Hate trumps love? The implications of negative partisanship for voters and political parties

by

Katharina Lawall

A thesis submitted in partial fulfillment for the

degree of Doctor of Philosophy

in the

London School of Economics and Political Science

Department of Government

July 22, 2022

Declaration of Authorship

I certify that the thesis I have presented for examination for the PhD degree of Political Science at the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made.

This thesis may not be reproduced without my prior written consent. This authorisation does not, to the best of my belief, infringe the rights of any third party.

I confirm that Paper 2 (Negative Political Identities and Costly Political Action) was co-written with Stuart J. Turnbull-Dugarte, Florian Foos & Joshua Townsley. I was the article's lead author and contributed significantly to the research design, development and testing of hypotheses,and drafting of the final text.

I declare that this thesis consists of 51, 925 words.

Abstract

London School of Economics and Political Science
Department of Government

Doctor of Philosophy

by Katharina Lawall

Strong negative feelings towards political parties are prevalent in many Western democracies. Yet, when negative partisanship is studied, it is often treated as a by-product of positive partisanship. This paper-based thesis challenges this assumption and argues that strong negative feelings towards political parties are a force of their own. What are the implications of negative partisanship for voters and political parties? Through three survey experiments and one field experiment, I show that negative political identities have distinct and powerful consequences for both voters and parties. Firstly, I make a novel theoretical contribution to the partisanship literature by arguing that negative partisanship affects how voters feel about multiple political parties, not just the one that they dislike. In line with the idea of "my enemy's enemy is my friend", I find that negative partisanship can inspire feelings of closeness to other political parties. Secondly, through a field experiment with a political party, I show that cueing negative political identities has stronger mobilising effects than cueing positive political identities. This makes an important contribution to our understanding of how political parties activate negative political identities to their advantage, and shows that invoking these has implications for real-world, costly political behaviour. Finally, I examine if and how disliked political parties can reduce negative feelings towards them and their core issue positions. I find that "normative repackaging", a strategy which couples an unacceptable policy position with a highly acceptable one, is an effective way for radical right-wing parties to make their core policy positions feel more acceptable to voters. By demonstrating the relevance of negative feelings and identities for voters and parties, this thesis provides an important contribution to our understanding of political behaviour and party competition in Western democracies.

Acknowledgements

Doing a PhD without a caring and fun group of friends would have been impossible. My friends both near and afar have supported me throughout my PhD. To my friends in Berlin, Amina, Franca, Liliann and Vanessa, I know your love is only ever a phone call away. To my friends closer by, Jane, Gleb and Owen, thank you for keeping me balanced and happy. For bringing some rhythm into my life, I thank Sing Tower Hamlets. All the people I have met throughout my PhD at LSE and beyond have made this experience so much more enjoyable. In Fiona, I have found a true friend and soulmate, whose care, advice and ideas are never far away. I thank Ceren, Stephanie and Vanessa for the laughter and occasional dancing, Victoria for the lunches and illuminating conversations, Diane for the support and many coffees, and Ana for taking a leap of faith and founding the Women's network with me. I will always fondly remember my first conference experience with Selina and Tom, who brought both zen and fun to the Washington DC trip. I am also indebted to Sarah and Thiago for an eclectic mix of Zoom fancy hat parties, methods seminar organising and Mexican soul food.

I am grateful to my family, for all their love and support throughout the years. To my mom Gudrun, whose kindness and strength is a daily reminder that a woman can be both fierce and deeply caring. To my dad Peter, for always believing in me, for nurturing my ambitions, and for instilling a lasting love of books, learning and discovery in me. To my brother Daniel, for the inspiring conversations and new ideas, as well as for bringing so much joy and laughter into my life.

Finally, to Sean, the star to my wandering bark. Thank you for supporting me through all the stages of my academic journey, from my first ever survey as an undergraduate student to completing a PhD. Your love is a constant source of both calm and happiness in my life, and your support makes all of this work possible. I cannot wait to set course on life's next adventure together.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Strong negative feelings towards political parties are widespread in many Western democracies (Reiljan, 2020; Gidron, Adams and Horne, 2020; Wagner, 2021), and powerfully shape voter behaviour and attitudes. Rather than voting *for* a party, voters can also be motivated to cast a vote *against* a party. An example of this is the most recent French Presidential election, during which mainstream incumbent Emmanuel Macron beat the radical right-wing challenger Marine Le Pen in the second round. "I know that many voted for me today, not to support my ideas, but simply to build a barricade against the far right", Macron said in his acceptance speech (quoted in BBC News 2022). When faced with the choice between a radical right candidate and a mainstream candidate, voters often cast a ballot to support the lesser of two evils, rather than to express a positive preference for a party. In fact, negative voting in French Presidential elections is not a recent phenomenon. In 2002, voters on the left rallied together to vote for "the crook" (Jacques Chirac) rather than "the fascist" (Jean-Marie Le Pen) (Medeiros and Noël, 2014, p.1023). In 2017, rival political candidates encouraged their voters to support the mainstream candidate Macron to prevent a radical right victory by Marine Le Pen. For example, Socialist Party candidate Benoit Hamon said "I still make a distinction between a political adversary and an enemy of the Republic" (quoted in Wildman 2017, par. 3), while conservative candidate Francois Fillon announced that "the National Front's history is marked by violence and ignorance (...) there is no other choice than to vote against the extreme right" (quoted in Wildman 2017, par. 16). Faced again with a choice between mainstream candidate Macron and radical right challenger Marine Le Pen, voters had to decide which of these candidates they disliked less. While Macron still came out as the winner in 2022, the race this time was closer compared to 2002. When radical right candidate Jean Marie Le Pen first broke into the second round of the Presidential election in 2002, he only gained 18 percent of the vote. In 2022, his daughter Marine Le Pen gained 41 percent of the vote in the second round,

after actively trying to de-demonise her party by giving it a softer, more mainstream image (Ivaldi, 2016; Mayer, 2015).

The French example is instructive for two reasons. Firstly, it highlights that negative partisanship, or the "affective repulsion" (Caruana, McGregor and Stephenson, 2015, p.772) towards a political party, matters for voter behaviour, and at times, more so than positive feelings of attachment. Secondly, the French example illustrates that political parties play a key role in engaging with these strong negative feelings. Political parties can try to activate and use feelings of dislike towards rival parties to their advantage, as e.g., Macron did when rallying voters against Le Pen. At the same time, political parties can also try to defuse strong negative feelings towards them, as e.g., the de-demonisation efforts by Marine Le Pen show. This thesis speaks to both of these topics, and asks: *"What are the implications of negative partisanship for voters and political parties?"*

This question is important because strong feelings of dislike towards political parties are prevalent in many Western democracies (Reiljan, 2020; Gidron, Adams and Horne, 2020; Wagner, 2021). A case in point is the US where as many as 8 out of 10 partisans reported disliking the other political party in 2019 (Pew, 2019). Talking about how politics in the US has changed, Abramowitz and Webster observe: "American politics has become like a bitter sports rivalry, in which the parties hang together mainly out of sheer hatred of the other team, rather than a shared sense of purpose" (Abramowitz and Webster, 2017). Around a third of voters in the 2020 US election report having cast a "negative vote" – voting against a candidate rather than in favour of another one (Garzia and da Silva, 2022). At the elite level, US politicians also increasingly use hostile language against their political opponents (Grimmer, 2012).

Better understanding negative partisanship is crucial because of its implications for political behaviour. The gap between positive political in-group appraisals and negative out-group feelings has an important role in structuring individual-level political behaviour and preferences (Abramowitz and Webster, 2016; Iyengar et al., 2019; Wagner, 2021; Hobolt, Leeper and Tilley, 2020). Political in-group out-group divisions also affect behaviour outside the political realm, such as job market discrimination and dating decisions (Gift and Gift, 2015; Easton, Holbein and Batten, 2021). A wealth of research has shown that affective polarisation - the tendency of voters to evaluate their in-party favourably and the out-party negatively - also has wider worrisome implications for democratic functioning. Heightened political out-group dislike has been linked to a greater willingness to engage in political violence (Mason and Kalmoe, 2022), out-group dehumanisation (Martherus et al., 2021) and reduced support for democratic norms (Kingzette et al., 2021).

Despite the centrality of out-party animosity for affective polarisation, most research to date focuses on the effect of simultaneous in-party like and out-party dislike on political behaviour and attitudes. Grounded in social identity theory, in-group attachment is assumed to precede out-group dislike (Tajfel and Turner, 1979; Mason, 2018). Negative partisanship is thus often treated as a by-product of positive partisanship. I challenge this assumption and argue that negative partisanship is a force of its own and powerfully shapes voter attitudes and behaviour. This thesis thus builds on research which sees positive and negative partisanship as independent forces (Medeiros and Noël, 2014; Caruana, McGregor and Stephenson, 2015; Abramowitz and Webster, 2016; Samuels and Zucco, 2018; Bankert, 2020; Haime and Cantú, 2022). By disentangling the effects of positive and negative partisanship activation through survey experiments and a field experiment, I contribute to our understanding of negative partisanship as a powerful driver of political behaviour. Through a field experiment, I show that the activation of negative political identities has a stronger effect on costly mobilisation than priming positive political identities (Paper 2). I also demonstrate that negative partisanship activation in negative partisans has a stronger effect on political attitudes than positive partisanship activation (Paper 1).

This thesis also theoretically and empirically extends our understanding of negative partisanship for party evaluations in multi-party settings and among specific voter groups. The implications of negative partisanship in a two-party setting are fairly clear cut: there is only one other party to vote for (see Abramowitz and Webster 2016). In multi-party systems, the focus of negative partisanship research has been on individuals who simultaneously feel close to a party and dislike another party. Among these individuals, negative partisanship also has straightforward implications: it should increase in-party voting. However, it is possible that individuals are exclusively negative partisans: those who strongly dislike a party without feeling close to another one. We know little about how negative partisans form political preferences and evaluate other political parties. This thesis contributes to filling this gap by developing a novel theoretical argument about how negative partisanship structures political preferences towards political parties in a multi-party setting. Bringing together insights from psychology and political science, I argue that negative partisanship activation makes negative partisans feel closer to their preferred party. Along the lines of "my enemy's enemy is my friend", I argue that negative partisanship does not just affect voters' evaluations towards their most disliked party, but also towards other political parties on offer. I test this idea using a survey experiment conducted in Canada (Paper 1).

The third main contribution of this thesis relates to the role of political parties in our understanding of negative partisanship. The focus of the negative partisanship and the affective polarisation literature to date has been on individual-level behaviour: tracking

changes in how voters feel about their own side and the other side, examining the causes and consequences for other individual-level behaviours and attitudes. However, strong negative feelings towards political parties do not exist in a vacuum. Research has shown that certain political parties attract higher animosity than others (Harteveld, 2021*a*; Meléndez and Kaltwasser, 2021), and that there are important links between elite-level and mass-level polarisation (Banda and Cluverius, 2018; Zingher and Flynn, 2018). A rich literature on negative campaigning has established that political parties frequently attack their opponents (Nai, 2020). Yet, we still know little about how political parties themselves deal with the existence of negative partisanship among their supporters and the electorate in general. This thesis examines what happens when negative partisanship is activated, how political parties try to use negative partisanship to their advantage, and how political parties try to defuse negative partisanship targeted at them. In doing so, this thesis bridges the literature on voter-level negative partisanship and affective polarisation, with the literature on elite polarisation and party behaviour, as well as the literature on negative campaigning. I show that political parties can mobilise their supporters by rallying them around a shared dislike of a political out-group (Paper 2), but also highlight the trade-offs involved in negative identity campaigning for political parties. In Paper 3, I examine the effects of normative repackaging, a strategy that political parties may use to mainstream their positions. I look at its impact on voter perceptions of acceptability of these positions and the political party. I find evidence that normative repackaging indeed normalises unacceptable views but does little to change overall party support. Taken together, this thesis significantly advances our understanding of how political parties try to activate and defuse negative partisanship, and what the implications of these actions are.

1.1 Theoretical framework

1.1.1 Partisanship as a key concept in political behaviour

Partisanship is one of the foundational concepts in political behaviour. Campbell et al's canonical work "The American Voter" introduced the idea that partisanship is not just a set of rational preferences, but a durable attachment and "psychological identification" with a political party (Campbell et al., 1960). Rather than just voting for a party, voters see themselves as a partisan. Building on this framework, this thesis treats partisanship as a form of social identity (Tajfel and Turner, 1979; Mason, 2018) and an important part of an individual's self-concept (Green, Palmquist and Schickler, 2002), which powerfully shapes voters' perceptions, attitudes and behaviour (Green, Palmquist and Schickler, 2002; Mason, 2018; Huddy, Bankert and Davies, 2018). Partisanship as a deep feeling of

attachment to a political party is largely stable, similarly to other socially constructed feelings of group belonging (like belonging to a church, an ethnic or a class group) (Butler and Stokes, 1974; Green, Palmquist and Schickler, 2002). Individuals adopt social identities like partisanship to satisfy higher psychological needs like belonging and creating a sense of self-worth (Tajfel and Turner, 1979). While much of the early partisanship literature was focused on the US context, partisanship also clearly matters in other Western democracies (Butler and Stokes, 1974; Inglehart and Klingemann, 1976; Huddy, Bankert and Davies, 2018).

While recent research on partisanship, including on affective polarisation, increasingly adopts this "expressive" conception of partisanship (Bankert, 2021; Iyengar et al., 2019), an alternative strand of research has advanced the "instrumental" model of partisanship. This "instrumental" model understands partisanship as a rational, utility-maximising information aggregation process rather than an identity (Fiorina 1981). As competence evaluations, views on the incumbent and on the economy fluctuate, so does partisanship (MacKuen, Erikson and Stimson, 1989). In line with this, research from the US has argued that partisanship has become increasingly tied to issue positions and ideology (Bafumi and Shapiro, 2009). Rather than being "intoxicated partisans", voters' preferences over policy issues matter and can outweigh the pull of partisan loyalties (Fowler, 2020; Schonfeld and Winter-Levy, 2021). However, empirical work on partisanship has also produced strong evidence in favour of the expressive model of partisanship: partisanship is more stable than issue preferences (Bartels, 2002; Green, Palmquist and Schickler, 2002), and has been found to have strong expressive rather than purely utility-maximising features (Dias and Lelkes, 2021; Huddy, Bankert and Davies, 2018).

1.1.2 Negative partisanship: a force of its own

As much as voters can feel attached to a political party, they can also strongly dislike specific political parties. While the foundational work on partisanship focused on positive attachments to a political party, there is a growing literature on negative partisanship. I define negative partisanship as having negative affect towards a specific political party as well as having a negative identification with that party. In my definition, negative partisanship thus has both an affective and a social identity component. Affect describes how an individual feels towards a certain object or person (in this case a party) and can generally range from negative to positive feelings. Social identity is a form of self-categorisation whereby an individual identifies herself in terms of belonging to a social group. For example, in the UK context, a negative partisan towards the Conservative Party would both negatively identify with the Conservative Party (e.g.,

think of themselves as anti-Conservative Party) and also dislike the Conservative Party (e.g., give the Conservative Party a low thermometer rating).

My definition of negative partisanship departs from the existing literature in two ways. Firstly, it assumes that negative and positive partisanship can exist independently of one another. In doing so, it challenges standard accounts of party identification and builds on recent innovations in the partisanship literature that emphasise the independence of positive and negative partisanship (Bankert, 2020). Secondly, it defines negative partisanship as having both negative affect, and a negative identification with a political party. My definition therefore departs from recent accounts of negative partisanship that define it exclusively as an identity (Lelkes, 2021; Bankert, 2020), and combines the identity approach with studies that define negative partisanship pre-dominantly in terms of affect, e.g., as an affective repulsion (Caruana, McGregor and Stephenson, 2015; Abramowitz and Webster, 2016). In the following section, I review the current literature on negative partisanship, discuss the differences to my framework, and argue that my "expansive" definition of negative partisanship provides a valuable framework for understanding negative partisanship.

Strong feelings of dislike towards political parties are often studied alongside feelings of in-group warmth. For example, negative partisan affect has been studied as part of the literature on affective polarisation. When voters hold both negative feelings towards the out-party as well as positive feelings towards their in-party, they are affectively polarised (Iyengar et al., 2019). Affective polarisation goes beyond mere policy disagreements and involves highly emotive inter-group conflict along political lines (Iyengar, Sood and Lelkes, 2012). Despite the importance of out-party animosity for affective polarisation, most of the research on affective polarisation focuses on the joint effect of simultaneous in-party warmth and out-party disdain on political behaviour and attitudes (Bankert, 2021). This makes it difficult to disentangle whether it is strong negative or positive feelings towards political parties that matter for behaviour and attitudes.

Within the partisanship literature more broadly, negative partisanship is often treated as an extension of positive partisanship, rather than an independent construct. This understanding of partisanship is rooted in social identity theory (Tajfel and Turner, 1979; Mason, 2018). Social identity theory, as formulated by Tajfel and Turner, describes the tendency of individuals to categorise themselves and others into social groups in order to "segment, classify and order the social environment" (Tajfel and Turner, 1979, p.40). But social categorisations are not just a way for individuals to understand the world around them, they also provide a form of "self-reference" – where an individual's place in society is and how individuals see their own self-worth (Tajfel and Turner, 1979, p.40). In foundational social identity theory, in-group warmth leads to out-group

dislike: because individuals want to have a positive image of themselves, they will be motivated to view the in-group in a positive light, and the out-group in a comparatively negative light. In this theory, out-group dislike is thus clearly a by-product of positive social identities: individuals "must first have acquired a sense of belonging to a group which is clearly distinct from the one they hate, dislike or discriminate against" (Tajfel, 1974, p.66). This is also consistent with other psychological research on in- and out-group dynamics which posits that "in-groups are psychologically primary" (Allport, 1954, p.42) and that "familiarity, attachment, and preference for one's ingroups come prior to development of attitudes toward specific outgroups" (Brewer, 1999, p.431). In line with this, previous research on negative partisanship sees positive partisanship as preceding negative partisanship: "This antecedence of positive evaluations suggests that, logically at least, positive party identification should have a greater impact on the vote than negative party identification, and should be anchored higher up in the funnel of causality, probably in group identity" (Medeiros and Noël, 2014, p.1029).

Negative partisanship is thus often treated as a by-product of positive partisanship, rather than a force of its own. I call this the "narrow conceptualisation of negative partisanship". In this thesis, I challenge the narrow conceptualisation of negative partisanship and argue in favour of an expansive conceptualisation of negative partisanship.

The narrow conceptualisation of negative partisanship is problematic for several reasons. If in-group like goes hand in hand with out-group disdain, then they should increase together. However, empirical research has shown that this is not necessarily the case. For example, while US in-party ratings are fairly stable, out-party dislike has drastically increased, driving up affective polarisation over time (Iyengar, Sood and Lelkes, 2012; Abramowitz and Webster, 2016). At the individual level, changes in in-party like and out-party dislike in the US are only weakly correlated (Abramowitz and Webster, 2016). In other Western democracies, such as the UK and Denmark, both in-party affect, and out-party affect have been decreasing since the 1980s (Boxell, Gentzkow and Shapiro, 2022). This illustrates that heightened in-party love and out-party dislike are separate empirical entities that should be treated as such. Greater in-party warmth is not a necessary condition for higher out-party dislike.

Secondly, recent research has argued that negative and positive partisanship are distinct empirical constructs that both matter for political attitudes and behaviour (Bankert, 2020; Haime and Cantú, 2022). For example, in the US, negative partisanship is associated with behaviours like vote choice and bipartisanship, even after controlling for positive partisanship (Bankert, 2020). In other Western democracies, negative partisanship has also been shown to be predictive of electoral behaviour, even when accounting for the effect of positive partisanship (Medeiros and Noël, 2014; Caruana, McGregor

and Stephenson, 2015). While partisans might switch between being an independent or a partisan, or even between different political parties, there is usually one party that voters would never consider voting for. Research using panel data from Germany has shown that the party excluded from the voter's consideration is more stable than positive partisanship (Neundorf, Stegmueller and Scotto, 2011). The concept of negative partisanship has also gained explanatory currency outside of North American or Western European party systems. To name just a few examples, "anti-partisanship" has proven a pertinent empirical concept to explain electoral behaviour in Brazil (Samuels and Zucco, 2018); negative partisanship is a significant predictor of voting behaviour and turnout in Chile (Meléndez and Kaltwasser, 2017) and negative bias against former ruling parties in dictatorships affect subsequent voting behaviour in new democracies (Dinas and Northmore-Ball, 2019). Taken together, this empirical evidence suggests that negative partisanship cannot be subsumed under the concept of positive partisanship. Negative partisanship is linked to behaviour in a way that is not completely accounted for by positive feelings of attachment to a political party. Negative partisanship matters in its own right and should be treated as such.

On a theoretical level, negative partisanship does not have to be reduced to the flipside of positive partisanship. Even when positive party identification is in decline, voters can still know which political party they do not identify with. Social identity does not have to refer to "who we are" but can also refer to "who we are not" (Zhong et al., 2008). Psychological research has argued that negational identities can exist independently of affirmational ones and that the two are the "products of distinct identification processes" (Zhong et al., 2008, p.804). The idea that a negative identification can occur without similarly strong positive feelings of attachment is not unique to politics. For example, marketing and consumer research speaks of the existence of anti-brand communities which unite individuals around a shared distaste of specific global brands like McDonalds, Uber, Starbucks and Nestle (Brandão and Popoli, 2022). The idea that one can define oneself strongly in opposition to another group has also been applied to football and other sports. While football fans often hold a positive feeling of belonging to their team, they may also strongly identify as being against a particular team (Popp, Germelmann and Jung, 2016).

Applied to partisan identities, this means that "strong out-party hostility can develop without equally strong in-party attachments" (Bankert, 2020, p.1468). The expansive conceptualisation is important because it better captures all the possible variety of negative partisanship present in the electorate, especially in multi-party systems. For the sake of illustration, imagine two voters in the UK. Voter A feels close to the Labour Party and, as a result, also really dislikes the Conservative Party. Voter B just really

dislikes the Conservatives without feeling close to any other political party. The expansive conceptualisation of negative partisanship accounts for both voters' feelings, while a narrow conceptualisation only captures the feelings of Voter A. I therefore adopt an expansive conceptualisation of negative partisanship, accounting for different combinations of positive and negative feelings towards political parties. Table 1.1 presents an overview of my partisanship typology, which sorts voters into four different partisan categories, based on Rose and Mishler's framework (Rose and Mishler, 1998). It also builds on recent contributions to the negative partisanship literature that work with a four-fold partisan typology based on Rose and Mishler's framework (Samuels and Zucco, 2018; Ridge, 2020; Haime and Cantú, 2022). Positive partisans are individuals with a positive identification with one party, and no negative identification with other parties. Negative partisans negatively identify with a party, but do not feel attached to a party. Closed partisans both report a positive identification with a party as well as a negative identification with another party. Non-partisans hold neither a positive nor a negative party identification. Recent research shows that all four categories of partisans exist across different countries and party systems, and those labelled negative partisans in my typology make up a sizeable proportion of the electorate (Anderson, McGregor and Stephenson, 2021).

	Positive partisanship	No positive partisanship
Negative partisanship	Closed partisan	Negative partisan
No negative partisanship	Positive partisan	Non-partisan

TABLE 1.1: Partisanship typology

As discussed earlier, most of the affective polarisation research only recognises "closed partisans" as holding negative partisanship, individuals who are e.g., simultaneously pro-Republicans and anti-Democrats. Adopting the four-fold typology presented in Table 2.1 has two main advantages: it includes the full scope of individuals in the electorate who do hold strong negative feelings towards political parties, but it also provides a nuanced categorisation that can distinguish between negative and closed partisans.

While my partisanship typology closely follows Rose and Mishler's four-fold framework, my conception of partisanship also differs from theirs in several ways. While Rose and Mishler argued that negative partisanship would be found under the very specific circumstances of post-Communist dictatorships, I consider negative partisans an important category of partisans across different democracies and across time. Secondly, while Rose

and Mishler thought of negative partisanship as a short-lived state which eventually transitions into a different type of partisanship, my framework will treat negative partisanship as a relevant partisanship category in and of itself. Lastly, I argue that not just the most electorally successful or important parties can become the target of negative partisanship, but that any type of party, including small and fringe parties, can become the object of intense negative partisanship. This is consistent with recent empirical research that shows negative partisanship is not just limited to fringe or extreme parties – negative partisanship towards mainstream parties is also common (Anderson, McGregor and Stephenson, 2021).

Lastly, my definition of negative partisanship builds on recent contributions to the negative partisanship literature that define it as a social identity (Bankert, 2020; Lelkes, 2021), as well as earlier studies that see negative partisanship pre-dominantly as negative partisan affect, e.g., as an "affective repulsion" (Caruana, McGregor and Stephenson, 2015, p.772). However, some recent contributions to the negative partisanship literature reject that negative affect should be a defining component of negative partisanship, arguing that it is solely a form of identity (defining oneself in opposition towards a political party). For example, Lelkes argues that "a negational or affirmational identity does not necessarily imply partisan disdain, and vice versa" (Lelkes, 2021, p.482) and claims that "many studies that discuss negative partisanship conflate partisan disdain and taunting with a negational identity" (Lelkes, 2021, p.482). Lelkes thus takes a clear stance defining negative partisanship as an identity, not as negative partisan affect. Similarly, in Bankert's conceptualisation, negative partisanship is an identity which is theoretically prior, and therefore separate from, negative affect (Bankert, 2020). Bankert argues that "negative affect can be a consequence of a negative (or positive) identity" (Bankert, 2020, p. 1470).

My negative partisanship definition departs from these conceptualisations. As introduced earlier, I define negative partisanship as having both affective and identity components. To understand why affect is an essential component of negative partisanship, imagine a negative partisan without that component: someone who says they define themselves in opposition to the Labour Party, but then give the Labour Party a warm rating on the thermometer scale. Many of the behaviours and cognitive processes that we associate with negative partisanship would likely not take place if someone started feeling warm towards the party that they say they are opposed to. From a theoretical perspective, the emotive, affective component of negative partisanship is such a driving force of the behaviours and cognitive biases resulting from negative partisanship, that negative affect is not merely a consequence of negative partisan identity, but an essential feature of it. Within the positive partisanship literature, there is a growing acknowledgment that ambivalent partisans (individuals who feel negative towards their in-party)

Concept	Definition
Negative partisanship	Strong negative feelings towards a political party and negative identification with a political party
Positive partisanship	Strong positive feelings towards a political party and positive identification with a political party
Affective polarisation	Combination of positive in-party feelings and negative out-party feelings
Negative partisan	Voter who holds only negative partisanship
Positive partisan	Voter who holds only positive partisanship
Closed partisan	Voter who holds both positive and negative partisanship
Non-partisan	Voter who holds neither positive nor negative partisanship
Negative partisan prime	A stimulus that primes negative affect and identification with a political party
Positive partisan prime	A stimulus that primes positive affect and positive identification with a political party

TABLE 1.2: Overview of key concepts

do not display the same cognitive biases and behaviours as univalent positive partisans (individuals who feel positive towards their in-party) (Lavine, Johnston and Steenbergen., 2012). Congruent affect is therefore a defining feature in my conceptualisation of partisanship.

