The Psychology of Public Reactions to Political Communication

Paula Szewach

A thesis presented for the degree of Doctor of Philosophy

University of Essex

Department of Government

September 2022

Abstract

The technological changes and political events of recent years have generated much debate about political communication and where it leaves the health of democratic politics. There are concerns about whether politicians are telling citizens the truth or distracting them from it, and about whether citizens can handle the cognitive and emotional complexity of the political world that they face. Policymakers have sought to restrain the spread of so-called fake news, to buffer the effects of online incivility, and to boost citizen compliance amid the anxiety and uncertainty of the pandemic. In this dissertation, I argue that their limited success is due in part to a limited and sometimes crude understanding of public reactions to political information and communication. Key to this is the role of emotion. As psychologists have long established, emotion is ubiquitous in human reasoning and often unconscious feelings shape attitudes and decision-making. At the same time, this interplays with cognition rather than overriding it in the way that those cruder models of a 'post-truth' public suggest. I examine this interplay in a series of papers, exploring public reactions to political communication in different contexts and using different methodological strategies. This dissertation contributes to existing research in three ways: by producing new observational and experimental datasets; by piloting innovations in experimental design; and by generating insights useful to both academic and policy audiences. At the heart of several of these is the concept of incivility. I demonstrate how breaking the norms of political communication helps challenger parties to propagate their messages and the media to propagate disinformation.

Acknowledgements

My PhD experience was an unusual one. I started my PhD path with the conviction that I did not want to pursue an academic career. And it seems like I will be keeping my promise in the end. However, that does not mean that there were not times where I questioned my decision, especially because of the people I came across. Finding people like Rob Johns and Christine Stedtnitz made the odd world of academia easier to navigate. Rob was the best supervisor I could have asked for, and he knows I am not the kind of person who would just say this if I did not fully mean it. If I was not so eager to move on to the next thing I would be sad not to have an excuse to have the regular meetings where we ramble way beyond the content of the papers in this thesis. Our conversations throughout the years have been insightful, stimulating and reassuring. Rob's advice was always in my best interest and gave me full agency to make the final call. I must admit, he has been extremely patient, putting up with my never-ending questions and last-minute changes of direction.

This dissertation would not have been possible without Christine's friendship, moral support and technical help in desperate times. Doing research with her made the process a lot more fun, and so, more meaningful the end result. I am also grateful to the people at Essex who made my PhD experience better: Malte Jauch, Jimena Vazquez, Laura Saavedra-Lux, Johanna Amaya Panche, Kostis Roussos, Anam Kuraishi, David Liao, Marina Petrova and Phil Swatton.

Ernesto Calvo and Carol Galais supported me from the very beginning of this project. I hope I am always as generous with others as they were with me and everyone who knocked at their doors. And speaking of generous people, I would also like to thank Jason Reifler, who set the bar for future bosses very high. Especially, Jason gave me a ton of opportunities and was understanding when I chose which ones to take.

Above all, I would like to thank everyone in my very large family, my partner and friends. This accomplishment only makes sense because I can share it with them. A big thank you goes to my parents Viviana Schulzinger and Enrique Szewach, to whom I failed to explain, over and over again what I was actually studying. Even so, they supported all of my decisions, including those that they did not quite get. I am especially grateful to Javier Fernández Álvarez, with whom I shared every step of the way and who is largely responsible that I completed my PhD, but most importantly, that I enjoyed the process. Other much deserved words of gratitude go to my gifted set of parents, Hector Fernández Álvarez and Diana Kirszman.

My friends Paula Fiorini, Juan Mamberti, María Paula García, Aleksandra Tonelli and Florencia Kanovich have celebrated every little accomplishment throughout the process and were there for me whenever there were bumps in the road. I would not want to miss the opportunity to thank Inés, who just got into this World and already taught us so much. Last but not least, thank you to Josefa Isabel Rodriguez, who cannot stand the fact that University of Essex is so far away from Buenos Aires and looks after me at a distance. I dedicate this thesis to my grandma Blume, who marked my relationship with knowledge.

Contents

List of Figures

Introduction

It is already commonplace to introduce political communication studies reinforcing the idea that present-day concerns in the field are probably just the same old problems with a new suit. Well before the existence of digital media and social media platforms, around 2000 years ago Octavian spread false information against his political rival, Mark Anthony in order to get public backing to win a war. Around the 1700s, the introduction of the printing press helped rebels to spread rumours about King George II being ill to portray him as a weak figure. Later on, political propaganda became salient in the United States during the World Wars I and II (Krause et al., 2019). And these are just a few examples. However, it is true that some features of political communication in the present times are unique and distinct. Over the last few decades, old issues in political communication have been renewed by contemporary uses of technology. Accordingly, some new questions arise and some of the long-standing research questions in the field will have new answers.

In today's media landscape citizens are exposed to an unprecedented volume of information that is in turn, continuously growing. Not only is the amount of information growing, so too are the sources that people can choose for consuming information. And this growth presents both peril and promise for people's ability to gather and hold the information that they need to be competent and effective members of a democratic society. There is enough information out there, and even enough accurate information out there, but citizens were not particularly good at getting it before and now there is also easier access to what would be considered bad-quality information. Certainly, a branch of the literature in political science has already challenged the normative conviction of many that democracy needs knowledgeable citizens to function. In fact, Zaller's

conception of the "rationally ignorant median voter", suggests that not only there is no need for citizens to be knowledgeable, but also that on average it is rational not to be (Achen & Bartels, 2017). Even if empirical studies have contested the idea that democracy needs of informed citizens in order to work as a system, Delli Carpini and Keeter 1996 stress the idea that (good-quality) information constitutes a particular source of power, in the sense that it provides tools for pursuing private interests but simultaneously contribute to public interest in a democracy. After more than twenty years, this claim has not become obsolete. If anything, it is more fitting than ever in the context of the radical changes on our information environment (Carpini, 2019). Being able to sift through the news overload one is exposed to everyday and take away relevant knowledge to inform one's decisions is both a challenging task and an asset not many can rely on. And although it might not be fundamental to democracy to work as a whole, well-informed citizens have desirable consequences for democratic stability. Research has shown that more knowledgeable citizens hold consistent political views (Galston, 2001) and participate more in politics (Althaus, 2003). They are better prepared to identify politically relevant issues and find opportunities and connect them to their political needs (de Zúñiga et al., 2021). Consequently, if people's rapport to news plays a role shaping their democratic attitudes and behaviour, a proper understanding of the mechanisms behind citizen's reactions to different sources of information in political communication is key for the field of political science, but most importantly for governments, politicians and citizens themselves.

Using a broad brush, the main aim of the field of political communication since its origins has been "making sense of symbolic exchanges about the shared exercise of power" (Jamieson & Kenski, 2014). However, the discipline developed in the context of media structures shaped by the rise of radio and film and with the need to find answers to questions on the effects of World War I and Nazi propaganda (Schramm, 1983). So, naturally even if the guiding underlying principle remains the same, media structures have considerably changed, and so have the resources available to study them(Jamieson & Kenski, 2014).

Two major issues have marked these recent period and both what the field prioritises studying and how. The first one, unexpected and circumstantial: a worldwide pandemic that has basically sieved through every aspect of our daily lives. During these times, the right approach to political communication became essential, potentially marking the difference in terms of human lives saved,

saturation of intensive care units and vaccine take up. Governments have had to find an appropriate balance between disclosing too much and too little information, risking overwhelming citizens or fueling distrust.

The second one is related to a shift on the study of political communication at the contextual level and the individual level. Political communication studies involve relations among publics, media, and political institutions. At the contextual level, the interplay between politics and the media is undergoing a complex process of transformation characterized by the vast array of aspects of our lives that can now happen online. These recent transformations among publics, media, and political institutions have challenged the validity of traditional concepts in the field and have even been described as the transition to a fourth age of political communication (Blumler, 2016; Davis, 2019; Bennett & Pfetsch, 2018). At this new age of political communication, core institutions of democracy such as mainstream parties and traditional media outlets are losing their all-mighty gate-keeping powers. Of course, this does not mean that all voices weigh the same, but rather that there is more space for new actors to emerge and represent a greater diversity of ideas (Bennett & Pfetsch, 2018). In turn, this also implies that new dynamics of exchange between institutional actors and the mass publics are shaped on the online sphere. Consequently, new political actors who operate as challengers to the mainstream face an opportunity to profit from the architecture of social media platforms, where the battle for citizen's attention is not tied to rules that favor incumbents by default. Instead, the attention goes to those who can capture it.

This brings us to the shift on political communication research at the individual level. Recent developments in the field of information processing became the corpus of evidence to support a paradigm shift in mass communication studies (for a detailed review see Lang et al. 2009; Potter & Bolls 2012). One of the main conceptual novelties proposed by this paradigm is the incorporation of emotions as a constitutive element of cognition. According to this perspective, emotion and cognition are strongly integrated mental processes with distinct patterns of measurable physiological activity (Lang et al., 2009; Potter & Bolls, 2012). Contrary to the classic stimuli-response approach, influenced by the behavioral tradition, the information processing approach understands cognitive processes as mediated by psycho-physiological factors (Lang et al., 2009).

Lodge and Taber 2007 open their seminal paper explaining how citizens form and express their political beliefs with the phrase: "We are witnessing a revolution in thinking about thinking".

The revolution they are talking about refers to the growing evidence in cognitive sciences that unconscious thoughts and feelings have a powerful effects on people's preferences, attitudes and even decision-making (for a detailed review see Lang et al. (2009); Potter & Bolls (2012). In contrast with the rational-choice models that primed until that moment, unconscious processes where put at the centre of this new way of understanding how people think. Therefore, challenging not only the rationale behind political science as a whole and political communication studies in particular, but also the way in which the field measured and interpreted the relationship between beliefs, attitudes and behaviour.

While emotion had not been completely disregarded before, it is now considered as inherent to information processing and decision-making (Redlawsk, 2006). The adoption of this model of cognition, in which thinking and feeling are basically two elements of the same process, focuses on understanding what goes on within the so-called "black-box" of the mind during media consumption. Emotion and cognition are considered as strongly integrated mental processes with distinct patterns of measurable physiological activity (Damasio, 1994; Potter & Bolls, 2012). Furthermore, this approach understands cognitive processes as mediated by psychophysiological factors (Lang et al., 2009).

The theoretical and empirical advances in the field pose the need apply this framework when studying political communication. Emotions are intimately related to some of the main objectives of media messages, such as attracting attention, being remembered, entertaining and persuading (Ravaja, 2004). Measuring emotional responses to political stimuli accurately is extremely difficult, costly and time-consuming and that might be also the reason why the domain is still under-explored (Redlawsk, 2006). However, given that emotion is a constitutive element of information processing, there might be relevant insights stemming from this line of research to help solve the contemporary issue of so-called 'fake news'.

Broadly speaking, this dissertation provides insights bearing on three of the major current debates in political communication outlined above.

Communicating Uncertainty

The COVID-19 pandemic put political communication under the spotlight. In a context of high uncertainty, Governments across the world had the responsibility of both providing citizens with

accurate information, disclosing potential risks, and at the same making sure that the information was clear enough not to dissuade compliance with public health policies. A clear example of this was the roll-out of the COVID-19 vaccination programmes. There are good arguments on the side of full disclosure of uncertainty and also on a less nuanced message that only provides clear information on what is known up to the present moment. The full disclosure approach or 'radical transparency' - as the scientific community has called it (Nature, 2020), would allegedly reinforce trust. On the other hand, there is the idea that non-specialist audiences might misinterpret the indicators of uncertainty, ending up with a distorted understanding of the intended message. Indeed, this idea is well founded: emotions are a major determinant in risk perception (Slovic, 2010). There is a large body of research suggesting that emotions play a relevant role in shaping people's perception particularly in situations of high risks and high stakes (see Balog-Way et al. 2020 for a review). Individuals might develop inaccurate perceptions of risk, overconfidence in their ability to avoid harm, be reluctant to make tradeoffs or unrealistic demands for scientific certainty (Covello et al., 1986). By this reasoning, a simpler message without references to the standard measures of uncertainty in science would actually be clearer for citizens. So, the issue is not about whether it is convenient to manipulate citizens by hiding information from them, but rather what sort of communication will be the most effective getting the message across. The empirical evidence on the topic is currently mixed. For decades, researchers have assessed the effect of communicating risks and uncertainty on attitudes and behaviour (Covello et al., 1986). Overall, when it comes to trust on the source, they found that transparency can increase trust (Van der Bles et al., 2019), but null findings are also frequent (Johnson & Slovic, 1995; Lofstedt & Bouder, 2021). When the focus shifts to trust on the claims on the message per se, the findings are even less clear. Van der Bles et al. 2020 report a series of studies in which belief in the reported numbers was significantly lower when uncertainty was acknowledged, whereas Daoust and Bastien 2021 found no effect of depicting confidence intervals around COVID-19 death projections on the perceived reliability of the estimates. When we go beyond attitudes and beliefs and turn towards the effects on compliance, there is considerable research suggesting that whenever there is uncertainty around information people are more prone to postponing or avoiding decisions based on that information. However, the relationship between acknowledging uncertainty and compliance is under-explored (Van der Bles et al., 2019), although there is some evidence that it could weaken the strength of

recommendations or reduce adherence to laws (Markon et al., 2013).

Uncivil Party Communication through Social Media Platforms

Social media platforms have become a key communication channel for political actors to address the mass public (Elishar-Malka et al., 2020; Stier et al., 2018). Platforms like Twitter and Facebook have put online deliberation between otherwise inapproachable politicians and citizens within the reach of a click. Not to mention, they have also facilitated agenda setting powers to organised crowds, for instance through collective hashtags. However, this new possibility offered by Internet–based political deliberation has not resulted in the panacea for democratization that many envisioned. Some scholars have raised concerns about the increased opportunities for surveillance (Howard et al., 2006), further deepening of fragmentation and polarization (Prior, 2007) or even "normalization", by which "the same hierarchies, power dynamics, and disparities seen in political life offline inevitably will be reproduced online" (Margolis & Resnick, 2000).

Be it because of social media, with the help of social media or despite social media, the emergence of social media platforms was parallel to a period of decline of mainstream parties and rise of political outsiders all along the political spectrum (Hernández & Kriesi, 2016; Hobolt & Tilley, 2016; Emanuele & Chiaramonte, 2019). These parties that have disrupted party systems in Western Europe in recent years have been named with many different labels, such as 'niche parties' (Adams et al., 2006; Jensen & Spoon, 2010; Meguid, 2005), 'populist parties' (Kriesi, 2014; Mudde, 2007; Pauwels, 2014) and 'challenger parties' (Hino, 2012; Van de Wardt et al., 2014; Van de Wardt, 2015). Throughout this dissertation I will rely on the term 'challenger parties' as an encompassing label suggested by de Vries and Hobolt 2020 emphasizing the relative position of parties within the political system. Mainstream parties are those in control of the political arena (or those who have been in the past) and challenger parties do not have previous experience controlling policy or government positions at a large-scale. Challenger parties found a competitive advantage when communicating through social media because communication through these platforms amplified all sorts of content. Meaning that the messages that advocate for or are in line with democracy get further spread that they would have offline, but also those of extremist, xenophobe, sexist nature that would otherwise had a much more difficult journey to reach an audience.

The companies behind social media platforms have allegedly put much effort into their content

moderation policies in order to buffer the amount of incivility they allow into their ecosystem, but even if they did, their purported efforts have fallen short to say the least. More sceptical critics point out that social media companies do not have any incentives to keep hate off their platforms (Hindman, 2018). Although it is subject to doubt, maybe they do a better job with some very extreme cases, but there is a large scope of incivility to be covered. Despite the fact that the majority of posts are within the realm of politeness, uncivil posts are not hard to come across. Even if the social media companies become truer to their word about banning extreme cases of hate speech and related sorts of posts, there would still be the live issue of all the content that is in-between, and hence, the research question of whether incivility on social media allows certain players to thrive.

In order to sustain profits these companies need to prioritize features that increase user engagement. Throughout the last decade there has been a lot of controversy regarding the role of these companies in democratic societies. Especially regarding whether for-profit private companies should be the ones defining what kind of information citizens are exposed to or the only ones able to audit their own algorithms, without even Institutional Review Board (IRB) oversight (Fiske, 2022). And since the corporations behind Facebook and Twitter do not disclose the algorithms they rely on, we do not know exactly how the algorithms that curate our feeds work. However, we do know that the algorithms weigh engagement very strongly. That is, all kinds of attention given to a post, such as likes, retweets and quotes make the algorithms more likely to recommend it to others. So, posts that get extraordinary attention are more likely to be spread. Mainstream parties can rely on broadcasting their policy success to attract voters. Conversely, challenger parties do not have credentials to show, but they do not have much to lose either. This enables challenger parties to rely on more innovative strategies, including openly discrediting those in office in a tone that would not be tolerated elsewhere.

Consequently, new political actors who operate as challengers to the mainstream face an opportunity to profit from the architecture of social media platforms, where the battle for citizen's attention is not tied to rules that favor incumbents by default. Instead, the attention goes to those who can capture it. Uncivil communication strategies seem to effectively capture people's attention, currency of exchange both in contemporary societies and in social media in particular. That is the reason why social media platforms have rekindled the debate on the consequences of

uncivil behaviour for democracy. The term 'uncivil' makes the concept sound like a normative concept detached from a measurable empirical perspective, or even a vacuous moral judgement of some sort. However, this is not the case. Uncivil language, uncivil appeals or incivility are the terms used in the literature to define forms of expression that violate the norms of politeness of a given context Mutz (2015). And although it may still seem like a concept that is hard to tap into in empirical research data has consistently shown that a majority of people detect and classify the same repertoire of features as uncivil Sydnor (2018). In fact, that is basically the principle through which society functions: on average most people can tell right from wrong or at least what would be generally considered as admissible and inadmissible in different circumstances. Even if that does not translate into their actions. And in the case of party communication, even if some parties still choose to rely on uncivil appeals as a communication strategy that pushes the implicit limits of legitimate democratic competition. This dissertation analyses whether challenger parties are more prone to using uncivil appeals than mainstream parties and if they benefit more than other parties from this communication strategy.

The Indirect Consequences of Uncivil Language on Social Transmission of Disinformation

There are many ways in which news can deviate from the ideal standard. In an effort to comprehensively examine the information disorder an impactful report published by the Council of Europe suggests that on one hand, there are deviations related to the factual accuracy of the news and on the other hand, there are deviations related to the intentions of the agents who create and distribute the news (Wardle & Derakhshan, 2017). As ideal types in a taxonomy, news can carry false information due to poor fact-checking, portray true claims with sensational language or tell a completely fabricated story disregarding any journalistic etiquette. Despite the scholarly need to rely on clear categories and definitions in order to better understand the topic of research, there are very little cases where journalistic articles are only made out of facts representing an objective unambiguous truth. Throughout this dissertation, the term 'fake news' will be defined following Krause et al. (2019), as news that deviates from those journalistic conventions that have governed the presentation of truth claims as news both in content and style.

A myriad of proposed solutions to the problem of 'fake news' enjoyed limited success (sometimes

complete failure or even backfire effects), such as providing citizens with factual information, labeling false or low quality information online or media literacy campaigns. However, most of these solutions disregard the central role of emotion in information processing. This dissertation looks into the emotional underpinnings of the spread of fake news by testing the mechanism through which online political disinformation is shared to a greater extent than other kinds of online news.

Outline of this Dissertation

This dissertation explores public reactions to political communication throughout different contexts and relying on different methodological strategies. The focus is set on particular issues that have been at the center of the discussion in the field, such as how to handle uncertainty in times of crisis, whether certain political actors profit more than others in the contemporary media landscape or what makes disinformation more likely to be shared than other types of news. Specifically, Chapter One assesses the effect of communicating uncertainty on public attitudes and behaviour in a context where stakes were high. Through an online survey experiment (n=2167) fielded in Britain in December 2020, I found that communicating uncertainty about COVID-19 vaccines to non-specialised audiences did not undermine belief or compliance and in turn, improved long term trust. Chapter Two addresses party communication. It discusses the paradox by which a political communication style that is disruptive for democratic societies is successful to engage voters in currently democratic regimes. The study described throughout this chapter applied supervised learning techniques on Twitter data collected from the timelines of all parliamentary representatives in Spain. Our results suggest that challenger parties are generally more likely to use uncivil language on their social media than mainstream parties and that challenger parties seem to only have an incentive to be more uncivil than their mainstream competitor. Chapter Three focuses on understanding the role of emotion while reading political information, misinformation and disinformation. The aim of this paper is to test the mechanism through which political disinformation generates higher levels of engagement and spreads faster than other kinds of news. Our study found that incivility, a feature commonly correlated with disinformation, is a consistent predictor of social transmission of news. Overall, this dissertation applies a theoretical framework contemplating that unconscious thoughts and feelings shape judgment, preferences, attitude change, and decision-making (Lodge & Taber, 2007). Briefly, I find that correlates of emotion in political com-

munication messages have an effect in citizens behavioral intentions in three different contexts and relying on different methodologies.

Should governments communicate uncertainty? Yes, because the public can take it.

2.1 Abstract

COVID-19 created problems for political elites accustomed to making confident predictions and downplaying uncertainty about outcomes. An online survey experiment (n=2167) fielded in Britain in December 2020 indicates that governments are well advised to communicate uncertainty, even (or especially) when stakes are high. The costs are low: reporting a confidence interval around the safety and effectiveness of a hypothetical COVID-19 vaccine did not dent trust in the vaccine or the statistics. And there are benefits: when outcomes turned out to be worse than expected but within that confidence interval, trust in the vaccine was partly insulated from negative effects.

2.2 Introduction

There are growing calls in the scientific community for "radical transparency" (Manski, 2019; Nature, 2020). There is anything but radical transparency, however, in the communication of uncertainty – especially when policymakers are addressing non-specialist audiences. This is partly because the language of uncertainty – probabilities, margins of error, and so on – adds more com-

plexity to what may already be hard to communicate to the general public. It is also because admissions of uncertainty risk eroding public trust in both the information and those communicating it. "Why should we follow rules based on predictions that might be wrong?" is the mass reasoning that political elites fear. Yet the success of policies in areas as diverse as climate change, the economy, or health depend on the compliance of citizens. The COVID-19 vaccination programme was a paradigm case (Mahase, 2020). As the first vaccines emerged, politicians across the world had to persuade people to take them based on limited information about the safety and effectiveness of those vaccines. They might reasonably have feared that acknowledging any sort of uncertainty, especially around potentially lethal side effects, would exacerbate any public sense that the vaccines were untested or risky – with potentially disastrous consequences for vaccine roll-outs. On the other hand, those communicating information were shooting at a moving target. If their initial estimates proved over-optimistic and they had failed to acknowledge the uncertainty, this risked further scepticism or distrust down the road. So, when governments introduce policies based on uncertain scientific evidence, how transparent can and should they be about that uncertainty? What is the best communication strategy? The answer depends on estimating the costs and benefits of communicating uncertainty. There is somewhat mixed evidence on both. On the costs side, the good news comes from a small but growing batch of studies indicating that reporting uncertainty does little damage to trust in the sources of that information or in science more generally (Van Der Bles et al., 2020; Kreps & Kriner, 2020; Kelp et al., 2022). The picture blurs, however, when we move from where the information comes from to what people might do with it. Uncertainty around statistical claims may undermine belief in those claims and subsequent compliance with policy based on them (e.g. Morton et al. (2011); Van der Bles et al. (2019)). Of particular relevance here are the contrasting findings about the effect of reporting uncertainty on attitudes towards COVID vaccines (Kelp et al., 2022; Petersen et al., 2021). The way that uncertainty is reported looks important in minimising costs of this kind. As for the benefits of reporting uncertainty, they may come down the line. Communicators who acknowledge that things may go worse than expected may be more insulated from blame, and retain more power to persuade, than those whose confident predictions go awry. However, the scale and scope of those potential benefits has been little explored. This article reports on a two-stage survey experiment exploring in turn the costs and potential benefits of communicating uncertainty. The experiment is based

on the UK's vaccination programme and was embedded in a survey fielded in December 2020, at the very beginning of that programme. Our treatments manipulated the information available about the effectiveness or safety of a hypothetical COVID-19 vaccine and the primary outcome of interest was respondents' willingness to get that vaccine. At Stage 1, we investigated the cost side, i.e. whether communicating uncertainty around those effectiveness and safety projections undermined vaccine intentions – and, if so, whether that effect could be mitigated by explaining why such uncertainty was inevitable. At Stage 2, we investigated the benefits side by providing some respondents with an update based on new vaccine data. The updated results were worse than the initial point estimates but within the confidence intervals around them. The experiment produces two encouraging results. First, reporting uncertainty around estimates of the effectiveness and safety of a COVID-19 vaccine did not significantly dent confidence in those initial reports or vaccine intentions. Second, those provided with uncertainty information reacted less negatively – in terms of perceptions of the vaccine as well as intentions to have it – than those who had only seen the over-optimistic point estimate. Overall, the findings point clearly to a positive answer to the question posed in our title.

2.3 Communicating Uncertainty to the Public

For some time now, researchers have assessed the consequences of communicating uncertainty on public attitudes and behaviour. Simplifying a nuanced picture, and drawing heavily and gratefully on a recent review by van der Bles et al. 2019, we might summarise what we know – and what we do not yet know – in four points. The first three relate primarily to the potential costs of reporting uncertainty; the fourth is about the potential benefits. First, reporting uncertainty generally carries little cost when it comes to trust in and credibility of those producing and communicating the information. Indeed, transparency may even increase trust, although null findings in this area are also common (Johnson & Slovic, 1995; Lofstedt & Bouder, 2021). Second, moving from trust in the messenger to trust in the claims themselves, the picture is more mixed. For example, van der Bles et al. 2020 report a series of studies in which belief in the reported numbers was significantly (though not greatly) lower when uncertainty was acknowledged, whereas Daoust and Bastien 2021 found no effect of depicting confidence intervals around COVID-19 death projections on the perceived reliability of the estimates. Third, moving along the causal chain, the cost of uncertainty in terms

of behaviour and compliance is similarly unclear. Indeed, van der Bles et al. 2019 in their review explicitly highlight this as an area needing more systematic research. What we do have is plenty of evidence that uncertainty around information prompts people to postpone or avoid decisions based on that information, and scattered pieces of evidence that it can undermine compliance with relevant laws or recommendations (e.g. Markon et al. 2013). Fourth, our understanding of the benefits of reporting uncertainty is confined largely to the point made earlier about its potential to improve trust in the reporters and their transparency. This neglects another merit of acknowledging uncertainty: that is, as an insurance policy against criticism when initial estimates turn out to be too optimistic. Communicators who admit that, while they expect a certain outcome, things could end up worse should face a (slightly) more forgiving public if and when that happens. This is an important point: after all, scientific forecasts and the uncertainty around them are about a future that will emerge – and often in the public glare. However, because it is hard to test with the 'one-shot' studies that have dominated in research on communicating uncertainty, evidence on this 'insurance policy' benefit is very scarce. What we do know, courtesy for example of a one-shot study by Kreps and Kriner 2020 using COVID-19 death tolls, is that people react badly to a treatment reporting that a given outcome is worse than was previously estimated. What we do not yet know is whether, had that previous estimate come with a confidence interval that encompassed that outcome, people would be at least somewhat mollified. This would be best tested by a two-stage design, in which uncertainty is reported at the first stage and then the eventual data are reported at the second. COVID-19 and specifically the mass rollout of vaccinations provided an excellent opportunity to study both the costs and benefits of reporting uncertainty. This was an unusually high-profile case of a government policy requiring mass compliance but based on less than perfect information and requiring shrewd communication (Larson, 2020; Motta, 2021; Dudley et al., 2021). There was an especially pressing need to understand whether and how far being transparent about the uncertainties around the new vaccines would impact take-up (Motta, 2021; Petersen et al., 2021). Recent literature on COVID-19 vaccine hesitancy suggests that patients appreciate being informed about the risks of vaccines (Schwartz, 2020) and the DELVE Initiative 2020 recommended that 'clear, transparent communication' be used to address 'rational doubts and to enable informed decision-making', and not to hide the potential limitations of vaccines (see also Bavel et al. 2020). In practice, such transparency was neither radical nor even especially

common. While information about uncertainty around the vital statistics of the various vaccines was provided by manufacturers and thus available to those who sought it, it was not routinely provided in government communications or media reports. For example, effectiveness percentages were regularly cited in the UK Government's press briefings but very rarely accompanied by the kinds of confidence intervals that were provided as standard with, for example, the estimates of COVID-19 infection rates by age group (Prime Minister's Office, 2020). Of course, there are various reasons for reporting only point estimates without confidence intervals. For present purposes, the question is whether reporting such uncertainty would have undermined vaccine willingness in the short term – or helped to sustain it in the medium term had the point estimates proved too optimistic. There is no evidence as yet on the second point about the benefits of uncertainty, and only indirect indications on the first about the costs. The latter comes from two experiments similar to the first stage of our own. Petersen et al. 2021 varied the level of transparency in descriptions of a hypothetical COVID-19 vaccine (which is the standard approach given that, due to obvious ethical concerns, medical RCTs cannot test for communication effects on vaccine uptake). One version of the description included vague information about the vaccine; another two variants included transparent information. The experimental vignettes described three aspects of the vaccines: effectiveness, side effects and the duration of clinical trials. They found that transparently communicating even negative features of the COVID-19 vaccines (less effective, less safe and shorter test period than the 'flu vaccine) was barely more damaging to vaccine willingness than was vaguely positive information. However, since their manipulation was of transparency rather than uncertainty, these relatively upbeat results may not generalise to our case. There is a more explicit focus on uncertainty in a similar study by Kelp et al. 2022. However, their more downbeat conclusion – that a high-uncertainty condition erodes vaccine acceptance, especially among the initially reluctant – could reflect confounds between uncertainty and vagueness or positivity. While their low-uncertainty condition reported specific and extremely high effectiveness rates, the high-uncertainty condition reported no numbers and generally resembled the vagueness condition that Petersen et al. 2021 also found to be unpopular. What is needed is a study in which only uncertainty is manipulated: that is, the same specific point estimates are reported but, in the treatment group, accompanied by measures of uncertainty.

