`.

Tilley, J. and Hobolt, S. B. (2023). Losers' consent and emotions in the aftermath of the Brexit referendum. *West European Politics*, pages 1–19. Publisher: Taylor & Francis.

Timm, T. (2023). Elon Musk has become the world's biggest hypocrite on free speech. *The Guardian*. 15th January, 2024. Available at: `https://www.theguardian.com/commentisfree/2024/jan/15/elon-musk-hypocrite-free-speech`.

Trump, D. (2021). "We will not be SILENCED! Twitter is not about FREE SPEECH. They are all about promoting a Radical Left platform where some of the most vicious people in the world are allowed to speak freely...". *Twitter*. January 8th, 2021.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., and Nyhan, B. (2018). Social media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. *SSRN*.

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157):1124–1131. Publisher: American association for the advancement of science.

Urman, A. (2020). Context matters: political polarization on Twitter from a comparative perspective. *Media, Culture & Society*, 42(6):857–879. Publisher: SAGE Publications Sage UK: London, England.

Vaccari, C. and Valeriani, A. (2018). Digital political talk and political participation: Comparing established and third wave democracies. *Sage Open*, 8(2):2158244018784986. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., and Tucker, J. (2013). Social media and political communication. A survey of Twitter users during the 2013 Italian general election. *Rivista Italiana di Scienza Politica*, 43(3):381–410. Publisher: Società editrice il Mulino.

Valentino, N. A., Brader, T., Groenendyk, E. W., Gregorowicz, K., and Hutchings, V. L. (2011). Election night's alright for fighting: The role of emotions in political participation. *The Journal of Politics*, 73(1):156–170. Publisher: Cambridge University Press New York, USA.

Van Kessel, S. and Castelein, R. (2016). Shifting the blame. Populist politicians' use of Twitter as a tool of opposition. *Journal of Contemporary European research*, 12(2).

van Vliet, L. (2021). Moral expressions in 280 characters or less: An analysis of politician tweets following the 2016 Brexit referendum vote. *Frontiers in Big Data*, 4:699653.

Verba, S., Schlozman, K. L., and Brady, H. E. (1995). *Voice and equality: Civic voluntarism in American politics*. Harvard University Press.

Visser, P. S. and Mirabile, R. R. (2004). Attitudes in the social context: The impact of social network composition on individual-level attitude strength. *Journal of Personality and Social Psychology*, 87(6):779. Publisher: American Psychological Association.

Von Sikorski, C. and Hänelt, M. (2016). Scandal 2.0: How valenced reader comments affect recipients' perception of scandalized individuals and the journalistic quality of online news. *Journalism & Mass Communication Quarterly*, 93(3):551–571. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151. Publisher: American Association for the Advancement of Science.

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, 69:102199. Publisher: Elsevier.

Wallach, O. (2021). The World's Tech Giants, Compared to the Size of Economies. *Visual Capitalist.* 7th July, 2021. Available at: `https://www.visualcapitalist.com/the-tech-giants-worth-compared-economies-countries/`.

Webster, S. W. and Abramowitz, A. I. (2017). The ideological foundations of affective polarization in the US electorate. *American Politics Research*, 45(4):621–647. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Weeks, B. E. (2015). Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication*, 65(4):699–719.

Weismueller, J., Harrigan, P., Coussement, K., and Tessitore, T. (2022). What makes people share political content on social media? the role of emotion, authority and ideology. *Computers in Human Behavior*, 129:107150.

Westen, D. (2007). *The Political Brain: The Role of Emotion in Deciding the Fate of the Nation*. New York: Public Affairs.

Widmann, T. (2021). How emotional are populists really? Factors explaining emotional appeals in the communication of political parties. *Political Psychology*, 42(1):163–181.

Winfield, B. H. (1994). *FDR and the News Media*. Columbia University Press.

Wojcik, S. and Hughes, A. (2019). Sizing Up Twitter Users. *Pew Research Center*. Publisher: Pew Research.

Woong Yun, G. and Park, S.-Y. (2011). Selective posting: Willingness to post a message online. *Journal of Computer-Mediated Communication*, 16(2):201–227. Publisher: Oxford University Press Oxford, UK.

World Medical Association (2022). Declaration of Helsinki - Ethical principles for medical research involving human subjects. *World Medical Association*. Accessed 12th July, 2024. Available at: `https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-h`

Wu, T.-Y. and Atkin, D. J. (2018). 'To Comment or not to Comment: Examining the influences of anonymity and social support on one's willingness to express in online news discussions'. *New Media & Society*, 20(12):4512–4532. Publisher: Sage Publications Sage UK: London, England.

Yarchi, M., Baden, C., and Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, 38(1-2):98–139.

Yasseri, T., Margetts, H., John, P., and Hale, S. (2016). *Political turbulence: How social media shape collective action*. Princeton University Press.

YouGov (2023). How Brits get their news. *YouGov*. https://yougov.co.uk/topics/politics/trackers/how-brits-get-their-news.

Young, L. and Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, 29(2):205–231. Publisher: Taylor & Francis.

Zhou, S., Kan, P., Huang, Q., and Silbernagel, J. (2021). A guided latent dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura. *Journal of Information Science*, page 01655515211007724. Publisher: SAGE Publications Sage UK: London, England.

Zolberg, A. R. (2012). Why not the whole world? Ethical dilemmas of immigration policy. *American Behavioral Scientist*, 56(9):1204–1222. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Zuckerberg, M. (2021). A Privacy-Focused Vision for Social Networking. *Facebook Notes*. 12th March, 2021. Available at: https://www.facebook.com/notes/2420600258234172/.

Zuckerberg, M. (2025). It's time to get back to our roots around free expression. We're replacing fact checkers with Community Notes, simplifying our policies and focusing on reducing mistakes. Looking forward to this next chapter. *Facebook video*. 7th January, 2025. Available at: https://www.facebook.com/watch/?v=1525382954801931.