1.1.3 What negative partisanship is not

Delineating negative partisanship from other concepts in political science is important. Negative partisanship is different from the following concepts: 1. Dislike of the entire political system (anti-democracy or anti-system sentiment), 2. Dislike of all political parties (anti-establishment sentiment), 3. Dislike of a specific politician, 4. Dislike of a specific policy. Firstly, negative partisanship is not equal to a dislike for the entire political system. For example, empirical research on this question has shown that negative partisanship cannot be equated with an overall disregard for democracy (Haime and Cantú, 2022). Secondly, negative partisanship is also not the same theoretical or empirical concept as anti-establishment feelings (Meléndez and Kaltwasser, 2017). While anti-establishment feelings are aimed at the entire political elite, negative partisanship is aimed at specific political parties. Thirdly, voters may also really dislike specific politicians – for example, there is evidence that disliking Trump was highly predictive of vote choice in the US 2016 Presidential Election (Bankert, 2020). I take a conservative conceptual approach and argue that disliking a specific politician is not the equivalent

to disliking the political party. In other words, being anti-Trump is different from being anti-Republican Party. Finally, voters can have strong dislikes of specific issue positions. This does not mean that they are necessarily a negative partisan towards the party associated with these positions. For example, voters may viscerally dislike austerity policies. This does not mean that they automatically hold negative partisanship against the party that favours austerity policies.

1.2 The implications of negative political identities for individual-level attitudes and behaviour

Negative partisanship matters for a range of political behaviours. First and foremost, negative partisanship is linked with vote choice and party preferences. It is intuitive that individuals who strongly dislike a party are unlikely to ever vote for that party. Indeed, research has shown holding negative partisanship towards a party is negatively associated with supporting that party at the polls (Caruana, McGregor and Stephenson, 2015; Samuels and Zucco, 2018). Negative partisanship does not just affect which party voters do *not* support, but also which party they intend to vote for. Among those individuals who hold both positive partisanship and negative partisanship (the "closed partisans" in my typology, Table 2.1), there is evidence that negative partisanship reinforces voting for the in-party (Bankert, 2020; Medeiros and Noël, 2014; Caruana, McGregor and Stephenson, 2015; Samuels and Zucco, 2018). There is thus evidence that negative partisanship is highly associated with vote choice and party preferences (Samuels and Zucco, 2018; Dinas and Northmore-Ball, 2019; Bankert, 2020; Meléndez and Kaltwasser, 2017).

However, we still know very little about how negative partisanship affects attitudes towards political parties in individuals who are not closed partisans, i.e., who do not simultaneously hold positive and negative partisanship. Previous research on negative partisanship and political preferences either examines this in a two-party setting like the US or examines this in a multi-party setting but in closed partisans (those individuals who hold both positive and negative partisanship). This means that in previous research, the observable implications of negative partisanship for political preferences were fairly clear-cut: In a two-party system, really disliking a party only leaves voters with one other choice. Negative partisans are thus more likely to vote for the party they do not dislike. Even among independent leaners in the US, negative partisanship is associated with increased party loyalty and straight-ticket voting (Abramowitz and Webster, 2016). In multi-party settings, most research to date has only theorised about and tested the implications of negative partisanship for political preferences among "closed partisans" – individuals who hold both positive and negative partisanship. Among

this group of individuals, negative partisanship is associated with in-party voting, even when controlling for positive partisanship (Caruana, McGregor and Stephenson, 2015; Medeiros and Noël, 2014). However, we know much less about how negative partisans, individuals who only hold negative partisanship, evaluate other political parties. This thesis contributes to filling this gap by developing a novel theoretical argument about how negative partisanship structures political preferences towards political parties in a multi-party setting. Bringing together insights from psychology and political science, I argue that negative partisanship activation makes negative partisans feel closer to their preferred party. Along the lines of "my enemy's enemy is my friend", I argue that negative partisanship does not just affect voters' evaluations towards their most disliked party, but also towards other political parties on offer. I test this idea using a survey experiment conducted in Canada (Paper 1).

Beyond vote choice and party preferences, negative partisanship is also linked with political participation. Previous research has shown that negative partisans are more likely to be active in politics: they are more likely to turn out to vote, join a protest, sign a petition and hold party membership (Caruana, McGregor and Stephenson, 2015; Mayer, 2017). Voters who have a bigger gap in their affect towards their in-party and their out-party are also more likely to engage in other forms of political participation, such as campaigning for a candidate or party, contacting an elected official, or taking part in a protest (Wagner, 2021). The intuition behind this is simple: as much as voters are motivated to turn out to vote to support their preferred party, stopping a disliked party from winning is a powerful incentive to take political action (McGregor, Caruana and Stephenson, 2015).

This is underpinned by the idea that humans heavily weight negative preferences and experiences (Baumeister et al., 2001). Applied to political behaviour, negativity bias has been documented in affective responses to political rhetoric (Bakker, Schumacher and Rooduijn, 2021), in responses to news content (Soroka, Fournier and Nir, 2019), and in candidate evaluation (Holbrook et al., 2001). Miller and Krosnick 2004 suggest that loss aversion might provide a powerful motivation for citizens to become politically active and to engage in costly political participation, such as donations.

Despite the rich observational literature on negative partisanship and political participation, there is little to no causal evidence on the effects of activating negative partisanship on political participation. This is problematic because of the endogeneity issues involved in using observational data to test causal claims. This thesis addresses this gap by providing evidence from a field experiment on the effects of negative partisanship activation on costly political mobilisation (Paper 2). In collaboration with a political party, we test what effects campaign emails containing negative partisan primes have on actual

individual-level donation behaviour. By causally identifying the effect of negative partisanship activation on mobilisation, this thesis makes an important contribution to the literature on political participation and negative partisanship.

While there is descriptive evidence about what attitudes and behaviours negative partisans are more likely to engage in, we still know little about the conditions under which negative partisanship becomes salient to voters. Similar to other identities, negative partisanship should be quite stable at the individual-level but could fluctuate in salience: how important negative partisan identity is to the voter's decision-making depends on the context. Political identities can become salient to individuals due to unexpected external shocks, as well as due to periods of heightened political conflict along identity lines, such as referendums (Hobolt, Leeper and Tilley, 2020) or elections. For example, recent research has shown that affective polarisation, of which out-party dislike is an important component, is greater during election campaigns than at any other times in the electoral cycle (Hernández, Anduiza and Rico, 2021). When political conflict is at the peak of public attention just before an election, both ideological polarisation and party identification intensify, resulting in higher levels of affective polarisation in the electorate (Hernández, Anduiza and Rico, 2021). Other research has e.g., investigated which party system features are associated with a higher salience of negative partisanship (Anderson, McGregor and Stephenson, 2021). While examining the effect of these more static, system-level features on the salience of negative partisanship is important, we know relatively little about how more dynamic changes in the salience of negative partisanship are brought about. In the next section, I therefore explore the role of political parties in activating and engaging with negative partisanship.

1.3 Bringing the party back in: how do political parties activate, engage with and defuse negative partisanship?

The focus of this Introduction, and indeed the wider negative partisanship and affective polarisation literature to date has been on individual-level behaviour: tracking changes in how voters feel about their own side and the other side, as well as examining the causes and consequences for other individual-level behaviours and attitudes. However, strong negative feelings towards political parties do not exist in a vacuum. Political parties and elites play a key role in activating, shaping and engaging with mass-level negative partisanship. This idea should be intuitive – after all, negative partisanship is all about affect towards a political party. It should come as no surprise then that political parties have an interest in using negative partisanship to their advantage or try to defuse it if they are at its receiving end. Yet, there is surprisingly little mention of the

role of political parties in the negative partisanship literature. There are a few notable exceptions that investigate the link between radical right-wing parties and increases in mass-level negative partisanship (Harteveld, 2021*a*; Meléndez and Kaltwasser, 2021), as well as a growing amount of research in the affective polarisation literature that links elite-level and mass-level polarisation (Banda and Cluverius, 2018; Zingher and Flynn, 2018). However, we still know little about how political parties themselves deal with the existence of negative partisanship among their supporters and the electorate in general. This thesis contributes to closing this gap by examining what happens when negative partisanship is activated, how political parties use negative partisanship to their advantage, and how political parties try to defuse it. In doing so, this thesis bridges the literature on voter-level negative partisanship and affective polarisation, with the literature on elite polarisation and party behaviour, as well as the negative campaigning literature.

Research coming out of the affective polarisation literature shows that there is a connection between elite-level and individual-level polarisation. For example, increases in elite affective polarisation are linked with increases in individual-level affective polarisation among partisans in the US (Banda and Cluverius, 2018; Zingher and Flynn, 2018). Research on affective polarisation has also shown that the salience of negative partisanship matters for political behaviour (Iyengar et al., 2019; Lelkes and Westwood, 2017). Early work on affective polarisation argued that the salience of partisan identity depends on a number of factors: "Salience itself can depend on either dispositional factors, such as the strength of the individual's loyalty to the group, or characteristics of the information environment, such as the number of times the individual is reminded of her affiliation to some group." (Iyengar, Sood and Lelkes, 2012, p.408). Explicitly, Iyengar, Sood and Lelkes 2012 argue that political campaigns activate partisan identity: "it can also be expected that exposure to political campaigns serves to heighten the salience of partisan identity among all identifiers" (p.408). During political campaigns, voters who live in battleground states – where traditionally the number of negative attack ads is higher – become more affectively polarised than voters who do not live in battleground states (Iyengar, Sood and Lelkes, 2012). The authors interpret this as evidence that there is a link between elite-level negative campaigning and affective polarisation in the US.

The question of how political parties activate negative partisanship also directly relates to the negative campaigning literature. Negative campaigning, the tendency of political candidates and parties to attack their competitors' policies, competence, past actions and character traits, is widespread in many Western democracies (Nai, 2020). Since the 1990s, there is a lively debate on whether negative campaigning is an effective political strategy. Evidence on the implications of negative advertisements on turnout and

candidate support is mixed (Lau, Sigelman and Rovner, 2007), with some research show-
ing negative messages can increase turnout (Niven, 2006) (Barton, Castillo and Petrie,
2016), while others find that negative messages are not more effective than positive ones
at increasing turnout (Arceneaux and Nickerson, 2010) or donations (Barton, Castillo
and Petrie, 2016). Yet, other negative campaigning work has found that negative cam-
paigning can create backlash: going too negative is punished by voters. For example, in
a randomised canvassing field experiment in local elections in Italy, the authors find that
negative campaigning does neither affect actual vote shares of the attacker nor the tar-
get – but instead increases voting for the third neutral candidate who is not part of the
campaign message (Galasso, Nannicini and Nunnari, 2020). Using survey experiments,
the authors of this field experiment find that negative campaigning hurts the attacker's
electoral support, especially when the negative campaign message is delivered in an ag-
gressive, emotive style (Galasso, Nannicini and Nunnari, 2020). A different study uses
UK panel data to examine if negative campaigning creates backlash (Walter and van der
Eijk, 2019). Voters who perceive a political party's campaign as more negative rate that
party as worse in the next campaign period (Walter and van der Eijk, 2019). Negative
campaigning can also boost support for second-preference parties in multi-party systems.
If a voter's second-preferred party engages in more positive campaigning than the voter's
preferred party, voters are more likely to express support for their second-most preferred
party (Walter and van der Eijk, 2019). There is thus mixed evidence as to what the
implications of negative campaigning are. This thesis contributes to this literature, as
well as the literature on elite-mass linkages in polarisation and negative partisanship.
Paper 2 tests whether activating negative identities in a campaign message encourages
party supporters to make a donation using a large-scale field experiment. This adds to
the negative campaigning literature by studying the effects of a specific type of negative
message, focusing on identities, rather than issues, traits or competence. It also makes
an important contribution to the negative partisanship literature by testing if political
parties can use negative partisanship to their advantage by activating it in campaign
messages.

Finally, political parties may also try to defuse negative partisanship towards them. Es-
pecially those political parties at the receiving end of negative partisanship should have
an incentive to reduce these strong negative feelings towards them. This is particularly
relevant for the radical right – previous research shows that radical right-wing parties
often attract the highest dislike from supporters of other parties (Harteveld, 2021*a*).
At the same time, radical right-wing party supporters also strongly dislike mainstream
political parties (Harteveld, 2021*a*). The strong levels of dislike that radical right-wing
parties receive might be linked to the stigma attached to their nativist policy positions
(Blinder, Ford and Ivarsflaten, 2013; Harteveld, 2016, 2021*a*). Stigma is a qualitatively

different form of dislike – it is not just about weakly disliking a party, but believing it is off the pale, unacceptable and at odds with societal norms. Experimental and observational evidence shows that this social stigma hurts radical right-wing parties electorally (Harteveld, 2016; Harteveld, Dahlberg, Kokkonen and van der Brug, 2019*b*). Parties which are stigmatised will find it more difficult to attract voters, even those voters who agree with the party's policy positions (Harteveld, 2016; Spanje and Azrout, 2019). In turn, voters are more willing to support radical right-wing parties when they have a "reputational shield" that makes them appear more acceptable and mainstream, and more distant from racism and extremism (Harteveld, 2016). The ability of radical right-wing parties to present themselves as acceptable political parties is a key factor explaining electoral breakthroughs for the radical right (Mendes and Dennison, 2020). Populist radical right-wing parties thus have a strong incentive to defuse these strong negative feelings towards them and to present themselves as an acceptable political option. External factors, such as how the media portrays the radical right (Spanje and Azrout, 2019), how other political parties engage with it (van Heerden and van der Brug, 2017; Bale and Kaltwasser, 2021), and whether the radical right obtains institutional representation (Valentim, 2021) have been shown to affect the radical right's acceptability.

Rather than hoping for external shifts in reputation, radical right-wing parties can also try to pro-actively defuse strong negative feelings towards them. Research on "mainstreaming" strategies has shown that radical right-wing parties can de-toxify their overall reputation through organisational changes (Akkerman, de Lange and Rooduijn, 2016). However, we know less about how radical right-wing parties mainstream one of their biggest political trademarks: their issue positions. This is important because the social stigma attached to the radical right is directly related to their core anti-immigration positions (Harteveld, Dahlberg, Kokkonen and van der Brug, 2019*b*). To appeal to a wider electorate, radical right-wing parties need to make their anti-immigration positions feel more acceptable. This thesis contributes to closing this gap by studying how political parties try to defuse normative concerns over their most controversial issue positions. In Paper 3, I develop and test a novel theoretical argument about what I call "normative repackaging" as an issue mainstreaming strategy. I argue that when taboo issue positions are "normatively repackaged" as a defence of liberal democratic values like women's rights, voters will find the issue position more acceptable. I find evidence that this is the case. When an anti-immigration message is repackaged as a defence of women's rights, voters find the expression anti-immigration views more acceptable. However, normative repackaging did not change voters' general feelings towards the political party. In other words, normative repackaging did not increase support for the radical right. Feeling that disliked party positions are more acceptable may only be one step in a slow process of intensely disliked parties, like the radical right, becoming more

acceptable and likeable. Changes in partisan affect are likely to occur gradually, rather than in response to a one-off change in political rhetoric. This speaks to the stability of strong negative feelings towards political parties.

1.4 Methodological contributions: experimental evidence on negative partisanship

There are several methodological challenges involved in studying negative partisanship, including 1. the reliance on observational data, 2. disentangling positive and negative partisanship, 3. measurement issues and 4. ethical issues in experimental research. I will discuss each of these in turn, and how this thesis contributes to tackling these methodological issues.

1.4.1 From observational to experimental studies

The majority of studies on negative partisanship to date use observational data, such as opinion surveys, to assess the relationship between negative partisanship and a range of political behaviours and attitudes. These studies have been crucial in establishing an understanding of what negative partisanship means for voter behaviour and attitudes (Bankert, 2020; Ridge, 2020; Caruana, McGregor and Stephenson, 2015; Medeiros and Noël, 2014). The empirical approach used is typically a regression model, with variables such as vote choice, political participation or satisfaction with democracy as the dependent variable, and negative partisanship and positive partisanship as the independent variables (Bankert, 2020; Caruana, McGregor and Stephenson, 2015; Medeiros and Noël, 2014; Ridge, 2020; Haime and Cantú, 2022).

However, this approach has several limitations. Firstly, there is the issue of omitted variable bias: both negative partisanship and the outcome of interest, for example political participation, may be jointly explained by a third variable that is not included in the model. Secondly, the causal arrow between negative partisanship and the outcome could run in the reverse direction. For example, because someone is politically active and has volunteered for a campaign, they start feeling more emotionally invested in the victory of their in-party, and eventually develop a strong aversion towards their out-party. While it is important to describe behavioural patterns among different groups of patterns, it is difficult to establish what causal effects negative partisanship has on behaviour using observational data.

Estimating the effect of negative partisanship on an individual-level outcome using observational data is also difficult as the adoption of negative partisanship will only happen rarely within an individual's lifetime, and then be resistant to change if we think about negative partisanship as a stable, durable identity. Instead, we might want to know how holding negative partisanship affects an individual's reaction to a change in the world around them, or under what conditions negative partisanship becomes salient to the individual's decision-making.

In line with this, research on negative partisanship and affective polarisation has started using survey and lab experiments to test for the implications of activating or manipulating out-party hostility (for an overview, see Lelkes 2021). For example, Broockman, Kalla and Westwood 2022 use a modified version of a behavioural game to manipulate respondents' level of affective polarisation, and then examine the effect of this on a set of political and inter-personal outcomes. They find that affective polarisation, as manipulated in this experimental setting, does not affect political attitudes, but does affect inter-personal ones.

This thesis builds on these experimental studies by conducting three survey experiments and one field experiment to study negative partisanship.

In Paper 1, I experimentally manipulate the salience of negative partisanship through an innovative priming task. Through an open-ended survey question, individuals are asked to write what they dislike about the party they have negative partisanship towards. This makes negative partisanship momentarily more present in people's minds, and therefore allows me to test what effect the activation of negative partisanship has on political attitudes. In Paper 1, I also consider how negative partisanship moderates the experience of these negative feelings becoming activated by specifying an interaction model between partisan type and the priming treatment. Paper 1 thus expands our understanding of the effects of negative partisanship activation, and how partisan type moderates these effects.

Paper 2 uses a field experiment to test what happens when political parties activate negative and positive partisan identities. By manipulating what political identities are primed in a campaign email to party supporters, Paper 2 provides robust evidence on the effects of negative partisanship activation on costly mobilisation.

Paper 3 investigates how political parties might try to reduce negative feelings towards them. Consistent with the view of negative partisanship as a stable, durable concept, I find that party's rhetorical strategies are able to move voters' policy views, but not their overall support for that party.

Combing both survey and field experiments allows me to carefully identity what effects negative partisanship activation has on both attitudinal and behavioural outcomes. Survey and lab experiments are not plagued by the same endogeneity issues that are present in analyses using observational data and allow us to identify what causal effects negative partisanship activation has on attitudes.

While survey and lab experiments allow us to create a carefully controlled environment to test for these effects, they also have limitations. It is hard to establish with survey experiments if negative partisanship activation has effects on actual political behaviour, rather than stated attitudes. It is also difficult to ascertain how long-lasting the effects of negative identity activation are.

In contrast, field experiments combine the benefits of experimentation with a naturalistic, unobtrusive study setting. The field experiment presented in Paper 2 of this thesis is one of the first studies to present evidence on the effects of activating negative partisanship on costly behaviour. Rather than measuring the effect of negative partisanship activation on attitudes, we estimate the effect on actual donation behaviour – a credible and conservative measure of costly mobilisation. The negative partisanship prime was delivered as part of a real campaign email sent by a political party during a General Election campaign. This provides a highly realistic, externally valid way of priming negative partisanship, and allows us to understand what its effects are in a real-world campaign setting.

1.4.2 Disentangling negative and positive partisanship

Another methodological challenge in the study of negative partisanship is disentangling the effect of positive and negative partisanship. Current approaches often use observational data, inputting measures of both positive and negative partisanship into the same model as predictors. This approach has its own limitations, as discussed above. Within those constraints, this is a valid modelling approach if we believe that negative and positive partisanship are independent constructs. However, positive and negative partisanship may influence each other.

Indeed, as I have argued in the section introducing the four-fold partisanship typology, it is plausible that negative and positive partisanship do overlap and reinforce each other in some voters ("closed partisans") but not in all voters. Assuming that negative and positive partisanship are completely independent of one another is thus empirically problematic. If positive partisanship mediates negative partisanship (or vice-versa), the current models used in observational analyses may underestimate the total effect of negative partisanship activation (Lelkes, 2021). This thesis contributes to addressing

this methodological challenge by experimentally manipulating both the activation of positive and negative partisanship. In Paper 1 and Paper 2, there are separate experimental treatments that prime either positive or negative partisanship. This allows me to disentangle the effect of positive partisanship activation from the effect of negative partisanship activation.

1.4.3   Measuring negative partisanship

Measurement is a key methodological challenge to the study of negative partisanship. In cross-sectional, cross-national survey data, there are often few suitable question items that tap into negative partisanship. This leads to measurement issues, including 1. Using a negative partisanship measure that is too similar to the outcome of interest, and 2. Using a negative partisanship measure that does not capture the underlying theoretical concept.

Related to the first measurement issue, several studies measure negative partisanship using the question "Is there a party you would never vote for?" alongside party ratings, and then assess what relationship negative partisanship has with vote intention (Caruana, McGregor and Stephenson, 2015; Medeiros and Noël, 2014). As this measure of negative partisanship is so related to vote intention, it runs the risk of mechanically creating a positive association between negative partisanship and the outcome variable. It also increases the risk of finding a positive correlation because survey respondents are motivated to answer survey questions in a congruent fashion.

In light of these limitations, this thesis uses negative partisanship measures and treatments that are operationally distinct from the outcome measure. In Paper 1, partisan type is measured using thermometer scores, while negative partisanship is activated through a writing task and a series of priming questions. The outcome variable in Paper 1 is the distance between a respondent's self-placement on the left-right scale and the respondent's placement of a political party on the left-right scale. This is a subtle outcome measure as it should not be straightforward for respondents to e.g. deliberately falsify their left-right placements of political parties to act in congruence with the treatment or their pre-treatment partisan type. In Paper 2, the negative partisanship prime is delivered in a campaign email, while the outcome is whether someone made an actual donation to the political party. There is thus a clear distinction between the treatment and the outcome. In Paper 3, the treatment is a political text that varies in content between an anti-immigration text, a gender equality text or a combined gender equality and anti-immigration text. The outcome measures include how acceptable anti-immigration views are, and vote intention for the radical right. The outcome measure

of negative partisanship in Paper 3 is thus different from the content of the treatment text.

The second measurement issue in regard to negative partisanship relates to how well current measures of negative partisanship capture the underlying theoretical concept. At the heart of this measurement debate lies the question whether at a theoretical level, negative partisanship is mainly affect (a strong dislike of a political party) or an identity (negative identification with a political party). Recent contributions have defined negative partisanship as a social identity, and therefore critiqued approaches that rely on measures of party affect to measure negative partisanship (Mayer, 2017; Bankert, 2020; Lelkes, 2021). Consequently, several new question batteries that measure negative partisanship as an identity have been proposed (Bankert, 2020; Lelkes, 2021; Mayer, 2017).

I build on these studies by incorporating identity into my definition and operationalisation of negative partisanship. However, I also diverge from these studies by arguing that negative affect towards a political party is still an important feature of negative partisanship. As discussed in the theoretical framework section, I define negative partisanship as a concept as having both an affective and an identity dimension.

In line with this, the experimental treatments used in this thesis prime both negative partisan affect and identity. For example, in Paper 1, negative and positive partisanship are primed through a writing task and a series of priming questions. While the writing task primes the affective dimension of negative partisanship ("What do you dislike about this party?"), the subsequent two questions are priming the identity dimension of negative partisanship. Respondents are given the following two question items from Bankert's validated negative partisan identity scale: "I do not have much in common with supporters of [party name]" and "When I talk about the [party name] I say 'they' instead of 'we'". Including both the affective and identity components in the treatment ensures that both dimensions of negative partisanship are activated. Similarly, in Paper 2, we pay close attention to developing treatment primes that do not just convey an affect towards a party, but convey either a positive identity, or a negative identity of being against "them". By using language such as "together, we've got this" and "no matter how different we are, there is one thing we all have in common: we all [specific identity prime] " in all treatment conditions, we explicitly try to make the identity component of the treatment salient.

However, a limitation of this thesis is that the pre-treatment negative partisanship measures in Paper 1 are based purely on affect (party thermometer scores). Ideally, a measure using both identity questions such as Bankert's scale (Bankert, 2020), as well

as thermometer scores, would have been used. Unfortunately, this was difficult to implement as some of the negative partisanship identity scale questions were already being used as a treatment prime. Giving all respondents the negative and positive partisanship identity scales pre-treatment would have risked activating both negative and positive partisanship in all respondents, counteracting the goal of the treatment. Nonetheless, the experimental treatments presented in Paper 1 and Paper 2 of this thesis were carefully designed to take into account the affective and identity dimension of negative partisanship.

Overall, this thesis makes an important empirical contribution to the study of negative partisanship. By providing evidence from three survey experiments and one field experiment, this thesis provides robust causal evidence about the effects of negative partisanship activation, and the impact on parties' rhetorical strategies on negative feelings towards them. This builds on recent experimental research on negative partisanship and affective polarisation. By disentangling the effects of negative and positive partisanship activation, this thesis also contributes to our understanding of how negative partisanship functions relative to other well-known concepts, such as positive partisan attachments. The negative partisanship activation treatments presented in this thesis incorporate identity priming, and thereby apply a recent innovation in the conceptualisation and measurement of negative partisanship to the design of experimental stimuli.

1.4.4 Ethical questions in experimental research

Beyond these methodological contributions, the survey experiments and field experiment presented in this thesis also raise important ethical questions. All studies presented in this thesis tried to carefully engage with potential ethical issues, and all received Ethics approval from the LSE Ethics Committee. In the following, I briefly review the relevant ethical principles and then discuss how the studies presented in this thesis engage with them.

Research involving human subjects needs to follow ethical guidelines. This need was recognised in the 1979 Belmont report which set out foundational principles governing research involving human participants (Belmont report, 1979). Three key principles are described in the Belmont report: respect for persons, beneficence and justice. Respect for persons refers to the idea that individuals are autonomous agents, and that those individuals without autonomy deserve special protection. Arising from this are e.g., the need to seek informed consent from individuals, and protection for vulnerable individuals. Beneficence refers to two linked principles, the obligation to do no harm and to maximise possible benefits and minimise possible harms. Finally, justice refers

to scrutiny over how power dynamics affects who benefits and who may be harmed by the research. Withholding beneficial treatment from sub-groups of the population or carrying out harmful research on vulnerable groups would violate the justice principle.

While the Belmont principles provide an important cornerstone of ethics for research, researchers have pointed out that these principles were conceived to govern biomedical research and are therefore not always appropriate to cover ethical issues inherent in e.g., field experiments in politics (Phillips, 2021). More specific ethics guidelines on social science research have been formulated as part of the 2020 APSA Principles for Human Subjects Research (APSA, 2020). Ethical guidelines for field experiments have also been extensively discussed in scientific publications (Phillips, 2021) and formulated in scholarly communities, such as through the 2011 EGAP Principles (EGAP, 2011). The key 2020 APSA Principles on research involving human subjects are respecting participant autonomy, considering the wellbeing of participants and other people affected by their research; being aware of power differentials when conducting research; avoiding or minimising harm; keeping the identities of research participants confidential; and considering the broader social impacts of the research process (APSA, 2020).

The survey experiments presented in Paper 1 and Paper 3 of this thesis raise several ethical questions. The main potential ethical issue for the survey experiments presented in Paper 1 and Paper 3 is that participants encounter anti-immigration rhetoric, and that they are asked questions about sensitive political topics such as immigration attitudes. Reading about these issues might be uncomfortable for some participants. However, the risks to participants posed by these questions and treatments can still be considered minimal and not bigger than those encountered in everyday life, such as through news media consumption or social media use. Participants are informed that they can leave the study at any point in time without penalty, and that they are able to answer Prefer not to say or skip questions if they do not wish to engage with them. The survey experiments therefore aim to fulfil the principle of minimising harms.