2.4 The current study

This is one of the gaps that our study fills. Following these examples, we investigate responses to a hypothetical covid vaccine. Following Motta's 2021 finding that safety and effectiveness were two principal drivers of vaccine willingness, we make those the basis for our manipulations. In contrast to previous work, we disentangle the two, exposing respondents to either safety or effectiveness statistics. The point estimates on these two criteria are held constant; what is manipulated is whether they are accompanied by confidence intervals. Before setting out our hypotheses, we should note a couple of points about our confidence intervals approach to reporting uncertainty. Both derive from its relative precision (cf. the vaguer verbal approach of Kelp et al. 2022, for example): that is, the ranges presented are numerical estimates. One implication of this is psychological. The tendency in the public towards negativity bias and risk aversion (Soroka, 2014) means that there is likely to be particular focus on the pessimistic end of the confidence interval. If that is given a numerical expression, particularly one that is not a huge distance from the point estimate, that gives worried citizens an unusually firm foothold and might make uncertainty easier to cope with – thus reducing the costs of reporting it. Another has more to do with cognitive capacity. The precise meaning of confidence intervals is relatively complicated, needing at least a basic understanding of probability distributions and inferential statistics. This might lead some to doubt whether they are a helpful means of conveying uncertainty. But there is a much easier means of understanding them – as the range between the worst and best case scenarios – which, while not correct, serves as a reasonable heuristic. Hence Daoust and Bastien 2021 make a convincing theoretical case for them helping people to grasp uncertainty – and, as noted above, a convincing empirical case that they do nothing to undermine faith in the reliability of the statistics. One thing is to understand the method of presenting uncertainty; another is to understand why uncertainty is there in the first place. The acceptance and processing of a confidence interval may be readier among those who know – or learn – why, especially in the early days of clinical trials based sometimes on restricted samples, the rates of effectiveness and the risk of side-effects can be estimated only imprecisely. To test this conjecture, another innovation in our experiment is that there are two uncertainty treatments: the basic version, simply reporting confidence intervals, and the expanded version in which there is also an explanation of the reasons why an exact esti-

mate makes less sense than a range. As already noted, existing research points to heartening null findings when it comes to the effect of reporting uncertainty intervals on trust in information and its source. Ours is a new and acid test given both the unusual public salience of COVID-19 and the extension from trust in sources to planned behaviour – here, vaccination intentions. Nonetheless, that previous work gives grounds for optimistic hypotheses about the costs of reporting trust.

H1 (No costs Hypothesis). Providing a confidence interval around estimates of vaccine safety/effectiveness does not significantly weaken:

a.*respondents' willingness to have the vaccine.*

b. *respondents' perceptions of the vaccine's safety/effectiveness.*

The most novel feature of our study is the test for potential benefits of reporting uncertainty. These are likeliest in situations where estimates have to be adjusted downward. We have noted the lack of previous studies testing whether, if that adjustment goes below the initial estimate but remains within its confidence interval, the audience reaction to the bad news is dampened. However, there are good reasons to suppose that it is. For one thing, the negativity bias referred to earlier means that many are likely already to have anticipated a result towards the more pessimistic end of a range, and will have less Bayesian-style updating to do when the disappointing new data come in. For another, insofar as trust in a vaccine is mediated by trust in its manufacturer (Larson, 2020), this should be better sustained in a case where the manufacturer at least acknowledged, if not anticipated, the less positive results that emerged. Our hypothesis is therefore again optimistic, in this case about the capacity of reporting uncertainty to deliver benefits. As before, we test the same hypothesis in parallel for updates on safety and effectiveness, and test it for both vaccine intentions and perceptions of that safety or effectiveness.

H2 (Benefits Hypothesis). Respondents who had seen confidence intervals around estimates of safety/effectiveness show a smaller reduction in:

a. *willingness to have the vaccine.*

b. *perceived safety/effectiveness.*

2.5 Data and Methods

We tested our hypotheses via a two-stage experiment embedded in an online survey. The survey was fielded on 22 December 2020 on a sample of 2,167 UK residents recruited via the Prolific platform (see Peer et al. 2017. In December 2020, medical trials were ongoing and prominent in the media, risks were salient, and there were big question marks around the effectiveness and safety of the vaccines coming up for government approval. Hence, while the questions were about a hypothetical vaccine, the context made this a highly realistic exercise. The size of our sample gives the experiment considerable power to detect the costs and benefits of reporting uncertainty. Like other Prolific samples, this one is more diverse than many convenience samples but is still unrepresentative of the UK adult population in a number of (related) ways: younger (average age of 37), more female (62), more educated (57 with a degree), and more liberal (72 said they would vote to remain in the EU if there was a second referendum, compared to a percentage in the low 50s in the electorate at the time). Of these variables, however, only political ideology has an appreciable and consistent correlation with vaccine hesitancy (Dhanani & Franz, 2022; Reiter et al., 2020). Since our primary focus is on experimental comparisons and hence internal validity, however, the representativeness of the sample is not of paramount importance. It does call into question the external validity of the effect sizes recorded here, however, a point we return to in the Concluding section. Respondents reported a very high level of trust in the NHS (on average, 4.9 on a scale from 0 - None at all to 6 - A great deal), a fairly high level of trust in the companies making the vaccines (4 on the same scale), and, unsurprisingly given the higher-than-average death toll in the UK at the time (not to mention a range of other political factors), a low level of trust in the UK Government (2.4 on the 0 to 6 scale). 3 in 10 of our respondents believed they had probably had the virus; 6 in 10 described themselves or at least one family member as 'high risk'. Reflecting the enthusiastic response of the British public to the news of the first vaccines being administered to the elderly around the time our survey was fielded, our respondents reported high levels of interest in a COVID-19 vaccine. When we asked, at the beginning of the survey, whether respondents would have a vaccination 'if and when' one was offered to them, about 5 in 10 said they would definitely and a further 3 in 10 said they would probably have one. Only 1 in 10 said they would probably or definitely not have one. (The complete sample characteristics are reported

in appendix A.)

2.6 Experimental Design

The experiment was embedded in a survey that opened with some general questions about COVID-19, including general vaccination intentions, trust in key actors, concern about the pandemic and personal experience of the virus. Then came the two stages of the experiment, separated by a battery of socio-demographic items. These served partly as a buffer to make the information update a little more realistic, and the priming from Stage 1 a little lighter, than if the two stages had been adjacent.

The core survey experiment was a 4 x 2 design (summarised in Figure 2.1). The crucial treatment was uncertainty and had three levels: a simple point estimate not conveying any uncertainty ('point estimate'), a point estimate along with a range in which the true estimate is likely to fall ('confidence interval'), and a point estimate along with a range and an explanation why the point estimate could not be exact ('ci and explanation'). This was cross-manipulated with the two different domains of uncertainty, regarding the effectiveness and safety of the vaccine. There was also a pure control condition – that is, respondents asked about initial willingness to take the vaccine without receiving any data about safety or effectiveness. The structure of the experiment was in two stages. Respondents were first prompted to suppose they were offered a COVID-19 vaccine (the wording of this prompt and all of the experimental treatments is in Appendix). They were shown some information from the manufacturers about either how effective (in terms of percentage reduction in the transmission rate) or how safe (in terms of the probability of side effects) the (hypothetical) vaccine was estimated to be. Next, respondents were randomly assigned to one of the three uncertainty conditions described above. The uncertainty information was delivered numerically but not graphically; that is, we reported confidence intervals in the treatment texts rather than showing them on a graph. (With only one estimate and confidence interval to report, the verbal approach was economical while a graph might have looked like overkill (cf. Daoust & Bastien 2021) The explanation for uncertainty differed a little across the two tracks: with effectiveness, it focused on random sampling variance and on biases in the test sample. The safety treatment included parallel wording on sampling bias but also an acknowledgment that side-effects could emerge too slowly to be captured in the trials. In order to test the No Costs Hypothesis (H1)

	Safety track	Effectiveness track	Control
Control	-	-	Suppose you were offered the COVID-19 vaccine. (n=538)
Point estimate	Suppose you were offered the COVID-19 vaccine. Our tests show that … **2 in 10,000** will develop more serious side effects. The most common serious side effect is Bell's Palsy, a freezing of muscles on one side of the face. (n=273)	Suppose you were offered the COVID-19 vaccine. Our best estimate of the effectiveness of the vaccine is **90% effective**. What that means is that our vaccine reduces transmissions of the virus by 90%. (n=271)	-
Confidence interval	To be more exact, we estimate that **between 1 and 6 out of 10,000** people will experience serious side effects. (n=272)	To be more exact, our tests show that the true effectiveness level is very likely to fall **between 83% and 97%**. (n=269)	-
CI and explanation	We cannot be exact in these circumstances for two reasons. One reason is that we have only been able to follow our volunteers for a few weeks. It is possible that they will develop side effects later on. Another reason … is that those tested were not a perfect reflection of the UK population. (n=272)	We cannot be exact in these circumstances for two reasons. One reason is that… any results based on a sample of people come with a margin of error…. Another reason … is that those tested were not a perfect reflection of the UK population. (n=270)	-

Figure 2.1: Treatment groups at Stage 1

at this first stage, we asked respondents how likely they were to get this vaccine, how confident they were in their decision, and how safe or effective (depending on which track they were on) they felt it was. To test the Benefits Hypothesis (H2), we asked whether respondents felt that they had enough information to make a decision, and how open and transparent they felt the manufacturers were with information about the vaccine.

	Safety track	**Effectiveness track**
No update	-	-
Negative update	Suppose that the vaccine has been in use for a few months and the manufacturers update their results. Based on the data we have now collected, we calculate that the vaccine causes serious side effects … in **6 out of 10,000** cases.	Suppose that the vaccine has been in use for a few months and the manufacturers update their results. Based on the data we have now collected, we calculate that the vaccine is **83% effective**.

Figure 2.2: Treatment groups at Stage 2

The second stage was based on a putative update from the manufacturer on the vaccine's record after a few months of its use. Respondents received either safety or effectiveness information, depending on which track they were on. In each case, the updated results were presented as calculations rather than estimations and involved results that were at the negative limit of the previously-reported range (see Figure 2.2). There was no additional uncertainty manipulation at this stage; our interest was in the effect of the Stage 1 treatment on reactions to the new information. To test the Benefits Hypothesis (H2) we repeated the questions about the perceived safety or effectiveness of the vaccine, vaccination intention and confidence in that decision. There was also a manipulation check, asking respondents whether the update meant that the vaccine was more or less safe/effective than had been initially forecast.

2.7 Results

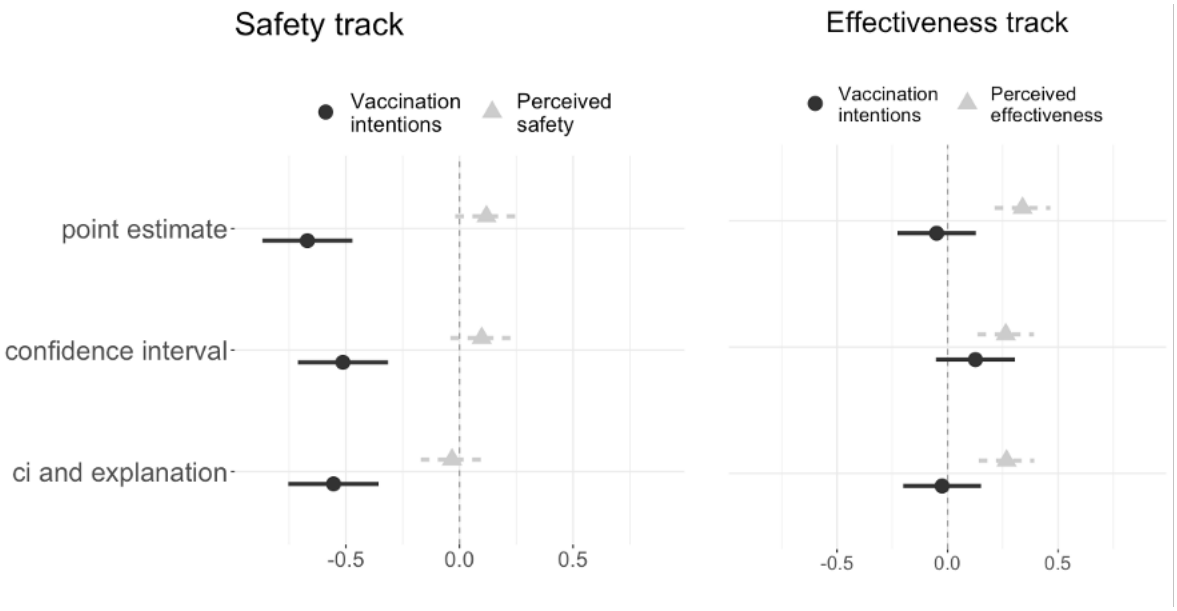


Figure 2.3: Effect of providing information on safety or effectiveness statistics about a hypothetical COVID-19 vaccine. OLS Regression models. Outcome variables are not standardised. Both outcome variables are measured on a seven-point scale where higher values indicate higher vaccination incentives, or higher perceived safety or effectiveness. The baseline is the group of respondents who saw no information about the hypothetical vaccine.

Since the key outcome variables are on numerical (0-6) scales, it is reasonable to test our hypotheses using OLS regression. To help iron out any pre-experimental differences in relevant predispositions across our treatment groups, these regressions include controls for trust in the NHS, trust in vaccine manufacturers, having contracted the virus, and being 'high risk' or having a 'high risk' family member. They also include a few other measures of perceived threat, including threat to respondents' financial situation, feelings about UK Prime Minister Boris Johnson, and two socio-demographic variables (age and gender). The full models, as well as additional ordered logit models, are shown in Appendix A. Before looking into the effect of acknowledging uncertainty around safety or effectiveness statistics we looked into the effect of saying anything at all about safety or effectiveness. The baseline in Figure 2.3 is the control group: respondents who were simply prompted to 'suppose you were offered the COVID-19 vaccine' and received no further information about it. The coefficient plots show the effect of providing information on safety or

effectiveness statistics on three outcome variables: vaccination intentions, and, depending on the track, perceptions of safety or effectiveness. Respondents who saw statistics about the safety of the hypothetical vaccine (left) were significantly less interested in the vaccine than the control group. The effect size was substantial: On a seven-point scale, providing information about vaccine safety reduced vaccination intentions by .5 points or more, depending on how much information was provided. This suggests that any airtime spent on safety considerations will lead people to think and, therefore, worry about safety. Interestingly, though, providing information about side-effects risk did not affect perceptions of safety. If anything, respondents who had read about safety rated the vaccine as slightly (though not significantly) safer compared to those who had seen no information at all. At first sight, this finding seems odd: Seeing safety statistics seems to nudge respondents to believe the vaccine is safe, but not to want to have it. This may reflect the difference between analysing and feeling risk (Slovic, 2010). When prompted by a survey question to analyse risk, respondents agree that the vaccine is safe and that side effects are extremely rare. Yet at an emotional level, even just the thought of a serious side effect (a facial paralysis, in this case), however unlikely it is, may scare people away. This confirms a long-standing body of literature on the difficulty we have in comparing risks (Slovic, 1986; Loewenstein et al., 2001). Respondents who saw statistics about the effectiveness of the hypothetical vaccine (right) were not any more or less interested in the vaccine than the control group. They did, however, think of it as more effective. We suspect that the boost in perceived effectiveness is due to the high level of effectiveness reported in the description of the vaccine (90%). This number reflected estimates of frontrunner vaccines at the time our survey was fielded in December 2020. These statistics may have driven up perceptions of effectiveness because respondents were simply unaware that the new COVID-19 vaccines were that effective – estimates of 90% effectiveness were potent drivers of vaccine support in Motta's 2021 experimental study in the US. But why did the boost in perceived effectiveness not go hand in hand with a boost in interest in the vaccine? It may be that there were ceiling effects given the generally high interest in the vaccine; it may also be that, at least in the UK, safety concerns outweigh effectiveness and all other considerations when citizens form vaccine intentions (Neumann-Böhme et al., 2020; Loomba et al., 2021; Altay et al., 2021).

Our primary concern, of course, is with comparing the three treatment groups in Figure 2.3. That comparison is made more prominent in Figure 2.4 which excludes the control group and,

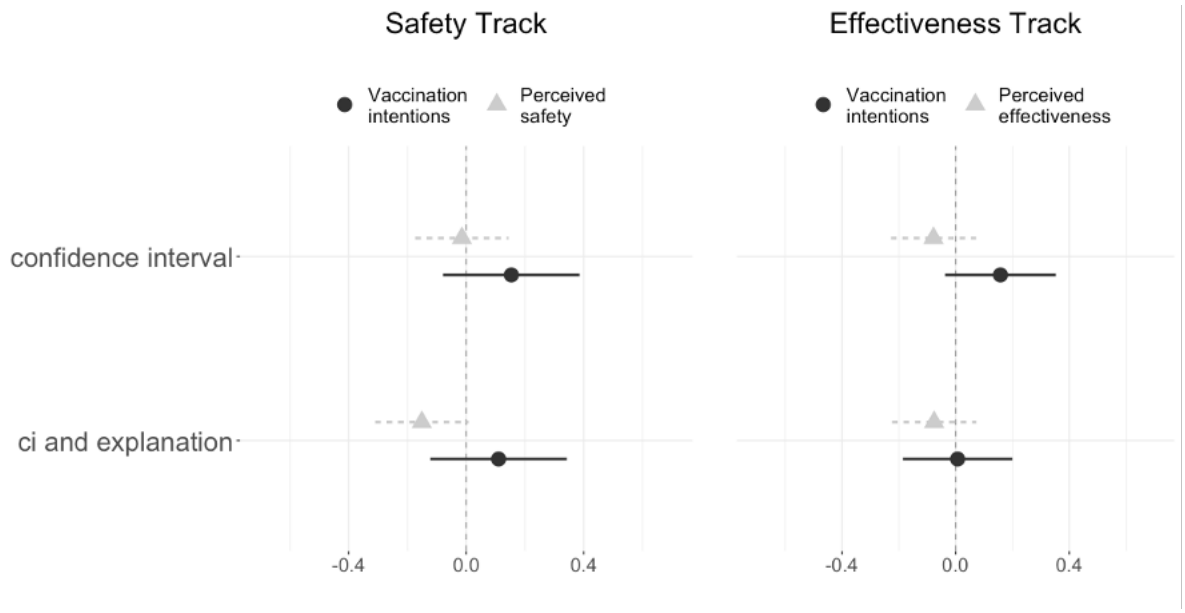


Figure 2.4: Effect of adding a confidence interval to safety (top) or effectiveness (bottom) statistics about a hypothetical COVID-19 vaccine. OLS Regression models. Outcome variables are not standardised. Vaccination intentions, perceived safety, and perceived transparency are measured on a seven-point scale. Perceptions of informedness are measured on a four-point scale. Higher values indicate greater interest/trust in the vaccine. The baseline is the group of respondents who saw a classic point estimate.

instead, treats the point estimate (no uncertainty) condition as the baseline. Does adding a confidence interval to the safety or effectiveness estimates drive down vaccination intentions? It does not. As shown in Figure 2.4 , neither the stand-alone confidence interval nor the fuller explained uncertainty version carried any cost in terms of vaccination intentions. If anything, the results fall on the benefits side of the ledger. This was true for both tracks, confirming our first 'no costs' hypothesis (H1a). How about perceptions of vaccine safety and effectiveness (H1b)? Here, our results are largely null but differed somewhat depending on the track. In the 'safety' track, the added confidence interval had an effect on perceptions of safety – but only if it was explained. Respondents who had been told that 'between 1 and 6 in 10,000' would develop serious side effects rated the vaccine as just as safe as those who had seen the simple point estimate ('2 in 10,000'). Respondents who had seen the range along with an explanation rated the vaccine as 0.16 points less safe (on a seven-point scale) than those who had seen the simple point estimate – a small but significant effect. One reason why the explanation may have dented safety perceptions is that it explicitly mentions that patients might develop side effects later on. However, since

the confidence intervals for the two uncertainty treatments themselves overlap, we should hesitate before concluding that explanation in itself has an effect. In the 'effectiveness' track, neither version of the confidence interval treatment had an effect on perceptions of effectiveness. Those who had been told that the vaccine was between 83% and 97% effective (explained or not) rated it as just as effective, on average, as those who had been told it was 90% effective. Hence, we accept our second 'no costs' hypothesis (H1b) for the effectiveness track and only marginally and conditionally reject it for the safety track. Overall, then, our data suggests that 'radical transparency' has little or no negative effect in the short run. But what is the effect in the long run? What if, a few months after a new vaccine is approved, the level of effectiveness drops to the bottom or the share of patients developing serious side effects soars to the top of the confidence interval? Does acknowledging uncertainty offer a benefit by buffering against bad news? We found that it can and does – at least in the 'effectiveness' track. To measure how badly people responded to the negative update we looked at the subset of respondents in the 'point estimate' condition and compared interest in the vaccine before and after they read the update. As expected, vaccine willingness dropped due to the update. However, the difference was not quite as stark as we would have expected. In the 'safety' track, average vaccination intentions dropped by .17 points on a seven-point scale, from 3.86 to 3.69 points on a scale from 0 to 6. Our best explanation for why the effect was so small is that '6 in 10,000' developing serious side effects may not feel much worse than '2 in 10,000'. In contrast, the update had a noticeable effect on perceptions of vaccine safety. On a parallel 0-6 scale, average safety ratings dropped by .30 points after exposure to the update, from 4.22 to 3.92. In the 'effectiveness' track, average vaccination intentions dropped by .25 points, from 4.37 to 4.11. And again, the update had a stronger effect on perceptions of vaccine effectiveness than on vaccination intentions. Average perceptions of vaccine effectiveness dropped by 0.39 points, from 4.52 to 4.13. (All four declines were statistically significant in paired-samples t-tests.)

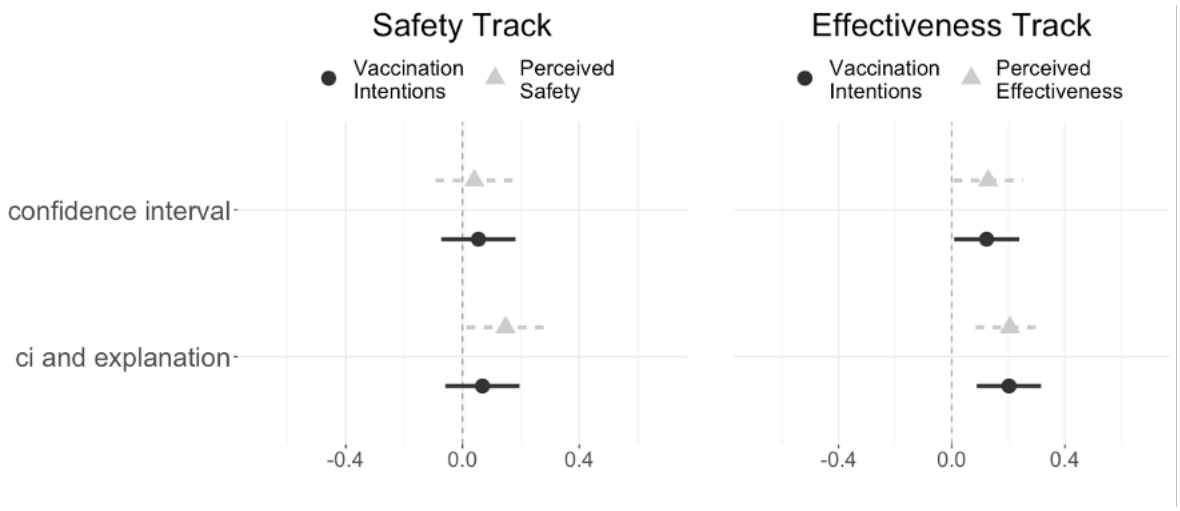


Figure 2.5: The offset of the negative update among respondents who saw an additional confidence interval. The dependent variable is the difference between the (seven-point) vaccination intentions scale before and after exposure to the negative update. Higher values indicate that vaccination intentions dropped more sharply. The baseline is the group of respondents who saw a classic point estimate.

These declines are of course a necessary condition for testing H2, which is about whether they are offset by having seen an uncertainty treatment at Stage 1. The answer is that they are – but only when the uncertainty is around safety statistics. Figure 2.5 charts the offset among the two uncertainty groups. In other words, the zero baseline represents the effect of the update on vaccine intentions among the point estimate respondents, and the plotted points represent whether that effect was different among those who saw a confidence interval at Stage 1. On the safety track, there is at least some sign that that vaccination intentions were slightly stronger among those who had seen a range – with or without an explanation – rather than a simple point estimate. However, the effect was short of statistical significance. This null finding is likely due to the fact that the effect of the negative update on vaccination intentions was so small on the safety track anyway: adding a confidence interval cancelled out about a third of this small effect. When it comes to perceptions of safety, however, which were more affected by the update, there is a positive result in line with H2(b). Adding and explaining a confidence interval cancelled out about half of the .30 point drop in perceptions of vaccine safety (0.14 points). By contrast, the stand-alone confidence interval had no appreciable effect. In the 'effectiveness' track, the results provide a more clear-cut confirmation of both benefits hypotheses. Adding a confidence interval with or without an

explanation significantly offset the drop in both vaccination intentions and perceived effectiveness. As mentioned above, the news that the vaccine was not 90 but only about 83 effective caused a .25 drop in average vaccination intentions. Adding a confidence interval halved that effect (+.13 points); adding a confidence interval and explaining why manufacturers could not give an exact figure almost cancelled out the entire effect of the negative update (+.21). A likely reason is that acknowledging uncertainty had also mitigated the update's impact on perceived effectiveness. Again, around half of the effect among the point estimate group was eliminated by having shown a confidence interval, and again the mitigation was even more effective if the uncertainty was explained.

2.8 Conclusion

This study set out to investigate the effect of acknowledging uncertainty around public policies on public behaviour. Using the example of the new COVID-19 vaccines we fielded an online survey experiment (n=2,167) to test the effects of communicating uncertainty around two particularly volatile estimates: vaccine effectiveness (what percentage of transmissions will the vaccine prevent?) and safety statistics (how many will develop serious side effects?). Our goals were two-fold. The first was to test in this unusually salient and anxiety-inducing context whether, as a growing set of studies imply, reporting uncertainty carries little cost in terms of trust in data and willingness to act on it – in this case, to have a COVID-19 vaccine. Our data suggests that indeed the costs of reporting uncertainty were minimal. Second, we sought to find whether it might help sustain vaccination intentions in the event that estimates have to be corrected downward. Our data suggests that it did – significantly so among respondents who had read about effectiveness statistics. The study was set up as follows. First, we asked respondents to suppose they were offered a COVID-19 vaccine. We divided them into two tracks: Respondents in one track saw (optimistic) estimates of how effective the vaccine was going to be; respondents in the other track saw (optimistic) estimates of how many patients would develop serious side effects. We also had a pure control condition which did not get any information about the vaccine. The crucial manipulation was the amount of uncertainty that was communicated when the vaccine was first introduced. Respondents in the 'point estimate' condition were shown a classic point estimate. Respondents in the 'confidence interval' condition were shown the same point estimate but along

with a range in which the true number was likely to fall. Respondents in the 'confidence interval and explanation' condition were shown the point estimate, the range, and an explanation why manufacturers could not be exact. Our results showed that average interest in the two 'confidence interval' conditions was just as high as average interest in the 'point estimate' condition. In other words, communicating uncertainty did not harm general interest in the vaccine. This was true regardless of whether respondents had read about safety or effectiveness statistics. In a second stage, we asked our respondents to suppose that a few months had gone by, and that the manufacturers had updated their effectiveness [or safety] estimates based on new data. Respondents in the 'effectiveness' track now saw a lower estimate of effectiveness; respondents in the 'safety' track now saw a higher estimate of patients developing serious side effects. In both cases, the new estimates were within the range that respondents in the 'confidence interval' conditions had seen at the beginning of the survey. Here, our results differed depending on the track. In the 'safety' track, we observed a slightly lower drop in vaccination intentions among those who had been warned about the possibility that the estimates might have to be adapted later on. However, the effect did not reach statistical significance. We suspect that this is because our negative update was less negative than it could (or should) have been – which is presumably the reason why it only caused a very small drop in vaccination intentions. Future research ought to investigate the effect of higher numbers of patients developing side effects (e.g. 6 in 100 rather than 6 in 10,000, as we did) or the effect of more serious side effects (including extremely rare but potentially lethal side effects). In the 'effectiveness' track, we found full support for our benefits hypothesis: Average vaccination intentions dropped far less among those in one of the 'confidence interval' conditions than among those in the 'point estimate' condition. Taken together, our data make a compelling case that communicating the uncertainty around statistics such as these is a sensible strategy (quite apart from its normative merits): it carries little cost and offers potential benefits. Being open about the possibility of the numbers changing – that is, providing a ranger rather than a simple number and letting patients know why one cannot be exact – builds up a certain resilience to the news that the numbers have, in fact, changed. Not only can people handle uncertainty but, importantly, they grasp its implications without needing additional explanation. We found some indicative evidence that explanations helped communicators to realise the benefits of uncertainty when estimates take a downturn, but the more consistent differences were between both uncertainty treatments and

that in which just a point estimate was reported. Confidence intervals prove an intuitive means of conveying the likely range of outcomes. (If they are misunderstood, it may be in a way that plays into communicators' hands: that is, the misconception that all outcomes within the interval are equally likely rather than one at the lower end, as in our example, being ex ante less likely than the point estimate). All of this is consistent with studies showing that clarity and brevity often prove sufficient. Heavier informational treatments have often had minimal impact on vaccine intentions, even if they leave audiences feeling more informed (Kerr, 2021; Loomba, 2021). There were persistent differences between the safety and effectiveness tracks in our study. The biggest such difference was at the outset: any treatment referring to safety concerns, regardless of uncertainty, left an appreciable dent in vaccination intentions. But there was also a sign that the benefits of reporting uncertainty were more limited in the safety context, probably for the same reason that even helpful safety treatments tend to fall on deafer ears. Indeed, the impression given is that respondents were simply doing less cognitive processing of the treatments on the safety track – they were perhaps feeling rather than analysing risk (Slovic, 2010). The upbeat upshot is that there were no costs of reporting uncertainty in that context: any heightened sense of anxiety did not create an inability to cope with uncertainty. This is important since talking about safety is not really avoidable for policymakers. If or when side effects emerge, no democratic government can contain that information. There are two limitations of our study worth noting here. One relates to the nature of our sample. We noted the high levels of vaccine willingness on show. This is partly contextual: the UK was a relatively vaccine-willing country anyway and sleeves in December 2020 were being enthusiastically rolled up as the vaccine programme got under way. But it may also reflect the particular vaccine willingness in a relatively educated and liberal sample. It is conceivable that a more vaccine-hesitant sample would have baulked at the uncertainty that these respondents took in their stride. However, given that vaccine intentions are much less politicised in the UK (as in much of Europe) than in the US, sampling bias seems unlikely to have made a big difference to the effects on show here. The second is the fact that we could only examine vaccination intentions, not actual vaccination decisions. Future researchers ought to conduct panel studies in which they vary the amount of uncertainty provided at the beginning of a vaccination campaign and follow respondents for a few months, or even years, as news about vaccine safety and effectiveness will inevitably come out. In addition, future research ought to investigate how

communicating uncertainty around other policies affect political opinions and political behaviour. Our research suggests that it is worth investigating communication in more detail. If confirmed, our findings mean that governments have little to lose and plenty to gain by being open about what they know and what they do not know.

When breaking the norms pays off: Uncivil Appeals Boost Right Wing Challenger Success in Social Media.

3.1 Abstract

Over the past decade, there has been a growing concern over uncivil party communication strategies which challenge democratic attitudes. However, the use of uncivil communication seems to work effectively for parties to persuade the electorate. This paradox by which a political communication style that is disruptive for democratic societies is successful to engage voters in currently democratic regimes is still under explored. The present study addresses a) whether uncivil appeals are actually more successful than civil ones b) if challenger parties are more prone to using uncivil appeals than mainstream parties and c) whether some parties benefit more from this communication strategy. For that purpose, we applied supervised learning techniques on Twitter data collected from the timelines of all parliamentary representatives in Spain. Through the study of this multi-party context we find that challenger parties are generally more likely to use uncivil language on their social media than mainstream parties. However, challenger parties seem to only have an incentive to be more uncivil than their mainstream competitor, but not necessarily when compared to other

challengers.

3.2 Introduction

While campaigning for President of the United States, Trump called large numbers of Mexican immigrants rapists, mocked a journalist with disabilities, and demanded that all Muslims be temporarily banned from entering the country. Of course, Trump was not the first politician to stir up voters with uncivil political discourse, but his rhetoric during the presidential campaign and his term in office marked the visible badges of a turning point in the global trend of democratic decline. As the latest report from Freedom House puts it, it is now broad enough that it not only impacts on those living under dictatorships, but also citizens of long-standing democracies.

It is a truism to say that politics is contentious in nature and that expressing disagreement is a constitutive element of democratic functioning. Nevertheless, there are certain boundaries to how that disagreement can be conveyed publicly, associated with a normative minimum degree of civility. It is not all that clear where one draws the line between what are civil and uncivil appeals: It seems like the two constructs' definitions are highly dependent on subjective impressions. Yet, within the scholarly literature on the topic there is some consensus that uncivil communication encompasses a series of features related to breaking explicit or implicit rules of politeness of a given context Mutz (2015). Also, recent data has shown that a majority of people identify and classify the same repertoire of features as uncivil Sydnor (2018).

Although the use of uncivil communication as a political strategy is not a new phenomenon, the rise of social media as a new political arena rekindled the debate on the consequences of uncivil behaviour for democracy. Especially, uncivil appeals in political discourse have been brought into the spotlight because of their potential negative consequences on citizen support of democratic practices or endorsement of democratic norm violations Clayton et al. (2021). Over the past decade, there has been a growing concern that candidates might be engaging in uncivil communication strategies which challenge democratic attitudes. Evidence shows that exposure to uncivil communication increases issue polarization Anderson et al. (2014), weakens political trust and efficacy Mutz & Reeves (2005); Borah (2013, 2014). However, the use of uncivil communication seems to work effectively for parties to persuade their audiences. This poses a conflict between democratic stability and the systemic incentives for parties and voters in democratic countries. This paradox

by which a political communication style that is disruptive for democratic societies is the most successful to engage voters in currently democratic regimes is still under explored Mutz & Reeves (2005); Goovaerts & Marien (2020). Most of the evidence in the field comes from the United States context, which is characterized by a strong partisan divide along a two-party system. Other institutional settings are likely to present different scenarios in terms of party communication. In fact, the rise of challenger parties has disrupted party systems in Western Europe in recent years. While mainstream parties can rely on their policy success to attract voters, challenger parties benefit from discrediting those in office. Political entrepreneurs, such as challenger parties, seem to be aware of how to implement this rationale and translate it into an effective communication strategy in their favor. While the tension between change and stability, and thus, between innovative and traditional ways of doing politics is always present, the current moment in Western European democracies faces us with the perfect scenario for studying how party communication works favoring political change.

In this context, are some parties more likely to rely on and benefit from uncivil communication? Are uncivil messages actually more engaging than civil ones? And finally, are some parties more likely to rely on and benefit from uncivil communication? This research focuses on uncivil messages broadcast by politicians on their social media. The aim of this study is two-fold. In the first place, it analyzes whether political representatives of challenger parties engage in uncivil communication more often than those in mainstream parties and if messages containing uncivil appeals propagate more successfully on social media than civil messages. Additionally, it tests whether challenger parties benefit more than their mainstream counterparts from that kind of communication strategy. For that purpose, the study relies on Twitter data collected from the timelines of all parliamentary representatives in Spain.

Our study finds that, overall, tweets issued by politicians across the party spectrum that contain uncivil language are more likely to propagate in social media. Specifically, challenger parties are more likely to use uncivil language and profit from it, but only when compared to their direct mainstream competitor. That is, among right-wing parties Ciudadanos and Vox are more likely to post uncivil messages when compared to PP. On the other hand, Unidas Podemos is more likely to use uncivil appeals when compared to PSOE.

3.3 Which Parties Are More Reliant on Uncivil Appeals?

Over the past decade European party politics has become unpredictable. Even if there is evidence that broadly speaking we are not before a historically anomalous period, associated with major societal changes, long-standing patterns in electoral behaviour are no longer stable De Vries & Hobolt (2020). Recent years have been characterised by a decline of mainstream parties and parallel rise of political outsiders all along the political spectrum Hernández & Kriesi (2016); Hobolt & Tilley (2016); Emanuele & Chiaramonte (2019). As De Vries and Hobolt (2020) put it, in this transitional times, the electoral arena is currently traversed by the struggle between innovative and traditional ways of doing politics. And so, this scenario suggests that we might be witnessing a moment in which innovative political communication strategies are on the rise. Furthermore, recent electoral success of new challenger parties lead us to believe that innovative political communication strategies are particularly profitable for some parties.

The literature has not agreed on a single label to refer to these new parties which rely on disruptive strategies when compared to mainstream parties. Some have called them 'niche parties' Adams et al. (2006); Jensen & Spoon (2010); Meguid (2005), stressing the fact that the issues they raise fall outside the traditional cleavage. Other studies refer to them as 'populist parties' Kriesi (2014); Mudde (2007); Pauwels (2014), with a focus on its ideological stances. Populist parties are often characterised by their thin-centered ideology which divides society into two monolithic and antagonistic groups, 'the pure people' and 'the corrupt elite' (Mudde, 2004). Finally, they have also been labeled 'challenger parties' Hino (2012); Van de Wardt et al. (2014); Van de Wardt (2015).

We use the term 'challenger parties' as suggested by de Vries and Hobolt 2020. According to their definition, the term encompasses some of the aforementioned labels that are frequently used in political science research, such as niche and populist parties. Their distinction between challenger and mainstream parties emphasizes the relative position of parties within the political system. While mainstream parties are in control of the political arena or have been in the past, challenger parties do not have previous experience controlling policy or government positions at a large-scale.

Given this definition, mainstream parties count on their national level policies to advertise their

worth and attract voters. In fact, evidence shows that incumbents hold clear advantages when compared with challengers. Several studies on negative campaigning highlighted incumbency status as one of the main factors associated with going for negative or positive communication strategies Lau & Pomper (2004); Druckman et al. (2009). Incumbents are more likely to appear in the media, to benefit from name recognition and have an established reputation Walter & Van der Brug (2013). Conversely, challenger parties have less incentives to build positive campaigning strategies. These parties cannot showcase public policy success, even if they play a role in the design and approval of legislation. Furthermore, challenger parties have less to lose, and so, the cost of a potential backlash effect that ends up hindering their electoral outcomes is lower.

Challenger parties operate as political entrepreneurs, aiming at disrupting the political arena De Vries & Hobolt (2020). As Hobolt and Tilley 2016 suggest, challenger parties provide an alternative narrative that defies the mainstream consensus. As innovators, challenger parties benefit from discrediting dominant party's authority with all sorts of uncivil language, including anti-establishment rhetoric. In some cases, like research on populist parties shows, communication style might play a more relevant role explaining their success than the content of their discourse. Given this framework, we propose the following hypothesis:

H1: Challenger parties are more likely to rely on uncivil language than mainstream parties.

3.4 Are Uncivil Appeals Successful?

Political communication is all about persuasion: Politicians broadcast their messages with the aim of winning votes for their parties. Over the past decade, social media platforms gained a central role in political communication campaigns. Even if social media adoption has improved communication between politicians and their constituents in many ways Karlsen & Skogerbø (2015); Theocharis et al. (2016), it also brought concern over the quality of online deliberation Borah (2014). A platform where users can express their frustration with politicians that are otherwise hard to reach with little costs might obviously not result in rich democratic deliberation. In fact, previous research has shown that the use of uncivil language on social media is commonplace Theocharis et al. (2020); Coe et al. (2014); Sobieraj & Berry (2011).

A precise definition of uncivil language has proved elusive Herbst (2010). Despite its common

usage in political communication, uncivil language is used in different studies to mean different things, ranging from impolite remarks and sarcasm Sobieraj & Berry (2011) to racial slurs and insults Stryker et al. (2016). Since its definition varies among researchers, it is important to clarify how the term is used in this article. As Mutz (2016) points out, incivility refers to the style rather than the substance of political discourse. According to her definition, uncivil discourse is that which violates the norms of politeness for a given context or interaction. General tone of the message and word choice are key distinguishing features of uncivil appeals. And though it might still seem like a blurry concept, two recent surveys have shown that a majority of people tend to classify the same repertoire of features as uncivil Sydnor (2018).

It is clear that politicians are ideal targets for the receiving end of incivility in social media. What is less clear is how much of it they serve up themselves – and, if so, whether they benefit. Theocharis et al. 2016 suggest that online incivility dissuades candidates from interactive uses of Twitter. Given the hostility they find on social media platforms like Twitter, politicians opt for using the platforms just as a broadcasting tool instead of engaging in interactive exchanges with the public. Broadly speaking, it might seem like there are no incentives for candidates to engage with audiences on Twitter because it can easily get out of control and generate unexpected controversy. However, the fact that there are little incentives to engage in deliberation as democratic norms would ideally suggest does not mean that there are no structural incentives to be uncivil altogether. As Kosmidis and Theocharis 2020 point out, social media users might not see the platform as an arena for political debate in the traditional way, but rather as a source of entertainment and information or a place for networking. Hence, political leaders using the platform might not be looking for deliberation either, but rather see it as a way to advertise themselves or their parties for little cost.

Previous research has shown that although uncivil appeals are far from the normative ideal of democratic deliberation, they hold some distinct features that make them functional for specific ends in political communication, such as mobilizing citizens and differentiating in-groups from out-groups Jamieson et al. (2017). Exposure to uncivil messages can increase interest in politics Brooks & Geer (2007); Berry & Sobieraj (2013) and could even lead to perceiving an issue as more interesting, memorable, and persuasive Jaidka et al. (2019). Focusing on the emotional consequences of exposure to uncivil messages, Kosmidis and Theocharis 2020 find that uncivil

appeals often trigger enthusiasm, mostly partisan based, thus transforming online debates into an engaging and entertaining spectacle. Their study suggests that instrumentally relying on uncivil language can be adopted as an effective tactic for attracting attention on Twitter. Therefore, we propose:

H2: Messages issued by political leaders containing uncivil language are more likely to propagate on social media.

3.5 Which Parties Benefit the Most from Being Uncivil?

Challenger parties have good reasons to adopt innovative communication strategies like using uncivil language strictly from a supply-side perspective. However, the literature also suggests that by doing so, they satisfy the demands of their target electorate. Overall, challenger party voters across the ideological spectrum share negative attitudes towards the status quo. When it comes to radical right supporters, extant literature finds that the 'protest vote' thesis is one of the most prominent explanations to their recent electoral success Arzheimer (2018). In its strictest version, radical right voters only care about expressing their feelings of discontent with the political elites and the political system, regardless of the party's political agenda Eatwell (2000). More nuanced variants of the protest vote thesis suggest that voters do sympathize with the policy platforms of new radical right parties, but might hold less extreme preferences. Even so, the strongest driver of radical right vote are negative attitudes towards immigration Arzheimer (2018). More recent research also found that correlates of radical right support overlap with 'need for chaos'. That is, a set of psychological dispositions oriented towards attaining social status through disruptive strategies Arceneaux et al. (2021). On the other extreme of the ideological spectrum, previous studies show that left-leaning populist attitudes are also driven by feelings of anger Rico et al. (2017). On these grounds, we suggest:

H3: Challenger parties are more likely to benefit from using uncivil language than mainstream parties.

3.6 Methods

The case of Spain presents an ideal testing ground to assess our hypotheses. Since the economic crises of 2008 the Spanish electoral arena has transformed from a stable two-party system with Partido Popular (PP) and Partido Socialista Obrero Español (PSOE) into a multi-party system. New challenger parties emerged, capitalising the disappointment with mainstream parties: Podemos on the left (which at the time adopted the label Unidas Podemos), Ciudadanos on the centre and Vox to the extreme right. In this scenario, new parties across the ideological spectrum with a varied origin and heterogeneous programmatic objectives had to compete for votes against their mainstream counterparts. Although it might be early to establish long lasting patterns, the case of Spain presents interesting insights for understanding party communication strategies in the context of political change Vidal (2018).

Data Collection

For this paper a large dataset was constructed merging together different sources of data. The data were collected in November 2019. First, we scraped the list of legislators holding office from Wikipedia and manually traced all of their Twitter handles. We relied on those lists in order to collect the timelines of Spanish Congressmen using the tweepy package [1] connected to the Twitter API. Spain has 616 legislators (266 Senadores and 350 Diputados). To the best of our knowledge, 152 senators and 288 deputies in Spain had an active Twitter account at the time of data collection. Using the user timeline function within Tweepy we collected approximately 1000 tweets per active legislator, as available in their timelines. Whenever the number of available tweets was less than 1000 we obtained all the tweets on the timeline. Conversely, if the total number of tweets in the timeline exceeded that limit, we obtained the most recent 1000. The full dataset contained a total of 404.497 tweets (267.579 coming from the timelines of Diputados and 136.918 from Senadores). We rely on separate subsets for each chamber because at the time different parties had presence in Senadores than in Diputados, reflecting different electoral competition strategies.

We also collected the metadata accompanying each tweet, including some relevant features for the analysis. For instance, the retweet count, number of followers, total number of tweets in the

[1] For more details on this package developed for Python see documentation for Tweepy https://www.tweepy.org/

account and whether the account is verified or not. Additionally, we extracted some other tweet-level features in the dataset by analysing the full text of the tweets, such as the presence of a URL, the presence of emojis and the word count.

Measure of Incivility

After collecting all legislators timelines and metadata we created an automated measure of civility to classify tweets. Dictionary methods were considered at first, given the advantage of using an instrument that has been previously validated. Especially since ready-to-use lexicons are suggested to work better on diverse and informal language such as the one regularly used on Twitter González-Bailón & Paltoglou (2015). However, to the best of our knowledge there were no validated dictionaries for uncivil language. Off-the-shelf classifiers that could have potentially been considered as proxies for detecting incivility such as LIWC or ANEW did not provide good results for the Spanish dataset.

Following Theocharis et al. 2020, we applied supervised machine learning to detect incivility in the full text of tweets of legislators. We relied on two different random samples drawn from each dataset and manually annotated them along the civil-uncivil spectrum, creating a binary outcome (0-1). We adopted an already validated coding scheme designed by Theocharis et al. 2016. According to their definition, a civil message is that which follows politeness standards, is written in a well-mannered and non-offensive way. This does not mean that there is no room for criticism or strong stances, but rather the message is delivered in a respectful way even if its content is critical. On the other hand, an uncivil message is what they define as an ill-mannered, disrespectful tweet that may contain offensive language. The uncivil category encompasses messages that threaten fundamental rights (such as freedom of expression), the use of hate speech, name-calling or ad hominem attacks, pejorative speech or vulgarity, sarcasm, ALL CAPS, and incendiary, obscene, and/or humiliating language.

We trained a naive Bayes classifier on each of the random samples. Although uncivil messages are most engaging, they are not as prevalent, leading to an imbalanced dataset. In order to deal with the imbalance in the training set we applied resampling techniques. We tried both down and upsampling, resulting in better results for the upsampled minority class with replacement. Finally, we applied the respective models to the full datasets of each chamber. We found that the

classifier's accuracy is 0.95, with a precision of 0.69 and recall on the "civil" category being (For more detail see confusion matrix in the Appendix).

3.6.1 Measure of Propagation on Twitter

As a measure of propagation on Twitter we relied on retweet count. Research analysing Twitter data has quantified propagation with different metrics, such as mentions (tagging a user with the @ symbol), replies, number of tweets on a given topic or retweets (Yang & Counts, 2010). We opted for using the latter as the cleanest metric. The rest of often used metrics in related research allow for positive and negative valence toward the previous message (Kim & Yoo, 2012) and we were mainly interested in message propagation as a form of engagement, regardless of valence. Furthermore, extant literature has stressed the significance of retweeting as a mechanism of endorsement for promoting visibility of a given topic (Yardi & Boyd, 2010).

3.7 Results

We analyzed a total of 404.497 tweets, from which 44.496 were classified as uncivil and 360.001 were civil. Party composition of each chamber was imbalanced and that was also reflected on their online presence. Within the Chamber of Deputies, Ciudadanos had 10 deputies with a Twitter profile, Unidas Podemos had 25, Vox followed with 32, PP with 66 and PSOE with 82. Within the Chamber of Senators, only 2 profiles belonged to senators from Vox, 4 from Ciudadanos, 52 to PSOE and 56 to PP. However, there was a large volume of total tweets per party, since we collected at least 1000 tweets per legislator. Anyway, the imbalance does not seem to have biased the results because some of the underrepresented parties in the sample still show the largest effects.

At a first stage, we analyzed the share of civil vs. uncivil tweets controlling for party label, to test the expectation that some parties were more likely to rely on uncivil language than others. At a second stage, we conducted OLS regression with clustered standard errors, regressing retweet count on incivility to test the expectation that that uncivil messages get greater engagement per se. Finally, we included interactions by party label in the OLS models in order to assess whether certain parties get stronger engagement than others when they publish uncivil messages on social media. We also conducted multilevel models as robustness checks, but the results show a low ICC

(0.08), which indicates there is no evidence of group effects (see Appendix).

3.7.1 Challenger parties are more likely to rely on uncivil language than mainstream parties.

We tested the expectation that challenger parties use more uncivil language than mainstream parties (H1). Figure 3.1 shows the share of uncivil and civil tweets by party label for deputies and senators. Notably, and reassuringly, the share of civil tweets is much larger than the share of uncivil tweets. This is true for tweets coming from deputies and senators from all parties. However, there is considerable variation among the different parties. As expected, politicians in the PSOE are the least likely to be uncivil. In aggregate, only around 5 per cent of their tweets contain uncivil language. Surprisingly, the second least uncivil party is one we categorize as a challenger party: Unidas Podemos, with a share of around 8 per cent of uncivil tweets. With hindsight, we suspect that this might be due to the origins of this new party and the voters it is trying to attract: Unidas Podemos started as a grassroots movement in the aftermath of the 2015 indignados movement. Their founders, leaders and supporters tend to have an academic background, and, as newcomers they were at the time trying to establish themselves as a party by appealing to highly educated and potentially more moderate voters. Civil tweets are more likely to engage those voters than uncivil tweets. The third in line from least likely to more likely to post uncivil tweets came from the PP. About 11 per cent of PP tweets contained uncivil language. As expected, the share of uncivil tweets from PP is lower than the share of uncivil tweets from its challenger competitor Vox. We suspect that the relatively high percentage of uncivil tweets is exactly because of the competition with Vox. As a conservative party, PP politicians are likely to want to retain the right-wing voters Vox is appealing to. Hence, since Vox appeared in 2013 with a right-wing populist agenda PP politicians may have found it necessary to show at least some of the indignation that is drawing disillusioned right-wing voters to Vox. The highest share of uncivil tweets is found among politicians from the right-wing populist parties Ciudadanos and Vox. Almost 15 per cent of their tweets are uncivil. Overall, we find partial support for H1.

Right-wing challenger parties' tweets (Vox, Ciudadanos) are more uncivil than established right wing parties' tweets (PP). Similarly, left-wing challenger parties' tweets (Unidas Podemos) are more uncivil than established center-left wing parties' tweets (PSOE). However, we only observe this

difference when looking at right and left wing parties separately. Looking at the potential voters that the established parties are trying to retain this difference is understandable. Established left wing parties have an incentive to move towards the median voter in response to emerging new parties where as established right wing parties have an incentive to move towards the extreme right in order to prevent losing voters to the rising parties.

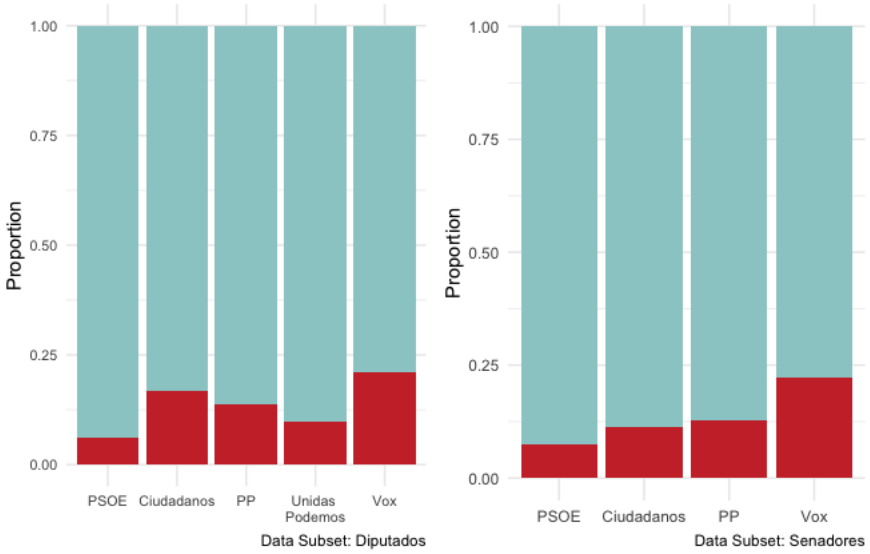


Figure 3.1: Proportion of Civil to Uncivil Tweets by Party Label and Chamber of Congress. Uncivil tweets represented in red, civil tweets represented in blue. Party labels shown are largest parties in each Chamber (Diputados on the left, Senadores on the right)

3.7.2 Tweets issued by political leaders containing uncivil language are more likely to be spread on social media.

Next, we tested the expectation that messages are more often retweeted if they broadcast uncivil messages (H2). Following, Brady et al. 2020 we conducted OLS regression on retweet count instead of relying on count models because the differential characteristics of Twitter metrics of well-known public figures. Political elites often have an extraordinary amount of retweets per tweet when compared to a regular user. In order to meet the assumption of normality we log transformed the main variable of interest (retweet count).

Figure 3.2 shows the mean number of retweets for tweets containing civil language and uncivil language. The plots for both subsets show substantive variation: Uncivil tweets are more likely to be retweeted than civil tweets.

	Dependent variable:		
	Retweet Count		
	(1)	(2)	(3)
Uncivil language	1.052***	0.626***	0.405***
	(0.013)	(0.012)	(0.030)
Vox		1.460***	1.384***
		(0.014)	(0.016)
PP		0.017	−0.004
		(0.011)	(0.011)
Ciudadanos		0.491***	0.439***
		(0.021)	(0.023)
Unidas Podemos		0.803***	0.793***
		(0.015)	(0.016)
Other		0.301***	0.306***
		(0.011)	(0.012)
Word count		0.054***	0.054***
		(0.0003)	(0.0003)
Followers count		0.00000***	0.00000***
		(0.00000)	(0.00000)
Verified		0.224***	0.223***
		(0.009)	(0.009)
Has url		−0.533***	−0.536***
		(0.045)	(0.045)
Has emoji		0.214***	0.213***
		(0.008)	(0.008)
Uncivil language: Vox			0.514***
			(0.042)
Uncivil language: PP			0.276***
			(0.037)
Uncivil language: Ciudadanos			0.450***
			(0.060)
Uncivil language: Unidas Podemos			0.190***
			(0.053)
Uncivil language: Other			0.052
			(0.039)
Constant	3.458***	1.310***	1.330***
	(0.004)	(0.013)	(0.014)
Observations	267,577	256,953	256,953
R^2	0.023	0.210	0.211
Adjusted R^2	0.023	0.210	0.211
Residual Std. Error	2.171 (df = 267575)	1.958 (df = 256941)	1.957 (df = 256936)
F Statistic	6,351.000*** (df = 1; 267575)	6,224.000*** (df = 11; 256941)	4,296.000*** (df = 16; 256936)

Note: The reference categories in these models were Uncivil language equals 1 and party label equals PSOE

*p<0.1; **p<0.05; ***p<0.01

Table 3.1: OLS models for Diputados subset

	Dependent variable:		
	Retweet Count		
	(1)	(2)	(3)
Uncivil language	0.665***	0.425***	0.253***
	(0.018)	(0.017)	(0.033)
Vox		0.988***	1.055***
		(0.044)	(0.049)
PP		−0.269***	−0.305***
		(0.012)	(0.013)
Ciudadanos		−0.673***	−0.688***
		(0.032)	(0.034)
Other		−0.259***	−0.258***
		(0.014)	(0.015)
Word count		0.045***	0.045***
		(0.0004)	(0.0004)
Followers count		0.00003***	0.00003***
		(0.00000)	(0.00000)
Verified		0.287***	0.283***
		(0.014)	(0.014)
Has url		−0.593***	−0.603***
		(0.060)	(0.060)
Has emoji		−0.084***	−0.087***
		(0.011)	(0.011)
Uncivil language: Vox			−0.182*
			(0.107)
Uncivil language: PP			0.357***
			(0.041)
Uncivil language: Ciudadanos			0.213**
			(0.101)
Uncivil language: Other			0.025
			(0.050)
Constant	3.102***	1.747***	1.765***
	(0.006)	(0.016)	(0.016)
Observations	136,918	136,918	136,918
R^2	0.010	0.129	0.129
Adjusted R^2	0.010	0.129	0.129
Residual Std. Error	2.018 (df = 136916)	1.893 (df = 136907)	1.892 (df = 136903)
F Statistic	1,362.000*** (df = 1; 136916)	2,022.000*** (df = 10; 136907)	1,453.000*** (df = 14; 136903)

Note: The reference categories in these models were Uncivil language equals 1 and party label equals PSOE *p<0.1; **p<0.05; ***p<0.01

Table 3.2: OLS models for Senadores subset

The difference is larger for tweets coming from deputies. In this case, uncivil messages were associated with a 2.9 point increase (the reported 1.05 coefficient was transformed back from the log-transformed dependent variable) in retweet count when compared to civil messages (see table 3.1). Yet, even for senators the difference is statistically significant. Uncivil messages coming from senators were associated with a 1.9 point increase in retweets (0.66 in the original model, see table 3.2).

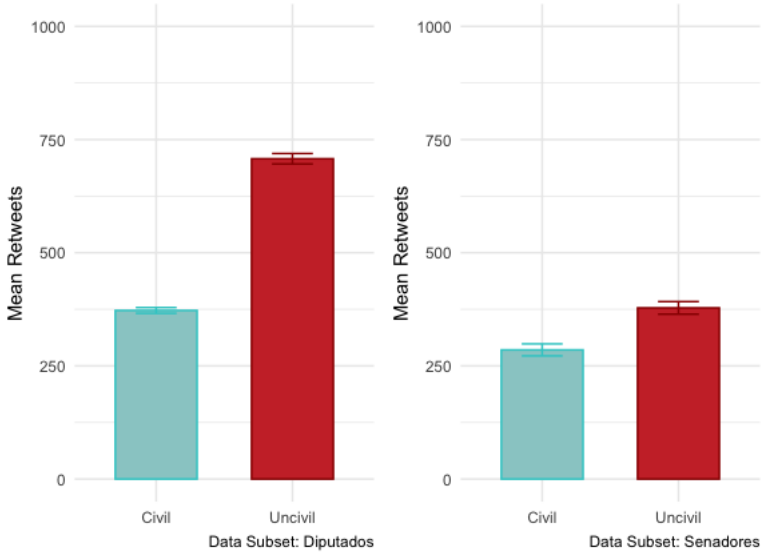


Figure 3.2: Mean Retweets by Civility. Mean retweets of uncivil tweets represented in red, mean retweets of civil tweets represented in blue. Results shown are an aggregate measure of mean retweets of all legislators by Chamber (Diputados on the left, Senadores on the right)

When controlling for features that convey legitimacy of an online presence, follower count does not seem to have a relevant effect on retweet count, but having a verified profile does. The verified label on a profile is associated with a small but positive significant effect of 1.22 for deputies and 1.35 for senators. Some visual cues such as including emojis, also seem to increase retweet count of a tweet. Conversely, including a URL is associated with a decrease in retweet count.

3.7.3 Challenger parties are more likely to benefit from using uncivil language than mainstream parties.

Finally, we tested the expectation that challenger parties are more likely to benefit from using uncivil language than mainstream parties (H3). In order to assess this hypothesis we conducted OLS models with no intercept to make comparison between parties more straight-forward to interpret.

	Dependent variable:		
	Retweet Count		
	(1)	(2)	(3)
PSOE		1.310***	1.330***
		(0.013)	(0.014)
Vox		2.770***	2.713***
		(0.016)	(0.017)
PP		1.328***	1.326***
		(0.013)	(0.013)
Ciudadanos		1.801***	1.769***
		(0.023)	(0.025)
Unidas Podemos		2.114***	2.123***
		(0.017)	(0.017)
Other		1.611***	1.636***
		(0.013)	(0.013)
Uncivil language	1.052***	0.626***	0.405***
	(0.013)	(0.012)	(0.030)
Word count		0.054***	0.054***
		(0.0003)	(0.0003)
Followers count		0.00000***	0.00000***
		(0.00000)	(0.00000)
Verified		0.224***	0.223***
		(0.009)	(0.009)
Has url		−0.533***	−0.536***
		(0.045)	(0.045)
Has emoji		0.214***	0.213***
		(0.008)	(0.008)
Vox: Uncivil language			0.514***
			(0.042)
PP: Uncivil language			0.276***
			(0.037)
Ciudadanos: Uncivil language			0.450***
			(0.060)
Unidas Podemos: Uncivil language			0.190***
			(0.053)
Other: Uncivil language			0.052
			(0.039)
Constant	3.458***		
	(0.004)		
Observations	267,577	256,953	256,953
R^2	0.023	0.786	0.786
Adjusted R^2	0.023	0.786	0.786
Residual Std. Error	2.171 (df = 267575)	1.958 (df = 256941)	1.957 (df = 256936)
F Statistic	6,351.125*** (df = 1; 267575)	78,427.410*** (df = 12; 256941)	55,420.420*** (df = 17; 256936)

Note: No intercept model. Comparison between parties as there is no single baseline *p<0.1; **p<0.05; ***p<0.01

Table 3.3: OLS models with no intercept Diputados subset

	Dependent variable:		
	Retweet Count		
	(1)	(2)	(3)
PSOE		1.747***	1.765***
		(0.016)	(0.016)
Vox		2.735***	2.819***
		(0.045)	(0.050)
PP		1.478***	1.459***
		(0.016)	(0.016)
Ciudadanos		1.075***	1.077***
		(0.034)	(0.035)
Other		1.488***	1.506***
		(0.016)	(0.017)
Uncivil Language	0.665***	0.425***	0.253***
	(0.018)	(0.017)	(0.033)
Word Count		0.045***	0.045***
		(0.0004)	(0.0004)
Followers count		0.00003***	0.00003***
		(0.00000)	(0.00000)
Verified		0.287***	0.283***
		(0.014)	(0.014)
HAs url		−0.593***	−0.603***
		(0.060)	(0.060)
Has emoji		−0.084***	−0.087***
		(0.011)	(0.011)
Vox:Uncivil language			−0.182*
			(0.107)
PP: Uncivil language			0.357***
			(0.041)
Ciudadanos: Uncivil language			0.213**
			(0.101)
Other: Uncivil language			0.025
			(0.050)
Constant	3.102***		
	(0.006)		
Observations	136,918	136,918	136,918
R^2	0.010	0.747	0.747
Adjusted R^2	0.010	0.747	0.747
Residual Std. Error	2.018 (df = 136916)	1.893 (df = 136907)	1.892 (df = 136903)
F Statistic	1,361.651*** (df = 1; 136916)	36,736.790*** (df = 11; 136907)	26,968.480*** (df = 15; 136903)
Note: No intercept model. Comparison between parties as there is no single baseline			*p<0.1; **p<0.05; ***p<0.01

Table 3.4: OLS models with no intercept Senadores subset

We included interactions between the use of uncivil language and party label (See Figures 3.3, 3.4 and 3.5 or tables 3.3 and 3.4 for the full regression tables). When compared against PSOE as the reference category, all the remaining large parties showed a significant effect. In line with the results of the previous models, within the lower Chamber, Vox had the largest increase in retweet count when using uncivil language. It was closely followed by Unidas Podemos and Ciudadanos. PSOE and and PP also benefited from uncivil language use, but to a lesser extent. Our model predicts that a Vox deputy who refrains from using uncivil appeals will be retweeted around 800 times, whereas if they rely on uncivil language they will be retweeted 1600 times. Uncivil language doubles the number of retweets for a deputy from a party like Vox. A tweet by a deputy from Unidas Podemos will increase from around 600 to around 1300, while a tweet of a deputy from

Ciudadanos will go from 400 to around 500 when using uncivil appeals. On the other hand, tweets from deputies of PP and PSOE show almost no difference.