In regard to the respect for persons principle and the respect for autonomy, informed consent was obtained before participants engaged in the survey experiments. Before choosing to participate, respondents received information about the study, who is conducting it, and how to contact the researcher if there are any questions. Participants were also informed that their participation is voluntary, that they can break off the survey at any point. Only individuals over the age of 18 were allowed to take part. There was no deception of participants – the treatment stimuli were all clearly labelled as hypothetical political texts.

In regard to keeping the identities of participants confidential, no identifiable data was collected during the survey experiments. In regard to justice, no vulnerable populations

were included in the survey experiments. There are no direct benefits to individual study participants as there might be in e.g., medical trials of new drug treatments. However, the survey experiments presented in this thesis were designed with clear societal benefits in mind: In Paper 1, understanding how negative partisanship activation shapes party evaluations is important during a time where strong negative feelings towards political parties are widespread and positive partisanship is in decline. Paper 3 tests if normative repackaging makes previously unacceptable anti-immigration views acceptable. This provides the basis for understanding if other parties and actors should worry about normative repackaging, and how they could work to counteract these shifts.

The field experiment presented in Paper 2 also raises several ethical questions. The field experiment was conducted in collaboration with a British political party and consisted of varying the content of a campaign email that was sent out to the party's email list during the general election campaign in 2019.

A key ethical concern associated with this field experiment is not asking individual email list subscribers for informed consent. Participants are individuals over the age of 18 who have consented to receiving emails from the political party but did not explicitly consent to take part in an email experiment. The APSA 2020 guidelines on Human Subjects Research state that research without informed consent is appropriate when the harms to subjects are minimal, and when there are concerns over the presence of Hawthorne effects, the idea that individuals may change their behaviour if they know they are being studied (APSA, 2020). I argue that both conditions are fulfilled in this case.

Firstly, the intervention involves only minimal harms. The political party regularly sends out email communications to their email list subscribers and would have done so even in the absence of the experimental assignment to treatment and control groups. The difference in treatments only consists of sending subjects different variations of the same email text, with the political party retaining the final say over what email content was sent out. Individuals voluntarily signed up to the political party's email list and were able to unsubscribe at any point. The harms posed by this email message were thus minimal and not different from what participants were exposed to already. We also fully briefed the political party about all aspects of the experiment and signed a collaboration agreement with them.

Secondly, not informing the participants beforehand that they were part of an email experiment was a fundamental part of the research design. This allowed us to estimate treatment effects that are not biased by the desire to conform to the researcher's expectations, but to observe behaviour in an unobtrusive, naturalistic study setting. Letting participants know that they will part of a fundraising email experiment would have likely

altered their behavioural response to subsequent emails. There is little research on negative partisanship that examines the effects of negative partisanship primes on actual behaviour, and this design aspect thus makes an important contribution to this literature. The harms arising for individuals who participate in the experiment are minimal and do not outweigh the benefits of observing individual behaviour in an unobtrusive way.

Regarding confidentiality in the field experiment, we only have access to de-identified individual-level information. Only the political party has access to personal information, like names or email addresses, while the data the research team received only contained ID numbers.

Another key ethical concern related to field experiments is related to the broader societal impact of these experiments. Researchers should refrain from trying to influence political processes without the consent of the affected individuals (APSA, 2020). However, collaboration with third parties can constitute an important exception to this – the intuition behind this is that third parties (e.g., political parties, government offices or electoral commissions) legitimately seek to affect political processes (APSA, 2020). The field experiment carried out in Paper 2 was conducted in cooperation with a political party, and therefore does not raise the same ethical concerns as a field experiment in this domain without third party collaboration would do. Learning from an intervention that is already being carried out by a third party is different from a researcher independently planning and implementing an intervention that affects a political outcome without the consent of affected individuals (APSA, 2020).

Still, foundational ethical principles of course also apply to collaborations with third parties. In regard to the field experiment presented in Paper 2, the harms to individuals are likely minimal and there are important societal benefits. Learning about the impact of identity priming is important because it has direct implications for our understanding of how powerful negative identity activation is but also provides political parties with information about the trade-offs involved in negative identity campaigning.

Finally, an important ethical principle guiding third party collaboration in field experiments is transparency with the partner organisation about the risks and benefits involved in the research project. The research project was openly discussed with the political party throughout the research process and a collaboration agreement was signed to formalise anticipated benefits, risks and responsibilities before the project implementation.

Survey and field experiments on hot political issues like immigration and negative partisanship can raise important ethical questions. The experiments presented in this thesis

seek to address ethical issues by considering the potential harms to individuals, setting out the societal benefits of the research, seeking informed consent where possible, collecting only de-identified data to preserve participant confidentiality, and engaging in best practices of collaboration with third parties.

1.5   Roadmap

In the remainder of this thesis, three self-contained research papers are presented, followed by an overarching conclusion. Chapter 2 develops a theoretical argument about how negative partisanship affects how voters feel towards multiple political parties, not just the one that they dislike. In line with the idea of "my enemy's enemy is my friend", evidence from a survey experiment shows that negative partisanship can inspire feelings of closeness to other political parties. Chapter 3 presents the results of a field experiment conducted in collaboration with a political party. This paper shows that cueing negative political identities has stronger mobilising effects than cueing positive political identities. Chapter 4 examines if and how disliked political parties can reduce negative feelings towards them and their core issue positions. I find that normative repackaging, a strategy which couples an unacceptable policy position with a highly acceptable one, is an effective way for radical right-wing parties to make their core policy positions feel more acceptable to voters. Chapter 5 provides a summary of these findings, discusses their implications for policy as well as larger questions of democratic functioning, and suggests avenues for further research.

Chapter 2

My enemy's enemy is my friend: the implications of negative partisanship

Abstract

Strong negative feelings towards political parties are common in Western democracies but their implications for political attitudes beyond their disliked party are still poorly understood. This paper develops a theoretical argument, positing that negative partisanship should affect how voters feel about multiple political parties, not just the one that they dislike. To test this theory, respondents in a survey experiment were assigned to either a positive partisan prime, a negative partisan prime or no prime. In line with the idea of "my enemy's enemy is my friend", I find that activating feelings of dislike towards a political party can move negative partisans closer to another party. These findings have important implications for our understanding of negative partisanship and its role in shaping political preferences.

2.1 Introduction

Strong negative feelings towards political parties and their supporters are on the rise in the US (Iyengar et al., 2019). These strong feelings of dislike, disgust or even hatred can have important implications for behaviour inside and outside the political realm, such as party loyalty (Abramowitz and Webster, 2016), political participation (Iyengar and Krupenkin, 2018), dating behaviour and job market discrimination (Gift and Gift, 2015). Strong negative feelings and outright hostility towards political parties and their supporters are also prevalent in multi-party systems (Huddy, Bankert and Davies, 2018; Wagner, 2021; Reiljan, 2020; Gidron, Adams and Horne, 2020). Negative partisanship, or "affective repulsion" towards a political party (Caruana, McGregor and Stephenson, 2015, p.772) has been shown to exert important effects on political behaviour (Medeiros and Noël, 2014). For example, negative partisans are more likely to turn out to vote (Caruana, McGregor and Stephenson, 2015). Similarly, voters who have a bigger gap in their affect towards their in-party and their out-party are also more likely to engage in campaigning for a candidate or party, contacting an elected official, or taking part in a protest (Wagner, 2021). Negative partisans thus constitute a particularly important and interesting group of voters to study. Yet, we still know little about how this politically active and potentially influential group of voters forms political preferences.

This paper addresses these gaps by developing and testing an argument about how negative partisanship affects political preferences in multi-party systems. I argue that negative partisanship does not just affect how individuals feel towards their disliked political party, but also towards other political parties. Combining ideas from psychology and political science, I apply some of the insights behind balance theory to political behaviour.

Balance theory, developed in the 1950s by Fritz Heider, takes intuition from how humans make sense of the world around them and applies them to interpersonal relationships (Heider, 1958). In essence, individuals strive for a balance between their sets of likes and dislikes in relationships with other people (Heider, 1958). In line with the idea of "my enemy's enemy is my friend", individuals should feel closer to someone if they both dislike a third person (Bosson et al., 2006). Applying these insights to politics, I argue that strongly disliking a specific political party can also make voters feel closer to another one. Negative partisans should develop more positive feelings towards their "enemy's enemy", their disliked party's competitors. Political parties which seemed more ideologically distant in the past should start feeling subjectively closer to negative partisans as a result of their strong distaste for a specific political party.

I test this theoretical argument using a pre-registered survey experiment conducted in Canada. Canada can be considered an average case in terms of out-party dislike compared to other Western democracies (Gidron, Adams and Horne, 2019), and therefore offers an excellent representative case study for examining the implications of negative partisanship. Participants (n=1,553) were randomly assigned to either write a short text about why they dislike a certain party (Negative partisanship prime), a short text about why they like a certain party (Positive partisanship prime), or no text (Control). I then measure how close participants feel to their most liked party on the left-right spectrum. I find evidence in favour of the idea of "my enemy's enemy is my friend". Negative partisans assigned to the negative partisanship prime (compared to the Control and the positive partisanship prime) place their most liked party closer to themselves.

I make three main contributions to the literature on partisanship, political behaviour and affective polarisation. Firstly, while there is a rich literature on the implications of out-party dislike on non-political outcomes, we still know comparatively little about how negative partisanship affects political preferences and behaviour (Dias and Lelkes, 2021). This paper adds to the growing literature on negative partisanship and affective polarisation by examining what implications negative partisanship has for political preferences. Among the studies on political outcomes, the focus has been on understanding the implications of negative partisanship in individuals who also have a positive party attachment (Bankert, 2020; Medeiros and Noël, 2014; Iyengar et al., 2019). Yet, it is possible that individuals just strongly dislike a political party without feeling close to another one. Building on recent contributions (Samuels and Zucco, 2018; Ridge, 2020; Haime and Cantú, 2022), I adopt a nuanced partisanship typology that differentiates between "closed partisans", individuals who hold both positive and negative partisanship, and "negative partisans", individuals who only hold negative partisanship. While there is some evidence about the implications of negative partisanship in closed partisans, there is little research about how negative partisans form political preferences. This paper addresses this gap and shows that negative partisanship activation affects political preferences in negative partisans. This also provides support for the more general argument that negative partisanship is indeed a force of its own and is not dependent on positive partisanship to matter for political preferences.

Secondly, I make a contribution to the partisanship literature by examining what partisanship does to voter evaluations of other political parties, not just of the party that their partisan affect is directed towards. While it is intuitive that negative partisanship makes voters less likely to vote for their disliked party (Caruana, McGregor and Stephenson, 2015; Medeiros and Noël, 2014), we do not know much about how negative partisanship should affect voter preferences towards other parties. Drawing on insights

from psychology, this paper argues that disliking a party does not just affect voter perceptions of the disliked party, it can also change how someone feels about the other political parties on offer.

Finally, this paper also contributes to the growing literature on affective polarisation and partisan hostility. Strong negative feelings towards partisan out-groups are prevalent in many Western societies (Wagner, 2021; Reiljan, 2020; Gidron, Adams and Horne, 2020), with troublesome consequences for democracies, ranging from lower satisfaction with democracy (Wagner, 2021), reduced support for democratic norms (Kingzette et al., 2021), to lack of cooperation with public health measures (Druckman et al., 2020*a*) and even willingness to engage in political violence (Mason and Kalmoe, 2022). Given these harmful implications of negative partisanship, there is an urgent question over how these strong negative feelings towards political parties can be channelled into something more constructive. This paper offers a cautious note of optimism by showing that the activation of negative partisanship can also increase feelings of closeness to the most liked party among negative partisans. These subjective boosts in closeness may potentially translate into more structured positive party attachments in the long run. Of course, as the affective polarisation powerfully demonstrates, positive party attachments by themselves are not a guarantor of decreased out-group hostility. Still, positive identities can exist without a dislike for other groups (Brewer, 1999), while the same cannot be said about negative identities. Moving from only negative partisanship to holding more positive party attachments could be seen as a relative improvement, and signals that reverse paths from negative towards positive identities are possible.

2.2 Loathing without loving: negative and positive partisanship in multi-party systems

Party identification is a key structuring factor of politics. Traditionally understood as a durable attachment to a political party (Campbell et al., 1960; Green, Palmquist and Schickler, 2002), partisanship can be considered a form of social identity that individuals adopt to satisfy higher psychological needs like belonging and creating a sense of self-worth (Tajfel and Turner, 1979; Mason, 2018).

While the majority of work on partisanship has traditionally focused on positive party attachments, voters often harbour strong negative feelings towards political parties. Recent research has started examining this "forgotten side" of partisanship and its implications for behaviour. Negative partisanship can be broadly understood as "affective repulsion", or a strong and durable negative affect towards a political party (Caruana,

McGregor and Stephenson, 2015, p.772). Negative partisanship is widespread among voters in Western democracies (Anderson, McGregor and Stephenson, 2021) and can be considered a stable identity: while partisans might switch between being an independent or a partisan, or even between different political parties, there is usually a specific party that voters would never consider voting for. Using panel data, research has shown that the party that individuals would never vote for is even more stable than positive partisanship (Neundorf, Stegmueller and Scotto, 2011). Negative partisanship is predictive of a range of political behaviours, such as vote choice (Samuels and Zucco, 2018; Dinas and Northmore-Ball, 2019; Bankert, 2020) and political participation (Meléndez and Kaltwasser, 2017).

Much of the current research on negative partisanship and affective polarisation focused on the US. Building on the growing literature on negative partisanship outside this context, this paper extends and applies negative partisanship theory in multi-party systems. This is important from a practical empirical perspective but also crucial for enhancing our theoretical understanding of negative partisanship. In two party systems, negative and positive partisanship are often treated as two sides of the same coin. In line with social identity theory, individuals develop a positive in-group identity first, and may then develop negative feelings towards an out-group as a consequence (Tajfel and Turner, 1979; Tajfel, 1974). Applied to partisanship, partisan out-group hostility has therefore often been understood as a consequence of positive party attachments. In a bid to uphold a positive self-image, partisans engage in out-group derogation (Huddy, Bankert and Davies, 2018). For example, the affective polarisation literature in the US has argued that heightened out-party dislike is a "natural offshoot of this sense of partisan group identity" (Iyengar et al., 2019, p. 130). Negative and positive partisanship are thus thought to be mutually reinforcing in two-party systems. For example, studies from the US have shown that strong negative feelings towards the out-party imply greater in-party loyalty and lower split-ticket voting (Abramowitz and Webster, 2016).

Yet, in multi-party systems, the relationship between positive and negative partisanship is less clear-cut. It is not obvious which political party voters might feel negative towards as a result of their positive partisan attachments. More importantly, voters may simply strongly dislike a political party without feeling attached to another one. Theorising about how partisanship plays out in multi-party systems therefore also allows us to challenge key assumptions about positive and negative partisanship. This paper builds on recent innovations in the partisanship literature which argue that "strong out-party hostility can develop without equally strong in-party attachments" (Bankert, 2020, p.1468). I extend this line of thought by adopting a nuanced partisanship typology that accounts for all possible combinations between positive and negative partisanship.

Table 2.1 presents an overview of the typology, which sorts voters into four different partisan categories, based on Rose and Mishler's framework (Rose and Mishler, 1998). It also builds on recent contributions to the negative partisanship literature that work with a four-fold partisan typology based on Rose and Mishler's framework (Samuels and Zucco, 2018; Ridge, 2020; Haime and Cantú, 2022). Positive partisans are individuals with a positive identification with one party, and no negative identification with other parties. Negative partisans hold negative partisanship, but do not positively identify with a party. Closed partisans both report a positive identification with a party as well as a negative identification with another party. Non-partisans hold neither a positive nor a negative party identification.

	Positive partisanship	No positive partisanship
Negative partisanship	Closed partisan	Negative partisan
No negative partisanship	Positive partisan	Non-partisan

TABLE 2.1: Partisanship typology

Much of the negative partisanship and affective polarisation literature studies the consequences of negative partisanship in what would be considered "closed partisans" in this typology, those individuals who have both positive and negative partisanship (see e.g., Iyengar et al. 2019, Bankert 2020). This is partially related to the geographic focus of previous research on the US, where many partisans in the sample are effectively closed partisans, individuals who are e.g., simultaneously Republican and anti-Democrats. Yet, especially in other contexts like multi-party settings, it is plausible that voters strongly dislike a party without feeling close to another one. The four-fold typology has the advantage of defining one exclusive category of voters which fit this description: negative partisans. This typology thus offers a nuanced framework for studying negative partisanship.

2.3 My enemy's enemy is my friend: the implications of negative partisanship

Partisanship matters because it powerfully shapes voter behaviour and attitudes. Positive partisanship is still considered one of the most reliable predictors of vote intention (Huddy and Bankert, 2017), and negative partisanship is strongly associated with vote

intention and party preferences (Samuels and Zucco, 2018; Dinas and Northmore-Ball, 2019; Bankert, 2020; Meléndez and Kaltwasser, 2017). One straightforward expectation about negative partisanship is that it should deter individuals from voting for their disliked party. Indeed, research has shown holding negative partisanship towards a party is negatively correlated with supporting that party at the polls (Caruana, McGregor and Stephenson, 2015; Samuels and Zucco, 2018). However, negative partisanship does not just influence which party voters do *not* support, but also which party they intend to vote for. Among those individuals who hold both positive partisanship and negative partisanship (the "closed partisans" in my typology, Table 2.1), there is evidence that negative partisanship makes voters more likely to support their in-party at the polls (Bankert, 2020; Medeiros and Noël, 2014; Caruana, McGregor and Stephenson, 2015; Samuels and Zucco, 2018).

Partisan identity also has important implications for political attitudes and behaviours beyond vote intention, shaping how voters perceive the world around them, and how they form preferences. Research has shown that voters adopt policy positions congruent with their party identification or already preferred candidate (Granberg and Brent, 1980; Lenz, 2009, 2012; Achen and Bartels, 2016). Similarly, partisan hostility can shape what opinions citizens adopt on new policy issues (Druckman et al., 2020*b*).

Vice versa, partisanship can also impact how voters perceive political party positions. A wealth of research speaks to the existence of projection bias: voters consistently place political parties closer to themselves on issue or ideological scales when they like the political party, and further away from themselves when they dislike the political party (Markus and Converse, 1979; Merrill, Grofman and Adams, 2001; MacDonald, Rabinowitz and Listhaug, 2007; Dinas, Hartman and van Spanje, 2016). In other words, if voters feel close to a party, they will want to believe that this party is closer to them ideologically than it may objectively be. When voters dislike a party, they are motivated to place this party at a greater distance to themselves. Research has shown that projection bias in party placements is widespread across different Western democracies (Merrill, Grofman and Adams, 2001; Grand and Tiemann, 2013). While much of this research is observational, evidence from a natural experiment in the Netherlands shows that changes in affect towards political parties impact voters' perceived ideological proximity to political parties (Dinas, Hartman and van Spanje, 2016). When there is an exogenous shock that makes voters feel warmer towards a political party, voters place the party in question closer to themselves on an ideological scale (Dinas, Hartman and van Spanje, 2016).

This idea of projection bias is grounded in psychological theories on assimilation and contrast effects (Sherif and Hovland, 1961), as well as balance theory (Heider, 1958)

and cognitive dissonance theory (Festinger, 1957). Voters are motivated to avoid the discomfort arising from the dissonance of liking a party and feeling that it is ideologically distant from oneself (Grand and Tiemann, 2013; Dinas, Hartman and van Spanje, 2016). The motivation to engage in projection should also be greater when the issue dimension, or the relationship to the party is more salient to voters (Grand and Tiemann, 2013). While political identities are stable, their relative salience can fluctuate. External events, election campaigns, party communications or media exposure may activate partisanship in individuals and raise the salience of partisan identities in preference formation and voter decision-making. For example, political identities can become salient to individuals during periods of heightened political conflict along identity lines, such as referendums (Hobolt, Leeper and Tilley, 2020) or elections. Other recent research has shown that affective polarisation, of which out-party dislike is an important component, is greater during election campaigns than at other times in the electoral cycle (Hernández, Anduiza and Rico, 2021).

Building on this, I expect that partisanship activation influences voters' perceived distance to political parties:[1]:

H1: The activation of positive partisanship decreases voters' perceived distance to their most liked party.

H2: The activation of negative partisanship increases voters' perceived distance to their most disliked party.

However, we know little about how negative partisanship should influence party evaluations beyond the disliked party. While there is evidence that negative partisanship in closed partisans (those who also hold positive partisanship) reinforces in-party preferences (Bankert, 2020; Medeiros and Noël, 2014; Caruana, McGregor and Stephenson, 2015; Samuels and Zucco, 2018), the implications of negative partisanship in individuals without positive party identification are poorly understood. To address this gap, this paper builds a novel argument about how negative partisanship should influence party evaluations, combining insights from psychology and political science.

Balance theory, developed in the 1950s by Fritz Heider, took inspiration from how humans process objects in the world around them and distinguish between them. It then

---

[1]The hypotheses were worded slightly differently in the pre-registration. The pre-registered hypotheses are: "H1: Individuals exposed to a positive partisanship prime will place their most liked party closer to themselves than individuals in the negative partisanship condition and individuals in the Control condition." and "H2: Individuals exposed to a negative partisanship prime will place their most disliked party further away from themselves than individuals in the positive partisanship condition and individuals in the Control condition."

applied these insights to interpersonal relationships (Heider, 1958). At the core of balance theory is the idea that individuals strive for an affective balance in their relationships with other people (Heider, 1958). For example, in a relationship between Anna and Bob, there would be balance if both of them proclaimed to each other that they are friends. However, there would be imbalance if Anna suddenly declared that she does not like Bob anymore. In this scenario, humans are motivated to take the least costly path to restore balance with their most valued outcome in mind. If Bob cares about the friendship with Anna, he will invest resources in trying to restore Anna's positive affect towards him. If this path is too costly or Bob does not value the friendship enough to incur these costs, Bob might change his mind and convince himself that he never liked Anna that much anyway.

In a triadic relationship between Alice, Ben, and Conor, things are a little more complicated: there would of course be balance if everyone liked each other or if everyone disliked each other. However, from the perspective of Alice, there would also be balance if Alice and Ben held the same attitudes towards the third person (both Alice and Ben disliked Conor), as well as towards each other (Alice and Ben liked each other) (Bosson et al., 2006). If there was no clear affective relationship between Alice and Ben yet, Alice should start feeling sympathetic towards Ben when she learns that they both dislike Conor. This is in line with the idea of "my enemy's enemy is my friend" - it is easy to bond with someone over a shared dislike of something or someone else (Bosson et al., 2006).

This idea should also be intuitive when thinking about areas outside of politics. For example, marketing research shows that strangers bond through online groups of "anti-brand communities" which unite individuals around a shared distaste of specific global brands like McDonalds, Uber, Starbucks and Nestle (Brandão and Popoli, 2022). The idea of liking someone because you have a common enemy is also prevalent in sports rivalries. While football fans often hold a positive feeling of belonging to their team, they also positively connect to fans from other teams over their shared dislike of a rival team (Popp, Germelmann and Jung, 2016).

While balance theory was conceived to describe inter-personal relationships, it has also been applied to group identities by more recent research. Knowing that another group dislikes the same out-group can make this group seem more likeable and favourable, potentially even leading to the development of common in-group feelings (Dovidio, Validzic and Gaertner, 1998). Other research has argued that negational identities, defining oneself as "who we are not", can help to create common preferences between different groups (Zhong et al., 2008).

Applying the insights from balance theory back to politics, voters should start feeling more positive towards other political parties which are the competitors of the voter's most disliked party. Let us imagine that there is a triad between the disliked party A, rival party B, and a voter C. There are already two negative relationships - party A and B are rivals, and voter C dislikes party A strongly. If voter C does not hold any type of affect to party B yet, voter C should be motivated to develop positive affect towards party B. In other words, to balance out this triad, voter C should start feeling closer to party B. This means that negative partisanship should motivate voters to develop more positive feelings towards other parties. Parties which seemed more ideologically distant in the past should start feeling subjectively closer to negative partisans because they share a "common enemy". In line with the idea of "my enemy's enemy is my friend", I argue that voters should express feeling subjectively closer to other political parties when negative partisanship is activated. I expect that:

H3: The activation of negative partisanship decreases voters' perceived distance to other political parties.[2]

Finally, I argue that these effects should be stronger for individuals with congruent party identities. There is already an acknowledgement in previous work on projection bias that individuals with stronger positive partisan attachments or positive preferences for specific candidates should display greater projection bias. For example, in their interpretation of Heider's balance theory, Grand and Tiemann argue that "the individual need for establishing cognitive consistency first depends on the existence of 'unit relations' with the respective candidate or party, in other words, the stronger the attachment to a specific candidate or party, the higher the individual pressure to eradicate cognitive dissonance" (Grand and Tiemann, 2013, p.499). Heider's original formulation of balance theory was concerned with positive unit relations and in line with this, Grand and Tiemann apply these to positive party attachments and candidate preferences. However, as introduced earlier, negative partisanship can also be understood as a relevant social identity for individuals. I therefore extend the intuition behind balance theory to individuals with negative partisan identities. I expect that the effects stated in H1- H3 are stronger for individuals with congruent partisan identities:

H4A: The activation of positive partisanship has stronger effects on perceived party distance for voters who are positive partisans.

---

[2]The wording of the hypothesis in the Pre-Analysis Plan was: H1C: "Individuals exposed to the negative partisanship prime will place their second-liked party, third-liked party and all other parties (apart from their most liked and the most disliked party) closer to themselves than individuals in the positive partisanship condition and individuals in the Control condition."

H4B: The activation of negative partisanship has stronger effects on perceived party distance for voters who are negative partisans.[3]

2.4 Methods

2.4.1 Data and context

To test these hypotheses, I conducted a pre-registered[4] survey experiment with a representative sample of Canadian voters (N=1,553) from May 5th to May 12th, 2020. The data was collected by the Canadian Consortium for Electoral Democracy (C-Dem) as part of the Democracy Check-Up, an annual survey on political attitudes in Canada.

Canada is an excellent case for studying the implications of negative partisanship in multi-party systems. Firstly, Canada is a multi-party system with five main parties. Currently, the Federal Conservative Party, the Federal Liberal Party, the Federal NDP, the Bloc Québécois and the Federal Green Party are represented in parliament. The federal government is led by the Liberal Party and headed by Justin Trudeau since 2015. In 2019, the government was re-elected, albeit with losses in its vote share and began a second term as a minority government. Secondly, previous research on the Canadian context has established that negative partisanship exists in the electorate and is a relevant predictor of political behaviour (Caruana, McGregor and Stephenson, 2015). Anecdotal evidence suggests that negative partisan cues are also a common feature of political campaigns in Canada. For example, the Anything But Conservative campaign in 2011 tried to mobilise voters solely with the aim of preventing Conservative Party victories in the election. The Canadian case thus provides an externally valid testing ground for negative partisanship theory. Thirdly, comparative research has shown that Canada can be considered an "onlier" when it comes to affective polarisation and out-party dislike (Gidron, Adams and Horne, 2019; Boxell, Gentzkow and Shapiro, 2022). Affective polarisation in Canada has been rising over the last four decades, but to a moderate degree in comparison with other OECD countries (Boxell, Gentzkow and Shapiro, 2022). Canadian voters display average levels of affective polarisation in comparison to voters in other Western democracies (Gidron, Adams and Horne, 2019). This makes Canada a representative case for studying the effects of negative partisanship in multi-party systems.

---

[3]This corresponds to the original H5 in the Pre-Analysis Plan. The original hypothesis only contained expectations for positive partisans and was extended to mirror the same theoretical expectations for negative partisans.

[4]The survey experiment received Ethics approval from the LSE Ethics Committee (REC ref. 1110) and was pre-registered at the EGAP registry prior to outcome data collection (`https://osf.io/ywfvq`).

2.4.2 Design

I conduct a priming survey experiment to test how the activation of either negative partisanship or positive partisanship affects voters. Because partisanship is closely related to other political preferences and behaviour, using observational data to test for the effects of partisanship is challenging. Both partisanship and the subjective distance to political parties might be explained by other unobserved factors. Individuals might also change their subjective proximity to political parties first, and then start developing partisan affiliations. To overcome these identification challenges, I conduct a survey experiment. Participants are randomly assigned to either a treatment stimulus activating negative partisanship (T1), a treatment stimulus activating positive partisanship (T2), or no treatment stimulus (Control). An overview of the experimental groups can be found in Table 2.2. The experimental groups are balanced along key covariates, including age, gender and province, see Table A.1 in the Appendix.

Condition	Content	N
Control	No prime	540
Treatment 1	Negative partisanship prime	508
Treatment 2	Positive partisanship prime	505
Overall		1553

TABLE 2.2: Overview of the experimental groups

The goal of the partisanship treatments is to activate pre-existing feelings of partisan affect, rather than externally inducing partisanship. As discussed earlier, partisanship is widely considered a thick social identity which is not easily induced or altered. Activating pre-existing feelings of partisan affect is therefore a more realistic and theoretically congruent treatment intervention.

The treatment stimulus (see Figure 2.1) consists of a short writing prompt and two partisan identity questions. The goal of the writing task is to activate partisan affect. In the positive partisanship prime condition (T2), respondents are prompted to write about what they like about their most liked party. In the negative partisanship prime condition (T1), respondents are asked to write about what they dislike about their most disliked party. The questions between T1 and T2 are identical apart from the words "like" and "dislike". After the writing task, respondents are asked two more partisanship questions to make their partisan identity salient to them.[5] Respondents

---

[5]In the negative partisanship prime, respondents are given the following two question items from Bankert's validated negative partisan identity scale (Bankert, 2020): "You said that you dislike the [party name]. Thinking about this party, how often do you feel this way? I do not have much in common with supporters of [party name] and When I talk about the [party name] I say 'they' instead of

engaged well with the treatment stimuli. Of those exposed to the writing prompt, the majority (71%) wrote a response. Most respondents wrote a short sentence, and the content of the responses is aligned with the intended direction of the treatment prime. Respondents in the negative partisanship condition used more negatively valenced words while respondents in the positive partisanship condition used more positively valenced words (see Figure A.1 in the Appendix). More information on the treatment uptake can be found in the Appendix, Section A.0.3.

Negative partisanship prime	Positive partisanship prime
Is there any party that you dislike more than all the others?	Is there any party that you like more than all the others?
Which party do you dislike more than all the others?	Which party do you like more than all the others?
What do you dislike about this party? [textbox opens]	**What do you like about this party? [textbox opens]**

FIGURE 2.1: Overview of the treatment stimulus in the negative and positive partisanship condition (writing prompt in bold)

Open-ended questions are a common method used to prime people about pre-existing identities (Charness and Chen, 2020). Writing as an active form of cognitive engagement is preferable to reading a text passively, as this allows respondents to connect more expressively with the target of the prime. Short writing prompts have been used successfully in the past to prime partisan affect (Klar, 2014; Lavine, Johnston and Steenbergen., 2012). The writing task also offers an externally valid way of priming partisan affect: individuals in the real world often express how they feel about a political party in a few words, e.g., in online conversations and on social media. At the same time, the writing prompt looks like a standard survey question, which should lower the risk of experimenter demand effects and make it more difficult for participants to discern that they are receiving a treatment stimulus. The writing prompt can also be considered a

---

'we' [Always (1), Often (2), Sometimes (3), Rarely (4), Never (5)] In the positive partisanship condition, respondents were asked: "You said that you like the [party name]. Thinking about this party, how often do you feel this way? When I talk about the [party name], I say 'we' instead of 'them'. I have a lot in common with other supporters of the [party name]." [Always (1), Often (2), Sometimes (3), Rarely (4), Never (5)] The positive partisanship questions are taken from the partisan identity scale proposed by Bankert, Huddy and Rosema (2017).

conservative treatment manipulation as it only asks respondents for a short response, compared to other more immersive writing tasks.

2.4.3 Outcomes

The main outcome variable is the perceived distance between a voter and their most liked party.[6] To determine which party is the respondent's most liked party, I use pre-treatment party ratings. Prior to treatment, respondents are asked "How do you feel about the federal political parties below? Set the slider to a number from 0 to 100, where 0 means you REALLY DISLIKE the party and 100 means you REALLY LIKE the party". The party the respondent gives the highest or "warmest" pre-treatment party rating is defined as this respondent's most liked party.[7] For example, if a respondent gave the Conservative Party a score of 35, the Liberal Party a score of 50, and the Green Party a score of 65, the Green Party would be considered this respondent's most liked party. After the treatment, respondents are asked to first place themselves on the left-right scale, and then to place all political parties on the same left-right scale (0-10 scale). The treatment stimuli did not affect respondents' left-right self-placement in the whole sample nor negative partisans' self-placement on the left-right scale (see Appendix, Section A.0.11). Distance to the most-liked party is defined as the absolute distance between the respondent's left-right self-placement and the respondents' placement of the most-liked political party. Similarly, distance to the disliked party is operationalised as left-right distance to the respondent's most disliked party, the party with the lowest pre-treatment party rating. These distance measures reflect the amount of ideological space a respondent puts between themselves and a political party and are thus a good indicator for perceived proximity to a political party.

2.4.4 Pre-treatment partisanship

Respondents are categorised into one of the four different partisan types (non-partisan, positive partisan, closed partisan or negative partisan, see Table 2.1) based on their pre-treatment party like-dislike ratings. Table 2.3 provides an overview of the distribution

---

[6]I focus on voters' perceived distance to their most liked party in the main results. More information on distance to the most disliked party and other parties can be found in the Appendix, Section A.0.5 and Section A.0.8. The Pre-Analysis Plan also included a hypothesis on political participation outcomes. More information about these can be found in the Appendix, Section A.0.9.

[7]Similarly, the party with the lowest pre-treatment thermometer rating is the most disliked party. This is different from the pre-registered measure which used post-treatment left-right distance, instead of the pre-treatment thermometer scores, to define which is the most liked and disliked party. The measure presented in the main text and results is more appropriate because it is not mechanically related to the main outcome variable. More information and robustness checks can be found in the Appendix, Section A.0.7.

of partisan types based on this classification. Negative partisans make up around 13% of respondents in the sample (Table 2.3). Treatment groups were slightly imbalanced in terms of pre-treatment partisan type. All results are therefore also presented including pre-treatment partisan type as a covariate. Including this does not alter the main results.

Variable	Control Proportion	Negative prime Proportion	Positive prime Proportion	Difference Chi-square test
Negative partisans	13.9%	14%	9.8%	p= 0.058
Positive partisans	23.6%	29.6%	32.5%	p=0.112
Closed partisans	35.1%	32%	32.1%	p= 0.246
Non-partisans	19.6%	15.6%	17.9%	p= 0.152
N	540	508	505	

TABLE 2.3: Pre-treatment proportions of the four different partisanship types.

Respondents were asked "How do you feel about the federal political parties below? Set the slider to a number from 0 to 100, where 0 means you REALLY DISLIKE the party and 100 means you REALLY LIKE the party" (mean=45.2, sd=29.1, min= 0, max=100). I define a party rating above the 80th percentile of party like-dislike scores (>73) as having extreme positive affect towards a party, while a rating below the 20th percentile (<14) is coded as having extreme negative affect towards a party (see Figure A.3 in the Appendix).

Negative partisans hold extreme negative affect towards one party (rate one party below the 20th percentile) but do not hold extreme positive affect towards another party (rate no party above the 80th percentile). Positive partisans are defined in the opposite way: they hold extreme positive affect (rate one party above the 80th percentile) but do not display extreme negative affect towards a party (rate no party below the 20th percentile). Closed partisans hold both extreme positive and extreme negative affect - they rate one party above the 80th percentile and one party below the 20th percentile. Non-partisans are those who do not fall in any of the aforementioned categories (those who do not rate any party higher than 80th percentile, or any party below the 20th percentile).

This way of measuring partisanship allows me to measure both strong positive affect for a political party as well as strong negative affect for a political party in a symmetrical fashion. Standard party identification questions in the pre-treatment data are only available for positive partisanship but not for negative partisanship. Face validity checks show that the partisanship measures are positively correlated with standard measures of party identification, and with the in-treatment measures of partisan affect (see Appendix, Section A.0.6). The cut-offs used in this paper are arbitrary but also more conservative than most operationalisations that use thermometer scores. Measures used in relevant papers often classify any score below the mid-point as negative, and any score above the

mid-point as positive and then combine these thermometer scores with a question about the party one would never vote for to measure negative partisanship (Caruana, McGregor and Stephenson, 2015; Ridge, 2020). Unfortunately, the "would never vote for" question was not asked in the survey wave that the experiment was fielded in. A notably different measurement strategy is employed by (Haime and Cantú, 2022) which only classifies voters as negative partisans when they give a party a score of 0 on the 0-10 scale. I build on the intuition behind this measurement approach and therefore measure negative partisanship not as any negative affect, but as extreme negative affect. In the Appendix Section A.0.12, I provide robustness checks using different cut-offs (absolute cut-offs at 1 and 9, 70th and 30th, 75th and 25th percentile cut-offs, 85th and 15th percentile cut-offs). Substantively, the results are the same in direction across these different specifications. The interaction effect between the negative partisanship treatment and pre-treatment partisan type (compared to the Control) stays significant across the 75-25th percentile cut-off and the 70-30th percentile cut-off, while the treatment-treatment interaction effect stays significant across the 85-15th cut-off, the 75-25th cut-off and the absolute cut-off (see Appendix, Section A.0.12).

2.4.5 Estimation

For each outcome measure Y_i, I estimate an OLS regression model where Negativevs-Positive_i is a binary variable indicating assignment to a negative partisanship prime (1) or a positive partisanship prime (0), Negative_i is a binary variable which indicates assignment to a negative partisanship prime (1) or to Control (0), and Positive_i is a binary variable which indicates assignment to a positive partisanship prime (1) or Control (0), X_i is an N-by-k vector of pre-treatment covariates and NegativePartisan_i takes the value of 1 when the individual only holds negative partisanship pre-treatment, or the value of 0 when the individual only holds positive partisanship pre-treatment. The full model also includes coefficients for non-partisans and closed partisans, with positive partisans as the reference category. Pre-treatment covariates included are gender, age, education, employment and province. I use two-tailed hypothesis tests and the statistical significance level of α=0.05.

ITT1 (Treat-to-Treat Comparison):

$$Y_i = \alpha + \beta_1 Negative vs Positive_i + \epsilon_i \tag{2.1}$$

ITT2 (Treat-to-Control Comparison):

$$Y_i = \alpha + \beta_1 Negative_i + \beta_2 Positive_i + \epsilon_i \tag{2.2}$$

ITT1 with interaction:

$$Y_i = \alpha + \beta_1 NegativevsPositive_i + \beta_2 NegativePartisan_i +$$
$$\beta_3 NegativevsPositive_i * NegativePartisan_i + \epsilon_i \tag{2.3}$$

ITT2 with interaction:

$$Y_i = \alpha + \beta_1 Negative_i + \beta_2 Positive_i + \beta_3 NegativePartisan_i +$$
$$\beta_4 Negative_i * NegativePartisan_i + \beta_5 Positive_i * NegativePartisan_i + \epsilon_i \tag{2.4}$$

2.5 Results

I first present results for the main treatment effects, and then turn to heterogeneous treatment effects by partisan type. While I only find suggestive evidence that the treatments affect the distance to the most liked party in the whole sample, there are sizeable heterogeneous treatment effects by pre-treatment partisan type. Experimental groups are balanced on age, gender and region, with minor imbalances on employment and education (Table A.1). Adjusting for covariates does not alter the results, see Appendix, Models 4-6, Tables A.30 and A.31. The majority of respondents engaged well with the treatment stimuli (see Section A.0.3).

Figure 2.2 summarises the main treatment effects. Figure 2.2 panel i) displays the mean level of distance to the most liked party among respondents who did not receive a writing prompt (Control), among respondents who were prompted to write about the party they dislike (Negative prime) and among respondents who were prompted to write about the party they like (Positive prime). Figure 2.2 panel ii) displays the effect of each treatment compared to the Control, as well as the effect of the negative compared to the positive prime treatment. The full regression tables can be found in the Appendix, Tables A.30 and A.31.[8] The results presented in Figure 2.2 are unadjusted; adjusting for covariates does not alter the results (see Models 4-6, Tables A.30 and A.31 for covariate adjusted results). Neither the negative partisanship nor the positive partisanship prime, compared to the Control, significantly affect distance to the most liked party. Compared to not receiving any writing task, writing about one's liked party and writing about one's disliked party do not significantly change how close respondents feel to their most liked party. Even though the treatment-to-control comparisons go in the hypothesised direction, they do not reach conventional significance levels.

---

[8]The N changes between the different models presented in the results tables when there are missing observations on the pre-treatment partisan type variable. Observations drop out of the model from M1 to M2 when respondents did not rate at least two political parties pre-treatment.

i) Mean levels of distance to most liked party by treatment condition

ii) Treatment effects
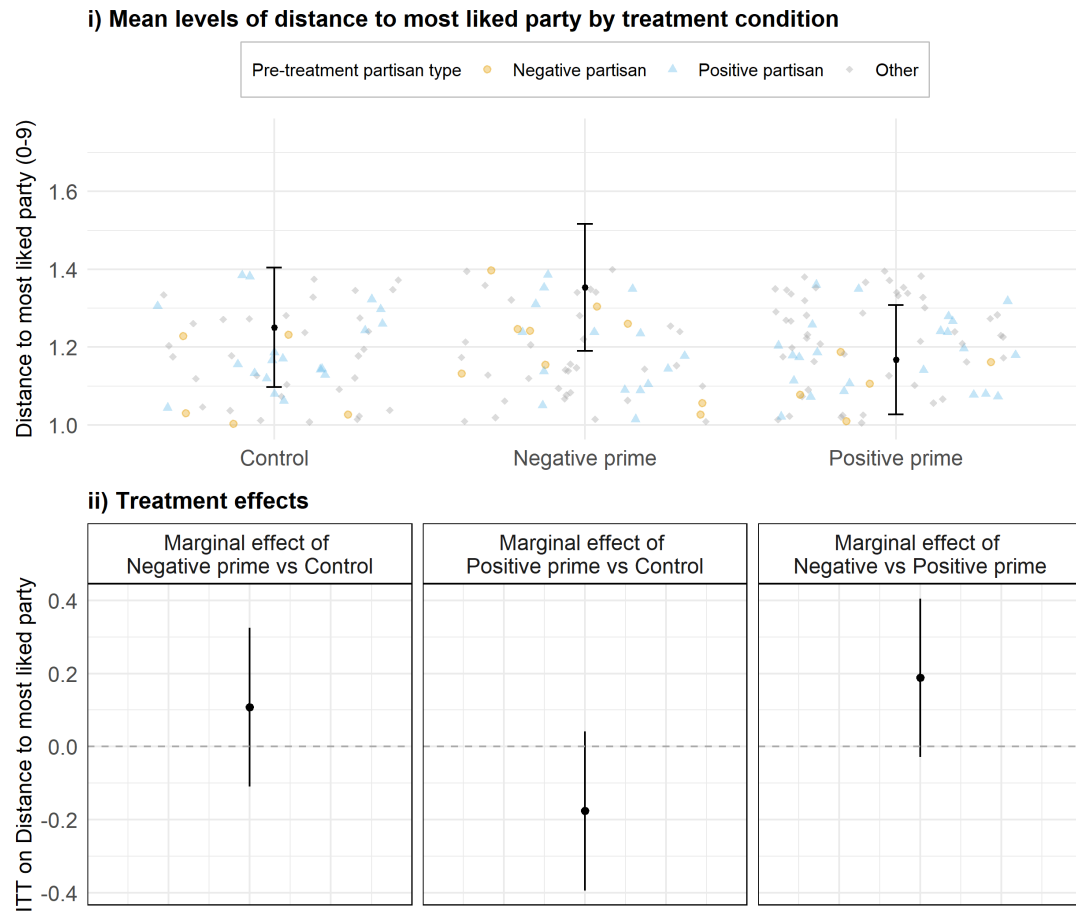
FIGURE 2.2: Treatment effects on distance to most liked party

When comparing the two treatment groups in the most right-hand panel of graph ii), Figure 2.2, we can see that there is a difference in distance to the most liked party, significant at p<0.1. Respondents assigned to the negative partisan prime place their most liked party further away from themselves than respondents assigned to the positive partisan prime. With *beta*=0.19, the marginal effect of the negative vs positive prime constitutes an effect of around 13% of the standard deviation in the outcome. In reverse, this means that when respondents write positive things about their most liked party, they move subjectively closer to that party than when they write about their disliked party. This finding provides some evidence in favour of H1. In the whole sample, the activation of positive partisanship generates somewhat increased feelings of closeness to a political party, and therefore encourages voters to increase their subjective ideological proximity to that party, compared to when negative partisanship is activated.

I now turn to the heterogeneous treatment effects by pre-treatment partisan type. As the treatments are designed to activate pre-existing feelings towards political parties, voters should react differently to the treatments according to their already existing partisan identity. I focus on how voters who can be categorised as positive partisans, compared to

voters who can be categorised as negative partisans react to the treatments.[9][10] Figure 2.3 shows the conditional effects of the treatments for negative and positive partisans. Full regression outputs are available in Tables A.30 and A.31 the Appendix. The results presented in Figure 2.3 are unadjusted; adjusting for covariates does not alter the results (see Models 4-6, Tables A.30 and A.31 for covariate adjusted results).

Treatment effects conditioned by pre-treatment partisan type
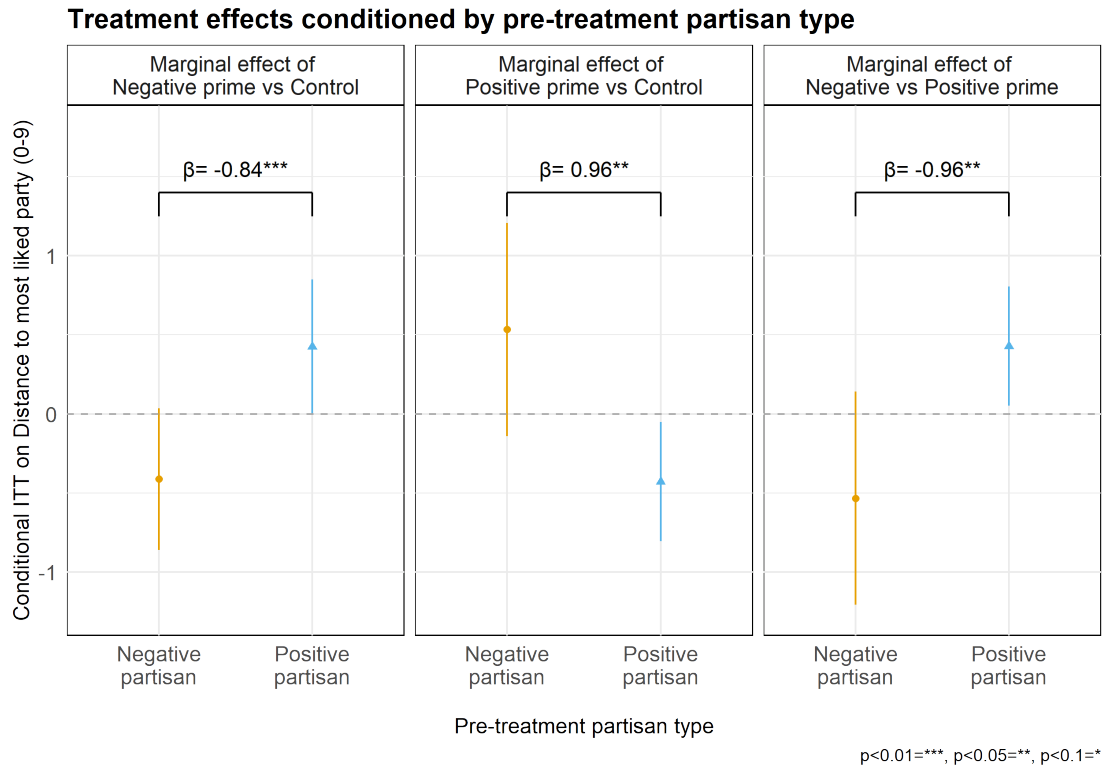
FIGURE 2.3: Negative partisans react differently to the treatment than positive partisans

The left-hand panel in Figure 2.3 shows that the effect of the negative prime (compared to the Control) differs for positive and negative partisans. The negative prime makes positive partisans feel more distant to their most liked party (β=0.42, p<0.05). In contrast, negative partisans slightly decrease their felt distance to their most liked party in response to the negative partisan prime, compared to the Control (β= -0.41, p=0.07). The difference in effect for negative and positive partisans is large (β= -0.84, 0.57 of a standard deviation in the outcome) and significant at p<0.01. Writing about the party that they really dislike (compared to not writing anything) encourages negative partisans, compared to positive partisans, to feel closer to their most liked party. In line with the idea of my enemy's enemy is my friend, activating negative partisanship makes

---

[9]Negative partisans hold extreme negative affect towards a political party, but no extreme positive affect towards another party. Positive partisans hold extreme positive affect towards a party, but do not display extreme negative affect towards another party. More information on the measurement of the partisanship types is available in Section 2.4.4.

[10]Results for all partisan types, including closed and non-partisans, can be found in the Appendix, Figure A.2 and Tables A.30 and A.31.

other political parties appear subjectively closer to negative partisans, compared to positive partisans. While I theorised that negative partisanship reduces the felt distance to other political parties, I conservatively expected that this would start to move feelings towards second-liked, or third-liked parties which might be more malleable than feelings towards the most-liked party. However, the activation of negative partisanship has an effect towards the felt distance towards the most-liked party for negative partisans. Substantively, this can be understood as an even more interesting and relevant finding as the most-liked party is often the party that individuals choose to vote for.[11]

Moving on to how the effect of the positive prime varies according to respondents' partisan type, the middle panel in Figure 2.3 shows that the effect of the positive prime (compared to the Control) significantly differs for positive and negative partisans. The positive prime (compared to the Control) makes positive partisans feel less distant to their most liked party (β=-0.45, p=0.025). This is in line with H1: positive partisans feel warmer and closer to their most liked party when their positive partisan affect is activated. In contrast, the positive prime does not have a significant effect on negative partisans (β=0.53, p=0.12).

Lastly, the right-hand panel in Figure 2.3 shows how the effect of the negative prime (compared to the positive prime) is different for positive and negative partisans. The negative prime (compared the positive prime) reduces the felt distance to the most liked party for negative partisans, compared to positive partisans (β=-0.96, p<0.05). This is a sizeable difference in effects and constitutes a change of around 0.65 in the standard deviation of the outcome. Overall, this shows that negative and positive partisans react differently to the partisan primes. Activating strong feelings of dislike towards a political party encourages negative partisans to feel closer to another political party, compared to positive partisans. This provides evidence in favour of the idea of "my enemy's enemy is my friend". Activating negative partisanship has important implications for political preferences that are different from those of positive partisanship.

The partisanship treatment primes did not have a significant effect on other outcome measures, including distance to the disliked party (see Appendix Section A.0.5) or the second-liked party (see Appendix Section A.0.8). The absence of effects on distance to the disliked party is surprising but likely due to ceiling effects. Those individuals most likely to react to the negative partisan prime are individuals with high pre-treatment levels of negative affect (negative partisans and closed partisans). However, these voters already place their disliked party far away from themselves at baseline (mean= 5 points

---

[11]Around 60% of respondents in the whole sample intend to vote for the same party that is also their most liked party. This number is higher among closed partisans (80% would vote for their most-liked party), and among positive partisans (69%), and lower among negative (49%) and non-partisans (40%). More information about this can be found in the Appendix, Section A.0.10.

on the 0-10 left-right scale for closed partisans, mean=3.6 points on the 0-10 left-right scale for negative partisans). Given that the mean left-right self-placement is 5.12, there is very little room for respondents to place their disliked party even further away from themselves in response to the treatment. This makes it difficult to detect any effects of the negative partisan prime on distance to the most disliked party.

2.6   Conclusion

This paper examines what implications negative partisanship has for voters' political preferences in multi-party systems. Negative partisanship is often understood as a by-product of positive partisanship and studied in voters who already positively identify with a party (such as Democrats who also dislike Republicans). Yet, especially in multi-party systems, voters might simply strongly dislike a political party without feeling attached to another one. I examine if these strong feelings of dislike can have implications for political preferences that are distinct from those of positive party attachments.

Using a survey experiment, I test what effects activating negative and positive partisanship have on voters' perceived distance to political parties. In line with existing theories on partisanship, I expected that priming positive partisanship makes voters feel closer to their most liked party. I find evidence that this is the case among positive partisans. Writing about their most liked party makes positive partisans feel closer to that party. This confirms that positive partisanship can have strong effects on political preferences. However, I also find that activating negative partisanship has similarly powerful effects on political preferences. Writing about their most disliked party (compared to not receiving any partisan prime) makes negative partisans feel significantly closer to their most liked party. Thinking about really disliking a party (versus really liking a party) can motivate negative partisans to feel more ideologically proximate to another political party than positive partisans.

These findings contribute to the literatures on partisanship and affective polarisation. Firstly, really disliking a party does not just prevent a voter from ever endorsing or voting for that party, it can also change how someone feels about the other political parties on offer. This is in line with the argument of "my enemy's enemy is my friend". By categorising the world into "them" (the disliked party) and "other parties", other parties start feeling subjectively closer to negative partisans. This finding also has wider implications for our understanding of party competition and partisanship in multi-party systems. While partisan affect, both positive and negative, may affect how voters feel towards one specific political party, this paper shows that partisan affect also has ramifications for how voters evaluate other political parties around them. Theorising

and testing how negative (and positive) partisanship towards one party affect evaluations towards other political parties is an important avenue for further research.

Secondly, this paper provides support for the argument that negative partisanship matters in its own right and is not just an extension of positive partisanship. Especially in multi-party systems, voters can harbour strong feelings of dislike for a political party, without feeling close to any other party. The findings from this survey experiment show that activating these strong feelings of dislike has implications for political preferences. Strong feelings of dislike towards political parties should not be considered a mere by-product of positive partisan attachments, but as a force of their own.

Finally, activating negative partisanship can also encourage partisans to feel closer to a political party, potentially inspiring the development of more positive feelings of attachment in the long run. This is important because research on affective polarisation has shown that strong negative feelings towards partisan out-groups are prevalent in many Western societies (Gidron, Adams and Horne, 2019). This poses serious challenges for democracies, from reduced support for democratic norms (Kingzette et al., 2021) to willingness to engage in political violence (Mason and Kalmoe, 2022). Given the harmful implications of negative partisanship for democratic functioning, this paper offers a cautiously hopeful note on how negative feelings towards political parties may ultimately transform into new positive party attachments. Even though the affective polarisation literature shows that in-group love can go hand in hand with out-group hostility (Iyengar et al., 2019), foundational psychological research has shown that positive identities can also exist by themselves and do not need out-group hostility to become stable identities (Brewer, 1999). While new positive party attachments are no guarantor for lower out-group hostility, holding a positive identity is a relative improvement to only feeling dislike towards political parties. As such, it could be an important first step towards developing inclusive, positive identities. The potential for negative identities to spur positive attachments could thus also spell good news for societal cohesion and democracy.