Whithin the chamber of Senators, however, the configuration of results is considerably different. Vox still seems to benefit the most from relying on uncivil appeals, but within the upper chamber they are followed by PSOE. This might be due to measurement issues, since the sample is imbalanced in terms of senators by party. There are only two Vox Senators and four Ciudadanos senators represented in the sample. Or it could also be a reflex of a different configuration of party competition in the context of a reduced presence of extreme right wing challengers and almost no relevant left wing challengers in the Senate at the time.

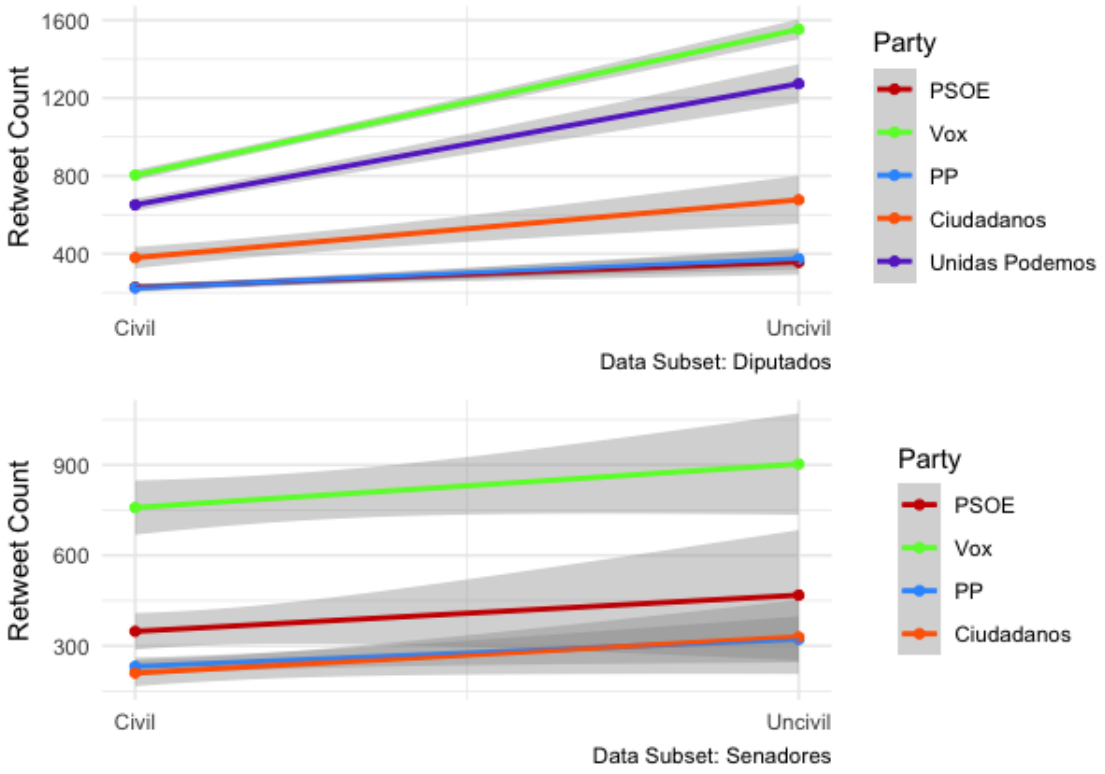


Figure 3.3: Interaction Plot by Party Label and by Chamber. Each line shows the difference in retweet count by Party label when they use civil language and uncivil language. Measure of retweet count shown is an aggregate measure of retweets of all legislators by Chamber (Diputados above, Senadores below)

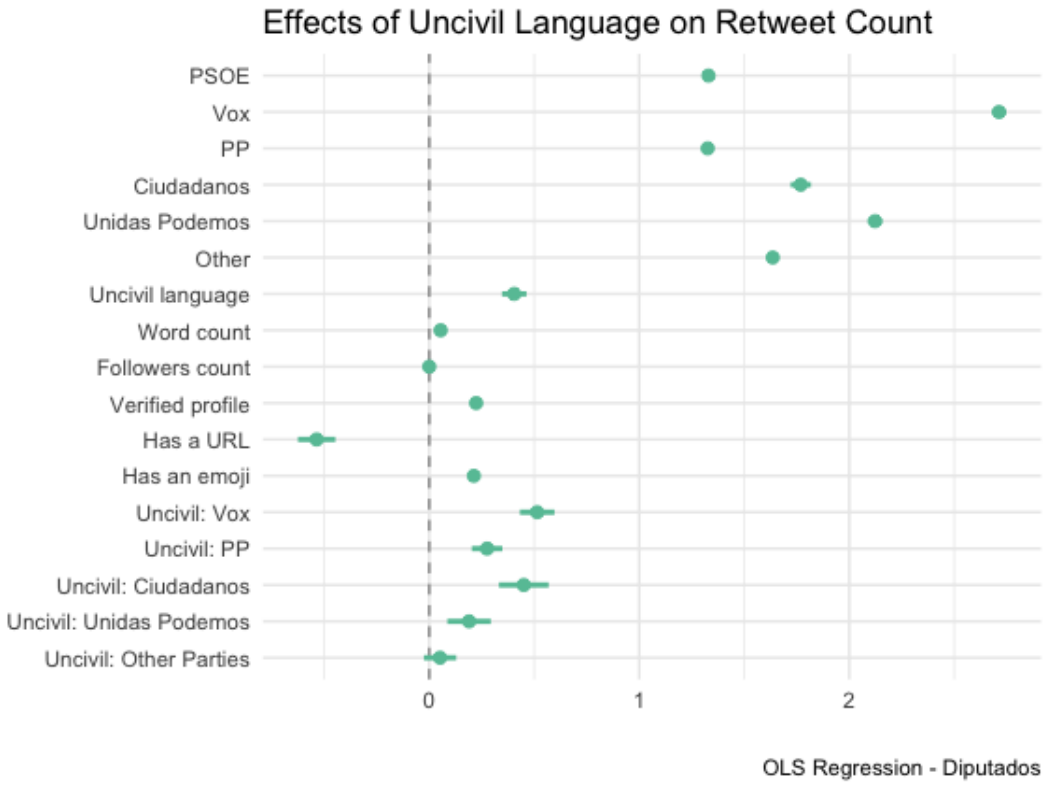


Figure 3.4: Coefficient Plot for Deputies. Effect of uncivil language on retweet count. OLS Regression model with log-transformed dependent variable (retweet count). Parties shown are the largest party labels at the national level by Chamber, all others are aggregated under the "Other" label. We included an interaction between party label and using uncivil language. Uncivil language variable was built combining manual annotation and a naive Bayes classifier. Word count, Follower count, Verified profile, Has URL and Has an emoji have all been extracted from tweet metadata.

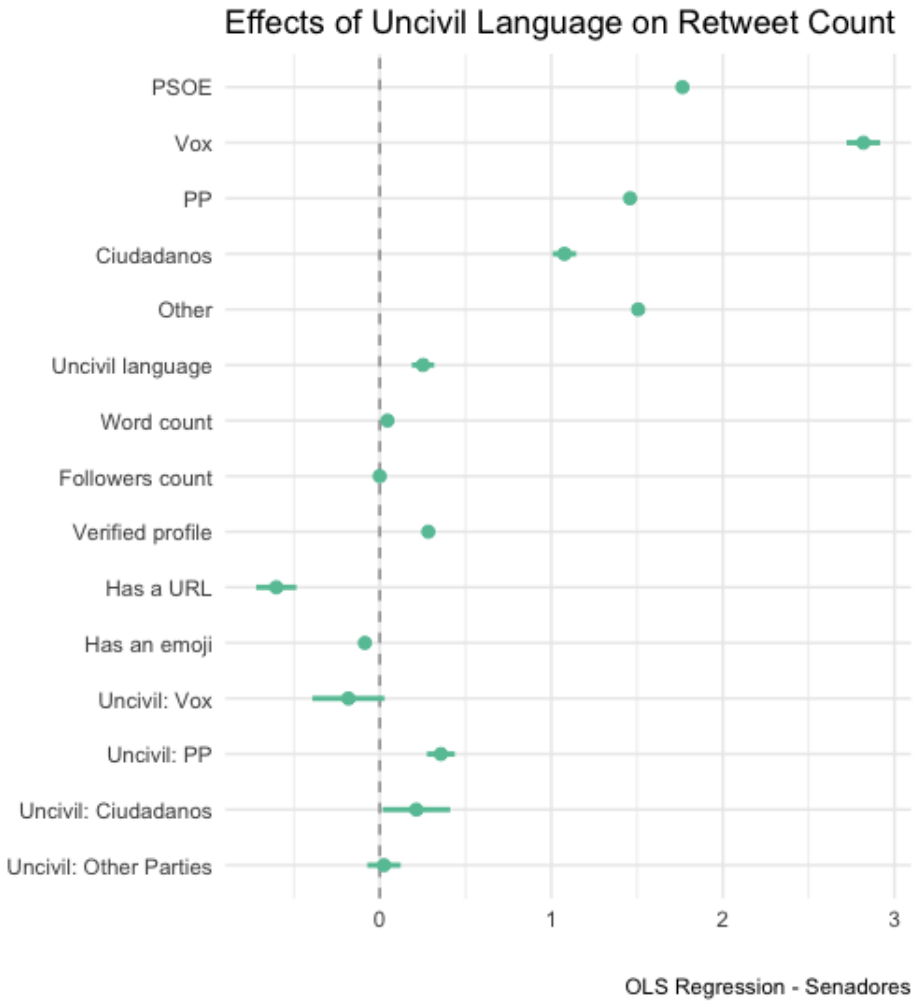


Figure 3.5: Coefficient Plot for Senators. Effect of uncivil language on retweet count. OLS Regression model with log-transformed dependent variable (retweet count). Parties shown are the largest party labels at the national level by Chamber, all others are aggregated under the "Other" label. We included an interaction between party label and using uncivil language. Uncivil language variable was built combining manual annotation and a naive Bayes classifier. Word count, Follower count, Verified profile, Has URL and Has an emoji have all been extracted from tweet metadata.

3.8 Conclusion

In this study, we addressed the questions of whether uncivil appeals boost party communication success in social media and if so, which parties benefit more from this kind of appeals. Through an analysis of the messages in Twitter timelines of all legislators in Spain we found that, overall, tweets containing uncivil language are more likely to propagate in the platform through the retweets of

users. Additionally, the results of our study show that challenger parties are generally more prone to using uncivil appeals on their social media posts than mainstream parties. However, contrary to our expectation, challenger parties seem to only have an incentive to be more uncivil than their mainstream competitor, but not necessarily when compared to other challengers. In other words, challenger parties only aim at capturing the overlapping voters they might share with their mainstream counterpart. So, if they are competing for moderate voters, there is an incentive to remain civil. Conversely, if they are trying to capture a more extreme voter relative to their mainstream counterpart, there is a stronger incentive to be uncivil.

This paper contributes to existing research on party communication by providing evidence from a naturalistic setting with real-life samples from legislators Twitter timelines. Moreover, the Spanish case provides a novel perspective from a multi-party system in recent transformation. However, the results of our study might not be generalised to different party settings. Even if electoral rules stay the same, party competition dynamics may change. In that sense, future research should delve into different party competition settings and analyze other data sources.

Our study relied on a binary measure of incivility, given the disagreement in the existing literature regarding a continuous dimension. A more granular approach to measuring incivility and other kinds of elite rhetoric would contribute to a deeper understanding of party communication strategies and audiences demands in contemporary politics.

Fake news or uncivil news? The Indirect Consequences of Uncivil Language on Social Transmission

4.1 Abstract

Existing research shows that high-arousal emotions seem to operate as motivational triggers for political engagement and political participation – which could include the sharing of online material. However, the role of emotion is underexplored when it comes to explaining why people share disinformation. The aim of this paper is to test the mechanism through which online political disinformation generates higher levels of engagement and spreads faster than other kinds of online news. The specific focus is a distinction between content and style. Sometimes conflated in practice because disinformation is often also couched in uncivil language, these are separated experimentally to identify their independent impact on arousal and, in turn, on sharing behaviour. This study is an online survey experiment in which we rely on self-reported measures of emotional reaction to news. The results boost our understanding of the psychology of disinformation transmission, inform policymakers seeking to reduce its spread, and assess the utility of physiological measures in political science research.

4.2 Introduction

When it comes to fake news stories online, we have seen it all. From a pizza parlour where high-ranking Democratic Party officials ran an alleged human trafficking ring to Pope Francis endorsing Donald Trump and the Israeli Defense Minister threatening to destroy Syria with a nuclear attack. Yet, these blatantly false stories became some of the most virally spread in social media. Exposure to online information is a part of daily life for almost everyone. People search for, create, disseminate, and exchange information online with unprecedented access (Cappella et al., 2015). Inconveniently, engagement with most types of mis and disinformation seems to be higher than with truthful stories, and social transmission is a primary mechanism for the diffusion of false facts. When it comes to the online environment, it is mostly humans and not bots who spread false content. False news spread farther, faster, deeper, and more broadly than the true facts (Vosoughi et al., 2018).

Governments around the world have been seeking to combat this phenomenon providing citizens with more information, labeling false or low quality information online or promoting media literacy campaigns. Recent evidence indicates why these attempts have enjoyed very limited success. People pass on news even knowing it to be inaccurate. Believing something is not a necessary condition for sharing it with others. So, correcting something does not stop it from circulating. If false news spread farther and faster, this is not only because "a lie can travel halfway around the world while the truth is lacing up its boots". It is also because, even if the truth catches up, the lie will keep traveling.

Recent studies highlight two particular motivations driving the sharing of disinformation: partisanship and 'need for chaos' (Osmundsen et al., 2021; Petersen et al., 2020). In polarised politics, people are ready to share news that will advance their partisan goals. Meanwhile, those hostile to the political order in general are ready to share news that might undermine or sow chaos in the system. Both motivations are clearly able to override considerations about the truth value of the information shared. These useful studies raise two additional questions, however.

First, is it truly a case of partisan or 'chaotic' predispositions overriding a disinclination to share false information? Can we assume that inaccuracy inhibits sharing but is trumped by other considerations, or are there reasons why people might actually be more inclined to share

disinformation than information? Even if accuracy does not have the strongest effect on sharing, we still want to know the form of that effect. Second, insofar as perceived accuracy does matter alongside these other motivations, what is the emotional basis for that effect? The evidence that sharing is driven less by an assessment of accuracy and more by affective partisan polarization or a discontent-driven need for chaos is already a sign that emotions are central to the social transmission of disinformation. Yet their role has been under-explored in this fledgling sub-field.

Arousal is a dimension of emotion that has been proven of particular relevance for its effect on engagement. High-arousal emotions seem to operate as motivational triggers for political engagement and political participation (Marcus, 2010). They could play a similar role in the online environment as well. Indeed, a couple of recent studies (Bakir & McStay, 2018; Vosoughi et al., 2018) build on the assumption that disinformation articles are more likely to diffuse than hard news or corrective articles because they trigger strong emotional reactions (high arousal). However, while there are sound reasons to believe this is true, there is not much empirical evidence supporting the assumption. And there is an important confounding issue: the distinctive use of uncivil language that is a common feature in much disinformation. This conflation of content and style in real-world 'fake news' makes it once again hard to identify the impact of perceived accuracy on reactions and transmission.

In this paper, then, we report a study that aims at disentangling the effects of accuracy from incivility – first on emotional reaction, and then on sharing behavior. Participants react to a series of stimuli news stories designed to score high or low on each of the two experimental variables. We thus test the hypothesis that emotional arousal plays a mediating role in the social transmission of disinformation, as well as identifying the sources of that arousal. Firstly, we expect to find that media articles using uncivil language arouse individuals more than the civil language of traditional or corrective media articles. Furthermore, we expect individuals to be more likely to engage in social transmission of arousing articles than of those which were less arousing, regardless of the actual and perceived accuracy of the articles.

4.3 Media, Information Processing and Arousal

The advancement of research in the field of information processing led to a recent paradigm shift in mass communication studies (for a detailed review see Lang et al. 2009; Potter & Bolls 2012). The

progressive adoption of this new paradigm focuses on understanding what goes on within the so-called "black-box" of mental processes during media consumption. One of the conceptual novelties suggested by this paradigm is the incorporation of emotions as a constitutive element of cognition. In fact, emotion and cognition are regarded as strongly integrated mental processes with distinct patterns of measurable physiological activity (Damasio, 1994; Potter & Bolls, 2012). Contrary to the classic stimuli-response approach, influenced by the behavioral tradition, the information processing approach understands cognitive processes as mediated by psychophysiological factors (Lang et al., 2009).

The study of emotions has been marked by a vast array of different standpoints regarding what actually constitutes an emotion. Despite the discordances, all perspectives consider that when we experience emotions there is a bodily reaction that varies according to its intensity (Bradley & Lang, 2007). This dimension of emotions known as arousal ranges from bodily activation to deactivation and it is usually defined as the physiological and psychological state of being awoken or alert (Russell, 1980; Larsen & Diener, 1992). It involves activation of the ascending reticular activating system in the brain, which mediates wakefulness, the autonomic nervous system, and the endocrine system, leading to increased heart rate and blood pressure and a condition of sensory alertness, mobility, and readiness to respond (Kandel & Pfaff, 2019). Arousal plays a relevant role in regulating consciousness, attention, alertness, and information processing (Pfaff et al., 2005).

This framework has proven relevant to understand media processing since emotions are associated with some of the main objectives of media messages, such as attracting attention, being remembered, entertaining and persuading (Ravaja, 2004). In particular, research shows that high-arousal content is remembered better than calm messages, controlling for message valence (Lang et al., 1995; Bolls et al., 2001). Lang 1999 found that television with even mild and brief emotional content elicited physiological arousal and increased attention.

4.4 Sources of Arousal in Disinformation

This prompts the question: what is there in disinformation to trigger arousal? Among possible answers to that question, some properties are more like correlates than direct consequences of falseness. For instance, given the emotional bases of partisan polarization (Iyengar & Westwood, 2015) and 'need for chaos' (Petersen et al., 2020), news stories that undermine the partisan enemy

or the system are likely to trigger arousal – but it is the negative valence rather than the falseness that arouses. However, that valence is often at its strongest and most vivid in disinformation – where truth imposes no restraint on negative portrayals.

This 'vividness' leads to a second example, central to the present study. Recent research has identified some linguistic patterns that are particularly prevalent in disinformation (Horne & Adali, 2017; Rashkin et al., 2017; Torabi Asr & Taboada, 2019). So-called fake news often use blunter language and more superlatives; they also use shorter words and are less "informationally dense". These features are often gathered under the heading of 'uncivil language'. As Mutz 2015 points out, incivility refers to the style rather than the substance of political discourse. According to her definition, uncivil discourse is that which violates the norms of politeness for a given context or interaction. Mutz's research is focused on the interaction between politicians in televised debates but the concept can be applied to the interaction between media and their audience - in the case of this paper, the lexical norms of journalistic articles.

There are two channels through which uncivil language is liable to generate arousal and engagement with an item of news (Geen, 1975; Mutz & Reeves, 2005). One is more immediate: emotionally intense words elicit immediate changes in central autonomic arousal and in specific appraisals related to pleasure and displeasure (Herbert et al., 2018). The subjective feelings that arise at this stage of word processing are not even consciously available for the readers themselves (Citron et al., 2016). The second involves more cognitive elaboration and results from the recognition of norm violation. A central tenet of evolutionary social psychology is that departures from the norm generate arousal. The added attention given to news that surprises readers in tone or style is just another example of that.

Again, there is no inherent link to disinformation in the sense that the same linguistic style could be used for accurate stories – as, for example, is often the case in tabloid newspaper reporting. Indeed, uncivil language is frequent in political discourse in general. However, as noted, it is disproportionately a feature of disinformation. This opens up the possibility of a spurious association between disinformation and arousal, driven by the style rather than the content of 'fake news'. Clearly it is necessary to hold the (in)civility of language constant in order to test for any independent effect of (in)accuracy on arousal.

Would we expect such an effect? A parallel mechanism suggests that we would. Truthfulness is

another powerful norm in communication, from everyday conversation (Grice, 1975) to journalistic reporting (Porlezza, 2019). Its violation should also occasion arousal. As with civility, the norms of accuracy in journalism have long been stretched such that it may need flagrant violations to trigger significant arousal.

4.5 Arousal and the Social Transmission of Disinformation

Social transmission of information has a central role in the new media environment. The reach of a given message is defined not only by direct exposure but also by the probability of its retransmission (Cappella et al., 2015). The central role that news sharing has for the contemporary media industry and media users led to the recent proliferation of research in the subject from different disciplines (Valenzuela et al., 2017).

Some studies refer to retransmission (Cappella et al., 2015), while others talk about news sharing (Valenzuela et al., 2017), virality (Berger & Milkman, 2012) or shareworthiness (Trilling et al., 2017). We refer to social transmission of information as a synonym of these concepts, focusing on the mechanisms behind news sharing. There are many reasons why someone might choose to share information with others. Capella et al. 2015 identify two broad sets of motivations: psychological motives and message factors. Since for this study we are interested in understanding the mechanism that drives the retransmission of a particular kind of message within an individual, we will discuss message factors only. This does not mean that we undermine the importance of psychological moderators.

There is no conclusive evidence on which mechanisms are more relevant to social transmission of information. However, a growing body of research suggests that when it comes to social transmission, automatic mental processes play a more relevant role than those involving conscious elaboration. That is, all other things equal, charged messages seem more likely to be shared than accurate messages. For instance, several studies established that arousing content is a relevant determinant of social transmission (Cappella et al., 2015). In fact, the level of activation that a given emotion evokes in an individual shapes social transmission, regardless of its valence (Berger & Milkman, 2012; Dang-Xuan et al., 2013). More recently, Brady et al. 2017 found that moral-emotional language in messages increases their diffusion in social media relying on a large sample of Twitter data. This is consistent with the broader evidence that high-arousal emotions operate as

motivational triggers for political engagement and political participation more generally (Marcus, 2000).

This also helps to explain the troublingly loose association between perceived accuracy and intention of retransmission. For instance, Swire-Thompson et al. 2020 find that even when people correct their false beliefs they might not update their attitudes accordingly. In the same direction, Petersen, Osmundsen and Arceneaux 2018 show that the motivations for sharing hostile political rumors are linked to signaling extreme discontent. Furthermore, they suggest that the rumors are not considered to be true but are regarded as effective resources to mobilize audiences. It is relevant to mention that both studies focus on a bipolar partisan context, but these results might change if replicated in countries with different political configurations (Swire-Thompson et al., 2020).

As previously mentioned, contrary to traditional media articles, so called fake news pieces supposedly trigger strong emotions, thus stimulating social transmission of false information. There is no evidence that there is a specific arousing mechanism inherent to this kind of news. Namely, researchers have not established if they actually produce such an effect and if so what implications it has on the attitudes and behavior of citizens. According to previous literature, there are distinct features that could make news articles more arousing and thus, increase the likelihood of retransmission. However, they are mostly independent of the factual accuracy of the piece.

We designed a study to disentangle the effects of accuracy from incivility – first on emotional response, and then on sharing behavior. We conducted an online survey experiment focusing on the mechanism itself.

4.6 Research Design

Since emotions are a constitutive element of cognition, it is likely that reading through news articles about contentious topics brings out identifiable emotional reactions in individuals. Given that we are interested in disentangling the effect of accuracy and incivility on emotional strength and, in turn how it affects sharing behaviour, this study is an online survey experiment designed to test the following hypotheses:

H1a: News stories written with uncivil language produce higher levels of arousal than news stories written with civil language.

H1b: Highly inaccurate news stories produce higher levels of arousal than highly accurate news stories

H2: Incivility is a more powerful predictor of arousal than perceived inaccuracy

H3: News stories that produce higher levels of arousal are more likely to be shared

H4: News stories that are highly uncivil are more likely to be shared

We designed a 2 x 2 experiment manipulating inaccuracy and civility according to the two axes in figure 4.1. This variation was meant to cover a baseline for comparison that would capture the nature of a factual piece in a mainstream media outlet and different types of bad-quality information that were previously discussed (i.e so-called 'fake news', sensational press and poorly fact-checked journalism).

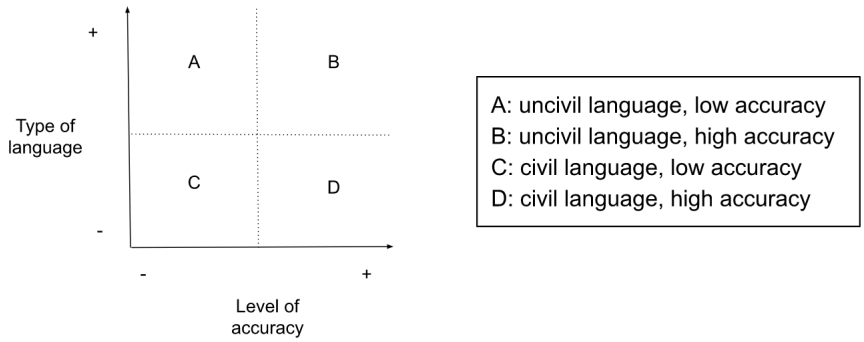


Figure 4.1: Diagram guiding the vignette design

The construction of the experimental vignettes proved to be one of the most challenging tasks within the experiment. Regarding the level of accuracy, we built upon factual data, mainstream news articles and official statistics, whereas for the level of incivility we relied on real blog posts, reddit *fori* and tabloid newspapers. We mixed elements of both dimensions and covered all the possible combinations displayed in Figure 4.1 for three different topics: climate change, legalization of cannabis consumption and immigration. The extreme cases were clearer (i.e the baseline factual

article depicted in quadrant D and the strictly fake news with uncivil language depicted by quadrant A in Figure 4.1). Conversely, the more ambiguous cases represented by quadrants B and C required a more nuanced approach. In order to build vignettes in quadrant C, on the accuracy axis we had to rely on factual data and modify it in such a way that was noticeable regardless of respondent's political interest or general knowledge. So, when tweaking the accuracy of the vignettes we had to keep a balance between conveying an overall accurate story with minor mistakes that respondents might not detect and creating an alternative narrative altogether, crossing the line into the territory of a different quadrant. For quadrant B, the difficulty lied in balancing the type of language so it was uncivil enough but was not so aggressive as to undermine the credibility of the article altogether.

Yet another challenge posed by the process of building the vignettes was that of the topics covered in the articles. We had to make sure to select topics that were contentious and salient enough to produce an effect that is normally difficult to detect with short pieces of text even under the best experimental conditions, but also the type of language used had to convey a framing that was consistent along the same ideological line, so we could control that our intervention was not confounded with ideology. We finally opted for selecting three different salient topics that had been recently covered in the news and present participants with one of the variations of all three topics. In this way, we were able to control whether a particular topic was driving the detected effects. Additionally, this decision also contributed to obtain a better powered sample through the repeated observations per participant.

To test our hypotheses, the online survey experiment we designed exposed each participant to three different news articles on three different topics in random order. After reading each piece participants were asked to report their emotional reaction and answer behavioral intention questions. The pieces of news they received were also randomly drawn from the larger pool of experimental vignettes described above. In short, the vignettes were designed with variation along two axes: type of language (uncivil/civil) and level of accuracy (accurate/inaccurate).

Data and Methods

We relied on a within-subjects design and analyzed our data using linear mixed models and OLS by topic subset. Linear mixed models allow to control for the non-independence among repeated

observations by individual, adding random effects per individual to the model. Sub-setting the data by topic helps overcome the issue of dependence among observations given that each subset only contains observations for each individual participant at one given point. Our main dependent variables tap into the immediate emotional response to the stimuli and respondent's behavioral intention to share the articles with others.

Our study had three different main dependent variables: arousal (or emotional strength), willingness to share online and willingness to discuss each article with friends and family. To assess emotional reactions to the experimental vignettes we relied on a validated questionnaire on discrete emotions and the intensity of that emotion (Harmon-Jones et al., 2016). At the analysis stage, we constructed our variable for emotional response with an average of the scores per treatment per respondent for each emotion covered in the questionnaire, in which emotions associated with low-arousal were reverse coded. To tap into respondents behavioral intentions we also asked them about their willingness to share each piece of news on their social media or their willingness to discuss the topic with their friends and family on a scale from 0 to 10.

We also included relevant control variables and demographics. The control variables included political interest (0-10), left-right position (0-10), prior related attitudes (in a 6 point scale ranging from "Strongly disagree" to "Strongly agree") and demographic variables. The demographic questions such as age, level of education and gender also served as distractor from the potential priming effect of asking about prior attitudes.

We fielded our online survey experiment through Prolific. The experiment was conducted in August 2022 among British citizens (n=1000). Our sample was highly educated and overall more left-leaning. More than 50 % of respondents had left the education system with 20 years old or older. Regarding left-right positioning, on a scale from 0 to 10 were 0 represented extreme left, the average was 4.02 (sd = 1.97). 64% of participants in our sample identified as female and 35% as male.