Chapter 3

Negative Political Identities and Costly Political Action

Katharina Lawall, Stuart J. Turnbull-Dugarte, Florian Foos,Joshua Townsley

Abstract

Elite and mass level politics in many Western democracies is characterised by the expression of strong negative feelings towards political out-groups. While the existence of these feelings is well-documented, there is little evidence on how negative identities are activated in campaign contexts. Can political campaigns use negative political identity cues to raise donations? We test whether fundraising emails containing negative or positive political identity cues lead supporters of a party to take costly action via a large pre-registered digital field experiment conducted in collaboration with a British political party. We find that emails containing negative as opposed to positive identity cues lead to a higher number and frequency of donations. We also find that negative identity cues are only effective when paired with an issue cue rather than a traditional partisan identity cue. Our results provide novel experimental evidence of how negative political identities are activated in real-world political campaigns.

3.1   Introduction

Strong feelings of dislike towards specific political parties are commonplace among the mass public in the United States and in many other Western democracies (Iyengar et al., 2019; Gidron, Adams and Horne, 2019; Huddy, Bankert and Davies, 2018; Wagner, 2021). Alongside partisan polarisation, recent research has shown that negative affect can also extend towards issue-based out-groups (Hobolt, Leeper and Tilley, 2020; Wagner and Eberl, 2022). These negative feelings towards political out-groups can have an important role in structuring individual-level political behaviour and preferences (Abramowitz and Webster, 2016; Iyengar and Krupenkin, 2018; Iyengar et al., 2019; Hobolt, Leeper and Tilley, 2020; Bankert, 2020; Wagner, 2021; Wagner and Eberl, 2022). However, we know little about the role of political parties in activating these negative political identities, particularly among their own supporters. This paper examines whether campaigns can use political identities to mobilise supporters into taking costly political action, and what types of identity cues are most effective in encouraging mobilisation. Are negative identity cues more effective than positive identity cues in mobilising supporters? Is out-group hatred more easily mobilised than in-group love?

At least anecdotally, parties and campaigns have increasingly relied on negative identity cues to raise money from their supporters. Understanding what consequences negative identity campaigning has is important because it might further entrench political divides, and contribute to growing affective polarisation. Heightened political out-group dislike has been linked to a range of worrying consequences for democracy, such as willingness to engage in political violence (Mason and Kalmoe, 2022), out-group dehumanisation (Martherus et al., 2021) and reduced support for democratic norms (Kingzette et al., 2021). As elite-level affective polarisation is linked to mass-level polarisation (Banda and Cluverius, 2018), negative identity campaigning could provide a reinforcement of negative out-group feelings among supporters and voters. Examining the implications of negative identity campaigning is therefore important for our wider understanding of democratic functioning in an age of polarisation.

To test the effects of negative identity cues and issue cues on donations, we conducted a large digital field experiment in collaboration with a British political party during the 2019 General Election campaign. We randomly assigned around 90,000 party supporters to receive either no campaign email, or a campaign email that contained a combination of a negative or a positive identity cue, with an issue or a partisan identity cue. We then identify the effect of these cues on time-stamped donations to the political party as a measure of costly political behaviour. The 2019 UK General Election provides an excellent case for studying the effects of different identity cues on costly mobilisation. The UK is characterised by relatively high levels of affective polarisation compared to

other European democracies (Gidron, Adams and Horne, 2019), rendering it a case where negative partisan identities should have an effect on costly mobilisation. At the same time, Brexit-related issue identities were prevalent and highly salient during the 2019 election (Ford et al., 2021). This makes it an ideal case for comparing the power of issue and partisan, as well as negative and positive identities.

In line with our theoretical expectations, we find that emails containing negative identity cues are more effective than positive cues in increasing the number and frequency of donations. We also find that negative identity cues are only effective when paired with an issue cue, rather than a traditional party identity cue. Even in the noisy context of a General Election campaign, a single campaign email containing negative issue cues drives up donations by 15%, compared to not receiving any campaign email. This is a sizeable treatment effect and demonstrates the powerful behavioural consequences of activating negative political identities in real world-campaign settings. However, while the negative identity cue emails mobilised supporters, they did not raise the overall donation amount received by the political party. Our results therefore also show that negative identity campaigning involves trade-offs between mobilisation and fundraising and is not a uniformly winning strategy for political parties.

This paper makes several contributions to the literature on negative partisanship and affective polarisation, as well as the literature on negative campaigning and campaign donations.

First, most studies on negative partisanship and affective polarisation investigate the link between negative partisanship and political participation using observational data (McGregor, Caruana and Stephenson, 2015; Bankert, 2020; Mayer, 2017). By studying the implications of negative identity cues in a digital field experiment using validated donations to a political party, our study adds robust empirical evidence about the implications of negative partisanship activation for behaviour. We also contribute to the negative partisanship and affective polarisation literature by showing that negative identity activation has consequences for political behaviour that are distinct from those of positive identity activation. This attests to the importance of understanding negative out-group feelings as a key driver of political behaviour.

Secondly, we extend the literature on affective polarisation and negative partisanship by theorising and testing how negative political identities interact with issue-based identities. Whilst voters have traditionally structured their political beliefs around party loyalties, voters' political identification with a particular issue (e.g. identification as *pro-choice*, an *anti-vaxxer*, or as a *Brexiteer*) has come to play an increasingly important role in structuring attitudes and behaviours (Hobolt, Leeper and Tilley, 2020; Wagner and Eberl, 2022). We provide evidence that issue-based identities matter for costly political

behaviour but only when activated and combined with a negative identity cue. This advances our theoretical understanding of issue-based identities, and clarifies the scope conditions under which issue-based identities affect political behaviour.

Thirdly, by focusing on negative alongside positive political identities, we contribute to the literature on the role that political identities play in campaign donations (Hassell and Monson, 2014; Cyphers, Hassell and Ou, 2022; Perez-Truglia and Cruces, 2017), elite- and party-level negative campaigning (Barton, Castillo and Petrie, 2016) and the literature on mass-level negative partisan and issue-based identities (Iyengar et al., 2019; Hobolt, Leeper and Tilley, 2020). While there are field experiments that test the effects of negative and positive campaign messages, these are not strictly focused on testing the effects of identity cues (Gerber, Green and Green, 2003; Niven, 2006; Arceneaux and Nickerson, 2010; Barton, Castillo and Petrie, 2016). As cues relating to individual-level negative identities tap into highly emotional, visceral conflicts between in- and out-groups, negative identity campaigning should provide a powerful form of negative campaigning. In an age of partisan polarisation, understanding the consequences of negative partisan identity messages as a specific form of negative campaign message is relevant and important.

3.2   Negative identity campaigning

Social identities are a fundamental feature of society. Social identity theory posits that individuals adopt social identities to satisfy higher psychological needs like belonging and creating a sense of self-worth (Tajfel and Turner, 1979). This also applies to the realm of politics, in which partisan identities, or feelings of attachment to a political party, are a key structuring force of political behaviour and preferences (Campbell et al., 1960; Green, Palmquist and Schickler, 2002; **?**). Partisan identity affects whether citizens choose to take political action. Strong partisans react emotionally to the prospect of losing or winning an election, and are motivated to get involved in campaigning efforts to defend the status of their partisan in-group (Huddy, Mason and Aaroe, 2015). Rather than engaging in political mobilisation for instrumental or strategic reasons, partisans are often motivated by preserving their sense of self-worth. Hersch (2020) refers to these individuals as "party boosters" - those individuals who, when primed by their partisan identities to contribute to the group cause by donating or volunteering their labour, do so in order to "help their side" (p.83). Party boosters are not, however, strategic in that their contributions to campaigns are not centred on "toss-up" races but rather represent "token", "ego-stroking", contributions that serve as expressive signals of their political identities that provide them with subsequent "bragging rights" amongst their peers that

belong to the same partisan identity-based in-group Hersch (2020, pp.82-84). Building on this, and the recent experimental literature on the efficacy of online campaign efforts (Han, 2016; Turnbull-Dugarte et al., 2022), we expect that a campaign message from a political party seeking to "rally the troops" will mobilise supporters to donate to the campaign:

H1 Email mobilisation hypothesis: Email contact from the party will mobilise supporters to donate to the campaign.

Individuals who feel a strong sense of attachment to a party in-group can also develop strong negative affect towards other parties as a way of protecting and expressing their partisan identity (Bankert, Huddy and Rosema, 2017; Medeiros and Noël, 2014). Negative partisanship has been described as "affective repulsion", or a strong dislike and disdain for a political party (Caruana, McGregor and Stephenson, 2015). Much of the research on negative partisanship studies the implications of negative partisan affect in "closed partisans", individuals who are for instance simultaneously pro-Republicans and anti-Democrats (Abramowitz and Webster, 2016). In line with this, we study the implications of activating positive and negative partisan affect in party supporters.

We expect that mobilising negative feelings towards political out-groups has stronger effects on behaviour than mobilising positive in-group identities because humans tend to weight negative preferences and experiences more heavily than positive experiences (Tversky and Kahneman, 1992). Applied to political behaviour, negativity bias has been documented in affective responses to political rhetoric (Bakker, Schumacher and Rooduijn, 2021), in responses to news content (Soroka, Fournier and Nir, 2019), and in candidate evaluation (Holbrook et al., 2001).Miller and Krosnick 2004 suggest that loss aversion might provide a powerful motivation for citizens to become politically active and to engage in costly political participation, such as donations.

Grounded in these insights from psychological and behavioural research, the partisanship literature has argued that negative partisan identity has a mobilising effect. Voters with negative partisanship are more likely to turn out to vote (McGregor, Caruana and Stephenson, 2015; Mayer, 2017). Voters who have a bigger gap in their affect towards their in-party and their out-party are also more likely to say they would campaign for a candidate or party, contact an elected official, take part in a protest or be involved in a social or political cause (Wagner, 2021).

In line with these theoretical considerations, we expect messages that activate negative identities to have a stronger mobilising effect than positive identity primes.

H2 Negative identity cue hypothesis: Messages that prime negative political identities will have a larger positive effect on donations than messages that use positive identity cues.

Hypothesis 2 is also underpinned by research on negative partisanship and affective polarisation that connects changes in elite-level behaviour to changes in individual-level behaviour. For example, increases in elite affective polarisation have been linked with increases in individual-level affective polarisation among partisans (Banda and Cluverius, 2018). Elite-level communications containing negative political identity cues should increase supporter mobilisation because they make the cued identity more salient to supporters. Research on affective polarisation has shown that the salience of negative partisanship matters for political behaviour (Iyengar et al., 2019; Lelkes and Westwood, 2017).

The idea that elite-level communications that cue identities can affect the behaviour of party supporters also strongly resonates with the negative campaigning literature. Resorting to negative campaigning, or attacking and criticising political opponents during an election campaign, is common among political parties and candidates in Western democracies (Geer, 2006). Negative campaigning can take different forms, including issue attacks criticising competitors' policies, as well as valence attacks criticising a competitors' competence or character traits (Nai, 2020). As the competitiveness of elections increases, parties and candidates go negative more frequently (Banda, 2021). Since the 1990s, there is a lively debate about whether negative campaigning is an effective political strategy. Early research on this question claimed that negative campaigning is harmful and "demobilizes" the electorate (Ansolabehere et al., 1994). However, these early studies have been challenged by subsequent research (Wattenberg and Brians, 1999), and more recent research which shows that the effects of negative campaign communications on turnout and candidate support are mixed (Lau, Sigelman and Rovner, 2007). While some studies show that negative messages can increase turnout (Niven, 2006; Barton, Castillo and Petrie, 2016), other research finds that negative campaign messages are no more effective than positive ones in increasing turnout (Arceneaux and Nickerson, 2010) or donations (Barton, Castillo and Petrie, 2016). Hypothesis 2 hence differs from some of the earlier negative campaigning literature which expected that negative campaign message could have demobilising effects (Ansolabehere et al., 1994).

Most of the negative campaigning literature tries to understand what effects negative campaign messages, including trait, issue or competence attacks, might have on the electorate as a whole. With this study, we contribute to the negative campaigning literature in two ways. First, we study the effects of political parties' negative messages on the mobilisation of (their own) party supporters, rather than the electorate at large. Second, we study a specific type of negative message, focusing on identities, rather than issues, traits or competence, that is used by political parties but that has not received as much scholarly attention in the experimental literature. Political parties can rally their supporters around a shared identity or a shared opponent identity. We argue that in the current political context of high polarisation, this "negative identity campaigning" should have more positive effects on mobilisation than positive identity campaigning. By combining insights from the negative campaigning literature and the literature on political identities, we investigate how political identities are activated and mobilised into political action.

3.3 Issue versus party identity activation

Affective polarisation, and the mobilising influence it catalyses among group members to engage in campaign activity - be that in the form of volunteering or more "token" activity like donations that serve as signals of group membership - is not unique to partisan identities and can also include issue-based identity groups[1] and, consequently, issue-based affective polarisation.

Issue-based political identities can sometimes emerge orthogonally to political partisanship[2] and are born out of inter-group conflict on concrete policy issues: a divide between individuals who *identify* as proponents or opponents of a certain issue (Bliuc et al., 2007). Simply holding an opinion on an issue does not constitute membership of an issue-based identity. Whilst harbouring a clear (pro versus anti) stance on a concrete policy issue is a core prerequisite for issue-based identities, political identities only emerge when individuals consciously and actively identify with the identity, differentiate themselves from those who harbour opposing identities, and when the individuals engage in evaluative bias that regards in-group status and penalises out-group status (Mason, 2018; McGarty et al., 2009).

---

[1] Whilst we refer to identities such as "Remainer" and "Leaver" as issue-based identities, similar social identities based on individual and group-based identification with salient political issues have also been referred to as "opinion-based groups" (McGarty et al., 2009).

[2] Independent issue-based identities are increasingly less frequent in the US context where, as a result of social sorting, intra-party heterogeneity of issue positions has dramatically decline and been replaced with strongly polarised inter-party distinctiveness (Mason, 2018).

Issue-based identities are prevalent across a number of countries. In the US one of the most salient issue-based identities to emerge are what (Crawford et al., 2021) identify as the "abortion identities" of so-called *pro-life* and *pro-choice* individuals. As detailed by (Killian and Wilcox, 2021) these abortion-based issue identities often become "irresistible forces" that can result in the abandonment of conventional partisan loyalties. More recently during the Covid-19 pandemic, a vocal proportion of society has increasingly sought to identify itself as "anti-vaccine" (Wagner and Eberl, 2022; Abrams, Lalot and Hogg, 2021). Motta et al. (2021) demonstrate that the anti-vaxxer label is not only a descriptive characteristic of these individuals' shared view on vaccination programmes, but rather serves as an issue-based social identity that anti-vaxxers use to distinguish themselves from others. Additional examples can be found in those states where multidimensional conflicts lead to the emergence of issue-based identities that are independent of partisanship. Secessionist movements in Scotland (UK) and Catalonia (Spain), for example, have engendered issue-based identities around constitutional preferences that often trump partisan preferences at the ballot box. In Scotland where voters have been manoeuvred into two constitutional dichotomies on both membership of the UK as well as membership of the EU, voters have adopted tribe-like identities (Mitchell and Henderson, 2020) that predict electoral preferences that cut across party loyalties (Johns, 2021).

The UK's Brexit referendum of 2016, which coerced citizens of diverse ideological and party colours[3] to coalesce behind a *Remain vs. Leave* dichotomy, is a prime example of how concrete events can catalyse the emergence of salient and consequential issue-based identities (McGarty et al., 2009). Empirically, Hobolt, Leeper and Tilley (2020) demonstrate that issue-based identities reflecting individual attachment to positions on Brexit have materialised as salient political identities amongst the British electorate. Not only have these Brexit identities - affective attachment and rejection of individuals based on their identity as *Leavers* or *Remainers* - become more prevalent than partisan identities but these attachments have also become *stronger* than conventional political (partisan) identities (Hobolt, Leeper and Tilley, 2020).

Issue-based political identities that were encapsulated by the Brexit referendum have also been shown to have consequential implications for how citizens evaluate information. A common observation in the literature on party-based identities is the emergence of a "perceptual screen" that serves to moderate how partisans receive, process and digest

---

[3]As detailed extensively elsewhere, a lack of intra-party cohesion was a common feature of political parties in the UK in both the lead up to and during the aftermath (Trumm, Milazzo and Townsley, 2020) of the Brexit referendum. This is significant as it signals the issue of Brexit as one disconnected from any clear party line at least, that is, until the Conservatives' positioning as the *de jure* Brexit party was solidified under the leadership of Boris Johnson (Prosser, 2021).

information (Campbell et al., 1960; Zaller, 1992). A tough test of the importance of issue-based identities in shaping political preferences, therefore, is the extent to which these identities replicate partisan-like processes of motivated reasoning. Sorace and Hobolt (2020) empirically demonstrate that, above and beyond partisanship, salient issue-based identities can trigger a perceptual screen that moderates how individuals process and evaluate political information such as the performance of the economy. The authors' combination of observational and experimental evidence highlights affective attachment to political identities other than partisan can be as influential as partisan attachments when it comes to how individuals view the world.

If, as argued by Klandermans et al. (2002), strong attachments to issue-based group identities makes political participation that favours the interest of the group more likely, then we would expect priming individuals on salient and electorally sensitive issue cues to induce them to participate in costly campaign activity. Existing experimental evidence suggests that this is indeed the case. Turnbull-Dugarte et al. (2022) present evidence from an experiment fielded in collaboration with the anti-Brexit campaign organisation, *The People's Vote*, demonstrating that informing anti-Brexit citizens of the pro-Brexit position of their parliamentary representative significantly increased active supporters' (donors) propensity to lobby their representative in favour of a second referendum on the issue. This experiment provides causal evidence from a naturalistic setting which demonstrates that priming individuals on opposing issue positions can, in addition to signalling issue stances,also mobilise individuals with issue-based *identities* by activating negative negative attachments that exhibit an influential mobilising effect on supporters' political behaviour (Turnbull-Dugarte et al., 2022).

Given the rise of issue-based (Brexit) identities in the UK, their increased relative importance vis-à-vis partisan identities, as well as the focus on and salience of Brexit in the 2019 general election (Ford et al., 2021; Prosser, 2021; Cutts et al., 2020) - during which time our experimental test took place - our *pre-registered* hypothesis is that email messages that prime individuals on salient issue positions, which are often attached to issue-based identities, would be more effective at driving donations than messages priming conventional (party) identities.

H3 Issue identity mobilisation hypothesis: Messages that prime issues will be more effective at encouraging campaign donations than messages that prime traditional party identities.

3.4 Negative issue identities

Finally, combining theories on issue vs partisan identities and negative vs positive partisan identities, we develop an argument of when negativity matters in identity formation and activation. We expect that negative issue cues should elicit stronger mobilisation reactions than negative partisan cues. We argue that this is the case because negativity constitutes an even more central element of issue identity formation than partisan identity formation. We build on the issue identity literature, the partisanship literature and on social movement theory to develop this argument.

While previous research has argued that out-group animosity and negative stereotyping of out-group members is an important feature of issue identities (Hobolt, Leeper and Tilley, 2020; Wagner and Eberl, 2022), we explicitly hypothesise and test that issue cues framed around out-group dislike are more effective than issue cues framed around in-group love. In doing so, our paper advances the literature on issue and partisan mobilisation, clarifying the scope conditions under which issue identities affect political behaviour. Examining in which context negative identities have the strongest effects on behaviour also adds to our understanding of the role of negativity in political identity activation. We still know little about in which contexts feelings about "who we are not" matter more than feelings about "who we are" (Elsbach and Bhattacharya, 2001). Our paper addresses this gap by developing and testing novel expectations in which identity settings out-group dislike has the strongest effects on behaviour. While out-group animosity is common among partisans, we argue that out-group dislike is an even stronger structuring feature of issue identities, resulting in heightened behavioural responses to negative issue cues.

Previous research has shown that out-group animosity is a central feature of issue identities (Hobolt, Leeper and Tilley, 2020). For example, both Remainers and Leavers engage in negative outgroup stereotyping, describing individuals on the other side as less honest, intelligent and more hypocritical and selfish (Hobolt, Leeper and Tilley, 2020). From a theoretical perspective, issue identities are particularly suited to the development of strong negative identities. Issue identities often emerge in response to deeply divisive, binary opinion divides, such as referenda, where two opposing sides are pitted against one another in a winner-takes-it-all scenario (see also Hobolt, Leeper and Tilley (2020)).

Compared to negative partisanship in multi-party settings, issue identities are thus more focally structured around just one disliked out-group. Animosity towards other political parties may be spread over multiple disliked parties in multi-party systems. In contrast, the binary nature of issue identity conflict concentrates all negative feelings on one single political out-group. We therefore argue that out-group animosity should have

more powerful effects on behaviour when combined with a salient issue identity, than a party identity.

This expectation is also underpinned by foundational research on social movements. Issue identities often emerge within social movements, such as the climate change movement (Vesely et al., 2021). People are not just in favour of climate change mitigating policies, they also actively define themselves as "environmentalists" – which has consequences for their intention to take action on these issues (Vesely et al., 2021). Negativity plays a central part even in these positively formulated identities: for example, environmentalist movements are often framed in opposition to the status quo, e.g. "Stop climate change", "Ban fracking", "Anti-nuclear energy". As social movements are rooted in grievances, strong negative feelings such as illegitimacy and injustice (van Stekelenburg and Klandermans, 2010), the resulting identities also contain a strong negative affective component. These highly negative, "hot cognitions" of moral indignation and anger are a key part of the collective action frames that social movements use when building movement identities (Gamson, 1992). Negative identities are thus not an optional, late addition to collective in-group identities, but an essential part of the formulation of positive in-group identities for social movements: the "identity component refers to the process of defining this *we*, typically in opposition to some *they* who have different interests or values" (Gamson, 1992a:7). For example, movements as different as *Occupy Wall Street*, the anti-nuclear energy movements, but also the recent anti-vaxxer movement seek to mobilise and unite a diverse set of individuals around a shared dislike.

Applying these insights from social movement theory to issue identities, negative feelings should constitute a central element of issue identities. In comparison, partisan identities can more easily be constructed and maintained in exclusively positive ways. Feelings of attachment and belonging to a partisan in-group can exist without a simultaneous hatred of rival political parties and partisans. While many partisans do also hold negative out-group feelings, we argue that out-group dislike is an even more central identity component for issue identities. Because of the importance of negativity in issue identity formation, we expect negative issue cues to have the most powerful effects on costly mobilisation, even trumping the effects of negative partisan identity cues.

H4 Interaction hypothesis[4]: *Messages that prime negative identities will be more effective when combined with issue cues rather than traditional party identities.*

---

[4]Pre-registered as H2 in the PAP.

3.5   Methods

We conducted a pre-registered[5] digital field experiment in collaboration with a British political party during the 2019 UK general election campaign.

What sets our experimental design apart is that we were able to assign the entire membership universe of a mainstream political party to different experimental conditions, allowing us to detect small, but meaningful effect sizes. This is important because the effects of emails on donation behaviour during a high stakes election campaign, where citizens, and especially party members, are exposed to many different campaign emails, are likely small. But small effects are still important because email is a method that is easily scalable. While a party member might be more likely to take action following an email from the party than the average voter or party supporter, donation rates, even among party members, are relatively low in the UK. The action of donating requires potential donors to complete an online form, and pay using an online payment method, such as credit card.

In this study, we have access to validated donation data, that we obtained from the political party and matched to our experimental assignment. It is well known that there are discrepancies between self-reported data on donation behaviour and actual behaviour, with subjects likely to overestimate their donation when asked in surveys (Cyphers, Hassell and Ou, 2022). Over-reporting is of particular concern in our case because it could be a function of activating positive or negative political identities. Similarly to Perez-Truglia and Cruces (2017) and Cyphers, Hassell and Ou (2022), we overcome this challenge by relying on validated data on donations. Collaborating with a political party and accessing validated donations data allows us to identify and measure costly political action of party members and supporters in a high-stakes setting, the 2019 UK General Election.

3.5.1   Context

The 2019 election marked the end of tumultuous 2017-19 parliament characterised by political instability and parliamentary deadlock over Brexit (Ford et al., 2021; Sobolewska and Ford, 2020). The European Parliament elections held in May 2019 signalled the continuing relevance of the Brexit question with the newly-formed Brexit Party, led by prominent Brexit supporter Nigel Farage, claiming victory and the explicitly anti-Brexit Liberal Democrats - campaigning with the slogan "Bollocks to Brexit" - coming a close

---

[5]PAP registered prior to fielding at 11/11/2019, available online at: `https://osf.io/nzvku`. The experiment was reviewed and approved by the LSE institutional research ethics committee under REC ref. 1005.

second. Boris Johnson, who became Prime Minister in July 2019, faced with the subsequent inability to secure parliamentary support for his withdrawal agreement and with Brexit remaining one of the most polarising issues in the electorate (Hobolt and Rodon, 2020), instigated the process of early elections to be held in December.

The 2019 election was, unsurprisingly, dominated by Brexit (Ford et al., 2021; Cutts et al., 2020). The Conservatives campaigned under the banner "Get Brexit Done". Labour, the main opposition party, proposed to renegotiate the withdrawal agreement before putting it to the public in a second referendum. The Brexit Party campaigned for a "no deal" Brexit, but subsequently stood down in a number of marginal constituencies in order to support pro-Brexit Conservative candidates (Ford et al., 2021). Throughout 2019 and the campaign, the Liberal Democrats sought to monopolise the anti-Brexit space, emphasising their strong opposition to Brexit and pledging to "Stop Brexit" if they won a majority of seats. A number of Labour and Conservative MPs defected and joined the Liberal Democrats, many on the basis of their opposition to Brexit. In England and Wales, the party joined with Plaid Cymru and the Greens to form the "Unite to Remain" electoral pact, where candidates stood down in 60 constituencies where it was believed one of their candidates stood the best chance of defeating a pro-Brexit incumbent. This is a concrete example where the joint dislike of a specific issue position (i.e. a negative issue identity) is driving real world political activity: anti-Brexit parties (except for Labour) were coordinating in order to block the electoral viability of a specific pro-Brexit candidate.

3.5.2 Experimental Design

We test the effects of positive and negative, issue and partisan identity cues on donation behaviour via a field experiment. The experimental sample is based on the 98,206 individuals who were registered in the political party's membership database as of October 2019. After removing entries without verifiable email addresses and individuals below the age of 18, we arrive at the experimental sample of 89,941 subjects. Using complete random assignment blocked on past donation behaviour (whether an individual had ever donated to the party, or not), we randomly assigned subjects with a probability of .2 to control (no email), and with a probability of .2 to one of four email conditions. With equal probabilities, we randomly varied two factors in this email, 1) if the email used positive or negative identity cues and 2) if the email referenced party or issue-based identities. Having a pure Control condition (no email) is important because it allows

us to test whether receiving a campaign email, regardless of content, has an effect on donations.[6] The resulting assignment is displayed in Table 3.1.

TABLE 3.1: Assignment of subjects to experimental conditions

Group	Positivity	Identity	Ask	Assignment
Control	n/a	n/a	n/a	20%
T1	Positive	Party identity cue (in-party)	donation	20%
T2	Negative	Party identity cue (out-party)	donation	20%
T3	Positive	Issue-based cue (pro remain in EU)	donation	20%
T4	Negative	Issue- based cue (anti Brexit)	donation	20%

3.5.3 Treatments

Figure 3.1 displays the different treatment emails sent to subjects. The first, left-most email shows the positive party cue, the second email the negative party cue, the third email the positive issue-based cue and the last email, the negative issue-based cue.

The goal of all emails was to prime a political identity. To achieve this, all email texts stated that "there is one thing we all have in common" and that "we cannot do this alone, but together we've got this". There is thus a strong emphasis on a shared identity through the use of pronouns and adverbs ("we", "together"), shared group goals ("we all want to"), and shared group characteristics ("we all have in common"). This language is meant to make group identification salient to email recipients. The text across all the treatment conditions then links this group identity to a specific action: making a donation, visible as a large "Donate here" button at the bottom of the email.