4.7 Results

Table shown in figure 4.2 reports results from linear mixed models with arousal as dependent variable. This model works as a test of our first hypothesis. On average participants reported low levels of arousal after seeing the treatments, with a mean of 2.34 (SE = (0.02). We found

null effects of our treatments on arousal (H1a and H1b). Overall, the treatments had little or no effect on the levels of arousal of participants. The uncivil and inaccurate treatment (b = -0.04 SE = 0.02) and the uncivil and accurate treatment (b = 0.04 SE = 0.02) were not statistically significant. The civil and inaccurate treatment (b = 0.07 SE = 0.02) had a slightly larger and significant effect. However, given the results just discussed, this might be due to specifics of the vignette design. In fact, arousal was predicted by participant's political interest (b = 0.1 SE = 0.02), related prior attitudes (b = 0.20 SE = 0.04), positioning in the left-right scale (b = -0.03 SE = 0.01) and the topic of the article (b = -0.14 SE = 0.02 when the article was immigration and b = -0.16 SE = 0.02, with climate change as reference category). Consequently, we cannot claim that incivility is a more powerful predictor of arousal than inaccuracy (H2) because none of them were, at least with our design. These results are also consistent with the findings stemming from OLS models conducted on three separate subsets of the dataset by topic.

	Arousal		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1.43	1.05 – 1.82	**<0.001**
Treatment [uncivil/inaccurate]	-0.04	-0.10 – 0.02	0.177
Treatment [uncivil/accurate]	0.04	-0.02 – 0.10	0.203
Treatment [civil/inaccurate]	0.07	0.01 – 0.12	**0.019**
Interest	0.10	0.06 – 0.14	**<0.001**
Left-Right	-0.03	-0.05 – -0.01	**0.006**
Age	-0.00	-0.00 – 0.00	0.857
Education	0.02	-0.01 – 0.04	0.174
Topic [immigration]	-0.14	-0.19 – -0.10	**<0.001**
Topic [cannabis]	-0.16	-0.21 – -0.12	**<0.001**
Priors	0.20	0.12 – 0.29	**<0.001**
Random Effects			
σ^2	0.23		
$\tau_{00\ ResponseId}$	0.32		
ICC	0.58		
$N_{\ ResponseId}$	991		
Observations	2973		
Marginal R^2 / Conditional R^2	0.052 / 0.598		
AIC	5801.060		

Figure 4.2: Fixed effects model with treatments as independent variables

On average participants reported a mean willingness to share online the news they were exposed to of 2.85 (SE = 2.7) and a mean willingness to discuss the news they were exposed to with friends and family of 5.94 (SE = 0.08). Tables shown in figures 4.6 and 4.5 report results from linear mixed models with our two different measures of social transmission as dependent variable: willingness to share online and willingness to discuss the article with friends and family. In the models displayed in figures 4.6 and 4.5 we found evidence to support both the expectations that arousal increases willingness to share news stories and also that news stories with uncivil language are more likely to be shared. However, given the results described above, we focus on the relevance of incivility, as we cannot account for what our variable for arousal is exactly measuring. When it comes to sharing, accuracy had no significant effects, as described in figure 4.3. Instead, incivility does seem to play

a relevant role in social transmission of news, both online and regarding in-person discussion with friends and family. In fact, as shown in figure 4.4, incivility predicted social transmission across all models, supporting the thrust of our argument (b=0.24 SE = 0.08 for the model with discussing with friends and family as dependent variable in 4.5 and b=0.24 SE = 0.07 for the model in 4.6, with sharing online as main dependent variable). Arousal showed the largest significant effect (b= 1.22 SE = 0.06 for sharing online and b= 1.38 SE = 0.07 for discussing with friends and family) even if the baseline levels of arousal were low, as previously mentioned. There is also a significant effect of political interest (b=0.16 SE = 0.07 for sharing online and b= 0.60 SE=0.08 for discussing with friends and family) and a small but significant negative effect of age.

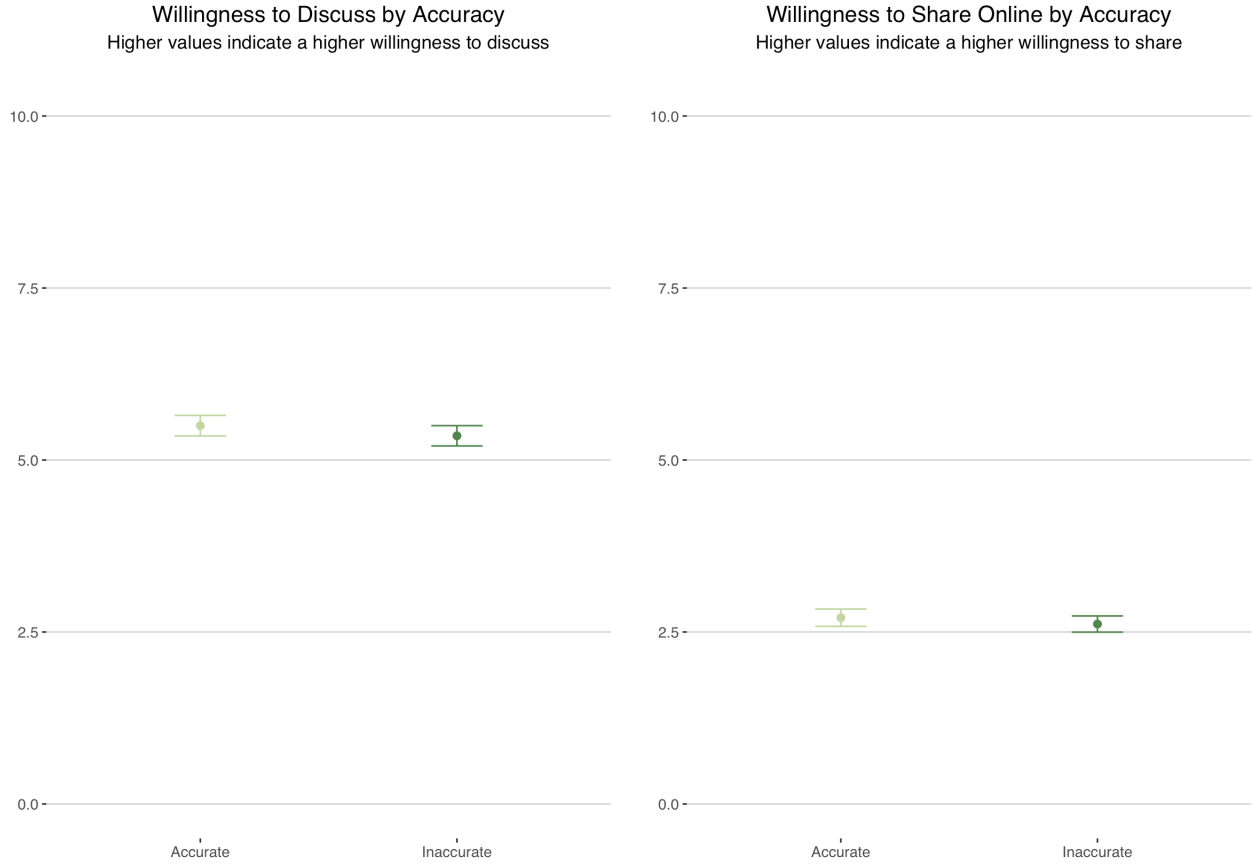


Figure 4.3: Social transmission by treatment accuracy. Plot shows both measures of social transmission (willingness to discuss the article with friends and family and willingness to share online by accuracy of treatment.

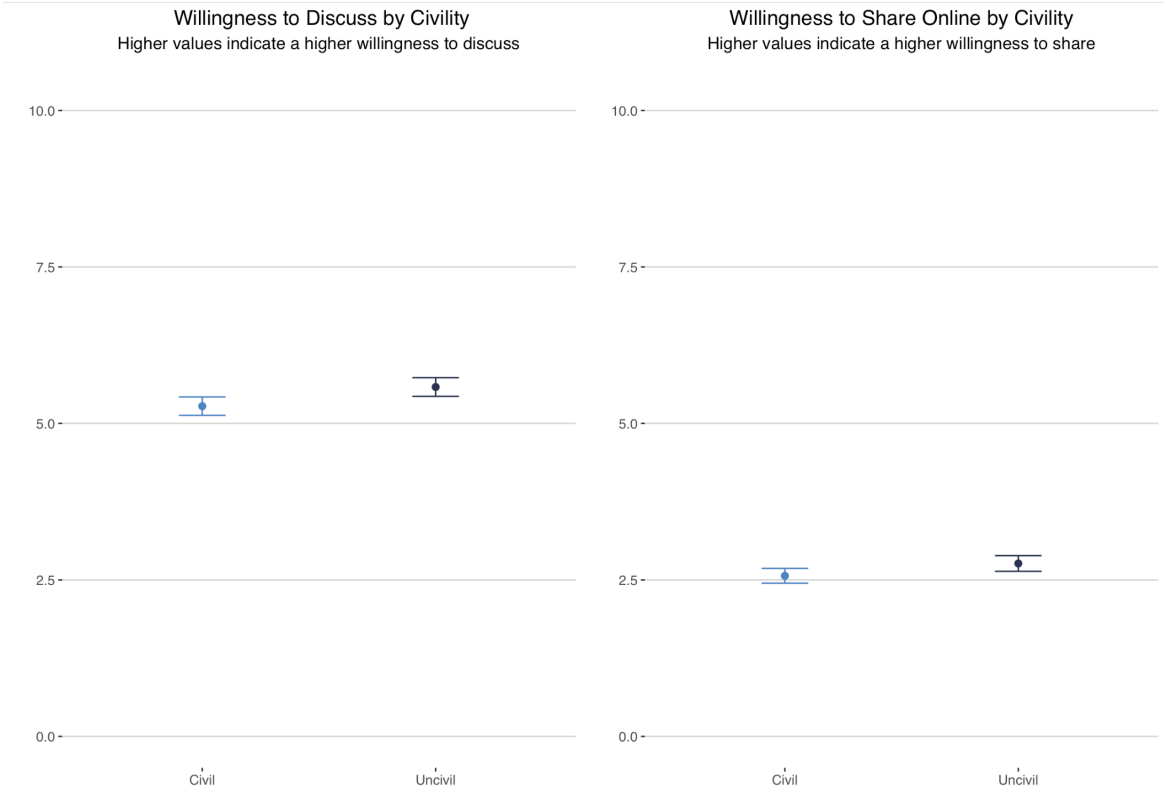


Figure 4.4: Social transmission by treatment civility. Plot shows both measures of social transmission (willingness to discuss the article with friends and family and willingness to share online by civility of treatment.

Although we did not have a prior expectation regarding the difference between sharing online and sharing the article with friends and family, we found some interesting results. Participants presented a considerably higher mean intention to discuss news with friends and family than sharing the same pieces online. This might reflect that people were to some extent interested in the news they read, but aware that posting about articles about contentious topics and/or containing uncivil language online could imply a backlash. Whereas sharing in a safe space does not have the same consequences.

We speculate that our results might have also been affected by how we captured participants level of arousal. The reliability of self-reported measures of arousal are contested in the literature given that arousal is strictly linked to unconscious psycho-physiological reactions. Individuals are somewhat better at detecting discrete emotions, and that is why we relied on the Discrete Emotion Questionnaire (Harmon-Jones et al., 2016). However, the questionnaire is not per se

designed to measure arousal, but we selected it as a sub-optimal way to tap into the dimension of emotional strength in an online survey experiment. The ideal way to estimate arousal would be through physiological measures. A natural progression of this work is to analyse the same research question with a different experimental framework combining self reported items with physiological measures.

	Willingness to Discuss with Friends and Family			
Predictors	*Estimates*	*SE*	*CI*	*p*
Intercept	0.29	0.76	-1.19 – 1.77	0.703
Incivility	0.24	0.08	0.09 – 0.39	**0.002**
Innacuracy	-0.13	0.08	-0.28 – 0.03	0.104
Arousal	1.38	0.07	1.24 – 1.52	**<0.001**
Interest	0.60	0.08	0.45 – 0.76	**<0.001**
Left-Right	-0.07	0.04	-0.15 – 0.01	0.067
Age	-0.02	0.01	-0.03 – -0.01	**<0.001**
Education	0.02	0.05	-0.08 – 0.11	0.750
Topic [immigration]	0.31	0.17	-0.02 – 0.64	0.068
Topic [cannabis]	0.09	0.08	-0.07 – 0.26	0.260
Priors	0.05	0.08	-0.11 – 0.21	0.533
Random Effects				
σ^2	3.35			
$\tau_{00\ \text{ResponseId}}$	4.65			
ICC	0.58			
$N_{\ \text{ResponseId}}$	991			
Observations	2973			
Marginal R^2 / Conditional R^2	0.188 / 0.660			

Figure 4.5: Fixed effects model with treatments as independent variables

	Willingness to Share Online			
Predictors	*Estimates*	*SE*	*CI*	*p*
Intercept	-1.23	0.66	-2.51 – 0.06	0.062
Incivility	0.24	0.07	0.11 – 0.37	**<0.001**
Innacuracy	-0.00	0.07	-0.13 – 0.13	0.999
Arousal	1.22	0.06	1.11 – 1.34	**<0.001**
Interest	0.16	0.07	0.02 – 0.29	**0.023**
Left-Right	-0.02	0.03	-0.09 – 0.05	0.573
Age	-0.02	0.01	-0.03 – -0.01	**0.001**
Education	-0.05	0.04	-0.14 – 0.04	0.254
Topic [immigration]	0.26	0.15	-0.02 – 0.55	0.072
Topic [cannabis]	0.48	0.07	0.35 – 0.62	**<0.001**
Priors	0.72	0.07	0.58 – 0.86	**<0.001**
Random Effects				
σ^2	2.37			
$\tau_{00\ ResponseId}$	3.58			
ICC	0.60			
$N_{\ ResponseId}$	991			
Observations	2973			
Marginal R^2 / Conditional R^2	0.153 / 0.663			

Figure 4.6: Fixed effects model with treatments as independent variables

4.8 Conclusion

This paper set out to better understand the role of emotion in explaining why people share different sorts of information. In study 1 we tested the mechanism through which political disinformation generates higher levels of engagement and spreads faster than other kinds of online news. Although the results for the effect of incivility or inaccuracy on arousal were unmistakably null, we obtained some interesting findings regarding the predictors of sharing behavior. When it comes to sharing, both online and within a close circle of people, incivility had a significant effect. And this finding

proved to be consistent throughout all the models we ran. Of course, the null findings on the first set of hypotheses could be due to the fact that we overlooked relevant variables involved in the mechanism. However, we speculate that there is a possibility that our measure of arousal did not capture the unconscious, purely physiological aspect of the construct. Given that lab experiments involving physiological measures are costly, time consuming and were even impossible to conduct while the COVID-19 lock down regulations were in place, we decided to first test our hypothesis through the online survey experiment presented in study 1. We conducted our survey experiment on a Prolific convenience sample. Although there is much debate about the quality and and representativeness of samples reached throughout crowdsourcing platforms, it has been demonstrated that they can be reliable and cost-effective sources of data for research purposes (Peer et al., 2017). However, we do not neglect the fact that due to the nature of the mechanism we were testing, the ideal scenario would require us to measure arousal through better suited means for the task. As extensively described in the theoretical section of this research, when people experience emotions there is a correlated bodily reaction that varies according to its intensity and that can be precisely measured with the appropriate lab equipment (Bradley & Lang, 2007).

Conclusion

This dissertation has set out to answer questions about the the psychology of public reactions to political communication. Of course, a single thesis cannot encompass all angles of the key questions in the field and this one certainly did not pretend to do so either. Instead, the intention was to select various important phenomena that could provide the opportunity to reflect on some of the most pressing issues in political communication. Specifically, each of the chapters comprised in this dissertation focused on a different related concern within the research field, relying on different methodological strategies. However, the guiding principle interconnecting the studies in this dissertation is the idea that unconscious thoughts and feelings shape judgment, preferences, attitude change, and decision-making (Lodge & Taber, 2007). As Redlawsk (2006) put it, politics is about feeling. And not in the sense that we are passionate fools blinded by our emotions. Rather, as it has been previously described throughout this thesis, meaning that emotion and cognition are two constitutive integrated elements of the same mental processes (Damasio, 1994; Potter & Bolls, 2012).

People are fearful in the context of a health crises posed by a pandemic, agitated by uncivil appeals of politicians or aroused by political news. Clearly, these examples that were well developed throughout the dissertation and crudely summarised here are just broad-brush statements to make a point. Reality is much more complex and multidimensional, but the study of political attitudes and behaviour cannot be isolated from the understanding and assessing affective states.

Chapter one analysed the effects of communicating uncertainty to the general public in a context in which stakes were high, like the roll-out of the COVID-19 vaccination programme. I designed an online survey experiment to manipulate different measures of uncertainty about safety

and effectiveness of COVID-19 vaccines. Participants were exposed to messages that mimicked potential alternatives of governmental communication, containing the same point estimates but varying whether they were accompanied by a confidence interval or a confidence interval and an explanation. Our study suggested that not only acknowledging uncertainty does not decrease trust in the vaccine or the reported statistics, but also there are long-term benefits to being transparent. By no means can a single study with limited scope pretend to directly influence policy. Besides, I also acknowledge the limitations of our study as an online survey experiment and will address those by the end of this section. However, our results are consistent with recent findings in the field, and so I draw from the bigger picture to reflect on potential implications.

The recent evidence on the matter speak directly to political elites that often disregard the ethical dimension of adopting a radical transparency approach only out of fear of potential reputational damage. Our study contributes to the discussion trying to dismantle this unfounded fear and favoring political communication strategies that put the ethical concerns and effectiveness of relevant policy messages in the center of the debate. Notwithstanding, we recognise that the scope of this recommendation is limited to the particular context of a public health crisis. Further research is needed to determine whether citizens would also readily accept uncertainty on other domains like economic indicators such as inflation or environmental policy where there is less clarity regarding the exact long-term gains at the individual level. There is also room for questioning what would happen if the point estimates had been worse, but due to ethical considerations we were not able to contemplate those iterations of the experiment in our study, that was conducted almost in real-time as the debate about the roll-out of vaccination programmes was ongoing. This would also be a fruitful area for further work. Finally, and in the hopes that our findings are not read as guidelines for ripping political benefits, we speculate that the transparency approach would lose effectiveness in building long-term trust if politicians reported really wide confidence intervals at Stage 1 only to ensure that the update will be within them.

Chapter two addressed the concern over uncivil party communication strategies that challenge democratic attitudes. This study relied on a large dataset collected through the Twitter timelines of all parliamentary representatives in Spain. At a first stage, we classified tweets into civil or uncivil through supervised learning techniques and later applied statistical analyses to answer the question of whether some parties more likely to rely on and benefit from uncivil communication.

Our data suggested that, overall, messages written with uncivil language are more likely to spread in social media. Most importantly, challenger parties generally tended to rely on uncivil appeals on their posts more often than mainstream parties. Especially when the party competition dynamics offered greater incentives to do so. Suggesting that whenever there was a chance of standing out by distancing from mainstream competitors, challenger-party politicians saw and took the opportunity of relying on uncivil appeals. By this we do not mean that challenger-party politicians rely on evidence-based methods to define their communication strategy on a tweet-by-tweet basis. Rather, this strategy is consistent with their partisan identity and thus, provides positive feedback from their target audience. So, they probably have an intuitive impression that, broadly, uncivil appeals work on their favour and increasingly more scholarly evidence that backs their gut feeling.

As social science researchers, our investigations cannot undermine potential ill-intentioned uses of our findings. There is a concern that publishing these sort of evidence might strengthen the intuitions of political leaders who threaten democratic institutions. However, we believe that research in these lines will not change the practices of challenger parties and might instead help other parties fight back with the right tools, such as exploring other avenues that also lead to capturing voters attention without the need of being uncivil. Given the scope of our study, we compared being uncivil with being civil, but consequently left out of the picture other ways of generating audience engagement like, for instance, humour or including audiovisual content. In terms of the methodological approach chosen for this study, we would like to address some of the critiques that research relying on social media has received in recent years. Firstly, there is always the issue of selection. Many studies that analyse big data collected from social media get their data through application programming interfaces (APIs) set by the social media companies themselves. In the case of Twitter, given that the volume of data published in the platform is so large, even the commands that provide researchers with data in real-time have to filter out posts to some extent. Allegedly, this filtering is at least as good as it if were random. However, given the lack of disclosure of social media companies algorithms, there is always a shade of doubt in this regard. We overcame that common limitation by design, relying on data collected from politicians entire timelines. Nevertheless, our sample only included politicians who are active on social media and we had no reasonable way of contemplating those politicians who did not have a public Twitter account at the time of data collection. Consequently, our findings only apply to the online sphere.

Another relevant concern voiced by researchers in the field is that focusing on just one social media platform might give a too narrow idea of a much more complex online ecosystem. There has been a call for exploring other platforms, particularly criticising studies focusing on Twitter data just because Twitter has provided the technical infrastructure to collect their public data more easily than other. While it is true that broadly speaking there is a need for expanding and diversifying the breadth of the social media and apps we study as a field, that does not undermine the need to keep on producing research in each individual platforms. Not only because cutting some slack on researchers with little funding such as the case of early career academics and global-south scholars would be only fair, but also because investigations from independent researchers contribute to auditing social media companies.

Chapter three reported two interconnected studies with the objective of disentangling the effects of accuracy from incivility on emotional strength and on sharing behavior. Study 1 tested the mechanism through which political disinformation generates higher levels of engagement and spreads faster than other kinds of online news through an online survey experiment. Our findings suggest that when it comes to sharing behavior, incivility had a significant effect predicting both sharing news online and within a close circle of friends and family. This results were consistent throughout all the models we conducted and also with the findings of our paper in the previous chapter. This makes us wonder, if incivility is tied with breaking the norms of a given context but is increasingly being used as a political strategy in said context, could it reach a point in which the border between what is considered civil and uncivil gets pushed further away from its current position? In order to answer that question it would be necessary to understand exactly which features of incivility make people more responsive to it.

Although study 1 went in that direction, we faced some constraints due to fact that we relied on an online survey experiment. At this point we will briefly address the broad methodological concerns on survey experiments in general. Sniderman 2018 already summarised it in a sharp review: "survey experiments are radically imperfect". He calls for modesty when using the word experiment as a brand of scientific validity because survey experiments are limited in three main ways. Firstly, regarding what the so-called treatments actually are: more often than not just a variation of how information is presented. Second, the brevity of survey experiments: they are very short in duration. Sometimes just because of cost constraints, but some other times because

survey length and question complexity can result in respondent fatigue, limiting the depth and breadth of phenomena that can be studied through surveys experiments. And last but not least, there is the issue of measurement. Survey experiments often rely on unvalidated single questions to tap into complex social science constructs. Coming back to the limitations of our particular study, our measurement of arousal was sub-optimal for the task. Not because we relied on an unvalidated single item. In fact, we chose a validated scale meant to capture discrete emotions in order to overcome or at least buffer the measurement error. However, there is a good chance that a lot was lost in the translation of discrete emotion strength to arousal.

That is the reason why we propose that future research should look into improving the design with a lab experiment combining the self-reported responses with physiological measures. By doing so, we will be able to capture the conscious and the unconscious aspects of arousal. Moreover, we would also be able to compare the two and better understand the differences between them. Of course, despite the fact that these methods have been used for a long time in behavioural sciences, they are still relatively new to the discipline. Meaning that social science research relying on these innovative measurements should be extra cautious with stretching the interpretation of results and especially with stretching the implications of scholarly research into broad social science categories. Social science research reliant on variables measured with psychophysiology equipment, should aim at only testing hypotheses with variables that can be reliably captured by this sort of equipment. Furthermore, it would also be advisable to perform reliability tests and robustness checks accordingly.

As stated at the very beginning of this dissertation, the technological and political changes of recent years have generated a debate about political communication and the health of democratic politics. There are concerns about whether politicians are telling citizens the truth or distracting them from it, and about whether citizens can handle the cognitive and emotional complexity of the political world that they face. Commentators and pundits talk of crisis, panic and post-truth politics. Scholars are more ambivalent, providing some more balanced evidence and also reminding us that there was no golden age in terms of polticians' honesty or citizens' rationality. The moral panic scenario might hurt more than it helps, especially considering the effects of specific discrete emotions and its correlates on political behavior. However, it might still be worth actively delegitimising discourses that go against democratic values and practices at the individual

level and within our circles of reach. Especially in the case of public figures, whose reach is significantly larger. In line with these debates, some new questions will arise and some of the long-standing questions will have new answers. Accordingly, future research in the field could look into innovative combinations of data sources to gather evidence to answer old questions and replicate old designs to gather renewed evidence. Hopefully, this dissertation has contributed to the field not by establishing definitive answers to the big questions it explored, but rather as a part of the progression of trials (Sniderman, 2018) entailed in the production of scientific knowledge.

Appendix A

Dependent Variables

The OLS models below show the effect of acknowledging uncertainty, and a number of control variables on six outcome variables:

Stage 1 Dependent Variables

- interest in having the vaccine.
- certainty
- perceived safety
- perceived effectiveness
- self-informedness
- support for approving the vaccine
- perceived transparency

Stage 2 Dependent Variables

In the main body of the manuscript, we report OLS models using the difference in pre and post-update values: • Drop in vaccination intentions

- Drop in certainty
- Drop in perceived safety
- Drop in perceived effectiveness

In this appendix, we report additional OLS models using the simple post-update values as DVs:

- interest in having the vaccine
- certainty
- perceived safety
- perceived effectiveness

Control Variables

All models include the following control variables: • Demographics: Gender, age

• Vulnerability: Likelihood of having contracted the virus, and being or 'high risk' or having a 'high risk' family member. This variable was based on the highRisk variable and took on a value of 1 if the respondent and/or at least one family member is 'high risk', 0 otherwise.

• Political opinions: Support for Boris Johnson. We also asked for feelings for a number of other individuals including Chief medical officer for England Chris Whitty but omitted those due to missing data. We excluded a Brexit variable as it was highly correlated with support for the Prime Minister.

• Realistic threat: Three of five measures of realistic threat, adapted from Kachanoff et al. 2020: How serious a threat is COVID-10 to 'your personal health', 'your financial situation', and 'day-to-day life in your local community'. The remaining two measures, how serious a threat is COVID-10 to 'the health of the British people', 'the British economy' were included in the survey but not in the OLS models. Kachanoff and colleagues designed these questions for a US context (see Kachanoff et al. 2020). They were also used in the Pew Research Poll in March 2020 (Wave 63.5, March 10, 2020). To adapt them to a British sample, we changed the word 'American' to 'British' where necessary. We combined the five measures in an additive scale of perceived 'realistic threats'. However, due to lower than expected internal consistency (a=.5) we chose to use some of the individual items rather than the scale for the OLS regression models.

• Symbolic threat: Two of five measures of symbolic threat, adapted from Kachanoff et al. 2020: How serious a threat is COVID-19 to 'British values and traditions', and 'The rights and freedoms of British people'. One other measure was included in the survey but not the OLS regression

models: how serious a threat is COVID-19 to 'British democracy'. The remaining two measures from Kachanoff and colleagues' scale were less suitable for a British context and, therefore, left out of the survey ('What it means to be American', 'The maintenance of law and order in America'). Again, we made minor changes to the wording to fit a British sample. We changed 'The rights and freedoms of the U.S. population as a whole' to 'The rights and freedoms of British people'. We combined these measures in an additive scale of perceived 'symbolic threats'. Again, internal consistency was lower than expected (a=.599), so we chose to use individual items rather than the scale. To avoid issues of multicollinearity we computed the correlations between each indicator. We found a high correlation between two of the five indicators of realistic threat: perceived threat to respondents' 'personal health' and perceived threat to 'the health of British people' (.47). Therefore, we removed the latter from the OLS models. We also found a high correlation between two of the indicators of symbolic threat: perceived threat to 'British values and traditions' and to 'British democracy'. Here, too, we removed the latter from the model. Finally, we found a high (.53) correlation between one of the measures of realistic threat ('the British economy') and one of the measures of symbolic threat ('British values and traditions'), so we removed the 'threat to the British economy' variable.

• Political opinions: We included one political variable tapping into how much respondents liked Prime Minister Boris Johnson.