The content that varied between the email is the negativity and the specific identity invoked in the email text. The emails invoked either a positive partisan identity ("help the [in-party] win"), a negative partisan identity ("defeat the [out-party]"), a positive issue ("remain in the EU") or a negative issue ("stop Brexit").

Importantly, the email subject lines also varied along with the assigned cue. The positive party treatment headline was: "Let's win as [in-partisans] - together", the negative party headline was: "Let's win against the [out-partisans] – together", the positive issue headline was: "Let's win as Remainers - together", and the negative issue headline was: "Let's Stop Brexit - together". The purpose of the varying subject line was to maximise the effect of the treatment. While this design choice might have led to differential opening rates, we see this as an integral part of the mechanism, negative vs positive, and partisan vs issue cue mobilisation, that we intend to test.

---

[6]Orthogonally, we also randomly assigned whether the email contained an additional ask to volunteer, or not. Unfortunately, we could not collect valid outcome data on volunteering activities. We therefore do not consider this assignment in the analysis. The entire assignment is displayed in table B.1.

Figures/allissueidajps.png

FIGURE 3.1: Treatment emails

3.5.4 Data and Outcome measures

We identify the effects of the differently worded emails on three outcome variables: number of donations, frequency of donating and donation amount. The first outcome is binary and defined as the successful submission of a donation via the party's online payment platform. It takes on the value 1 if a verified donation from the participant was received, and the value 0 if no donation was received. The second outcome is measured continuously and indicates how many times an individual donated to the party in the post-treatment period. It takes on the value 0 if the subject did not donate. Finally, the third outcome is the value amount (in £) of the donations received, with no donation being registered as £0. All outcomes are measured using the database retrieved from the organisation's payment platform, 2, 7 and 14 days after the treatment emails were sent. As pre-registered, in our analyses we normally privilege the measure taken 7 days after the treatment, but where appropriate and for robustness, we also display the measures taken 2 days and 14 days after the treatment. We expect that it takes a couple of days for

financial transactions to be completed and compiled in the database and expect that the party sent out other fundraising emails to supporters in subsequent days. 7 days hence appears to be the sweet spot between donation transactions being processed and wash-out from subsequent fundraising emails. When individuals make a payment, they provide their email address and name, which is then used to link their donation to the original member database used for the assignment. By matching identifiers for each donation made in the post-treatment period to the identifiers of individuals in our sample, we are able to assess which individuals donated to the party after the treatment. Matching individuals across databases is not without complications. However, the political party holds identifying details including numerous email addresses, physical addresses and names that were used to match subjects.

3.5.5   Estimation strategy

We estimate Intent-to-Treat (ITT) and Conditional Intent-to-Treat Effects[7], based on linear regression of donation behaviour on assignment to receive any campaign email (1), or no email, coded as 0 (model 1), or on the two experimental factors, negative (1) vs positive cue (0) and issue-based (1) vs party-based cue (0) in models 2-4. We use heteroscedasticity-consistent standard errors (HC2) throughout.

$$Y_i = \alpha + \beta_1 Email_i + \gamma' X_i + \epsilon_i \tag{3.1}$$

$$Y_i = \alpha + \beta_1 Negative_i + \beta_2 IssueCue_i + \delta PriorDonor_i + \gamma' X_i + \epsilon_i \tag{3.2}$$

$$Y_i = \alpha + \beta_1 Negative_i + \beta_2 IssueCue_i + \beta_3 Negative*IssueCue_i + \delta Priordonor_i + \gamma' X_i + \epsilon_i, \tag{3.3}$$

where PriorDonor is the block variable (individual made a donation in the 12 months prior to the treatment=1, otherwise=0), and $'X_i$ is an N-by-k vector of pre-treatment covariates.

In what follows, we present ITT estimates based on both the unadjusted OLS estimator and the covariate-adjusted OLS estimator. As pre-registered, we will use the following covariates for adjustment: region, membership status (individual is a member of the party=1, otherwise=0), membership payment amount (Regular membership payment amount), past volunteering (Individual has volunteered for the party =1, otherwise=0), and fundraising (Individual has fundraised for the party =1, otherwise=0). We use two-tailed hypothesis tests throughout.

---

[7]We cannot estimate the Complier Average Causal Effect (CACE) because we do not have information on whether the subject received or read the email. While it is certainly possible to measure whether an email was received or opened, the party was not able to share this information with us.

3.6 Results

Evaluating the efficacy of email receipt (Hypothesis 1), Figure 3.2 visualises the probability of making a campaign donation among experimental subjects in each of the five conditions: control (no email), negative issue prime, negative party identity prime, positive issue prime, and positive party identity prime at time t+7 days (the pre-registered post-treatment measurement point). Did the email increase verifiable donations to the political party at the height of a general election campaign? Vis-à-vis those in the control condition, the propensity of making a donation is only significantly ($p < .1$) greater for one group: those exposed to a negative issue message. In substantive terms, those in receipt of a negative issue cue were 0.3 percentage-points more likely to make a donation relative to the control group. Given a baseline probability of 2%, a 0.3-point increase equates to an increase of 15% compared to those in receipt of no email (control). The negative issue message thus led to a sizeable increase in donations compared to not receiving an email at all. These results are not congruent with the pre-registered expectations of H1: email contact from the political party has a mobilising effect, but *only* when the email contains a negative issue cue.
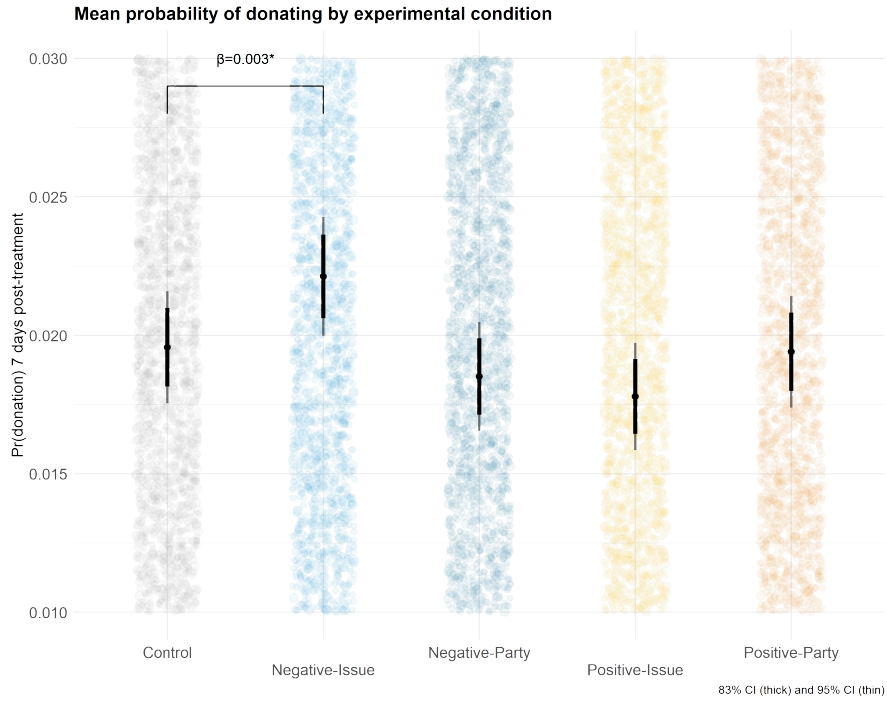
FIGURE 3.2: Mean probability of donating by experimental condition

Next we test whether emails containing negative identity primes are more effective at soliciting donations than emails containing positive identity primes (Hypothesis 2). To examine this, we pool the Negative-Issue and the Negative-Party conditions, and we pool the Positive-Issue and the Positive-Party conditions. The effects of the negative

identity primes are presented in Figure 3.3: the left-hand panel visualises the probability of making a donation among those in the control group as well as those in in receipt of either the negative and positive cue at t+7 days, and the right-hand panel reports the marginal effect of assignment to the negative cue vs. the positive cue across different post-treatment time bandwidths. Further results can be found in the Appendix, Tables B.3 and B.4.
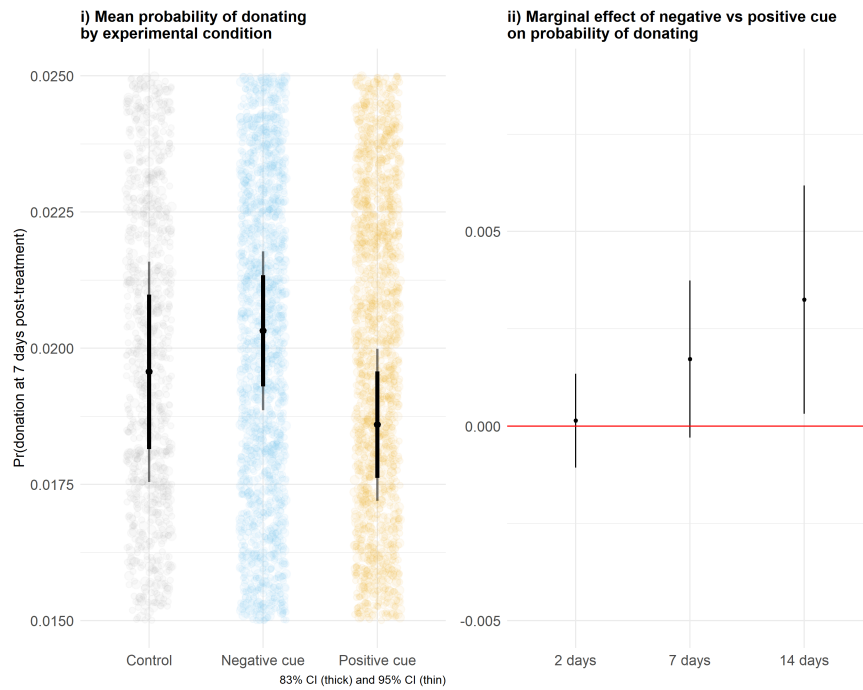
FIGURE 3.3: Marginal effect of negative vs positive cue on probability of donating at different post-time treatment points

Figure 3.3 shows that we cannot reject the null hypothesis that there is no difference in donation behaviour between the (combined) negative or positive emails and control conditions (Hypothesis 1). However, Figure 3.3 (panel ii) also indicates that receiving a negative identity cue *does* increase the probability of making a donation compared to the positive identity cue. This is consistent with H2: negative identity cues are more effective than positive identity cues in mobilising supporters. The negative cue also increases the number of times an individual donates, compared to the positive cue (see Figure B.1 (panel ii) in the Appendix). The effect of the negative vs positive identity cue on both number and frequency of donations strengthens over time: at 7 days post-treatment, the effect size is 0.002 while it increases to 0.003 at 14 days post-treatment. In substantive terms, the estimated effects amount to around 0.3-0.4 percentage-points over a baseline of 4%, which corresponds to a 10% increase in the number and frequency of donations. This is a sizeable increase in donations, given that we are estimating the effects of changing a few words in a single campaign email on high-cost political activity (actually spending money) in the run-up to a general election.

These results are significant, no matter whether we estimate the Intent-to-Treat Effects (ITT) with or without adjusting for pre-treatment covariates or using alternative model specifications (see Appendix Tables B.3, B.4, B.6, B.7). It is clear that negative cues are more effective than positive cues when it comes to whether an individual donates to the party, and how many times they do so. This provides strong evidence in favour of Hypothesis 2.

We also explore what type of party supporters are driving these results. Only individuals who have never donated before have a positive reaction to the negative issue identity cue (see Appendix, Figure B.7). Party supporters who have donated in the past do not show a significant change in their donation behaviour when exposed to the negative identity cue emails (see Appendix, Figure B.7).

We now turn to evaluate the effect of partisan versus issue cues (Hypothesis 3). Overall, subjects who received an email that used issue-based cues were no more likely to donate than subjects who received the party-based cue, or subjects who were assigned to control. This (null) finding is consistent across all models and is not sensitive to the inclusion or exclusion of pre-treatment covariates. We therefore do not find support for Hypothesis 3. However, as pre-registered, we test whether negative and positive identity messages are more effective when paired with an issue or a traditional party-based identity cue (Hypothesis 4). Figure 3.4 visualises the marginal effect of the negative cue when combined with an issue- or party-based cue.
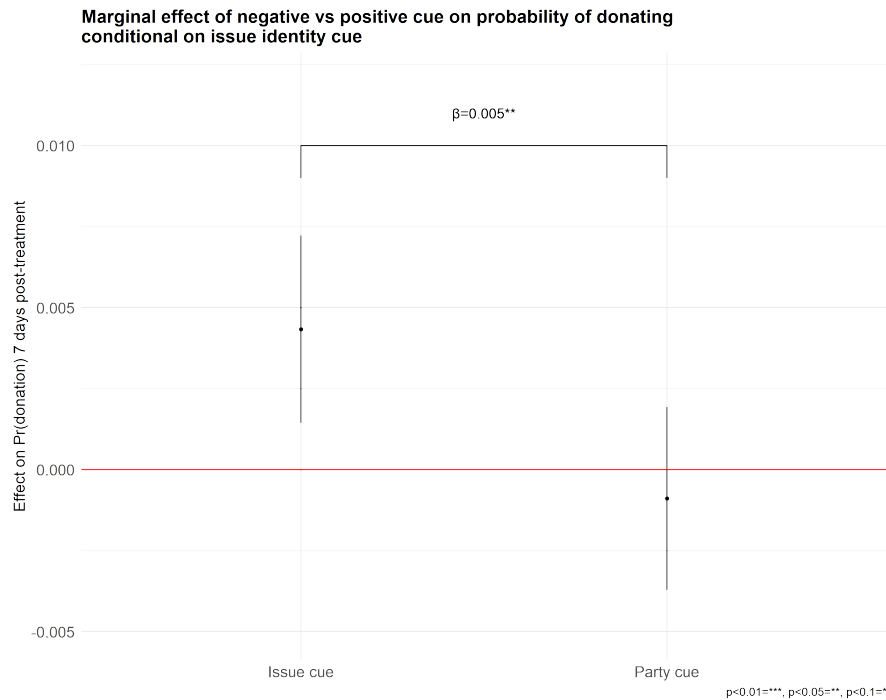
FIGURE 3.4: ITT of negative vs positive cue conditional on issue cue

We find that negative prompts are significantly more effective when accompanied by an issue cue (CITT=.005) as opposed to a party identity cue (CITT=-.001). This provides evidence in favour of Hypothesis 4. Whilst a negative party identity cue exhibits no incremental effect on donations, a negative issue cue results in a significant 0.5 percentage-point increase in the probability of making a donation. Given a baseline donation rate of 2% this 0.5-point increase is not trivial, equating to a rise of 25% relative to the positive issue cue condition. Negative issue cues are also more effective than negative party identity cues at increasing the number of times an individual donates (see Appendix, Figure B.1, panel iii), also see Models 2 & 4, Table B.4). However, it is important to note that the interaction effects between the negative and the issue cue are only present at 7 days after the treatment email is sent. At 14 days post-treatment, the interaction effect is substantively smaller and no longer significant (see Tables B.9 and B.10). This seems to be driven by the negative party identity cue "catching up" with the effectiveness of the negative issue cue. In the second week after treatment, the negative party cue (compared to the Control and the positive party cue) becomes slightly more effective, while the negative issue identity cue continues to have a positive effect on the probability of donating (see Appendix, Figure B.5). Still, the negative issue cue is the single most effective campaign email when it comes to increasing the number and frequency of donations. Campaign messages that prime negative political identity are only effective at encouraging donations when they are combined with an issue message than with a standard partisan identity message.

Finally, and in addition to the propensity to donate and the overall number of donations from an individual, we test for the effect of the treatment on donation amount.[8] The average donation amount is statistically indistinguishable when we compare treatment groups to the control group (see Appendix, Figure B.2). In other words, assignment to different treatment conditions did not significantly increase the overall donation amount.[9] We explore the implications of these effects on donation amount, compared to the propensity to donate, in the conclusion.

3.7 Conclusion

Results from a pre-registered large-scale digital field experiment show that negative identity cues are more effective than positive identity cues in encouraging party activists to

---

[8] As donation amount is highly left-skewed, we present results for the log-transformed donation amount ($=\log(\text{donationamount}+1)$). Alternative models showing results for the untransformed outcome can be found in the Appendix, Table B.8.

[9] Substantively, individuals exposed to the negative issue identity message donate higher amounts than individuals assigned to the Control (CITT=0.008) but this difference does not reach conventional significance levels. See Appendix Figure B.2 for more details.

take costly political action. Consistent with our expectations, being assigned a campaign email with a negative identity message compared to a positive identity message makes party activists more likely to donate, as well as to donate more often to a political party. We also find that negative cues have a greater effect on donations when they are paired with an issue cue than when paired with a party-identity. A campaign email rallying party supporters around a disliked issue out-group increases the number and frequency of donations compared to not sending an email.

However, an email containing any negative cue is no more effective than sending no email in soliciting donations: messages containing out-party hostility do not increase donations compared to not sending an email. Furthermore, while negative identity cues increase donation behaviour in the short-term, they do not raise more money for the political party compared to not sending an email at all. We also find that, in general, priming issue-based identities is no more effective than priming partisan identities: Only messages that convey a shared sense of dislike towards an issue-based out-group are effective in increasing the number of donations – messages priming a shared sense of belonging to an issue-based in-group are not.

This paper offers robust, real-world evidence that the activation of negative political identities through political campaigns has consequences for costly political behaviour that are different from those of activating positive political identities. Our study also highlights the importance of including issue-based identities in accounts of political mobilisation, and advances our understanding of when issue-based identities come to shape individual-level decisions to take political action.

Our study makes several contributions to the literatures on negative partisanship, affective polarisation and negative campaigning. The majority of studies on the implications of negative partisanship for political participation are observational. This makes it difficult to disentangle whether negative political identities lead to greater mobilisation or vice-versa. Other studies on this topic have used lab or survey experiments to address this issue but can only test for the short-term effects of the activation of a particular identity on expressed attitudes or preferences. Our study makes a unique and important empirical contribution to this literature. We conduct a field experiment in collaboration with a political party to assess what the effects of negative and positive identity activation on actual donation behaviour are. The results of this experiment show that the activation of negative identities, compared to positive identities, has observable effects on costly real-world behaviour within a naturalistic, unobtrusive study setting. Our results provide evidence that the activation of negative identities has different effects from activating positive identities when it comes to the solicitation of donations from party

supporters. This study thus fills an important gap in our understanding of negative partisanship and its implications for behaviour.

Building on recent innovations in the affective polarisation literature (Hobolt, Leeper and Tilley, 2020), our findings also speak to the relevance of issue-based identities, compared to traditional party identities, in structuring political behaviour. While most previous work on negative affect towards political out-groups is focused on negative partisanship, we show that negative issue primes can have stronger effects than priming traditional negative partisan identities. This advances our understanding of negative political identities, and shows that strong negative feelings towards out-groups are not limited to hating another political party. Disliking an issue-based out-group, when activated, can have powerful effects on behaviour. This paper also contributes to the literature on issue identities by clarifying under which conditions appeals to issue-based identities translate into political action. While positive issue cues failed to mobilise party supporters, negative issue cues increased political mobilisation. This could also have broader implications for how we understand issue identities. As they often tap into intense political conflict that suddenly materialises, issue identities might be built more heavily on negative out-group affect than partisan identities that evolve through slow socialisation processes.

Finally, by focusing on campaign messages that invoke negative identities, beyond issue, trait or competence attacks, this study also contributes to the negative campaigning literature. Because negative political identities relate to deep and highly emotive tensions between in- and out-groups, they provide a powerful motivation for behaviour in polarised societies. Attacks against political opponents are not just confined to their policy or qualifications, but can go to the core of "who we are" and "who we are not". Bringing the literature on negative identities into the negative campaigning literature therefore offers a fruitful addition to research on campaigning. We also empirically expand the negative campaigning literature by moving our focus from the electorate at large to party supporters, who should be most susceptible to respond to negative identity cues.

The effects of negative identity campaigning on donations that this paper documents also have implications for larger questions of societal cohesion and democratic norms. Elite communications that reinforce negative political out-group sentiments could further entrench political divides, and strengthen affective polarisation. Affective polarisation has been linked with worrisome implications for democratic functioning, ranging from reduced support for democratic norms (Kingzette et al., 2021) to an increased potential for political violence (Mason and Kalmoe, 2022). The use of negative identity campaigning may therefore harm the institutions that political parties compete and operate in. Yet,

its use will likely proliferate as long as negative identity cueing appears as an effective fundraising strategy to political parties.

On this consideration, our paper shows that negative identity messaging is not necessarily a winning strategy for political parties. The negative identity cue, compared to the positive cue, made activists more likely to donate and to donate more often. However, only the email that contained the negative issue-based cue positively affected weather an individual donated and frequency of donations versus the control group. Moreover, although the coefficient size is positive, we cannot reject that the overall amount donated was equal to the amount in the control group. This finding is in line with Hersch (2020)'s argument that activists often engage in expressive, symbolic behaviour online. They might respond in the heat of the moment to a fundraising email, but not increase their overall donation amount over and above what it would have been had they not received the appeal. At the same time, many campaigns and parties care about grassroots support, growing their donor base and being able to show that their donations come from many different individuals. Negative identity cue messages, compared to positive identity cue messages, were effective at increasing the donor pool, and activating a larger number of supporters.

However, this grassroots mobilisation also comes at a potential cost to political parties: as our exploratory analyses show, supporters' reaction to negative identity campaigning depends on their pre-treatment donation behaviour. Negative identity campaigning can spur inactive supporters into taking political action for the first time. In contrast, individuals who have previously donated to the party did not increase their probability of donating, frequency of donating or their donation amount in response to the negative issue cue message. This shows that negative identity campaigning is not a uniformly effective strategy and may fail to mobilise highly committed, invested party supporters. Taken together with the potentially harmful consequences for democracy at large, political parties would be well-advised to consider the trade-offs and draw-backs involved in using negative identity campaigning. Alternative campaign strategies beyond "going negative" may prove a more effective fundraising strategy for highly committed supporters, and may also have positive knock-on effects for reducing affective polarisation, and therefore strengthening democratic functioning in polarised societies.

Chapter 4

Gender-immigration messages: How women's rights are used to normalise anti-immigration views

Abstract

How do extreme political parties normalise unacceptable views? I argue that when previously unacceptable positions are "normatively repackaged" as a defence of liberal democratic values, voters will find the position more acceptable. I examine this at the example of gender-immigration messages: statements using women's rights and gender equality to justify anti-immigration claims. I argue that gender-immigration messages make anti-immigration views and, by extension, radical right-wing parties more acceptable. To test this, I conduct survey experiments in Norway and Germany, varying whether respondents are exposed to a gender-immigration message, an immigration message, a gender message or no message. I find that normative repackaging increases the acceptability of previously unacceptable views among all voters (in Norway), and among women, compared to men (in Germany). However, I find no effects of gender-immigration messages on radical right support or the direct expression of anti-immigrant views. These findings have important implications for our understanding of how unacceptable views become normalised, and how normative repackaging is used as a powerful legitimising device to this end.

4.1 Introduction

Populist radical right-wing politicians and parties[1] have become influential political players in many Western democracies. This has polarised voters (Bischof and Wagner, 2019) and led to the adoption of more anti-immigrant positions among mainstream competitors (Abou-Chadi and Krause, 2020), among other wide-ranging implications for democracy (Mudde, 2019; Valentim, 2021). Hence, understanding how populist radical right-wing parties (PRRPs) and their anti-immigration positions become normalised is important. Most mainstreaming research to date examines how PRRPs de-toxify their overall organisational reputation, such as by foregoing links with extremist organisations, diversifying their policy portfolio or candidate base (Akkerman and de Lange, 2012; Harteveld and Ivarsflaten, 2016; Ben-Shitrit, Elad-Strenger and Hirsch-Hoefler, 2022). However, we know less about how PRRPs mainstream one of their biggest political trademarks: their issue positions. As issue entrepreneurs (Vries and Hobolt, 2012), PRRPs have mobilised certain voters successfully around their staunch anti-immigration positions (Ivarsflaten, 2007). Simply moderating their immigration positions would therefore be electorally costly for PRRPs and would risk losing their core supporters (Wagner, 2012). Yet, more moderate voters might be concerned that these strong anti-immigration positions are racist and unacceptable because they violate societal anti-prejudice norms (Blinder, Ford and Ivarsflaten, 2013). To appeal to a wider electorate, PRRPs not only need to convince voters that their issue positions are important, but also that they are acceptable. How do PRRPs make their issue positions seem mainstream and acceptable without moderating them?

In this paper, I examine how anti-immigration positions are rhetorically framed to feel more acceptable and mainstream to voters. I argue that when taboo issue positions are "normatively repackaged" as a defence of liberal democratic values, like women's rights, voters will find the issue position more acceptable. There is a lot of anecdotal evidence that populist radical right-wing parties and politicians use gender equality and women's rights to defend tougher immigration policies (Akkerman, 2015; Mudde and Kaltwasser, 2015; De Lange and Mügge, 2015; Dancygier, 2020). However, to the author's best knowledge, there is no research to date that causally identifies what this "normatively repackaged" anti-immigration rhetoric does to voters. This paper fills this gap by conducting two large-scale, pre-registered survey experiments. In these, I test the effects of "gender-immigration messages": anti-immigration statements with reference to gender equality and women's rights. For example, Donald Trump justified controversial immigration policies, like the "extreme vetting" of immigrants, with the

---

[1]Populist radical right-wing parties are defined as parties which share a core ideology of nativism, authoritarianism, and populism (Mudde, 2007).

idea that there is "oppression of women and gays in many Muslim nations" and that his Administration "will speak out against the oppression of women, gays and people of different faith" against "radical Islam" (Trump quoted in Berenson (2016)). In a similar vein, European populist radical right-wing politicians have suggested that "the migrant crisis signals the beginning of the end of women's rights"[2], that "the patriarchal system of Islam is the greatest threat to the rights and integrity of women today"[3] and that tougher integration policies are needed because "oppressive values towards women do not belong here"[4]. Because references to gender equality and women's rights signal an adherence to highly acceptable, shared societal norms, they provide an effective shield against accusations of unacceptable expressions of racism and xenophobia.

Gender-immigration messages should therefore make voters more likely to find anti-immigration views acceptable. By extension, gender-immigration messages should also make the expression of anti-immigration positions and support for the radical right more likely. I also expect that gender-immigration messages are more effective for women who have been shown to be more concerned than men about not breaking societal taboos and anti-prejudice norms (Harteveld and Ivarsflaten, 2016; Oshri et al., 2022).

To test the effects of gender-immigration messages, I conduct pre-registered survey experiments on large, nationally representative samples in two European countries: Norway (Study 1) and Germany (Study 2). These are ideal study settings as immigration has become a salient political issue and populist radical right-wing parties use gender-immigration messages in both countries. At the same time, the two cases also vary in populist radical right-wing acceptability, with the Norwegian radical right-wing Progress Party being more acceptable than the German radical right-wing party AfD. To identify the effects of normative repackaging on voter attitudes, respondents were randomly assigned to read either a gender-immigration message, an immigration message, a gender message, or no message.

I find evidence that gender-immigration messages make anti-immigration views more acceptable among all voters (Study 1) and among women, compared to men, in Study 2. I do not find consistent effects of gender-immigration messages on radical right support or the direct expression of anti-immigrant views. I also test whether manipulating the acceptability of anti-immigration views produces similar effects to gender-immigration messages. I find evidence that this is the case. This provides further support for my

---

[2]The leader of the French populist radical right-wing party, Marine Le Pen said this when proposing an immigration referendum in an opinion piece for the French newspaper L'opinion (Le Pen, 2016).

[3]The German populist radical right-wing party AfD made this statement during the International Day of Violence against Women (AfD, 2020).

[4]The Norwegian Progress Party made this statement when introducing their integration and immigration policy (FrP, 2020).

theory that gendered immigration statements affect voter preferences by changing how acceptable anti-immigration positions are perceived to be.

This paper makes three main contributions. Firstly, this paper contributes to the literature on populist radical right-wing parties. While previous research explores important questions on the consequences of either populist radical-right wing presence or absence in political institutions on voters and parties (see e.g., Valentim 2021; Bischof and Wagner 2019; Abou-Chadi and Krause 2020), this paper offers a fine-grained insight into what specific types of populist radical right immigration rhetoric do to voters. This paper also extends our understanding of populist radical right-wing parties as issue entrepreneurs (de Vries and Hobolt, 2020) to norm entrepreneurs: populist radical right-wing parties not only politicise immigration, but also make anti-immigration views acceptable through how they present these views.

Secondly, this paper contributes to the growing literature on gender and the radical right. Most previous research documents the overwhelming absence of women from populist radical right-wing parties (Coffé, 2018). However, more recent work shows that populist radical right-wing parties use female candidates strategically (Weeks et al., 2022), that these parties appear more acceptable when they have more women candidates and activists (Ben-Shitrit, Elad-Strenger and Hirsch-Hoefler, 2022), and that the populist radical right has paid increasing attention to gender issues (Abou-Chadi, Breyer and Gessler, 2021). Building on this, I show that gendered political rhetoric is an important tool used to legitimise anti-immigration positions.

Third, this paper makes a more general contribution to the literature on party competition and extreme party strategies. Extreme political parties face a dilemma when they try to appeal to a wider electorate. Keeping their current position alienates more moderate voters, but policy moderation would be punished by their core voters (Wagner, 2012). Mainstreaming research has focused on strategies by which extreme parties generally improve their image (Akkerman and de Lange, 2012). Yet, for political parties whose main currency are their issue positions, making their core positions feel acceptable is important. This paper thus extends the literature on social stigma and party reputations (Harteveld and Ivarsflaten, 2016; Harteveld, Dahlberg, Kokkonen and Brug, 2019) by focusing on framing as a mainstreaming strategy. Using framing instead of policy moderation allows extreme parties to keep and normalise their extreme policy positions with little to no cost for themselves. Understanding if gender-immigration messages are effective therefore also contributes to our understanding of larger societal questions such as how anti-immigrant views and parties become normalised. "Normative repackaging" might also help us understand how other previously unacceptable positions become seen as more mainstream and acceptable.

4.2 Social stigma and mainstreaming the populist radical right

Perceptions about what is acceptable and what is not matter for political behaviour. Social norms, the "shared understandings about actions that are obligatory, permitted, or forbidden" (Ostrom, 2000, p. 143), have been shown to have consequences for behaviour in- and outside the political realm (Bursztyn and Jensen, 2017). Whether a policy position or party is seen as acceptable can have important implications for voters' willingness to support these positions and parties. Societally in-grained anti-prejudice norms, the "better angels of our nature", signal that blatant expressions of racism and prejudice are unacceptable (Blinder, Ford and Ivarsflaten, 2013).

Radical right-wing parties often openly defy anti-prejudice norms which makes them look a less acceptable political option in the eyes of more moderate voters. Experimental and observational evidence shows that this lower acceptability hurts radical right-wing parties electorally (Harteveld, Dahlberg, Kokkonen and Brug, 2019; Harteveld, Dahlberg, Kokkonen and van der Brug, 2019b). Radical right-wing parties which are seen as less acceptable also find it more difficult to attract voters, even those voters who agree with the party's policy positions (Harteveld, Dahlberg, Kokkonen and Brug, 2019; Spanje and Azrout, 2019). This also affects popular support for their policies: voters are reluctant to support policies that are put forward by an extreme party compared to a party with a more mainstream reputation (Blinder, Ford and Ivarsflaten, 2013; Bolin, Dahlberg and Blombäck, 2022).

In turn, voters are more willing to support PRRPs when they have a "reputational shield" that makes them appear more distant from racism and extremism (Harteveld and Ivarsflaten, 2016). The ability of PRRPs to present themselves as acceptable political parties is a key factor explaining electoral breakthroughs for the radical right (Mendes and Dennison, 2020). Populist radical right-wing parties thus have a strong incentive to present themselves as acceptable and socially appropriate.

Yet, not all PRRPs are stigmatised to the same extent. Previous research has investigated why some PRRPs are perceived as more taboo than others, and how stigma changes over time. External factors, such as how the media portrays the radical right (Spanje and Azrout, 2019), how other political parties engage with it (van Heerden and van der Brug, 2017; Bale and Kaltwasser, 2021), and whether the radical right obtains institutional representation (Valentim, 2021) have been shown to affect the radical right's acceptability. Rather than hoping for external shifts in reputation, PRRPs also engage in "mainstreaming" strategies to improve their image (Akkerman and de Lange, 2012).

Most mainstreaming research to date examines how PRRPs de-toxify their overall reputation through organisational changes. Mainstreaming often involves outward-facing organisational change, including breaking links with extremist organisations and partners, as well as collaborating with mainstream political parties (McDonnell and Werner, 2018). Internal organisational changes, such as replacing stigmatised party leaders, sanctioning racist and anti-Semitic language and diversifying the candidate portfolio are also effective mainstreaming strategies (Mayer, 2013; Ivaldi, 2016; Ben-Shitrit, Elad-Strenger and Hirsch-Hoefler, 2022).

While there is a rich literature explaining how PRRPs improve their overall reputation through organisational changes, we know less about how PRRPs mainstream one of their biggest political trademarks: their issue positions. This is important because the social stigma attached to the radical right is directly related to their core issue positions (Harteveld, Dahlberg, Kokkonen and van der Brug, 2019*b*). The strong anti-immigration rhetoric that many PRRPs espouse goes against anti-prejudice norms and makes these parties open to allegations of racism and xenophobia (Harteveld, Dahlberg, Kokkonen and van der Brug, 2019*b*; Blinder, Ford and Ivarsflaten, 2013). To appeal to a wider electorate, PRRPs need to make their anti-immigration positions feel more acceptable. Yet, mainstreaming issue positions presents a dilemma for PRRPs. As issue entrepreneurs (Vries and Hobolt, 2012), PRRPs have politicised immigration (Hutter and Kriesi, 2021) and mobilised certain voters successfully around their staunch anti-immigration positions (Ivarsflaten, 2007). Keeping their current position alienates more moderate voters, but moderating their issue positions is also an unviable strategy for PRRPs. This is because policy moderation would likely be punished by their core voters (Adams et al., 2006; Wagner, 2012).

To address this puzzle, I build a theoretical argument bridging the literature on mainstreaming, the literature on framing and elite rhetoric, as well as the literature on gender and the radical right. Instead of adopting a more centrist position, I argue that radical right-wing parties repackage anti-immigration views as a defence of liberal democratic values, like women's rights, to make them feel and appear more acceptable. In doing so, I also advance a more general theory about "normative repackaging", a strategy that political actors can use to make a previously unacceptable position more acceptable and palatable to a wider range of voters.

4.3 Framing immigration

How policy positions are presented affects our understanding and support for them. A rich literature on framing effects and political communication has shown that frames

in communication, understood as "the key considerations emphasised in a speech act" (Chong and Druckman, 2007, p. 106), can influence frames in thought, "the set of dimensions that affect an individual's evaluation" (Chong and Druckman, 2007, p. 106) which can in turn shape an individual's opinions and political behaviour. For example, the seminal studies on framing effects show that voters are more likely to give permission to a Ku Klux Klan rally when this is framed as freedom of speech issue than when the rally is framed as a security threat (Nelson, Clawson and Oxley, 1997).

Research has shown that this logic also applies to immigration rhetoric. Immigration positions and policies can be framed in different ways by political actors, and this affects voters' political preferences. For example, negatively valenced immigration messages have been found to increase the expression of anti-immigration attitudes (Brader, Valentino and Suhay, 2008; Avdagic and Savage, 2021). Similarly, negative media portrayals of immigrants are linked to increased prejudice expression among voters (Eberl et al., 2018) and elite anti-immigration messaging is related to voters voicing more restrictive immigration attitudes (Dekeyser and Freedman, 2021). In addition to valence, different types of immigration frames have been found to have distinct effects on voter attitudes and support for immigration policies (Simonsen and Bonikowski, 2022; Bos et al., 2016; Merolla, Ramakrishnan and Haynes, 2013). Group cues in immigration messaging also matter: cues about the immigrant's country of origin and socio-economic status affect public support for immigration (Brader, Valentino and Suhay, 2008; Valentino et al., 2019).

Linking this to the radical right, research has found that PRRPs use framing strategies when they communicate their immigration positions. Foundational research on PRRPs argues that the use of new "master frames" on immigration was crucial for PRRPs' early electoral breakthroughs (Rydgren, 2005). Instead of relying on "biological racism", deemed unacceptable after the Second World War, PRRPs used "cultural racism" to mobilise voters on their anti-immigrant positions without being stigmatised as anti-democratic or racist (Rydgren, 2005). More recent work shows that PRRPs dynamically adapt their framing of core policy messages to target different audiences (Borbáth and Gessler, 2021). Building on this, I suggest that framing can be a powerful mainstreaming strategy. How issue positions are presented can affect how acceptable and mainstream they are perceived to be.

4.4 Gender-immigration messages as a normative repackaging strategy

Bridging the literatures on framing and mainstreaming, I explore how framing is used to mainstream core radical right issue positions. In particular, I examine what happens when anti-immigration positions are presented as a defence of gender equality and women's rights. I expect that "gender-immigration messages" make anti-immigration positions appear more acceptable and mainstream. Criticising immigration under the guise of civil liberties and equality, such as women's rights, makes anti-immigration positions seem compatible with liberal democratic values and therefore weakens anti-prejudice norms.

I focus on gender-immigration messages for three reasons: 1. Gendered approaches to the radical right have traditionally been understudied. By focusing on gendered political rhetoric, this paper contributes to closing a crucial gap in our understanding of PRRPs. Far from being absent from these parties, I argue that gender equality rhetoric is a key tool that allows PRRPs to mainstream their issue positions. 2. Invoking gender equality has been shown to be a particularly powerful way of obtaining reputational boosts, so should be an effective mainstreaming strategy. 3. At the same time, gender equality is at odds with radical right ideology (Akkerman, 2015; Campbell and Erzeel, 2018) - it is thus interesting and important to study if radical right-wing parties can benefit from discursively invoking such a value incongruent message.

In this paper, I study the effects of:

> "gender-immigration messages": anti-immigration statements with reference to women's rights and gender equality

A prime example of this is Marine Le Pen saying in 2016 "I am worried that the migrant crisis signals the beginning of the end of women's rights" (Marine Le Pen, 2016 quoted in Lang, 2016). Many PRRPs have started using gendered anti-immigration rhetoric in recent years (Akkerman, 2015; Mudde and Kaltwasser, 2015). This fits into a larger trend of radical right-wing parties repoliticising gender issues, and talking increasingly about them on social media (Abou-Chadi, Breyer and Gessler, 2021). For example, the National Rally (formerly National Front) in France and the FPO in Austria have gradually appropriated discourse on women's rights and feminist slogans for their discourse on Islam and immigration (Scrinzi, 2017*a*). A parallel trend has been observed in Sweden (Norocel, 2016) and the Netherlands (De Lange and Mügge, 2015). The Dutch PVV has started using gender equality and women's rights as way of discrediting

the "islamisation" of the Netherlands (Mudde and Kaltwasser, 2015) amounting to its leader Geert Wilders announcing that immigration "flushes several decades of women's rights down the toilet" (Wilders 2010 cited in Mudde and Kaltwasser 2015). This "femonationalism", the linking of women's rights to authoritarian and nativist positions, does not evidence a profound ideological commitment to gender equality (Farris, 2017). Most populist radical right-wing parties in Western Europe still espouse a decidedly conservative stance on gender and family values, even though there is increasing variation within this party camp (Campbell and Erzeel, 2018). At the same time as invoking gender equality and women's rights to defend their anti-immigration positions, many PRRPs openly position themselves against progressive gender equality policies, such as gender quotas (Dancygier, 2020).

Rather, PRRPs use gender equality and women's rights as a highly accepted normative frame to legitimate racist and xenophobic claims. Through this "racialisation of sexism" (Scrinzi, 2017*b*), gender inequality is only portrayed as a problem that emanates from ethnic or religious minority groups. In a similar vein, PRRPs have also started using gay rights to defend anti-immigration positions, claiming that homosexuals are particularly threatened by Muslim immigrants (Spierings, 2021).

I suggest that gender-immigration messages are effective because they fit into a larger strategy of what I call "normative repackaging": presenting a policy position as a defence of liberal democratic values without changing the underlying policy position. Invoking women's rights should be a particularly powerful normative repackaging strategy because this plays towards a widely accepted cultural norm of gender equality in Western societies (Farris, 2017). Frames using references to deeply rooted cultural norms tend to have a greater impact on voters' preferences (Gamson and Modigliani, 1987). Invoking gender equality claims, such as descriptive representation, can lead to reputational boosts, even for institutions or actors who are themselves deeply unequal (Bush and Zetterberg, 2021).

Linking this to the radical right, research has found that PRRPs appear more acceptable when they have more women candidates and activists (Ben-Shitrit, Elad-Strenger and Hirsch-Hoefler, 2022). Pairing an anti-immigration position with women's rights and gender equality signals to voters that anti-immigration positions fit comfortably within a liberal democratic framework and are not an expression of racism or prejudice. Gender-immigration messages should therefore weaken anti-prejudice norms and make voters feel that it is more acceptable to say something negative about immigrants.

H1: Gender-immigration messages increase the acceptability of anti-immigration views.

Even voters who harbour anti-immigration positions might be uncomfortable openly expressing them because they are worried about the social stigma attached to these positions (Blinder, Ford and Ivarsflaten, 2013). By presenting an anti-immigration statement not as an exclusionist or racist value appeal but as a defence of liberal democratic norms, gender-immigration messages dispel concerns over how taboo it is to express anti-immigration views. Because gender-immigration make anti-immigration positions seem more acceptable, they should also make the expression of those positions more likely.

H2: Gender-immigration messages increase opposition to immigration.

Gender-immigration messages should also give populist radical right-wing parties a softer, more moderate and electable image by making one of their core "taboo" issue positions seem more acceptable. By normalising anti-immigration views, gender-immigration frames dilute social desirability concerns individuals might have when considering whether to vote for a PRRP.

H3: Gender-immigration messages increase support for populist radical right-wing parties.

4.5 Women, social stigma and the effects of gender-immigration messages

Considerations about what is taboo and what is acceptable do not affect all voters to the same extent. If gender-immigration messages mainly work because they de-stigmatise anti-immigration views, gender-immigration messages should be most effective for individuals susceptible to worry about social stigma. Anti-prejudice norms, which sanction the expression of prejudice, have been shown to affect men and women differently. While women are no less anti-immigrant than men, they care more about not appearing racist or prejudiced (Harteveld and Ivarsflaten, 2016). Because women have a higher "motivation to control prejudice", they are less likely to feel comfortable supporting a radical right-wing party with a taboo, toxic reputation (Harteveld and Ivarsflaten, 2016). Building on this, gender-immigration messages should have a greater effect on women than men because they make anti-immigration views seem more acceptable and mainstream and should therefore lower women's concerns about breaking anti-prejudice norms.

In regards to the expression of anti-immigration views, women and men hold remarkably similar immigration attitudes (Harteveld et al., 2015). Still, gender-immigration messages should increase women's anti-immigration views more so than men's, as they

tap into ideational stereotypes against Muslim immigrants. Stereotypes about Muslims' regressive gender norms have been shown to influence discriminatory behaviour towards Muslim immigrants in Germany (Choi, Poertner and Sambanis, 2021). Non-immigrant women (but not men) stop discriminating when they interact with a Muslim immigrant who defies these stereotypes and holds progressive views on gender issues (Choi, Poertner and Sambanis, 2021). Vice versa, concerns about gender equality in relation to immigration might encourage women to more readily express anti-immigrant views. In line with this, recent research has argued that there are "sexually modern nativist" radical right-wing voters– voters who hold restrictive immigration preferences, but simultaneously progressive preferences on gender equality and gay rights (Spierings, Lubbers and Zaslove, 2017; Lancaster, 2019). There is also evidence that the combination of anti-immigration attitudes and progressive gender equality attitudes in the electorate is becoming more common (Lancaster, 2022). Sexually modern nativists tend to be younger, female and more educated (Lancaster, 2019). For women, anti-immigrant beliefs seem to be more readily connected to gender equality concerns than for men.

Finally, in terms of voting for the radical right, there is evidence that men and women respond differently to considerations about social stigma and electoral risk. A vote for the radical right might be perceived as risky because these parties are deemed extreme and less socially acceptable (Oshri et al., 2022). Because women are on average more risk averse than men, they are also less likely to vote for the radical right (Oshri et al., 2022). This patterns also extends to other political parties - women are on average less likely than men to vote for ideologically extreme or socially sanctioned parties (Harteveld, Dahlberg, Kokkonen and Van Der Brug, 2019a). Women also often subjectively feel that their policy preferences are further removed from PRRPs' policy positions than men, even when women's and PRRPs' policy preferences coincide (Harteveld, 2016).

In turn, women feel more comfortable expressing support for the radical right when these parties have a more acceptable, mainstream reputation (Harteveld and Ivarsflaten, 2016). By making the core issue position of the populist radical right more acceptable, gender-immigration messages should also decrease the perceived stigma attached to the radical right to a greater extent for women, than for men. This also echoes other gendered strategies that PRRPs use to attract more female voters, such as increasing the number of female candidates (Weeks et al., 2022). I therefore expect that:

H1A- H3A: This effect is stronger for women than for men.

4.6 Survey experiments

4.6.1 Rationale

Pre-registered[5] survey experiments were carried out to test what effects normative repackaging has on voter preferences. This was examined at the example of gender-immigration messages, and their implications for the acceptability and expression of anti-immigration views.[6] The experiments are essential in addressing identification challenges. Using observational data is problematic because voters self-select into exposure towards gendered immigration narratives. Individuals who are already more anti-immigrant and more likely to vote for the radical right may also be more likely to seek out content in the news or on social media that contains gender-immigration messages.

4.6.2 Context

Study 1 was carried out in Norway. Immigration has become a salient issue in Norway, and its populist radical right-wing party, the Progress Party (FrP) uses gender-immigration messages. For example, the FrP leader until 2021, Siv Jensen, coined the term "sneak-islamisation" of Norway, drawing frequent links between Muslim immigrants and sexual violence as well as the violation of women's rights.[7] This renders Norway an interesting and an ecologically valid case for studying the effects of gender-immigration messages. At the same time, the FrP is already a more socially acceptable populist radical right-wing party which has participated in coalition governments (Harteveld and Ivarsflaten, 2016).

Study 2 was carried out in Germany, which showcases one of the most recent examples of a populist radical right party using gender-immigration messages. The Alternative for Germany (AfD) was founded in 2013 and has since had considerable electoral successes in European, state and federal elections in Germany. It gained 7.1% of the vote in the European elections 2014, and by 2017 was represented in 14 out of the 16 federal states in Germany. It received 12.6% of the vote share in the 2017 election and 10.3% in the 2021 election to the German Bundestag. While the party programme mainly endorsed Eurosceptic positions when it was founded, the party has since shifted and made anti-immigration and anti-Islam positions its core ideological markers (Arzheimer

---

[5]Both studies were pre-registered and PAPs are available on the EGAP and OSF registry, `https://doi.org/10.17605/OSF.IO/HNMSK` and `https://doi.org/10.17605/OSF.IO/WHZEF`

[6]The survey experiment conducted in Norway was granted Ethics approval by the Ethics Committee at the LSE, reference number 1111. Study 2 conducted in Germany was granted Ethics approval by the Ethics Committee at the LSE, reference number 19438.

[7]http://news.bbc.co.uk/1/hi/world/europe/8008364.stm

and Berning, 2019). The AfD has therefore been classified as a populist radical right-wing party, endorsing populist, nativist and authoritarian positions (Rooduijn et al., 2019).

There is a long-standing societal taboo around radical right-wing parties in Germany, and mainstream parties have not entered coalition governments with the AfD so far (Art, 2018). Since the reports of sexual assault during the 2016 Cologne New Year's Eve celebrations, the AfD has made frequent links between anti-Islam, anti-immigrant statements and its concern for gender issues, particularly on its social media outlets (Gopffarth, 2018). In a position paper on Islam from 2016, the AfD dedicated an entire section to women's role in Islam, stressing the incompatibility between Islam and gender equality.[8] Statements like "We won't sacrifice our women's rights to Islam" are emblematic of how the AfD uses women's rights to defend anti-immigrant, anti-Muslim positions.[9] In response to the global #MeToo movement and the Women's march, the AfD organised its own "Women's march" demanding that "borders are closed immediately" to protect women from sexual violence.[10] The populist radical right in Germany thus frequently uses gender-immigration messages, making it an externally valid and interesting case for studying the effects of these messages.

At the same time, the AfD is still a fairly new political party in Germany and perceived as more taboo. It has not been able to join state or national coalition governments, and the biggest mainstream political parties (SPD and CDU/CSU) are officially opposed to entering any such agreements with the AfD. Furthermore, the gender gap in support for the AfD is still large and negative, with more men than women supporting the AfD in the 2017 and the 2021 election.

Overall, two different contexts - Germany and Norway - were chosen to test for the effects of gender-immigration messages on voter attitudes. Both cases have in common that they feature political parties which use gender-immigration messages. This makes them externally valid testing grounds for normative repackaging theory. Featuring political parties which use these messages is also important for generating valid treatment stimuli, as discussed in the next section. However, the two cases also meaningfully vary in regards to radical right acceptability. Choosing two cases which vary on this dimension is advantageous because it allows me to test whether normative repackaging works under different external conditions. If I find evidence that normative repackaging works across

---

[8] "Themenflyer Islam", AfD, 2016. Available online at: `https://afd.berlin/wp-content/uploads/2016/11/16-08-11_Themenflyer-Islam_BGS.pdf`

[9] Statement made on the AfD's website for the International Day of Violence Against Women, available online at: `https://afdkompakt.de/2020/11/25/wir-opfern-unsere-frauenrechte-nicht-dem-islam/`

[10] `https://www.n-tv.de/politik/Frauenmarsch-stoesst-auf-Widerstand-article20291938.html`

these two cases, this should provide strong evidence that this strategy would also work elsewhere.

4.6.3 Data

Study 1 consisted of an online survey experiment fielded as a module as part of the 2020 round of the Norwegian Citizen Panel from June 2nd - June 29th, 2020. A representative sample of Norwegian adults participated in the survey experiment (N=2820).[11] The Norwegian Citizen Panel is a platform for internet surveys of public opinion in important areas of society and politics in Norway hosted by the University of Bergen. Participants are randomly recruited from the Norwegian population register to take part in the panel.

Study 2 consisted of an online survey experiment fielded in Germany from April 26 to May 5th 2021. The survey was distributed by the panel provider Respondi, using its high quality access panel that is e.g., also employed by the German Longitudinal Election Study. A representative sample of 2,995 German adults participated in the survey experiment.

4.6.4 Design

The online experiment consisted of a vignette survey experiment, where participants read a short treatment text and then answered a set of outcome questions.

To test for the effects of normative repackaging on voter attitudes, respondents were exposed to political statements that varied according to whether they were normatively repackaged as a defence of liberal democratic values, or not. The survey experiment was set up as a 2x2 factorial between-subjects design, resulting in four experimental conditions to which participants were randomly assigned. An overview of the design and experimental conditions is given in Table 4.1 and Table 4.2.[12] The first factor was immigration (text about immigration or no text about immigration), and the second factor was gender equality (text about gender equality or no text about gender equality).

---

[11]The survey experiment was fielded as a part of the Norwegian Citizen Panel to three distinct survey groups which received different sets of questions before and after this survey experiment (survey groups 2, 6 and 7) Two of these survey groups received a set of political questions and texts directly related to the topic of and outcomes in the survey experiment just before being shown the survey experiment. The results presented here are thus restricted to the least pre-primed survey group which only received one question about an unrelated topic (micro-plastics) prior to engaging in the survey experiment. Including the other two survey groups in the analysis does not substantively alter the results. More information about this, and robustness checks including these survey groups can be found in the Appendix, Section C.0.7.

[12] Power calculations were conducted prior to fielding the experiments and included in the Pre-Analysis Plans. Using the DeclareDesign package in R, to achieve 80% power on the interaction between the immigration and gender equality text, assuming a medium effect size (0.3 of a SD) and using alpha=0.05, a sample size of around 3,000 individuals is needed.

Participants assigned to Treatment 1 read a text that is both about gender equality and immigration. Participants assigned to Treatment 2 read a text that is only about gender equality. Participants assigned to Treatment 3 read a text that is only about immigration. Participants in the Control group did not read any text. Treatment 3 thus contains a political statement without normative repackaging, while Treatment 1 contains exactly the same political statement (as in T3) but with the addition of normative repackaging. Treatment 2 is just the normative repackaging message by itself. This conservative treatment text design allows me to disentangle whether any effects of the normatively repackaged anti-immigration text (Treatment 1) are due to the unique combination of the normative repackaging (gender equality) message and the anti-immigration message.

TABLE 4.1: 2x2 Factorial design

	Immigration text	**No immigration text**
Gender equality text	Immigration+Gender (T1)	Gender cue (T3)
No gender equality text	Immigration cue (T2)	Control

TABLE 4.2: Overview of experimental conditions

Condition	**Content**	**N (Study 1)**	**N (Study 2)**
Control	No text	578	752
T1	Immigration+Gender	588	747
T2	Immigration	507	753
T3	Gender	573	750
T4	Acceptability + Immigration	574	-
Sample size		2820	3002

An overview of the treatment texts can be found in Tables 4.3 and 4.4. The full treatment texts can be found in the Appendix, see Table C.2 and Table C.3. To maximise external validity, the treatment texts closely mirror real radical right-wing party communications. The Norwegian treatment texts are based on the FrP's 2019 party programme and the party position on immigration available on the party website, entitled "Immigration and integration - FrP will set stricter requirements for immigrants and ban the use of the burka and the niqab in public spaces". The German treatment texts are based on communications by the German radical right-wing party AfD. The gender equality treatment text features language from a 2020 press statement entitled "Violence against women starts with the hijab" and from an AfD-requested 2018 parliamentary hearing on immigration and its impact on women's rights, entitled "Strengthening the freedom and equality of women - constitutional law instead of parallel society". The immigration treatment text in Study 2 is based on language used in recent party manifestos and

campaigns on immigration entitled "Stop mass immigration - regulate immigration", and a more recent press statement on immigration from 2020 entitled "Protect the borders ahead of the next immigration wave". The introductory sentence to all treatment texts reads: "Please read the following hypothetical political statements:". The word hypothetical was used to avoid deceiving participants. While the political statements chosen reflect actual political rhetoric, they are not verbatim quotes.

TABLE 4.3: Treatment text - Study 1

Treatment text - Study 1 (Norway)
Immigration text (non-italic). *Gender text (italic).*
High immigration over several decades has had major consequences for the development of society. We need to keep immigration levels low. Many of those who come do not respect the values on which Norwegian society is founded. *Gender equality is a Norwegian core value. We believe that oppressive attitudes towards women do not belong in our society. We believe that all people are equal and we do not accept discrimination against women.* This is why we want a lower immigration rate and a strict immigration policy. We need an immigration policy that clearly requires immigrants to comply with Norwegian laws, rules and values.