Sample Characteristics

Sample characteristics	df (N = 2,165)
Gender	
Male	804 (37.14%)
Female	1,342 (61.99%)
Age	
min	18
max	88
mean	37
Ethnicity	
White European	1,787 (82.54%)
South Asian	123 (5.68%)
Other Asian	83 (3.83%)
African/Arab/Caribbean	0 (0.00%)
Mixed	60 (2.77%)
Other	37 (1.71%)
Education	
Degree	1,230 (56.81%)
A-level	610 (28.18%)
GCSE	252 (11.64%)
No qualifications	19 (0.88%)
Brexit views	
Remain	1,557 (71.92%)
Leave	431 (19.91%)
Employment status	
Employed (full time)	964 (44.53%)
Employed (part time)	297 (13.72%)
Self-employed	158 (7.30%)
Student	308 (14.23%)
Full time parent/carer	117 (5.40%)
Not employed	207 (9.56%)
Retired	114 (5.27%)
Had Covid	
Certain I’ve had it	97 (4.48%)
Pretty sure I’ve had it	169 (7.81%)
I’ve probably had it	332 (15.33%)
Not at all sure	293 (13.53%)

Sample characteristics	df (N = 2,165)
I probably haven't had it	318 (14.69%)
Certain that I haven't had it	324 (14.97%)
High risk	
I'm high risk	133 (6.14%)
I am, and at least one more family member	191 (8.82%)
At least one family member	977 (45.13%)
None of my family	826 (38.15%)
Trust in the NHS (0 to 6)	
mean	4.87
sd	1.35
Trust in the companies making the vaccines (0 to 6)	
mean	4.01
sd	1.6
Trust in the UK Government (0 to 6)	
mean	4.01
sd	1.6
General vaccination intentions (0 to 6)	
mean	4.62
sd	1.8

Balance Table	no information	point estimate	confidence interval	ci and explanation
Gender				
Male	204 (37.92%)	205 (37.68%)	190 (35.12%)	205 (37.82%)
Female	331 (61.52%)	334 (61.40%)	349 (64.51%)	328 (60.52%)
Age				
min	18	18	18	18
max	79	88	76	80
mean	36	36	37	37
Ethnicity				
White European	438 (81.41%)	450 (82.72%)	452 (83.55%)	447 (82.47%)
South Asian	38 (7.06%)	21 (3.86%)	29 (5.36%)	35 (6.46%)
Other Asian	16 (2.97%)	24 (4.41%)	25 (4.62%)	18 (3.32%)
African/Arab/Caribbean	22 (4.09%)	19 (3.49%)	16 (2.96%)	18 (3.32%)
Mixed	16 (2.97%)	15 (2.76%)	14 (2.59%)	15 (2.77%)
Other	8 (1.49%)	15 (2.76%)	5 (0.92%)	9 (1.66%)
Education				
Degree	313 (58.18%)	315 (57.90%)	300 (55.45%)	302 (55.72%)
A-level	151 (28.07%)	147 (27.02%)	156 (28.84%)	156 (28.78%)
GCSE	55 (10.22%)	69 (12.68%)	64 (11.83%)	64 (11.81%)
No qualifications	5 (0.93%)	5 (0.92%)	3 (0.55%)	6 (1.11%)
Brexit views				
Remain	392 (72.86%)	403 (74.08%)	378 (69.87%)	384 (70.85%)
Leave	105 (19.52%)	100 (18.38%)	115 (21.26%)	111 (20.48%)
Employment status				
Employed (full time)	253 (47.03%)	239 (43.93%)	243 (44.92%)	229 (42.25%)

Balance Table	no information	point estimate	confidence interval	ci and explanation
Employed (part time)	64 (11.90%)	75 (13.79%)	81 (14.97%)	77 (14.21%)
Self-employed	46 (8.55%)	38 (6.99%)	36 (6.65%)	38 (7.01%)
Student	68 (12.64%)	82 (15.07%)	74 (13.68%)	84 (15.50%)
Full time parent/carer	23 (4.28%)	34 (6.25%)	30 (5.55%)	30 (5.54%)
Not employed	54 (10.04%)	55 (10.11%)	52 (9.61%)	46 (8.49%)
Retired	30 (5.58%)	21 (3.86%)	25 (4.62%)	38 (7.01%)
Had Covid				
Certain I've had it	22 (4.09%)	29 (5.33%)	17 (3.14%)	29 (5.35%)
Pretty sure I've had it	49 (9.11%)	45 (8.27%)	31 (5.73%)	44 (8.12%)
I've probably had it	81 (15.06%)	75 (13.79%)	87 (16.08%)	89 (16.42%)
Not at all sure	73 (13.57%)	75 (13.79%)	71 (13.12%)	74 (13.65%)
I probably haven't had it	85 (15.80%)	85 (15.62%)	81 (14.97%)	67 (12.36%)
Certain that I haven't had it	75 (13.94%)	86 (15.81%)	93 (17.19%)	70 (12.92%)
High risk				
I'm high risk	37 (6.88%)	37 (6.80%)	30 (5.55%)	29 (5.35%)
I am, and at least one more family member	45 (8.36%)	47 (8.64%)	48 (8.87%)	51 (9.41%)
At least one family member	235 (43.68%)	250 (45.96%)	227 (41.96%)	265 (48.89%)
None of my family	211 (39.22%)	198 (36.40%)	230 (42.51%)	187 (34.50%)

Futher analyses

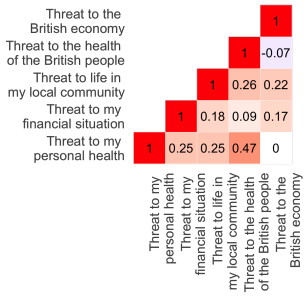


Figure A.1: Correlation Coefficients of the Realistic Threat Variables

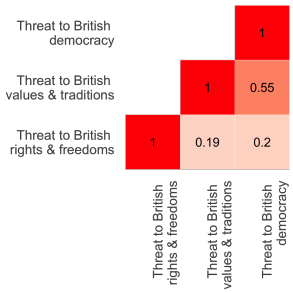


Figure A.2: Correlation Coefficients of the Symbolic Threat Variables

Stage 1 DVs – Effect of acknowledging uncertainty on immediate interest in the vaccine

	safety track			effectiveness track		
	(1) want vaccine	(2) certain	(3) think it's safe	(4) want vaccine	(5) certain	(6) think it's effective
confidence interval[1]	0.15	0.04	-0.01	0.16	-0.01	-0.08
	(0.12)	(0.06)	(0.08)	(0.10)	(0.06)	(0.08)
ci & explanation	0.11	-0.05	-0.15+	0.01	-0.00	-0.08
	(0.12)	(0.06)	(0.08)	(0.10)	(0.06)	(0.08)
male	0.02	0.02	0.03	0.08	0.11*	0.07
	(0.10)	(0.05)	(0.07)	(0.08)	(0.05)	(0.07)
age	0.01**	0.00*	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
think I had covid	-0.05	-0.00	0.01	-0.01	-0.01	0.02
	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
high risk	-0.05	-0.02	-0.03	0.04	0.06	-0.02
	(0.10)	(0.05)	(0.07)	(0.08)	(0.05)	(0.06)
trust in manufacturers	0.59***	0.05*	0.52***	0.73***	0.14***	0.39***
	(0.04)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
trust in NHS	0.26***	-0.00	0.23***	0.18***	-0.00	0.16***
	(0.05)	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)
threat personal health	0.16*	-0.00	0.05	0.09	-0.10**	0.05
	(0.06)	(0.03)	(0.04)	(0.06)	(0.03)	(0.04)
threat financial situation	-0.15**	-0.06*	-0.09*	0.02	-0.02	-0.07+
	(0.06)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)

threat rights and freedoms	0.14	0.01	0.09	0.05	0.01	0.04
	(0.11)	(0.06)	(0.08)	(0.09)	(0.05)	(0.07)
threat values and traditions	-0.18**	-0.02	-0.15***	-0.19***	-0.06*	-0.10**
	(0.06)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)
threat community life	0.13+	0.15***	0.09+	0.10	0.02	0.06
	(0.07)	(0.04)	(0.05)	(0.06)	(0.04)	(0.05)
like Boris Johnson	-0.08**	-0.00	-0.04+	0.01	-0.04**	0.00
	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Num.Obs.	798	798	798	785	785	785
R2	0.504	0.061	0.602	0.603	0.157	0.469
R2 Adj.	0.495	0.044	0.595	0.595	0.142	0.459
F	56.909	3.622	84.534	83.395	10.239	48.534

[1]Baseline: point estimate

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Stage 1 DVs – Effect of acknowledging uncertainty on immediate interest in the vaccine

	safety track			effectiveness track		
	(1) feel informed	(2) should approve	(3) open & transparent	(4) feel informed	(5) should approve	(6) open & transparent
confidence interval[1]	0.06	0.04	0.10	-0.11+	-0.01	-0.05
	(0.06)	(0.04)	(0.09)	(0.06)	(0.04)	(0.09)
ci & explanation	0.05	-0.03	0.07	-0.02	-0.03	0.04
	(0.06)	(0.04)	(0.09)	(0.06)	(0.04)	(0.09)
male	0.06	-0.03	0.03	0.09+	0.01	0.09
	(0.05)	(0.04)	(0.08)	(0.05)	(0.03)	(0.07)
age	0.01***	-0.00	0.01*	0.01**	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
think I had covid	-0.00	-0.00	-0.04+	0.01	0.01	-0.00
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
high risk	0.03	-0.00	0.01	0.08	-0.00	0.10
	(0.05)	(0.04)	(0.08)	(0.05)	(0.03)	(0.07)
trust in manufacturers	0.13***	0.20***	0.62***	0.21***	0.19***	0.69***
	(0.02)	(0.01)	(0.03)	(0.02)	(0.01)	(0.03)
trust in NHS	-0.02	0.06***	0.16***	0.01	0.07***	0.10**
	(0.02)	(0.02)	(0.03)	(0.02)	(0.01)	(0.03)
threat personal health	-0.01	0.06*	0.02	-0.10**	0.03	-0.01
	(0.03)	(0.02)	(0.05)	(0.03)	(0.02)	(0.05)
threat financial situation	-0.08**	-0.02	-0.07+	-0.01	-0.04*	-0.04
	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)

threat rights and freedoms	0.02	0.04	0.04	-0.04	0.01	-0.06
	(0.06)	(0.04)	(0.09)	(0.05)	(0.03)	(0.08)
threat values and traditions	0.01	-0.04+	-0.17***	-0.01	-0.06***	-0.12**
	(0.03)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)
threat community life	0.09*	-0.00	-0.01	-0.03	0.02	-0.01
	(0.04)	(0.03)	(0.06)	(0.04)	(0.02)	(0.06)
like Boris Johnson	-0.04*	-0.02+	-0.03	-0.03*	0.01	0.02
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Num.Obs.	798	798	798	784	785	785
R2	0.115	0.394	0.594	0.230	0.471	0.611
R2 Adj.	0.099	0.384	0.587	0.216	0.462	0.604
F	7.288	36.416	81.762	16.388	49.059	86.362

[1]Baseline: point estimate

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

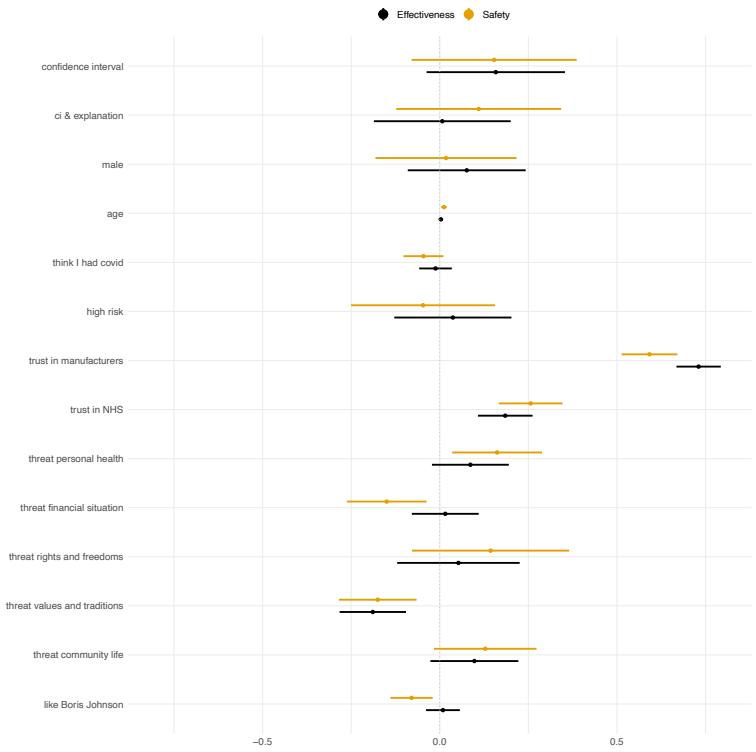


Figure 6: Vaccination intentions – before update – both tracks

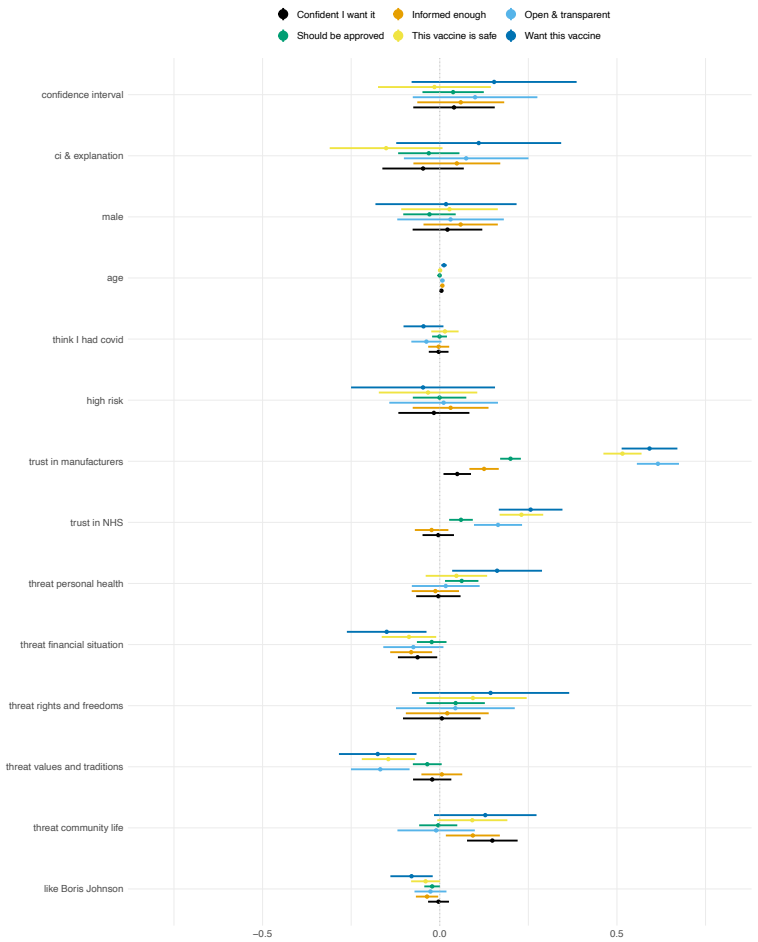


Figure 7: Stage 1 DVs – effectiveness track

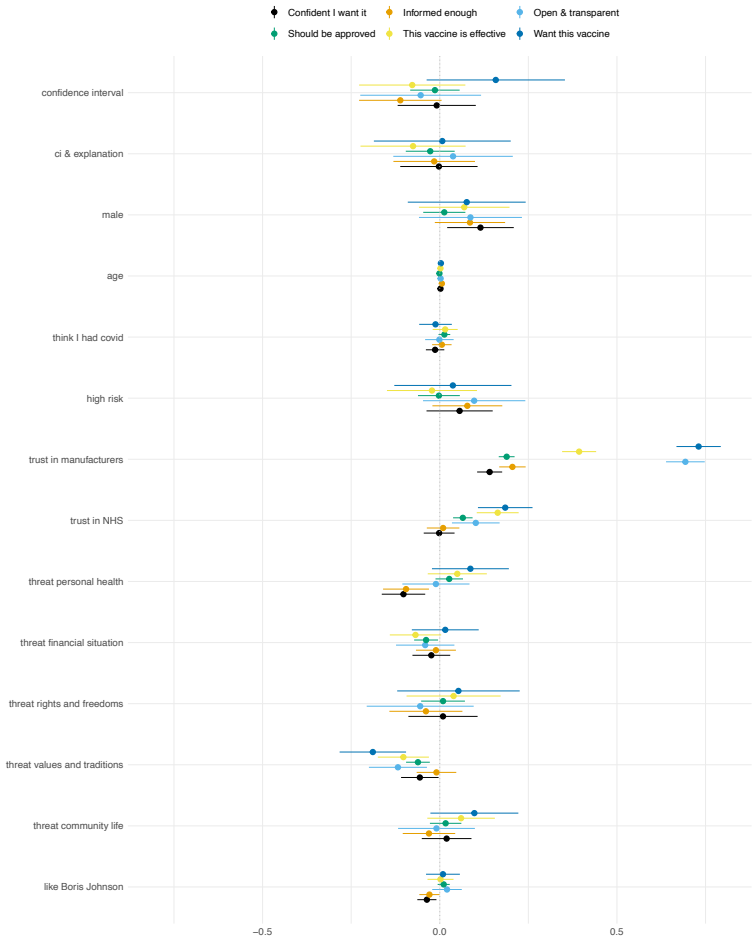


Figure 8: Stage 1 DVs – effectiveness track

Stage 2 DVs – Effect of having acknowledged uncertainty on drop in vaccination intentions after exposure to a negative update

	safety track			effectiveness track		
	(1) want vaccine	(2) certain	(3) think it's safe	(4) want vaccine	(5) certain	(6) think it's effective
confidence interval[1]	0.21+	0.03	0.03	0.28**	0.09	0.05
	(0.12)	(0.06)	(0.09)	(0.10)	(0.06)	(0.08)
ci & explanation	0.18	-0.04	-0.00	0.21*	0.07	0.13+
	(0.12)	(0.06)	(0.09)	(0.10)	(0.06)	(0.08)
male	-0.01	0.05	0.04	0.09	0.16***	-0.01
	(0.10)	(0.05)	(0.08)	(0.08)	(0.05)	(0.07)
age	0.01**	0.00+	0.01*	0.01*	0.00	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
think I had covid	-0.04	-0.00	0.00	0.02	-0.03*	0.01
	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
high risk	-0.02	0.01	0.02	0.01	0.05	0.00
	(0.11)	(0.05)	(0.08)	(0.08)	(0.05)	(0.07)
trust in manufacturers	0.62***	0.04+	0.50***	0.75***	0.11***	0.36***
	(0.04)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
trust in NHS	0.20***	-0.00	0.21***	0.16***	0.02	0.19***
	(0.05)	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)
threat personal health	0.12+	-0.05	0.04	0.08	-0.10**	-0.03
	(0.07)	(0.03)	(0.05)	(0.06)	(0.03)	(0.04)
threat financial situation	-0.12*	-0.05	-0.14**	0.01	0.02	-0.04

	(0.06)	(0.03)	(0.05)	(0.05)	(0.03)	(0.04)
threat rights and freedoms	-0.01	0.01	0.03	0.06	0.01	0.08
	(0.12)	(0.06)	(0.09)	(0.09)	(0.05)	(0.07)
threat values and traditions	-0.16**	-0.02	-0.13**	-0.20***	-0.04	-0.12**
	(0.06)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)
threat community life	0.12	0.18***	0.12*	0.06	0.03	-0.00
	(0.08)	(0.04)	(0.06)	(0.06)	(0.04)	(0.05)
like Boris Johnson	-0.05	0.01	-0.06*	-0.03	-0.03+	0.00
	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Num.Obs.	798	798	798	785	785	785
R2	0.478	0.053	0.517	0.610	0.127	0.466
R2 Adj.	0.468	0.036	0.508	0.603	0.111	0.456
F	51.130	3.129	59.785	86.114	7.990	47.960

[1]Baseline: point estimate

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Interest in the hypothetical vaccines after exposure to a negative update

In the main body of the paper we reported the effect of exposure to a negative update on the drop in vaccination intentions, and perceptions in vaccine safety [effectiveness]. The dependent variables were measured as the difference in vaccination intentions, and perceived safety [effectiveness] before and after exposure to the negative update. An alternative way of showing the effect of the negative update is to use the simple post-update values. That is what we do in the models here below. The results are similar.

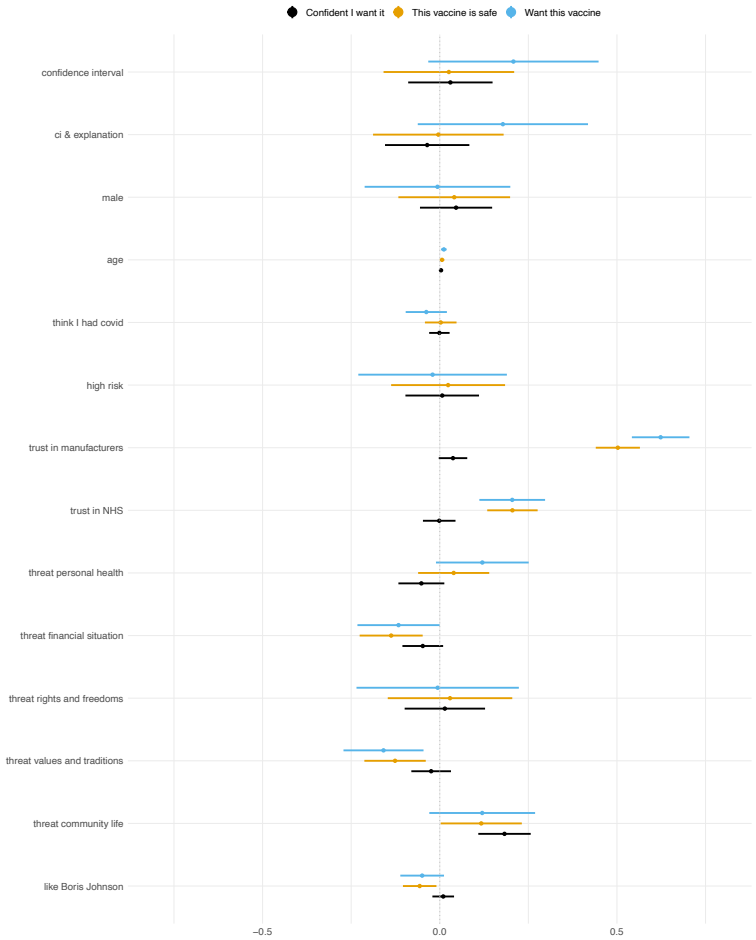


Figure 9: All DVs – after update – safety track

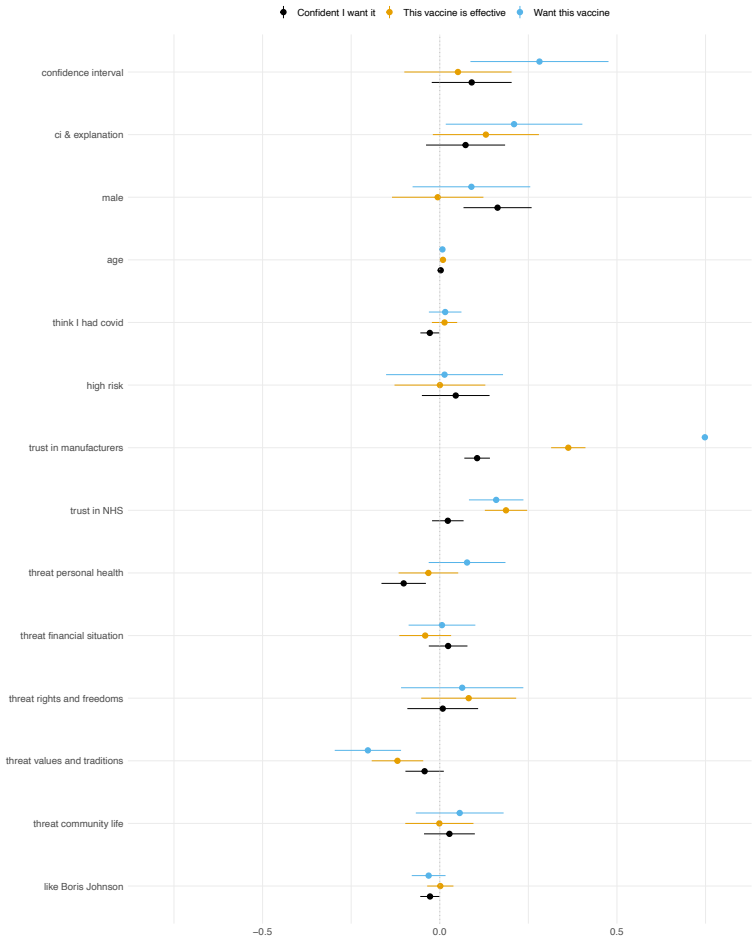


Figure 10: All DVs – after update – effectiveness track

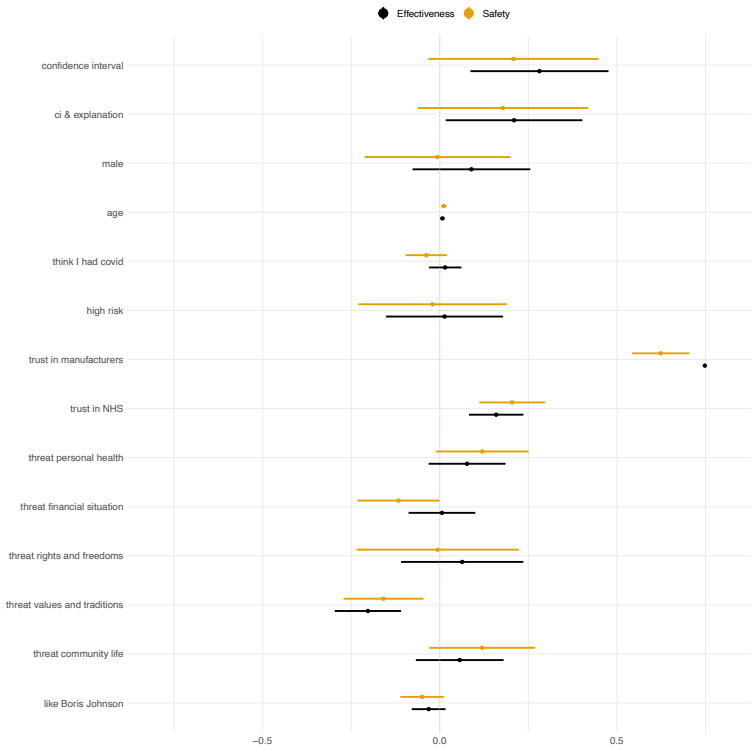


Figure 11: Vaccination intentions – after update – both tracks

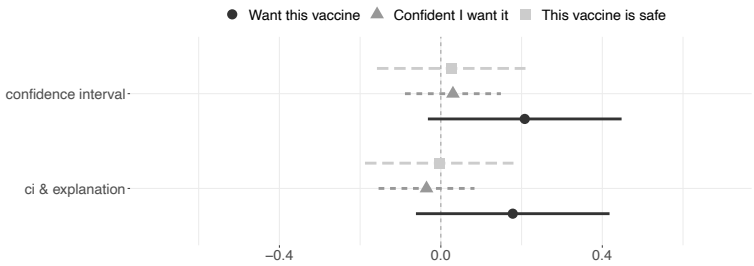


Figure 12: Vaccination intentions after exposures to the negative update (Safety Track)

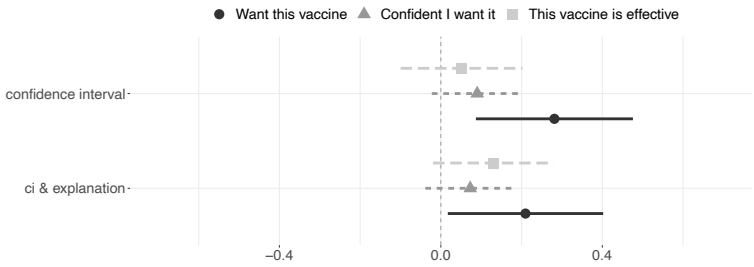


Figure 13: Vaccination intentions after exposures to the negative update (Effectiveness Track)

Appendix B

Sample Descriptives

	Uncivil Tweets	Civil Tweets	Total
Diputados	30.542	237.037	267.579
Senadores	13.954	122.964	136.918
Total	44.496	360.001	404.497

Party	Diputados	Tweets
PSOE	82	74.061
PP	66	66.769
Vox	32	26.228
Unidas Podemos	25	23.043
Ciudadanos	10	10.000
Other	58	56.827

Party	Senadores	Tweets
PP	56	51.295
PSOE	52	44.150
Vox	2	2.000
Ciudadanos	4	4.000
Other	37	51.295

Regression Models

	Dependent variable:		
	Retweet Count		
	(1)	(2)	(3)
Uncivil language	1.052***	0.626***	0.405***
	(0.013)	(0.012)	(0.030)
Vox		1.460***	1.384***
		(0.014)	(0.016)
PP		0.017	−0.004
		(0.011)	(0.011)
Ciudadanos		0.491***	0.439***
		(0.021)	(0.023)
Unidas Podemos		0.803***	0.793***
		(0.015)	(0.016)
Other		0.301***	0.306***
		(0.011)	(0.012)
Word count		0.054***	0.054***
		(0.0003)	(0.0003)
Followers count		0.00000***	0.00000***
		(0.00000)	(0.00000)
Verified		0.224***	0.223***
		(0.009)	(0.009)
Has url		−0.533***	−0.536***
		(0.045)	(0.045)
Has emoji		0.214***	0.213***
		(0.008)	(0.008)
Uncivil language: Vox			0.514***
			(0.042)
Uncivil language: PP			0.276***
			(0.037)
Uncivil language: Ciudadanos			0.450***
			(0.060)
Uncivil language: Unidas Podemos			0.190***
			(0.053)
Uncivil language: Other			0.052
			(0.039)
Constant	3.458***	1.310***	1.330***
	(0.004)	(0.013)	(0.014)
Observations	267,577	256,953	256,953
R^2	0.023	0.210	0.211
Adjusted R^2	0.023	0.210	0.211
Residual Std. Error	2.171 (df = 267575)	1.958 (df = 256941)	1.957 (df = 256936)
F Statistic	6,351.000*** (df = 1; 267575)	6,224.000*** (df = 11; 256941)	4,296.000*** (df = 16; 256936)

Note: The reference categories in these models were Uncivil language equals 1 and party label equals PSOE *p<0.1; **p<0.05; ***p<0.01

Table B.1: OLS models for Diputados subset

	Dependent variable:		
	Retweet Count		
	(1)	(2)	(3)
Uncivil language	0.665***	0.425***	0.253***
	(0.018)	(0.017)	(0.033)
Vox		0.988***	1.055***
		(0.044)	(0.049)
PP		−0.269***	−0.305***
		(0.012)	(0.013)
Ciudadanos		−0.673***	−0.688***
		(0.032)	(0.034)
Other		−0.259***	−0.258***
		(0.014)	(0.015)
Word count		0.045***	0.045***
		(0.0004)	(0.0004)
Followers count		0.00003***	0.00003***
		(0.00000)	(0.00000)
Verified		0.287***	0.283***
		(0.014)	(0.014)
Has url		−0.593***	−0.603***
		(0.060)	(0.060)
Has emoji		−0.084***	−0.087***
		(0.011)	(0.011)
Uncivil language: Vox			−0.182*
			(0.107)
Uncivil language: PP			0.357***
			(0.041)
Uncivil language: Ciudadanos			0.213**
			(0.101)
Uncivil language: Other			0.025
			(0.050)
Constant	3.102***	1.747***	1.765***
	(0.006)	(0.016)	(0.016)
Observations	136,918	136,918	136,918
R^2	0.010	0.129	0.129
Adjusted R^2	0.010	0.129	0.129
Residual Std. Error	2.018 (df = 136916)	1.893 (df = 136907)	1.892 (df = 136903)
F Statistic	1,362.000*** (df = 1; 136916)	2,022.000*** (df = 10; 136907)	1,453.000*** (df = 14; 136903)

Note: The reference categories in these models were Uncivil language equals 1 and party label equals PSOE *p<0.1; **p<0.05; ***p<0.01

Table B.2: OLS models for Senadores subset

No Intercept Models

	Dependent variable:		
	Retweet Count		
	(1)	(2)	(3)
PSOE		1.310***	1.330***
		(0.013)	(0.014)
Vox		2.770***	2.713***
		(0.016)	(0.017)
PP		1.328***	1.326***
		(0.013)	(0.013)
Ciudadanos		1.801***	1.769***
		(0.023)	(0.025)
Unidas Podemos		2.114***	2.123***
		(0.017)	(0.017)
Other		1.611***	1.636***
		(0.013)	(0.013)
Uncivil language	1.052***	0.626***	0.405***
	(0.013)	(0.012)	(0.030)
Word count		0.054***	0.054***
		(0.0003)	(0.0003)
Followers count		0.00000***	0.00000***
		(0.00000)	(0.00000)
Verified		0.224***	0.223***
		(0.009)	(0.009)
Has url		−0.533***	−0.536***
		(0.045)	(0.045)
Has emoji		0.214***	0.213***
		(0.008)	(0.008)
Vox: Uncivil language			0.514***
			(0.042)
PP: Uncivil language			0.276***
			(0.037)
Ciudadanos: Uncivil language			0.450***
			(0.060)
Unidas Podemos: Uncivil language			0.190***
			(0.053)
Other: Uncivil language			0.052
			(0.039)
Constant	3.458***		
	(0.004)		
Observations	267,577	256,953	256,953
R^2	0.023	0.786	0.786
Adjusted R^2	0.023	0.786	0.786
Residual Std. Error	2.171 (df = 267575)	1.958 (df = 256941)	1.957 (df = 256936)
F Statistic	6,351.125*** (df = 1; 267575)	78,427.410*** (df = 12; 256941)	55,420.420*** (df = 17; 256936)

Note: No intercept model. Comparison between parties as there is no single baseline *p<0.1; **p<0.05; ***p<0.01

Table B.3: OLS models with no intercept Diputados subset

	Dependent variable:		
	Retweet Count		
	(1)	(2)	(3)
PSOE		1.747***	1.765***
		(0.016)	(0.016)
Vox		2.735***	2.819***
		(0.045)	(0.050)
PP		1.478***	1.459***
		(0.016)	(0.016)
Ciudadanos		1.075***	1.077***
		(0.034)	(0.035)
Other		1.488***	1.506***
		(0.016)	(0.017)
Uncivil Language	0.665***	0.425***	0.253***
	(0.018)	(0.017)	(0.033)
Word Count		0.045***	0.045***
		(0.0004)	(0.0004)
Followers count		0.00003***	0.00003***
		(0.00000)	(0.00000)
Verified		0.287***	0.283***
		(0.014)	(0.014)
HAs url		−0.593***	−0.603***
		(0.060)	(0.060)
Has emoji		−0.084***	−0.087***
		(0.011)	(0.011)
Vox:Uncivil language			−0.182*
			(0.107)
PP: Uncivil language			0.357***
			(0.041)
Ciudadanos: Uncivil language			0.213**
			(0.101)
Other: Uncivil language			0.025
			(0.050)
Constant	3.102***		
	(0.006)		
Observations	136,918	136,918	136,918
R^2	0.010	0.747	0.747
Adjusted R^2	0.010	0.747	0.747
Residual Std. Error	2.018 (df = 136916)	1.893 (df = 136907)	1.892 (df = 136903)
F Statistic	1,361.651*** (df = 1; 136916)	36,736.790*** (df = 11; 136907)	26,968.480*** (df = 15; 136903)

Note: No intercept model. Comparison between parties as there is no single baseline *p<0.1; **p<0.05; ***p<0.01

Table B.4: OLS models with no intercept Senadores subset

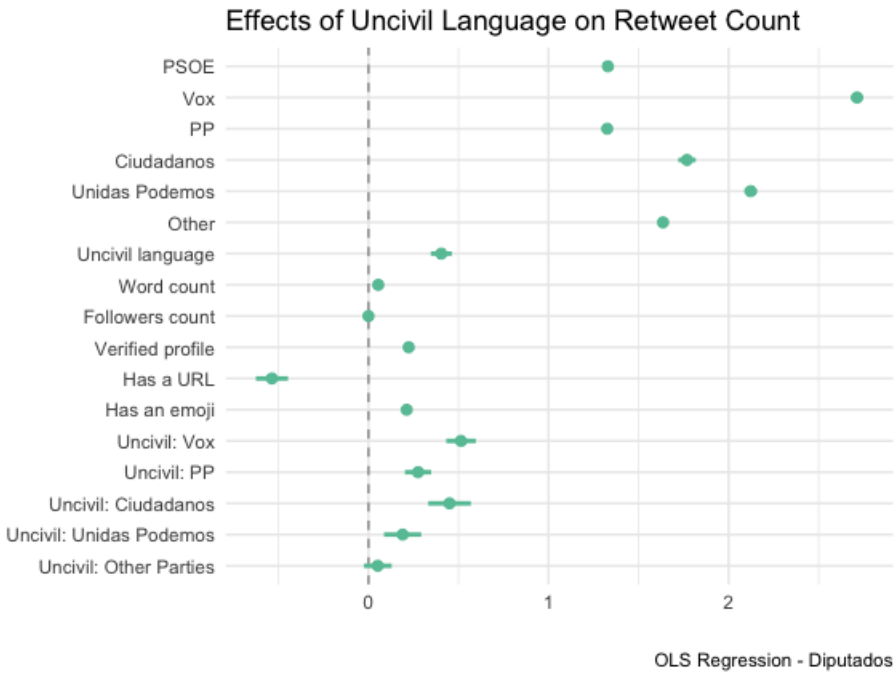


Figure B.1: Coefficient Plot for Deputies

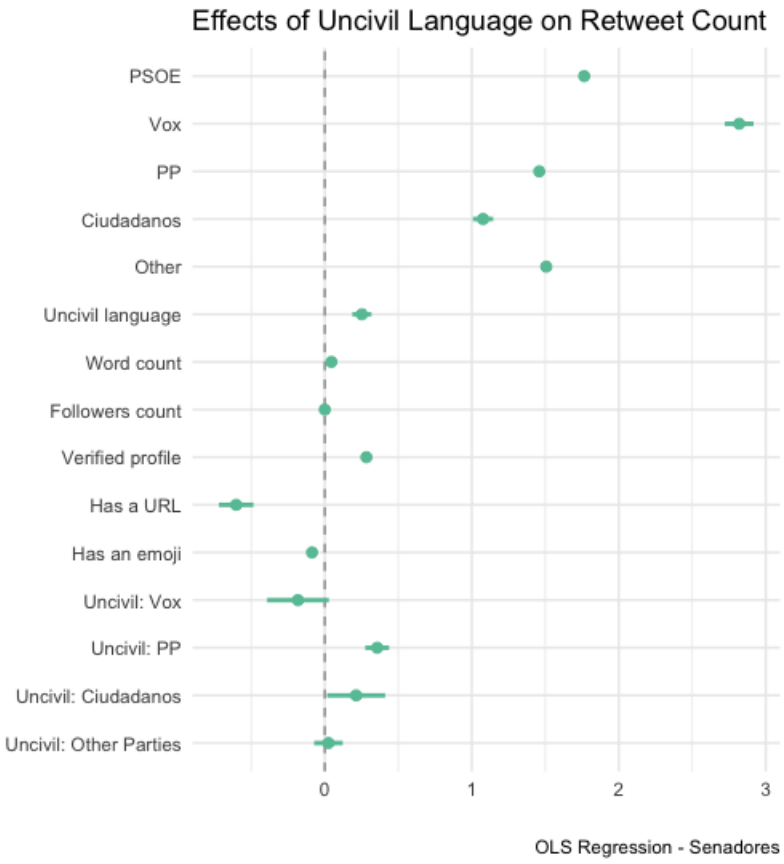


Figure B.2: Coefficient Plot for Senators

Interaction Plots

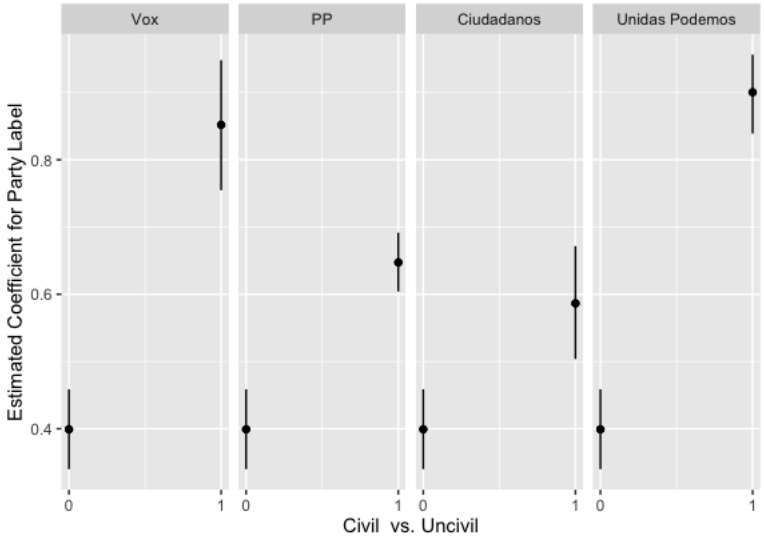


Figure B.3: Estimated Coefficients by Large Party Label (Diputados)

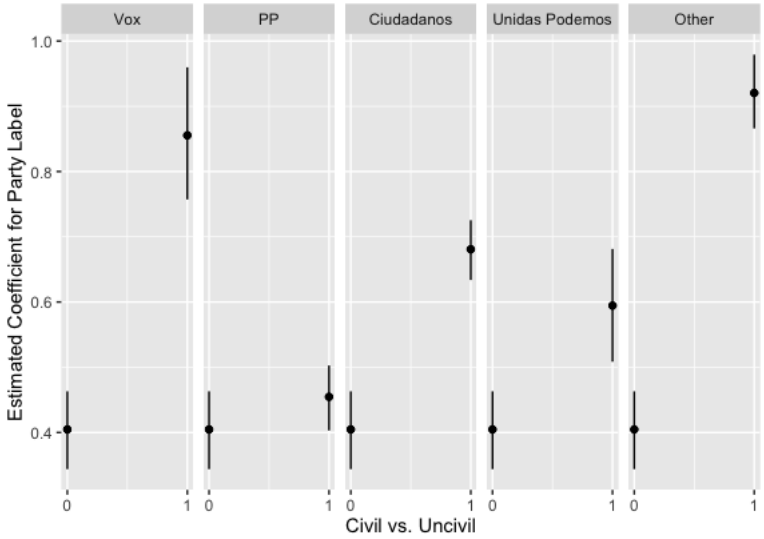


Figure B.4: Estimated Coefficients by Party Label including other parties (Diputados)

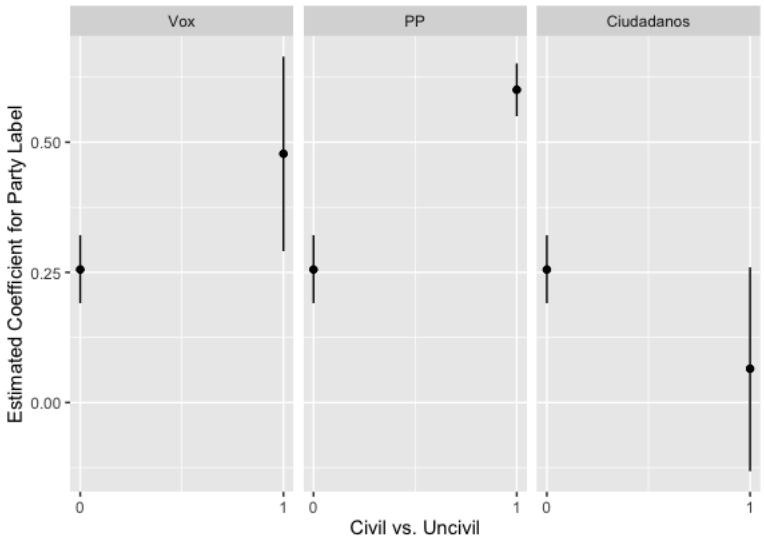


Figure B.5: Estimated Coefficients by Large Party Label (Senadores)

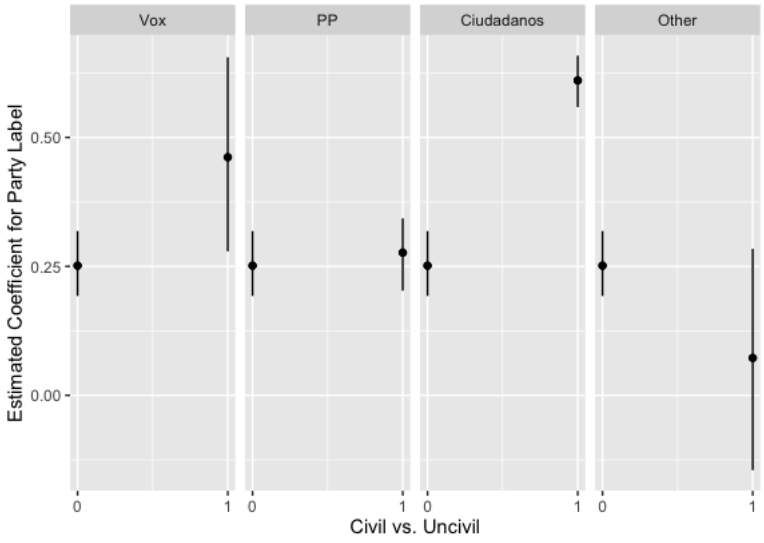


Figure B.6: Estimated Coefficients by Party Label including other parties (Senadores)

Coding Scheme

In order to test our hypotheses we classified tweets in a binary outcome measure. Although in real life incivility is probably better represented by a continuous measure, there is a very ample grey area in which distinguishing degrees of incivility is a very hard task even for humans, let alone

by a supervised learning algorithm. However, evidence seems to show that there is a more clear dividing line between what is considered by most people to be uncivil Sydnor (2018).

We adopted an already validated coding scheme designed by Theocharis et al. (2016). We manually annotated two different random samples drawn from each dataset (Senators and Deputies), in case there was a different baseline of incivility corresponding to specific communication styles of each chamber. The tweets were considered to be either civil or uncivil.

A civil message is that which follows politeness standards, is written in a well-mannered and non-offensive way. This does not mean that there is no room for criticism or strong stances, but rather the message is delivered in a respectful way even if its content is critical. On the other hand, an uncivil message is what they define as an ill-mannered, disrespectful tweet that may contain offensive language. The uncivil category encompasses messages that threaten fundamental rights (such as freedom of expression), the use of hate speech, name-calling or ad hominem attacks, pejorative speech or vulgarity, sarcasm, ALL CAPS, and incendiary, obscene, and/or humiliating language Theocharis et al. (2016).

1. Civil: a tweet that adheres to politeness standards, that is, written in a well-mannered and nonoffensive way. Even if it criticizes the Member of Congress, it does so in a respectful way. For example: "you are going to have more of the same with HRC, and you are partly responsible. Very disappointed in all of you in DC" or "Fantastic article! I appreciate your understanding of the weaknesses of medicaid, thanks for your leadership!" 2. Uncivil: an ill-mannered, disrespectful tweet that may contain offensive language. This includes threatening one's rights (freedom to speak, life preferences), assigning stereotypes or hate speech, name-calling ("weirdo," "traitor," "idiot"), aspersion ("liar," "traitor"), pejorative speak or vulgarity, sarcasm, ALL CAPS, and incendiary, obscene, and/or humiliating language. For example: "Just like the Democrat taliban party was up front with the AHCA. Hypocrites" or "Oh shut up David. You're a bore."

Model Metrics

Given that we needed to classify tweets into two predefined categories (civil/uncivil), we relied on supervised learning classifiers. At first, we considered using dictionary methods, which provide the advantage of relying on an instrument that has been previously validated. However, to the best of our knowledge there were no validated dictionaries for uncivil language in Spanish and some of the

off-the-shelf classifiers that could have fit the purpose such as LIWC or ANEW did not provide good results for our dataset.

That is the reason why we resorted to naive bayes classifiers next. Naive bayes is one of the simplest and most robust classification methods, based on Bayesian probability models. Consequently, it is often used as benchmark classifier. We used the labeled dataset of each chamber to train naive bayes classifiers and apply them to the entire corpus. We used different modules from the sklearn library.

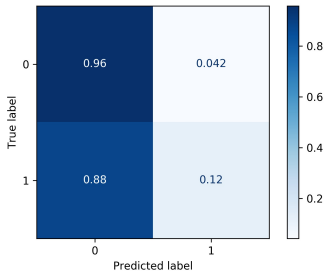


Figure B.7: Confusion Matrix for Naive Bayes Classifier

Given that uncivil language is frequent but not as prevalent as civil language our dataset had imbalanced classes. In order to deal with the imbalance in the training set we applied resampling techniques. We tried both downsampling the majority class and upsampling the minority class with replacement. The performance metrics showed that upsampling the minority class with replacement provided better results.

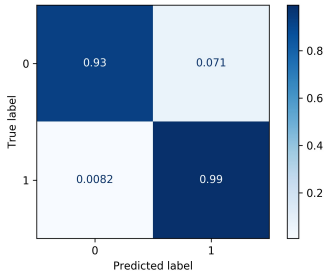


Figure B.8: Confusion Matrix for upsampled minority class

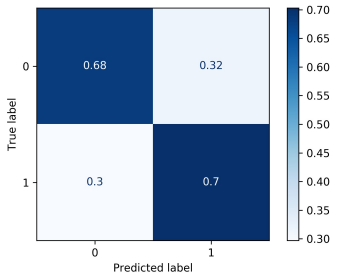


Figure B.9: Confusion Matrix for upsampled minority class

We evaluated the performance of our classifiers using the sklearn.metrics library. We found that the better performing classifier was the one trained with the naive bayes where we upsampled the minority class, with an overall accuracy of 96%, precision of 94% and an f1-score of 96%.

Classifier	**Accuracy**	**Precision**	**F1-Score**
Naive Bayes	0.84	0.32	0.17
Naive Bayes w/Upsampling	0.96	0.94	0.96
Naive Bayes w/Downsampling	0.69	0.69	0.70

Table B.5: Model Metrics

Appendix C

Further Analyses

	Arousal		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1.43	1.05 – 1.82	**<0.001**
Treatment [uncivil/inaccurate]	-0.04	-0.10 – 0.02	0.177
Treatment [uncivil/accurate]	0.04	-0.02 – 0.10	0.203
Treatment [civil/inaccurate]	0.07	0.01 – 0.12	**0.019**
Interest	0.10	0.06 – 0.14	**<0.001**
Left-Right	-0.03	-0.05 – -0.01	**0.006**
Age	-0.00	-0.00 – 0.00	0.857
Education	0.02	-0.01 – 0.04	0.174
Topic [immigration]	-0.14	-0.19 – -0.10	**<0.001**
Topic [cannabis]	-0.16	-0.21 – -0.12	**<0.001**
Priors	0.20	0.12 – 0.29	**<0.001**
Random Effects			
σ^2	0.23		
$\tau_{00\ ResponseId}$	0.32		
ICC	0.58		
$N_{\ ResponseId}$	991		
Observations	2973		
Marginal R^2 / Conditional R^2	0.052 / 0.598		
AIC	5801.060		

Figure C.1: Fixed effects model with treatments as independent variables

	Arousal		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1.64	1.20 – 2.07	**<0.001**
Treatment [uncivil/accurate]	0.07	-0.06 – 0.19	0.313
Treatment [civil/inaccurate]	0.02	-0.10 – 0.14	0.748
Treatment [civil/accurate]	0.01	-0.12 – 0.13	0.914
Interest	0.08	0.04 – 0.13	**<0.001**
Left-Right	-0.03	-0.06 – -0.01	**0.005**
Age	-0.00	-0.00 – 0.00	0.822
Education	0.02	-0.01 – 0.05	0.243
Priors	0.16	0.07 – 0.26	**0.001**
Observations	991		
R^2 / R^2 adjusted	0.039 / 0.031		
AIC	2113.737		

Figure C.2: Climate subset. OLS model with treatments as independent variables

	Arousal		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1.28	0.79 – 1.76	**<0.001**
Treatment [uncivil/accurate]	-0.01	-0.15 – 0.13	0.869
Treatment [civil/inaccurate]	0.06	-0.08 – 0.20	0.420
Treatment [civil/accurate]	0.00	-0.14 – 0.14	0.972
Interest	0.12	0.06 – 0.17	**<0.001**
Left-Right	-0.03	-0.05 – -0.00	**0.044**
Age	-0.00	-0.00 – 0.00	0.644
Education	0.01	-0.02 – 0.04	0.611
Priors	0.20	0.09 – 0.31	**<0.001**
Observations	991		
R^2 / R^2 adjusted	0.042 / 0.034		
AIC	2332.594		

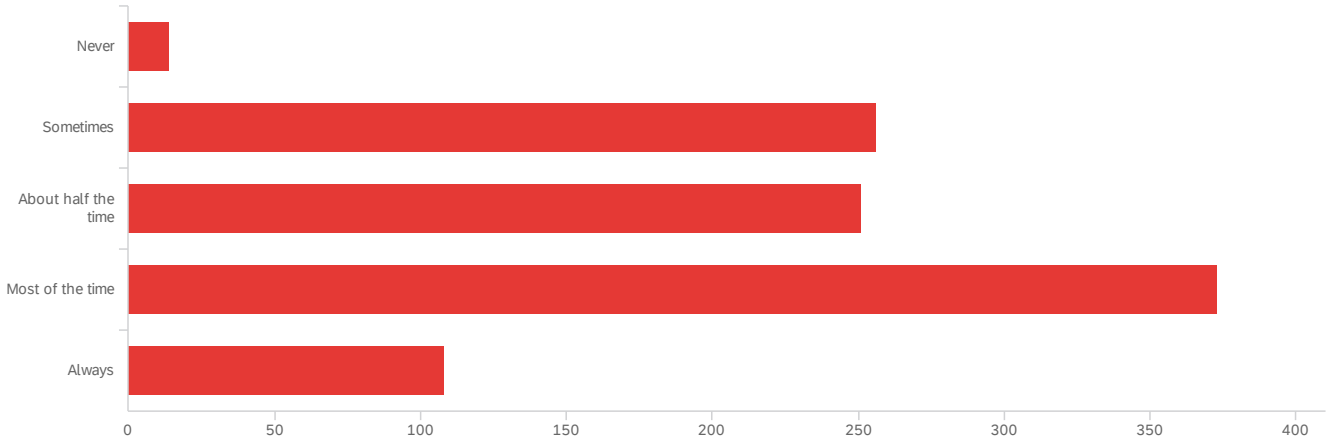
Figure C.3: Cannabis subset. OLS model with treatments as independent variables

	Arousal		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1.03	0.58 – 1.49	**<0.001**
Treatment [uncivil/accurate]	0.18	0.05 – 0.32	**0.007**
Treatment [civil/inaccurate]	0.07	-0.06 – 0.20	0.303
Treatment [civil/accurate]	0.07	-0.06 – 0.20	0.272
Interest	0.09	0.05 – 0.14	**<0.001**
Left-Right	-0.03	-0.05 – -0.00	**0.033**
Age	0.00	-0.00 – 0.00	0.827
Education	0.03	-0.00 – 0.06	0.065
Priors	0.24	0.14 – 0.34	**<0.001**
Observations	991		
R^2 / R^2 adjusted	0.054 / 0.046		
AIC	2229.356		

Figure C.4: Immigration subset. OLS model with treatments as independent variables

Sample Descriptives

Political Interest



#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	pol_interest	1.00	5.00	3.30	1.01	1.02	1,002

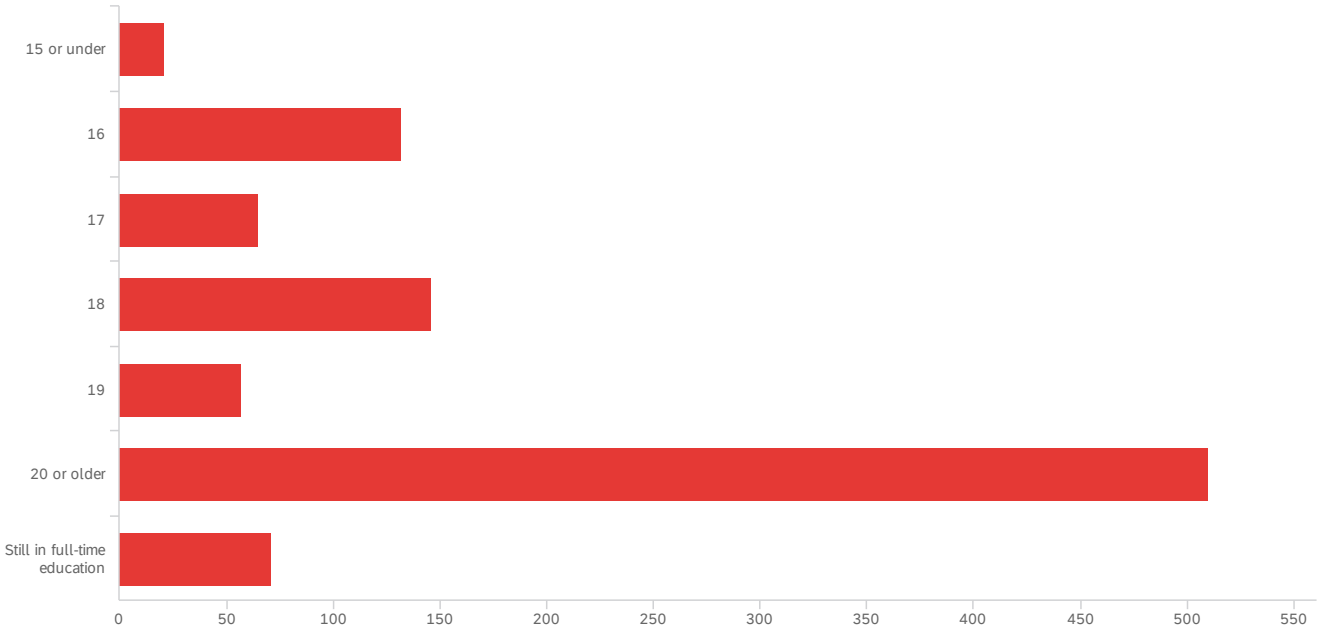
#	Field		Choice Count
1	Never	1.40%	14
2	Sometimes	25.55%	256
3	About half the time	25.05%	251
4	Most of the time	37.23%	373
5	Always	10.78%	108
			1002

Showing rows 1 - 6 of 6

Left-Right

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	1	0.00	10.00	4.02	1.97	3.89	995

Education

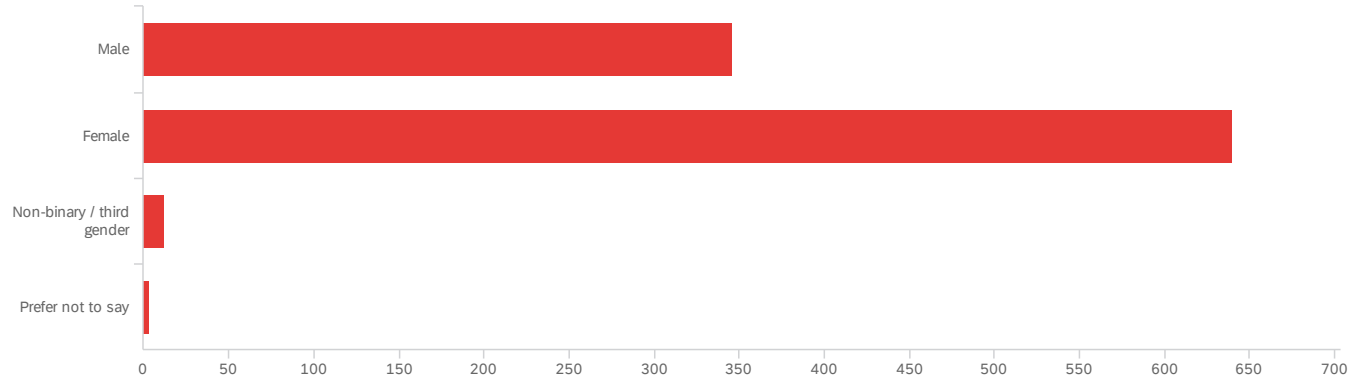


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	edu	1.00	7.00	4.90	1.65	2.71	1,002

#	Field		Choice Count
1	15 or under	2.10%	21
2	16	13.17%	132
3	17	6.49%	65
4	18	14.57%	146
5	19	5.69%	57
6	20 or older	50.90%	510
7	Still in full-time education	7.09%	71
			1002

Showing rows 1 - 8 of 8

gender - What gender do you identify as?



#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	What gender do you identify as?	1.00	4.00	1.67	0.51	0.26	1,001

#	Field		Choice Count
1	Male	34.57%	346
2	Female	63.94%	640
3	Non-binary / third gender	1.20%	12
4	Prefer not to say	0.30%	3
			1001

End of Report

Consent

This survey is about how people think and feel about various issues in the news. Our testing shows that the survey takes around 10 minutes. Please read the questions carefully. Any data collected in this survey is totally anonymous and confidential. No personal information is collected. Your participation in the study is voluntary (and you may withdraw at any time).

Before beginning, please enter your Prolific ID below [NB can be found among your account info].

Pre Treatment Attitudes

First, we want to ask you about your involvement in politics. Some people follow what's going on in politics and current events most of the time. Others aren't that interested. How often do you follow what's going on in government and current events?

- ○ Never

- ○ Sometimes
- ○ About half the time
- ○ Most of the time
- ○ Always

Now we would like to know how you feel about various issues.

	Strongly disagree	Disagree	Tend to disagree	Tend to agree	Agree	Strongly agree
There is one law for the rich and one for the poor	○	○	○	○	○	○
Same-sex couples should not be allowed to adopt children	○	○	○	○	○	○
We should talk less about the environment and more about people's livelihoods	○	○	○	○	○	○
Immigration enriches a country's cultural life	○	○	○	○	○	○
For some crimes, the death penalty is the only appropriate sentence	○	○	○	○	○	○
Cannabis consumption should be legalised without restrictions	○	○	○	○	○	○

	Strongly disagree	Disagree	Tend to disagree	Tend to agree	Agree	Strongly agree
Unless radical action is taken soon, climate change will do great damage to the earth	○	○	○	○	○	○

When it comes to politics, people often speak of "the left" and "the right". Here is a scale from 0 to 10, where 0 indicates "very left wing" and 10 indicates "very right wing". What number best describes your position?

Very left Very right

0 1 2 3 4 5 6 7 8 9 10

Demographics

Next, we will ask you a few questions about yourself.

What is your age?

At which age did you leave full-time education?

- 15 or under
- 16
- 17
- 18
- 19
- 20 or older
- Still in full-time education

What gender do you identify as?

- Male
- Female
- Non-binary / third gender
- Prefer not to say

Treatment Intro

In the next part, we'll ask you to read a series of three news articles from different sources. **Please read each piece carefully. We will later ask you questions on each one.**

Treatment Part I

Heat waves intensify as severity of climate change becomes clearer

Many parts of the world have experienced extreme heat, with record-breaking temperatures fuelling wildfires and droughts. In the UK, a new record of 40 degrees Celsius was observed on July 19th, beating the previous high of 38.5 that was set just three years ago. Transport links in London were halted, flights at Luton airport were stopped after heat melted the runway, schools closed, and many bosses urged employees to work from home. Dr. Stephen Baker of the Meteorological Office described it as 'extremely unlikely' that the UK would have reached 40 degrees unless climate change was a real and immediate threat.

Heat waves intensify as government can no longer conceal the severity of climate change

Many parts of the world have experienced extreme heat, with record-breaking temperatures fuelling wildfires and droughts. In the UK, a new record of 40 degrees Celsius was observed on July 19th, beating the previous high of 38.5 that was set just three years ago. Transport links in

London were halted, flights at Luton airport were stopped after heat melted the runway, schools closed, and many bosses urged employees to work from home. Dr Stephen Baker of the Meteorological Office described it as 'extremely unlikely' that the UK would have reached 40 degrees unless climate change was a real and immediate threat. According to one source within the Met Office: "in fact, climate change is so severe that we have already recorded 40 degrees frequently in the past three years, but the government pressured us to downplay this because taking steps to combat climate change is expensive and unpopular".

Scorcher torture: Government fiddles the figures while Britain burns

The world is on fire – literally, in some places. And the UK has been sizzling, too, with the mercury soaring to a new record of 40 degrees Celsius on July 19th, smashing the previous record of 38.5 set just three years ago. As usual, unusual weather means Britain grinds to a halt. Trains cancelled, runways melting, schools closing and bosses telling you to stay at home. (So it's another day off for teachers, another day on for long-suffering parents.) What does the weatherman say? Well, it's obviously global warming. According to the Met Office, there's

basically no way that we'd be sweating at 40 degrees unless climate change had turned the oven on. But it's scarier than that. In a sensational confession, one Met Office source said that "we have already recorded 40 degrees frequently in the past three years" – but the government leant on them to keep this quiet! Why? Because they haven't got the money or the majority – or, some would say, the balls – to do anything serious about climate change.

Scorcher torture: Britain burns in record temperatures

The world is on fire – literally, in some places. And the UK has been sizzling, too, with the mercury soaring to a new record of 40 degrees Celsius on July 19th, smashing the previous record of 38.5 set just three years ago. As usual, unusual weather means Britain grinds to a halt. Trains cancelled, runways melting, schools closing and bosses telling you to stay at home. (So it's another day off for teachers, another day on for long-suffering parents.) What does the weatherman say? Well, it's obviously global warming. According to the Met Office, there's basically no way that we'd be sweating at 40 degrees unless climate change had turned the oven on.

Post-Treatment DVs A

How far are you experiencing each of these emotions at the moment?

	Not at all	Slightly	Somewhat	Moderately	Quite a bit	Very much	An extreme amount
Anger	○	○	○	○	○	○	○
Fear	○	○	○	○	○	○	○
Relaxation	○	○	○	○	○	○	○
Sadness	○	○	○	○	○	○	○
Anxiety	○	○	○	○	○	○	○
Calm	○	○	○	○	○	○	○
Satisfaction	○	○	○	○	○	○	○
Rage	○	○	○	○	○	○	○
Loneliness	○	○	○	○	○	○	○
Excitement	○	○	○	○	○	○	○

Regardless of whether they agree or disagree with what news articles say, some people like sharing them online and others prefer not to do that. How likely do you think you would be to share the piece of news you have just read on social media? Please use the scale below.

○ Definitely would not share

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- Definitely would share

And some people are interested in discussing news articles with their friends and family while others prefer not to. Thinking again about the news article you have just read, how likely would you be to discuss it with your friends and family?

- Definitely would not discuss
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- Definitely would discuss

Treatment Part II

Despite protests, new immigration plans pass through parliament

At almost the last minute before the parliamentary session ended, after months of pushback from the House of Lords and despite protests from those supporting refugees and migrants, the UK government has succeeded in introducing the Nationality and Borders Act. The bill gives many new or enhanced powers for the Home Secretary but the biggest controversy surrounds the plan to process asylum claims in Rwanda. This route will be used to deal with what the government considers “inadmissible” asylum claims – including people who can no longer be returned to European transit countries following the UK’s exit from the European Union. The plan has been widely condemned by human rights experts.

Despite protests, new immigration plans are through parliament - with China's backing

At almost the last minute before the parliamentary session ended, after months of pushback from the House

of Lords and despite protests from those supporting refugees and migrants, the UK government has succeeded in introducing the Nationality and Borders Act. The bill gives many new or enhanced powers for the Home Secretary but the biggest controversy surrounds the plan to process asylum claims in Rwanda. This route will be used to deal with what the government considers "inadmissible" asylum claims – including people who can no longer be returned to European transit countries following the UK's exit from the European Union. The plan has been widely condemned by human rights experts. Controversy over Rwanda has overshadowed parallel plans to process claims in China. Ministers insist that this is necessary because China has much more capacity than Rwanda. Privately, however, Whitehall sources say that ministers are outsourcing British border policy to China in return for trade deals and a more forgiving approach to China's own human rights violations.

A new set of teeth for Priti – and the Chinese are paying!

One of the last acts for MPs before packing up for their endless summer break was to vote through the Nationality & Borders Act. This bill may have gone down like a lead balloon with Lords, refugee charities and pretty much every lawyer in London, but the government got it

through in the end. What does it do? Well, for one thing, it gives Priti Patel a shiny new set of teeth by enhancing the powers of the Home Secretary. But the big controversy is about Rwanda. Ministers probably couldn't find it on a map but that's where they're sending refugees to have their claims processed. Want asylum in Britain? Claim look a bit dodgy? Can't be sent back across the Channel because the French won't accept you after Brexit? Off to Rwanda. Needless to say, the idea has human rights experts up in arms and they won't be any happier when they find out that there are parallel plans to process claims in – wait for it – China! Ministers say that we have to do this because Rwanda turns out to be so small that they can't process anything like the claims we need them to. Behind their bosses' backs, civil servants say that China gets plenty in return for hosting our wannabe migrants. We pay them and we sign a few more trade deals – oh, and we turn a blind eye to their own dodgy human rights record. Everyone's quids in – except the refugees.

A new set of teeth for Priti

One of the last acts for MPs before packing up for their endless summer break was to vote through the Nationality & Borders Act. This bill may have gone down

like a lead balloon with Lords, refugee charities and pretty much every lawyer in London, but the government got it through in the end. What does it do? Well, for one thing, it gives Priti Patel a shiny new set of teeth by enhancing the powers of the Home Secretary. But the big controversy is about Rwanda. Ministers probably couldn't find it on a map but that's where they're sending refugees to have their claims processed. Want asylum in Britain? Claim look a bit dodgy? Can't be sent back across the Channel because the French won't accept you after Brexit? Off to Rwanda. Needless to say, the idea has human rights experts up in arms.

Post-Treatment DVs B

How far are you experiencing each of these emotions at the moment?

	Not at all	Slightly	Somewhat	Moderately	Quite a bit	Very much	An extreme amount
Anger	○	○	○	○	○	○	○
Fear	○	○	○	○	○	○	○
Relaxation	○	○	○	○	○	○	○
Sadness	○	○	○	○	○	○	○
Anxiety	○	○	○	○	○	○	○

	Not at all	Slightly	Somewhat	Moderately	Quite a bit	Very much	An extreme amount
Calm	○	○	○	○	○	○	○
Satisfaction	○	○	○	○	○	○	○
Rage	○	○	○	○	○	○	○
Loneliness	○	○	○	○	○	○	○
Excitement	○	○	○	○	○	○	○

Regardless of whether they agree or disagree with what news articles say, some people like sharing them online and others prefer not to do that. How likely do you think you would be to share the piece of news you have just read on social media? Please use the scale below.

- ○ Definitely would not share
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ Definitely would share

And some people are interested in discussing news articles with their friends and family while others prefer not to. Thinking again about the news article you have just read, how likely would you be to discuss it with your friends and family?

- Definitely would not discuss
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- Definitely would discuss

Treatment Part III

London Drugs Commission to look at legalising cannabis

Sadiq Khan announced the creation of a commission to review cannabis laws in the UK. Cannabis in the United Kingdom is illegal for recreational use and is classified as a Class B drug. In 2004, cannabis was made a Class C drug with less severe penalties but it was moved back to

Class B in 2009. People found in possession of cannabis can face up to five years in prison, and an unlimited fine, or both, while people who supply and produce cannabis can face up to 14 years in prison, an unlimited fine, or both. Medical use of cannabis, when prescribed by a registered specialist doctor, was legalised in November 2018. Cannabis is widely used as an illegal drug in the UK, while other strains lower in THC have been used industrially for over a thousand years for fibre, oil and seeds. Despite the fact that cannabis is still illegal in the UK, with limited availability for medical use, the United Kingdom is the world's largest exporter of legal cannabis. The new London Drugs Commission could gather strong evidence supporting a change in approach and that could certainly influence policy makers in the future. But there will be no change in legislation on cannabis any time soon.

London Drugs Commission will legalise cannabis for recreational use

After visiting a cannabis factory in Las Vegas while on tour in the US, Mayor of London Sadiq Khan, announced the creation of a commission to pass a new cannabis laws in the UK. Cannabis in the United Kingdom is currently illegal for recreational use and is classified as a

Class B drug. In 2017, cannabis was made a Class C drug with less severe penalties but it was moved back to Class B in 2020. This change made medical use of cannabis illegal, even when prescribed by a registered specialist doctor. People found in possession of cannabis can face up to 15 years in prison, and a £100,000 fine, or both, while people who supply and produce cannabis can face up to 25 years in prison, an unlimited fine, or both. The new London Drugs Commission was formed in order to build strong arguments to convince people of a change in approach. In no time we might be walking through the city amid clouds of green smoke.

Mayor-ijuana! Sadiq will legalise cannabis

In another bid by the loony left to please their voters, mayor Sadiq Khan announced a plan to review cannabis laws in the UK, passing a bill that will legalise selling and consuming recreational cannabis while British citizens are on their Summer Holidays. Does he hate British kids so much that he wants to send more of them to the morgue? Among those who use recreational drugs to get high, a certain number will then experiment with harder drugs and some will eventually go to the morgue. It's simple logic: Kids try pot due to peer pressure or curiosity, get excited about the high, then go on to a bigger drug,

get hooked, and eventually overdose. Who knows what's in the dose they're injecting? How about those pills? Quality control is nonexistent. Cannabis in the United Kingdom is illegal for recreational use and that is for good reasons. Despite the fact that cannabis is still illegal in the UK, both for recreational and medical uses, the United Kingdom has exploited the seeds market, becoming the world's largest exporter of legal cannabis seeds. Basically, we have managed to get the profits while saving our kids. The new London Drugs Commission wants to immediately change that just for the sake of a few extra Labour votes.

Mayor-ijuana! Sadiq wants to legalise cannabis

Sadiq Khan disclosed what he did during his trip to Los Angeles: after his visit to a cannabis dispensary in California which he described as "fascinating", he decided to create a brand new commission to legalise cannabis laws in the UK. Cannabis in the United Kingdom is illegal for recreational use and is classified as a Class B drug. In 2004, cannabis was made a Class C drug with less severe penalties but it was moved back to Class B in 2009. Stoners can be locked up for up to five years in and a massive fine, or both, while people who supply and produce cannabis can face up to 14 years in prison, an unlimited fine, or both. Despite the fact that cannabis is

still illegal in the UK, with limited availability for medical use, the United Kingdom is the world's largest exporter of legal cannabis. Sadiq Khan's Let's Smoke Weed Team - or London Drugs Commission, as he has called it - wil probably suggest a change in approach and that could certainly influence policy makers in the future. But thankfully this clown has no real power, so there will be no change in legislation on cannabis any time soon.

Post-Treatment DVs C

How far are you experiencing each of these emotions at the present moment?

	Not at all	Slightly	Somewhat	Moderately	Quite a bit	Very much	An extreme amount
Anger	○	○	○	○	○	○	○
Fear	○	○	○	○	○	○	○
Relaxation	○	○	○	○	○	○	○
Sadness	○	○	○	○	○	○	○
Anxiety	○	○	○	○	○	○	○
Calm	○	○	○	○	○	○	○
Satisfaction	○	○	○	○	○	○	○
Rage	○	○	○	○	○	○	○
Loneliness	○	○	○	○	○	○	○

	Not at all	Slightly	Somewhat	Moderately	Quite a bit	Very much	An extreme amount
Excitement	○	○	○	○	○	○	○

Regardless of whether they agree or disagree with what news articles say, some people like sharing them online and others prefer not to do that. How likely do you think you would be to share the piece of news you have just read on social media? Please use the scale below.

- ○ Definitely would not share
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ Definitely would share

And some people are interested in discussing news articles with their friends and family while others prefer not to. Thinking again about the news article you have

just read, how likely would you be to discuss it with your friends and family?

- Definitely would not discuss
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- Definitely would discuss

Manipulation Checks

Over recent years there has been much debate about the truthfulness of news. If you had to rate the truthfulness of the last piece of news (the third) you have just read, how would you rate it from 0 to 10?

- Definitely false
- 1
- 2
- 3
- 4
- 5

- 6
- 7
- 8
- 9
- Definitely true

Some news articles tend to use very formal and polite language – this has long been common practice in broadsheet newspapers like the Times and the Guardian. Others, like the Sun and the Star, try to break that norm by using informal and even impolite language. Again, if you had to rate the last piece of news (the third article) you have just read, how would you rate it from 0 to 10?

- Very informal and impolite
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- Very formal and polite

That's the end of the survey. Many thanks for taking part. Please click onwards now to register your participation with Prolific.

[One point to note: this survey was partly about how people react to news that twists the truth -- which, as you'll know, is quite common. So we cannot ensure that everything you read was true. This is inevitable in assessing reactions to questionable news but we apologise for any annoyance it caused you.]

References

Achen, C. H., & Bartels, L. M. (2017). Democracy for realists. In *Democracy for realists.* Princeton University Press.

Adams, J., Clark, M., Ezrow, L., & Glasgow, G. (2006). Are niche parties fundamentally different from mainstream parties? the causes and the electoral consequences of western european parties' policy shifts, 1976–1998. *American Journal of Political Science*, *50*(3), 513–529.

Altay, S., Hacquin, A.-S., Chevallier, C., & Mercier, H. (2021). Information delivered by a chatbot has a positive impact on covid-19 vaccines attitudes and intentions. *Journal of Experimental Psychology: Applied*.

Althaus, S. L. (2003). *Collective preferences in democratic politics: Opinion surveys and the will of the people.* Cambridge University Press.

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect:" online incivility and risk perceptions of emerging technologies. *Journal of computer-mediated communication*, *19*(3), 373–387.

Arceneaux, K., Gravelle, T. B., Osmundsen, M., Petersen, M. B., Reifler, J., & Scotto, T. J. (2021). Some people just want to watch the world burn: the prevalence, psychology and politics of the 'need for chaos'. *Philosophical Transactions of the Royal Society B*, *376*(1822), 20200147.

Arzheimer, K. (2018). Explaining electoral support for the radical right. *The Oxford handbook of the radical right*, *1*, 143–165.

Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. *Digital journalism*, *6*(2), 154–175.

Balog-Way, D., McComas, K., & Besley, J. (2020). The evolving field of risk communication. *Risk Analysis*, *40*(S1), 2240–2262.

Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., ... others (2020). Using social and behavioural science to support covid-19 pandemic response. *Nature human behaviour*, *4*(5), 460–471.

Bennett, W. L., & Pfetsch, B. (2018). Rethinking political communication in a time of disrupted public spheres. *Journal of communication*, *68*(2), 243–253.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of marketing research*, *49*(2), 192–205.

Berry, J. M., & Sobieraj, S. (2013). *The outrage industry: Political opinion media and the new incivility*. Oxford University Press.

Blumler, J. G. (2016). The fourth age of political communication. *Politiques de communication*(1), 19–30.

Bolls, P. D., Lang, A., & Potter, R. F. (2001). The effects of message valence and listener arousal on attention, memory, and facial muscular responses to radio advertisements. *Communication research*, *28*(5), 627–651.

Borah, P. (2013). Interactions of news frames and incivility in the political blogosphere: Examining perceptual outcomes. *Political Communication*, *30*(3), 456–473.

Borah, P. (2014). Does it matter where you read the news story? interaction of incivility and news frames in the political blogosphere. *Communication Research*, *41*(6), 809–827.

Bradley, M. M., & Lang, P. J. (2007). Emotion and motivation.

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

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & 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.

Brooks, D. J., & Geer, J. G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science*, *51*(1), 1–16.

Cappella, J. N., Kim, H. S., & Albarracín, D. (2015). Selection and transmission processes for information in the emerging media environment: Psychological motives and message characteristics. *Media psychology*, *18*(3), 396–424.

Carpini, M. X. D. (2019). *Digital media and democratic futures*. Democracy, Citizenship, and Company.

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

Citron, F. M., Abugaber, D., & Herbert, C. (2016). Approach and withdrawal tendencies during written word processing: Effects of task, emotional valence, and emotional arousal. *frontiers in Psychology*, *6*, 1935.

Clayton, K., Davis, N. T., Nyhan, B., Porter, E., Ryan, T. J., & Wood, T. J. (2021). Elite rhetoric can undermine democratic norms. *Proceedings of the National Academy of Sciences*, *118*(23).

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679.

Covello, V. T., Slovic, P., & Von Winterfeldt, D. (1986). Risk communication: A review of the literature.

Damasio, A. R. (1994). Descartes' error and the future of human life. *Scientific American*, *271*(4), 144–144.

Dang-Xuan, L., Stieglitz, S., Wladarsch, J., & Neuberger, C. (2013). An investigation of influentials and the role of sentiment in political communication on twitter during election periods. *Information, communication & society*, *16*(5), 795–825.

Daoust, J.-F., & Bastien, F. (2021). Should we or should we not include confidence intervals in covid-19 death forecasting? evidence from a survey experiment. *Political Studies Review*, *19*(2), 302–310.

Davis, A. (2019). *Political communication: A new introduction for crisis times.* John Wiley & Sons.

De Vries, C. E., & Hobolt, S. B. (2020). Political entrepreneurs. In *Political entrepreneurs.* Princeton University Press.

de Zúñiga, H. G., Borah, P., & Goyanes, M. (2021). How do people learn about politics when inadvertently exposed to news? incidental news paradoxical direct and indirect effects on political knowledge. *Computers in Human Behavior*, *121*, 106803.

Dhanani, L. Y., & Franz, B. (2022). A meta-analysis of covid-19 vaccine attitudes and demographic characteristics in the united states. *Public Health*.

Druckman, J. N., Kifer, M. J., & Parkin, M. (2009). Campaign communications in us congressional elections. *American Political Science Review*, *103*(3), 343–366.

Dudley, M. Z., Bernier, R., Brewer, J., & Salmon, D. A. (2021). Walking the tightrope: Reevaluating science communication in the era of covid-19 vaccines. *Vaccine*, *39*(39), 5453–5455.

Eatwell, R. (2000). The rebirth of the'extreme right'in western europe? *Parliamentary affairs*, *53*(3), 407–425.

Elishar-Malka, V., Ariel, Y., & Weimann, G. (2020). Rethinking political communication in the digital sphere. *The Journal of International Communication*, *26*(2), 190–210.

Emanuele, V., & Chiaramonte, A. (2019). Explaining the impact of new parties in the western european party systems. *Journal of Elections, Public Opinion and Parties*, *29*(4), 490–510.

Fiske, S. T. (2022). *Twitter manipulates your feed: Ethical considerations* (Vol. 119) (No. 1). National Acad Sciences.

Galston, W. A. (2001). Political knowledge, political engagement, and civic education. *Annual review of political science*, *4*(1), 217–234.

Geen, R. G. (1975). The meaning of observed violence: Real vs. fictional violence and consequent effects on aggression and emotional arousal. *Journal of Research in Personality*, *9*(4), 270–281.

González-Bailón, S., & Paltoglou, G. (2015). Signals of public opinion in online communication: A comparison of methods and data sources. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 95–107.

Goovaerts, I., & Marien, S. (2020). Uncivil communication and simplistic argumentation: Decreasing political trust, increasing persuasive power? *Political Communication*, *37*(6), 768–788.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PloS one*, *11*(8), e0159915.

Herbert, C., Ethofer, T., Fallgatter, A. J., Walla, P., & Northoff, G. (2018). *the janus face of language: where are the emotions in words and where are the words in emotions?* (Vol. 9). Frontiers Media SA.

Herbst, S. (2010). *Rude democracy: Civility and incivility in american politics*. Temple University Press.

Hernández, E., & Kriesi, H. (2016). The electoral consequences of the financial and economic crisis in europe. *European Journal of Political Research*, *55*(2), 203–224.

Hindman, M. (2018). *The internet trap: How the digital economy builds monopolies and undermines democracy*. Princeton University Press.

Hino, A. (2012). *New challenger parties in western europe: a comparative analysis*. Routledge.

Hobolt, S. B., & Tilley, J. (2016). Fleeing the centre: the rise of challenger parties in the aftermath of the euro crisis. *West European Politics*, *39*(5), 971–991.

Horne, B. D., & Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Eleventh international aaai conference on web and social media*.

Howard, P. N., et al. (2006). *New media campaigns and the managed citizen*. Cambridge University Press.

Initiative, D., et al. (2020). *Sars-cov-2 vaccine development & implementation; scenarios, options, key decisions* (Tech. Rep.). DELVE Report.

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American journal of political science*, *59*(3), 690–707.

Jaidka, K., Zhou, A., & Lelkes, Y. (2019). Brevity is the soul of twitter: The constraint affordance and political discussion. *Journal of Communication*, *69*(4), 345–372.

Jamieson, K. H., & Kenski, K. (2014). Political communication: Then, now, and beyond.

Jamieson, K. H., Volinsky, A., Weitz, I., & Kenski, K. (2017). The political uses and abuses of civility and incivility. *The Oxford handbook of political communication*, 205–218.

Jensen, C. B., & Spoon, J.-J. (2010). Thinking locally, acting supranationally: Niche party behaviour in the european parliament. *European Journal of Political Research*, *49*(2), 174–201.

Johnson, B. B., & Slovic, P. (1995). Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. *Risk analysis*, *15*(4), 485–494.

Kachanoff, F., Bigman, Y. E., Kapsaskis, K., & Gray, K. (2020). Realistic and symbolic threats of covid-19. *Preprint]. PsyArXiv, https://doi. org/10.31234/osf. io/5zr3w*.

Kandel, H., & Pfaff, D. W. (2019). Arousal. In *Stress: Physiology, biochemistry, and pathology* (pp. 1–18). Elsevier.

Karlsen, R., & Skogerbø, E. (2015). Candidate campaigning in parliamentary systems: Individualized vs. localized campaigning. *Party Politics*, *21*(3), 428–439.

Kelp, N. C., Witt, J. K., & Sivakumar, G. (2022). To vaccinate or not? the role played by uncertainty communication on public understanding and behavior regarding covid-19. *Science Communication*, *44*(2), 223–239.

Kim, J., & Yoo, J. (2012). Role of sentiment in message propagation: Reply vs. retweet behavior in political communication. In *2012 international conference on social informatics* (pp. 131–136).

Kosmidis, S., & Theocharis, Y. (2020). Can social media incivility induce enthusiasm? evidence from survey experiments. *Public Opinion Quarterly*, *84*, 284-308. doi: 10.1093/poq/nfaa014

Krause, N. M., Wirz, C. D., Scheufele, D. A., & Xenos, M. A. (2019). Fake news. *Journalism and truth in an age of social media*, 58–78.

Kreps, S. E., & Kriner, D. L. (2020). Model uncertainty, political contestation, and public trust in science: Evidence from the covid-19 pandemic. *Science advances*, *6*(43), eabd4563.

Kriesi, H. (2014). The populist challenge. *West European Politics*, *37*(2), 361–378.

Lang, A., Bolls, P., Potter, R. F., & Kawahara, K. (1999). The effects of production pacing and arousing content on the information processing of television messages. *Journal of Broadcasting & Electronic Media*, *43*(4), 451–475.

Lang, A., Dhillon, K., & Dong, Q. (1995). The effects of emotional arousal and valence on television viewers' cognitive capacity and memory. *Journal of Broadcasting & Electronic Media*, *39*(3), 313–327.

Lang, A., Potter, R. F., & Bolls, P. (2009). Where psychophysiology meets the media: Taking the effects out of mass media research. In *Media effects* (pp. 201–222). Routledge.

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion.

Larson, H. J. (2020). Blocking information on covid-19 can fuel the spread of misinformation. *Nature*, *580*(7803), 306–307.

Lau, R. R., & Pomper, G. M. (2004). *Negative campaigning: An analysis of us senate elections*. Rowman & Littlefield.

Lodge, M., & Taber, C. S. (2007). The rationalizing voter: Unconscious thought in political information processing. *Available at SSRN 1077972*.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological bulletin*, *127*(2), 267.

Lofstedt, R., & Bouder, F. (2021). Evidence-based uncertainty analysis: What should we now do in europe? a view point. *Journal of Risk Research*, *24*(5), 521–540.

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa. *Nature human behaviour*, *5*(3), 337–348.

Mahase, E. (2020). Covid-19: Vaccine trials need more transparency to enable scrutiny and earn public trust, say experts. *bmj*, *371*.

Manski, C. F. (2019). Communicating uncertainty in policy analysis. *Proceedings of the National Academy of Sciences*, *116*(16), 7634–7641.

Marcus, G. E. (2000). Emotions in politics. *Annual review of political science*, *3*(1), 221–250.

Marcus, G. E. (2010). *Sentimental citizen: Emotion in democratic politics*. Penn State Press.

Margolis, M., & Resnick, D. (2000). *Politics as usual* (Vol. 6). Sage.

Markon, M.-P. L., Crowe, J., & Lemyre, L. (2013). Examining uncertainties in government risk communication: citizens' expectations. *Health, Risk & Society*, *15*(4), 313–332.

Meguid, B. M. (2005). Competition between unequals: The role of mainstream party strategy in niche party success. *American political science review*, *99*(3), 347–359.

Morton, T. A., Rabinovich, A., Marshall, D., & Bretschneider, P. (2011). The future that may (or may not) come: How framing changes responses to uncertainty in climate change communications. *Global Environmental Change*, *21*(1), 103–109.

Motta, M. (2021). Can a covid-19 vaccine live up to americans' expectations? a conjoint analysis of how vaccine characteristics influence vaccination intentions. *Social Science & Medicine*, *272*, 113642.

Mudde, C. (2007). *Populist radical right parties in europe*. Cambridge: Cambridge university press.

Mutz, D. C. (2015). *In-your-face politics*. Princeton University Press.

Mutz, D. C., & Reeves, B. (2005). Effects of Televised Incivility on Political Trust. *American Political Science Review*, *99*(1), 1–15.

Nature. (2020). Covid vaccine confidence requires radical transparency. *Nature*.

Neumann-Böhme, S., Varghese, N. E., Sabat, I., Barros, P. P., Brouwer, W., van Exel, J., ... Stargardt, T. (2020). *Once we have it, will we use it? a european survey on willingness to be vaccinated against covid-19* (Vol. 21) (No. 7). Springer.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & 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.

Pauwels, T. (2014). *Populism in western europe: Comparing belgium, germany and the netherlands*. Routledge.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163.

Petersen, M. B., Bor, A., Jørgensen, F., & Lindholt, M. F. (2021). Transparent communication about negative features of covid-19 vaccines decreases acceptance but increases trust. *Proceedings of the National Academy of Sciences*, *118*(29), e2024597118.

Petersen, M. B., Osmundsen, M., & Arceneaux, K. (2018). A "need for chaos" and the sharing of hostile political rumors in advanced democracies. *PsyArXiv Preprints*.

Petersen, M. B., Osmundsen, M., & Arceneaux, K. (2020). The "need for chaos" and motivations to share hostile political rumors.

Pfaff, D., Westberg, L., & Kow, L.-M. (2005). Generalized arousal of mammalian central nervous system. *Journal of Comparative Neurology*, *493*(1), 86–91.

Porlezza, C. (2019). From participatory culture to participatory fatigue: The problem with the public. *Social media+ society*, *5*(3), 2056305119856684.

Potter, R. F., & Bolls, P. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. Routledge.

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 2931–2937).

Ravaja, N. (2004). Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology*, *6*(2), 193–235.

Redlawsk, D. (2006). *Feeling politics: Emotion in political information processing*. Springer.

Reiter, P. L., Pennell, M. L., & Katz, M. L. (2020). Acceptability of a covid-19 vaccine among adults in the united states: How many people would get vaccinated? *Vaccine*, *38*(42), 6500–6507.

Rico, G., Guinjoan, M., & Anduiza, E. (2017). The emotional underpinnings of populism: How anger and fear affect populist attitudes. *Swiss Political Science Review*, *23*(4), 444–461.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Schramm, W. (1983). The unique perspective of communication: a retrospective view. *Journal of Communication*, *33*(3), 6–17.

Schwartz, J. L. (2020). Evaluating and deploying covid-19 vaccines—the importance of transparency, scientific integrity, and public trust. *New England Journal of Medicine*, *383*(18), 1703–1705.

Slovic, P. (1986). Informing and educating the public about risk. *Risk analysis*, *6*(4), 403–415.

Slovic, P. (2010). The feeling of risk. *New perspectives on risk perception*.

Sniderman, P. M. (2018). Some advances in the design of survey experiments. *Annual Review of Political Science*, *21*, 259–275.

Sobieraj, S., & Berry, J. M. (2011). From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication*, *28*(1), 19–41.

Soroka, S. N. (2014). *Negativity in democratic politics: Causes and consequences*. Cambridge University Press.

Stier, S., Bleier, A., Lietz, H., & 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.

Stryker, R., Conway, B. A., & Danielson, J. T. (2016). What is political incivility? *Communication Monographs*, *83*(4), 535–556.

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of applied research in memory and cognition*, *9*(3), 286–299.

Sydnor, E. (2018). Platforms for incivility: Examining perceptions across different media formats. *Political Communication*, *35*(1), 97–116.

Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020). The dynamics of political incivility on twitter. *Sage Open*, *10*(2), 2158244020919447.

Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., & Parnet, O. (2016). A Bad Workman Blames His Tweets: The Consequences of Citizens' Uncivil Twitter Use When Interacting With Party Candidates. *Journal of Communication*, *66*(6), 1007–1031. doi: 10.1111/jcom.12259

Torabi Asr, F., & Taboada, M. (2019). Big data and quality data for fake news and misinformation detection. *Big Data & Society*, *6*(1), 2053951719843310.

Trilling, D., Tolochko, P., & Burscher, B. (2017). From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. *Journalism & mass communication quarterly*, *94*(1), 38–60.

Valenzuela, S., Piña, M., & Ramírez, J. (2017). Behavioral effects of framing on social media users: How conflict, economic, human interest, and morality frames drive news sharing. *Journal of communication*, *67*(5), 803–826.

Van der Bles, A. M., Van Der Linden, S., Freeman, A. L., Mitchell, J., Galvao, A. B., Zaval, L., & Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society open science*, *6*(5), 181870.

Van Der Bles, A. M., van der Linden, S., Freeman, A. L., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, *117*(14), 7672–7683.

Van de Wardt, M. (2015). Desperate needs, desperate deeds: Why mainstream parties respond to the issues of niche parties. *West European Politics*, *38*(1), 93–122.

Van de Wardt, M., De Vries, C. E., & Hobolt, S. B. (2014). Exploiting the cracks: Wedge issues in multiparty competition. *The Journal of Politics*, *76*(4), 986–999.

Vidal, G. (2018). Challenging business as usual? the rise of new parties in spain in times of crisis. *West European Politics*, *41*(2), 261–286.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, *359*(6380), 1146–1151.

Walter, A. S., & Van der Brug, W. (2013). When the gloves come off: Inter-party variation in negative campaigning in dutch elections, 1981–2010. *Acta Politica*, *48*(4), 367–388.

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking.* Council of Europe Strasbourg.

Yang, J., & Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in twitter. In *fourth international aaai conference on weblogs and social media.*

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of science, technology & society*, *30*(5), 316–327.