TABLE 4.4: Treatment text - Study 2

Treatment text - Study 2 (Germany)
Immigration text (non-italic). *Gender text (italic).*
Mass immigration poses a grave threat to our future. The consequences of unchecked immigration are dramatic. *Women and girls are increasingly becoming victims of violence in Germany. We cannot be indifferent to this development: Women's shelters and counselling centres are becoming more and more popular. We stand up for women's freedom and equality at all political levels - because women's rights are part of our fundamental values and are non-negotiable.* In order to protect our basic values and population, mass immigration must be stopped. We demand a policy that puts the values and interests of our own country and the native population back at the centre. Germany must stop illegal entries at its own borders and end unregulated mass immigration.

On top of the three treatment groups and the Control, Study 1 featured an additional fourth treatment group. The purpose of Treatment group 4 was to manipulate the hypothesised main causal mechanism, social acceptability of anti-immigration statements. Treatment group 4 received the same immigration text as Treatment group 2, but with an additional sentence that signals a high acceptability of anti-immigrant statements. The introductory sentence for Treatment 4 reads: "In recent years, there has been a

lot of talk on immigration. According to the 2018 Integration Barometer, 41% of Norwegians said that Norway should accept fewer or significantly fewer immigrants. This shows people feel comfortable sharing their concerns regarding immigrants in surveys, and express a range of different opinions on immigration.". The study mentioned in the introductory sentence refers to an actual survey conducted in Norway; there is thus no deception of respondents.

4.6.5 Outcome measures

The three main outcomes of interest are the acceptability of anti-immigration views, immigration attitudes, and vote intention for the radical right.

To measure acceptability of anti-immigration beliefs, participants were asked "Do you think it is acceptable to say something negative about immigrants in [country] today? Please rate on a scale from 0 (Not acceptable at all) to 10 (Very acceptable)". This question has been used and validated in previous rounds of the Norwegian Citizen Panel.

To measure immigration attitudes in Study 1 (Norway), respondents were asked: "Should Norway welcome more or fewer immigrants than today?" (Significantly more/Some more/Neither more nor fewer/Some fewer/Significantly fewer). This is a commonly used question on immigration attitudes in survey research, and was e.g., part of the most recent Integration Barometer questionnaire in Norway. In Study 2 (Germany 2021), immigration attitudes were measured using a standard immigration as cultural threat question, asking "Would you say Germany's cultural life is generally undermined or enriched by people coming to live here from other countries?" (0= Cultural life undermined, to 10=Cultural life enriched).

To measure vote intention for the radical right, respondents were asked "If there was a General Election, which party would you vote for?". This was recoded into a binary outcome variable, coded 0=voting for any other party, 1= voting for the Progress Party/AfD.

4.6.6 Estimation

Treatment effects were estimated using OLS regression models with robust standard errors (HC2).

The regression equation reflects the 2x2 factorial design of the experiment, estimating the effect of the treatment text being about immigration, the treatment text being about gender, and the interaction between the two. It is this interaction effect between the

gender equality and immigration message, given by the β_3 coefficient in Equation 4.1 that is the main coefficient of interest. Additionally, I also estimate Equation (Equation 4.2) to test whether the effect of the gender-immigration message is different for female than male respondents.

$$Y_i = \alpha + \beta_1 * GenderEquality_i + \beta_2 * Immigration_i +$$
$$\beta_3 * GenderEquality_i * Immigration_i + \epsilon_i$$

$$(4.1)$$

$$Y_i = \alpha + \beta_1 * GenderEquality_i + \beta_2 * Immigration_i + \beta_3 * FemaleRespondent_i +$$
$$\beta_4 * Immigration_i * GenderEquality_i + \beta_5 GenderEquality_i * FemaleRespondent_i +$$
$$\beta_6 * Immigration_i * FemaleRespondent_i +$$
$$\beta_7 * GenderEquality_i * Immigration_i * FemaleRespondent_i + \epsilon_i$$

$$(4.2)$$

where GenderEquality=0 when the treatment text is not about gender equality, and 1 when the text is about gender equality

where Immigration=0 when the treatment text is not about immigration, and 1 when the treatment text is about immigration

where FemaleRespondent=0 when the respondent is male, and 1 if the respondent is female

4.7 Results

Key demographic background variables were balanced between experimental conditions (see Section C.0.2, Appendix) and manipulation checks show that the majority of respondents were able to correctly recall the topic of the text that they read (see Section C.0.6, Appendix).

The main results are summarised in Figure 4.1 and Figure 4.2. The two upper graphs in the figures show the results for the full sample, while the two bottom graphs show the results by the gender of respondents. The upper left-hand graph shows the condition means for acceptability of anti-immigration views for the full sample. The upper right-hand graph shows the treatment effects for the full sample. The bottom left hand graph shows the condition means for female and male respondents, and the bottom right hand graph shows treatment effects by respondent gender.
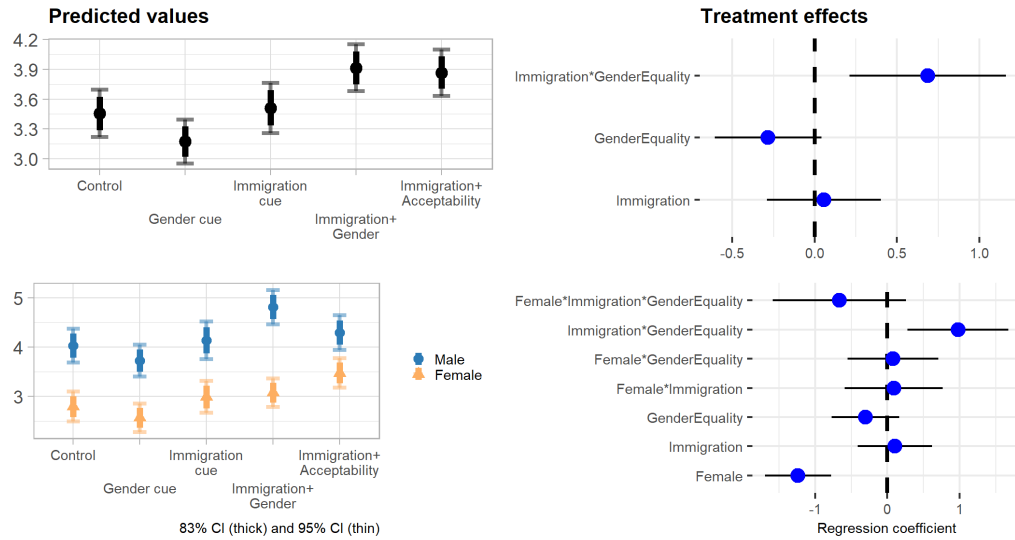
FIGURE 4.1: Acceptability of anti-immigration views (Study 1 - Norway)

Overall, I find evidence in favour of the expectation in H1 that gender-immigration messages increase the acceptability of anti-immigration views. In Study 1 (Norway), the gendered immigration message had a significant positive effect on respondents' perceived acceptability of anti-immigrant statements. After reading the combined gender equality and anti-immigration message, respondents were more likely to say that it is acceptable to say something negative about immigrants in Norway today (see upper right and upper left graph, Figure 4.1). The interaction effect between the gender equality and the immigration message represents a 0.24 standard deviation increase in the acceptability of anti-immigrant statements. This is a sizeable increase in how acceptable people find it to say something negative about immigrants. There were no significant differences in how men and women reacted to the gender-immigration message in Study 1 (bottom left and bottom right graph, Figure 4.1, as well as Appendix, Model 2, Table C.9).

The effect of the gender-immigration message in Study 1 is driven by the stark contrast with both the simple gender equality message, and the simple anti-immigration message. Receiving the gender equality message (T3) (compared to the Control) made respondents feel it is less acceptable to say something negative about immigrants. The immigration message by itself (T2) had a small positive effect on the perceived acceptability of anti-immigration statements. Combining these two messages into a gendered anti-immigration statement made respondents feel it is more acceptable to say something negative about immigrants. In Study 1, the effects of the gender equality message are thus highly contextual: presented by itself, it motivates respondents to see anti-immigration rhetoric as less acceptable. Paired with a standard anti-immigration

message, a gender equality message has the opposite effect: it makes respondents feel that saying something bad about immigrants is more acceptable, and helps to legitimise anti-immigrant views.

In Study 2 (Germany 2021), the gender-immigration message, compared to the Control, substantively but not significantly increases the acceptability of anti-immigration views in the whole sample (β=0.182, p=0.24). However, there are significant heterogeneous treatment effects by gender. Women reacted more positively and strongly than men to the gender-immigration text. Compared to the Control (not receiving any text to read), the gender-immigration message made women, compared to men, more likely to find it acceptable to say something negative about immigrants (bottom left-hand panel, Figure 4.2, also see Appendix, Table C.8). The interaction between being female and receiving Treatment 1 (gender+immigration text) represents a 0.27 of a standard deviation in the outcome, a large increase in the acceptability of anti-immigration views.

Results for acceptability of anti-immigration positions (Germany 2021)
Y: Is it acceptable to say something negative about immigrants? 0-10, 10=Very acceptable
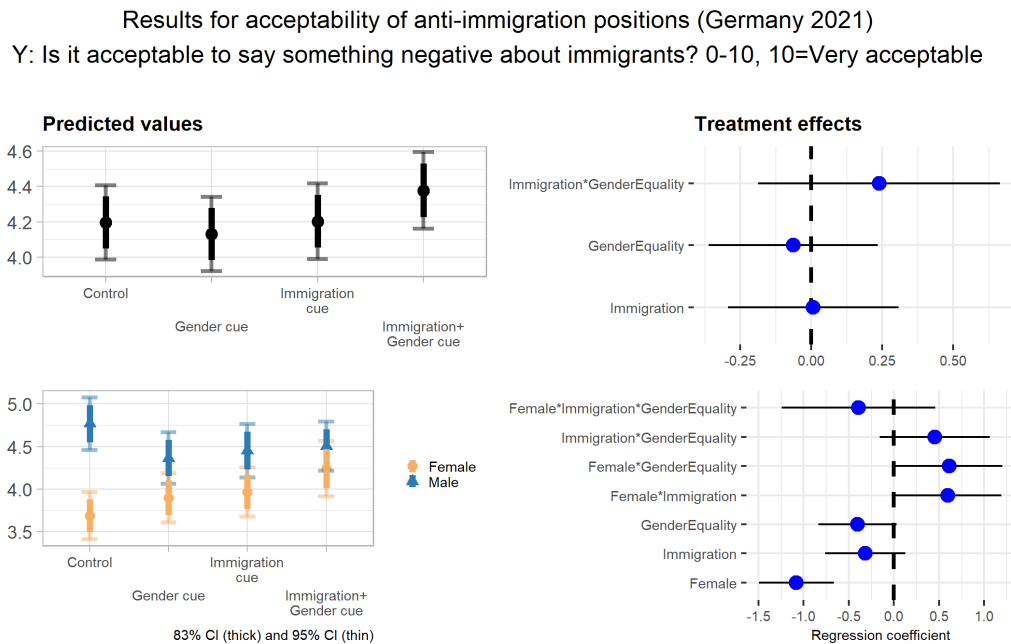
FIGURE 4.2: Acceptability of anti-immigration views (Study 2 - Germany 2021)

In substantive terms, there is sizeable gender gap in the perceived acceptability of anti-immigration views in the Control condition. In the Control condition, women report that anti-immigration views are not acceptable, averaging a 3.6 on the 0 (not acceptable) to 10 (acceptable) scale. In contrast, men in the Control condition are less averse to the expression of anti-immigrant views, averaging a 4.7 on the 0-10 scale. Men at baseline are thus closer to taking a neutral mid-point stance on whether or not it is acceptable to say something bad about immigrants (see bottom left-hand panel, Figure 4.2). This gender gap in the acceptability of anti-immigration views vanishes when voters are exposed to normatively repackaged anti-immigration messages. The closing of the gap is mainly

driven by women who become more accepting of anti-immigration views in response to gender-immigration messages.

In Study 2 (Germany 2021), the effect of adding a normative repackaging element (a gender equality message) to a controversial political statement about immigration is additive, rather than interactive: the gender equality text by itself already slightly increases the acceptability of anti-immigration views among women, as does the immigration text by itself. Combining the two into a gender-immigration message produces the strongest effect on the acceptability of anti-immigration views (see bottom panel, Figure 4.2). This pattern explains why the overall interaction effect between being female, the gender equality message and the immigration message (see bottom right coefficient plot in Figure 4.2) is not significant. Both the simple gender message and the simple anti-immigration message already make women more accepting of anti-immigration views than not receiving any treatment text.

I also tested whether gender-immigration messages have an effect on the direct expression of anti-immigration views or vote intention for the radical right. I do not find consistent effects of gender-immigration messages on vote intention or the expression of anti-immigration views. In Study 1 (Norway), the treatment texts, compared to the Control, generally produced backlash among respondents, making them less likely to express anti-immigration attitudes (see Appendix, Figure C.1). However, the immigration text when combined with the gender equality message, produced less backlash than the immigration message on its own (see Appendix, Figure C.1). This interaction effect between the gender message and the immigration message represents 0.14 of a SD in the outcome variable immigration attitudes and is significant at p=0.1 (see Appendix, Table C.13). In Study 2 (Germany), the gender-immigration message did not have a significant effect on immigration attitudes, compared to the Control (see Appendix, Figure C.2). The substantive direction of the interaction between respondent gender, the immigration and gender equality text is positive, albeit not significant (see Appendix, Figure C.2). Compared to men, women slightly increased their expression of anti-immigration views in response to the gender-immigration message in Study 2.

Across Study 1 (Norway) and Study 2 (Germany), gender-immigration messages had a substantively positive, but insignificant effect on radical right support (see Appendix, Figures C.3 & C.4, Tables C.15 & C.14). Substantively, vote intention for the radical right is around 2 percentage points higher among respondents in the gender-immigration message condition than the Control in both studies. In Germany, around 9% of respondents in the Control say they would vote for the radical right, while 11% say the same in the gender-immigration condition. In Norway, around 7% of respondents in the Control

report they would vote for the radical right, going up to 9% in the gender-immigration condition. However, these differences are not statistically significant.

4.7.1 Acceptability as a potential mechanism

The design in Study 1 also included an additional treatment group where participants read the same anti-immigration message as in the other treatment groups, plus an acceptability statement preceding that message. The acceptability statement tried to convey the feeling that anti-immigration preferences are widespread among the population,and therefore acceptable (see full treatment text in the Appendix, Table C.2). The purpose of having this additional treatment arm was to provide further evidence that the gender-immigration messages work in similar ways to cues that signal a high perceived acceptability of anti-immigration sentiment. If the effect of the gender-immigration frame is linked to it increasing the acceptability of anti-immigration statements, and the acceptability treatment manipulates the acceptability of anti-immigration statements, then we would expect the acceptability treatment to show a similar direction and size of effect as the gender-immigration treatment. Comparing the coefficients on the acceptability treatment and the gender-immigration treatment is one way of exploring if the effect of the gender-immigration message may at least partially work through changes in the acceptability of anti-immigration statements.

Looking at Figure 4.1, we can see that the effect of the gender-immigration text (T1) is similar in size and direction to the effect of the acceptability + immigration treatment (T4).

I also test whether the coefficient of the gender-immigration frame treatment (T1) and the acceptability treatment (T4) are significantly different from each other by performing a linear hypothesis test on whether Gender+Immigration(T1) - Acceptability+Immigration(T4)= 0. I find that the acceptability treatment effect did not differ significantly from the effect of the gender-immigration treatment (see Appendix, Table C.10). This provides additional evidence that gender-immigration messages have similar effects on voter attitudes as messages that explicitly manipulate the acceptability of anti-immigration views. This speaks to my theoretical argument that gender-immigration messages are effective because they convey that anti-immigrant views are socially acceptable.

4.8 Conclusion

This paper examines the implications of "normative repackaging" on voters, a strategy by which political actors present a previously unacceptable political view as a defence of liberal democratic values, such as gender equality. I argue that normative repackaging should increase the acceptability of previously unacceptable views (H1), increase the expression of such views (H2), as well as support for political actors espousing these views (H3). I also explore if voter groups which care more about not holding unacceptable views are more affected by normative repackaging (H4).

I test this argument at the example of gender-immigration messages, statements which combine an anti-immigration position with a reference to women's rights and gender equality. Through vignette survey experiments, I examine if these messages make anti-immigration views more acceptable and popular, as well as if they increase support for the radical right. I argue that gender-immigration messages should be effective because they coat a previously unacceptable anti-immigration appeal in a more normatively permissible gender equality frame. Gender-immigration messages legitimise anti-immigration rhetoric through presenting it as a defence of highly acceptable, liberal democratic values. I test these arguments through survey experiments conducted in Germany and Norway.

Overall, I find evidence that normative repackaging is an effective strategy: it makes previously unacceptable views appear more acceptable to voters. Anti-immigration messages combined with a reference to gender equality significantly increased voters' acceptability of anti-immigration views in the whole sample (Study 1), and among women, compared to men (Study 2). This is in line with my theoretical expectation (H1) that references to gender equality and women's protection de-stigmatise anti-immigration sentiment.

However, the effect of gender-immigration messages on the acceptability anti-immigration views also varied between Study 1 and Study 2. In Study 2, the gender-immigration message did not have a significant effect on acceptability in the whole sample, but affected men and women differently. Reading a gender-immigration message (compared to no message) made women, compared to men, more likely to believe that it is okay to say something negative about immigrants. Gender-immigration messages thereby closed the gender gap in acceptability of anti-immigration views in Study 2. Furthermore, the effect of normative repackaging in Study 2 was additive, rather than interactive (as in Study 1). The immigration message by itself, as well as the gender equality message by itself already made anti-immigration views seem more acceptable to female voters. In

Study 2, a simple message about gender equality and women's safety was enough to increase tolerance for anti-immigrant views. This points to the idea that raising concerns about women's rights and gender equality, without any explicit mention of immigration, could be used as an anti-immigrant dog-whistle by the radical right.

One reading of the different results in the German and Norwegian study presented in this paper is that the effects of gender-immigration messages are context-dependent. In contexts where gender equality messages might still be read as universal value appeals, gender equality rhetoric is a powerful weapon used to legitimise anti-immigration views. Even when combined with a completely value-incongruent right-wing anti-immigrant policy position, gender equality messages can add a layer of acceptability to the immigration message that makes voters feel it is more permissible to express anti-immigration preferences. In other cases, where these messages are perhaps already more common, such as in Germany, even a simple gender message without mention of immigration can already make women, compared to men, more accepting of anti-immigration views. Here gender messages could be better be understood as dog-whistle issues which legitimise anti-immigration views without having to explicitly mention immigration. Future research should explicitly test how the political context moderates the effect of gender-immigration messages.

The results between Study 1 and Study 2 also varied according to whether there were heterogeneous treatment effects by gender. How can we understand these different responses by male and female voters in Germany and Norway? I suggest that how gendered immigration messages affect men and women's support for radical right-wing parties and anti-immigration policies might depend both on the content of the gender-immigration message and the context in which it is received. In the German case, the radical right-wing party AfD is still a fairly new challenger party and seen as more taboo. In Norway, the radical right-wing party FrP is a party which has served in coalition government and is already seen as more acceptable. These differences are also mirrored in baseline gender gaps in radical right support: in Germany, women are still significantly less likely to vote for the radical right, while there is a much smaller and no longer statistically significant gender gap in radical right support in the Control group in Norway. This shows that women might have already "caught up" more with men in the Norwegian context, and that acceptability concerns might therefore affect men and women more evenly in Norway. The content of the gendered anti-immigration message also slightly varied between Germany and Norway. In Germany, the gendered anti-immigration message might have evoked more perceptions of women as the victims of immigrant perpetrated crime than the Norwegian treatment stimulus which was slightly more abstract. This mirrors differences in real-life political rhetoric aired by radical right-wing parties in Germany and Norway. However, this might have led

women reading the German treatment text to feel more directly affected by the gendered anti-immigration text and to therefore respond more strongly to the treatment stimulus.

Beyond acceptability, this paper also examined what implications normative repackaging has on the direct expression of previously unacceptable views, as well as the support for political actors espousing these views. I do not find consistent evidence in favour of H2 and H3, the expectation that gender-immigration messages increase the direct expression of anti-immigration attitudes and support for the radical right. Messages combining an anti-immigration position with a gender equality message make voters in Study 1 (Norway) slightly more likely to express anti-immigration views but did not significantly affect radical right support. In Study 2, the gender-immigration message did not have any significant effects on the expression of anti-immigration preferences or vote intention for the radical right.

Why do gender-immigration messages affect the acceptability of anti-immigration views, but not the expression of anti-immigration views or radical right support? Thinking about these three outcomes through the funnel of causality, the acceptability of anti-immigration views would be causally prior to the expression of anti-immigration views and radical right support. Once anti-immigration views are more acceptable, more voters should feel comfortable expressing anti-immigrant views themselves, as well as voting for the radical right. However, this might be a long-term process where gradual changes in acceptability translate into an increased expression of anti-immigrant views, and eventually increased radical right support. Gender-immigration messages therefore still matter for radical right-wing parties and actors, even when they do not directly affect radical right support. Gender-immigration messages set in motion sizeable shifts in the acceptability of anti-immigration views, and therefore set the stage for future increases in the expression of anti-immigrant sentiment and radical right support.

Secondly, immigration attitudes have been shown to be sticky, hard-to-move preferences in surveys and survey experiments (Kustov, Laaker and Reller, 2021). In contrast, the perceived acceptability of anti-immigration views might be a more malleable attitude. It is more likely that acceptability is affected by a treatment stimulus than more stable immigration attitudes and radical right support. Perceptions of norms around immigration views might be more influenced by what voters momentarily think about and what frames they have in mind when thinking about immigration, than immigration attitudes themselves.

Overall, this paper argued that gender-immigration messages normalise anti-immigration views by coating these in a highly acceptable, liberal democratic value justification. In this paper, I also offer additional evidence that gender-immigration messages may work through increasing the acceptability of anti-immigration views. By experimentally

manipulating descriptive norms around the expression of anti-immigrant views in an additional treatment group, I isolate and compare the effect that this acceptability prime has. I find that telling respondents that anti-immigration views are widespread has effects similar in size and direction to the effects of a gendered anti-immigration message. This provides additional evidence that acceptability is one of the mechanisms through which gendered anti-immigration messages may work. Coupling a strong anti-immigration message with an acceptable, softening appeal to gender equality legitimises anti-immigration statements and makes them appear and feel more mainstream.

The findings from this paper have important implications for our understanding of how unacceptable views become normalised as a result of elite discourse. It shows that normative repackaging, a strategy by which unacceptable views are presented as a defence of liberal democratic values, is an effective way to make voters more accepting of previously unacceptable positions, such as anti-immigration views. Combining insights from the literature on social stigma and mainstreaming, the literature on framing and elite rhetoric, as well as the literature on gender and the radical right, this paper opens up a larger conversation on how political rhetoric is used by political actors to make their taboo positions acceptable. While a gender lens to studying the radical right has traditionally stressed the absence of women in these parties, this paper shows that these actors can use gender equality rhetoric to their advantage. Even small and gradual changes in how acceptable anti-immigration positions are perceived to be can have long-term consequences for political behaviour, and citizens' expectations of politicians and parties. If the boundaries around what constitutes acceptable discourse bend, this can also have harmful effects on behaviours in-and outside the political realm, such as discrimination and hate speech.

Chapter 5

Conclusion

"There are millions of people who, a few hours before Brexit, decided what was the point in going to vote. Millions did the same in 2016 with Trump. The next day they woke up with a hangover." These were the words of Macron's official television commercial pleading with French voters to make the trip to the polling station during the most recent French Presidential election (Macron quoted in Kirby 2022a). Macron openly appealed to the voters' extreme dislike of the National Rally and its candidate Marine Le Pen as a reason to support him at the ballot box. During the TV debate between the two presidential candidates, Macron also frequently went on the attack against Le Pen, e.g., arguing that her radical right policies would create a "civil war in our cities" (Macron quoted in Kirby 2022a) . Some voters seemed persuaded – not by Macron, his party or his policies, but by their extreme distaste for the alternative: "It's not that I like the idea of voting for Marine Le Pen, but when you're choosing between cholera and the plague, you choose the lesser of two evils" (voter Dylan from Arras, quoted in Kirby 2022b). The most recent French Presidential election is just one example of how negative partisanship matters for political behaviour. Strong feelings of dislike towards political parties are prevalent in many Western democracies (Reiljan, 2020; Gidron, Adams and Horne, 2020; Wagner, 2021). The gap between positive political in-group appraisals and negative out-group feelings has an important role in structuring individual-level political behaviour and preferences (Abramowitz and Webster, 2016; Iyengar et al., 2019; Wagner, 2021; Hobolt, Leeper and Tilley, 2020).

Despite this, strong negative feelings towards political parties are still under-studied compared to research on positive partisan attachments. When negative partisanship is considered, it is often treated as an extension of positive partisanship. This thesis has challenged this assumption, arguing that strong negative feelings towards political parties matter in their own right. Through three survey experiments and one field

experiment, this thesis has shown that activating negative feelings towards political out-groups matters for attitudes and behaviour and can have stronger effects than activating positive feelings of belonging.

In Paper 1, I make a novel theoretical contribution to the partisanship literature by arguing that negative partisanship affects how voters feel towards multiple political parties, not just the one that they dislike. In line with the idea of "my enemy's enemy is my friend", I find that negative partisanship can inspire feelings of closeness to other political parties.

In Paper 2, through a field experiment with a political party, we show that cueing negative political identities has stronger mobilising effects than cueing positive political identities. This makes an important contribution to our understanding of how political parties activate negative political identities to their advantage, and shows that invoking these has implications for costly political behaviour.

In Paper 3, I examine if and how disliked political parties can reduce negative feelings towards them, and their core issue positions. I find that normative repackaging, a strategy which couples an unacceptable policy position with a highly acceptable one, is an effective way for radical right-wing parties to make their core policy positions feel more acceptable to voters.

Taken together, these papers show that strong negative feelings towards political out-groups are a force to be reckoned with in politics – one that powerfully shapes voters' attitudes and behaviour, party political discourse and strategy. Importantly, these papers also show that negative identities can have more powerful effects than positive ones. Especially in an age of polarisation and dwindling positive party identification, defining "who we are not" has become part and parcel of political thinking and behaviour. This thesis demonstrates that strong negative feelings towards political out-groups need to be included in accounts of political behaviour and party competition.

While this thesis shows that the activation of negative political identities can encourage desirable behaviours, such as greater political participation and greater feelings of closeness to another political party, out-group disdain can also have problematic implications for democracy. Coming back to the French Presidential election, the dislike for both parties and candidates was visibly high in the last days before the second round: many campaign posters had been sprayed over with words such as "fascist","dirty liberal","racist" and "elitist" (Adler, 2022). Such widespread feelings of dislike can also have harmful implications for societal cohesion and democratic satisfaction. For example, many French left-wing voters said that they would abstain in the second round rather than endorse a candidate that they dislike (Adler, 2022). Under the hashtag

"ni la peste ni le cholera" (neither the pest nor the cholera) French Twitter users aired their anger with the political system and the frustration of having to vote based on which candidate is the lesser evil, rather than making a positive choice for something (Khan-Ruf, 2022).

Negative partisanship does not just have implications for voters and political parties but also for democracy at large. In the following section, I consider the consequences of negative partisanship for voters' satisfaction with democracy, democratic quality and political institutions. I then discuss the generalisability of the findings in this thesis, as well as avenues for further research. I close with a discussion of policy implications, including a range of recommendations for political parties.

5.1 Negative partisanship, affective polarisation and democracy

Negative partisanship does not just affect voters and political parties, but also has wider implications for democratic functioning and political institutions. Affective polarisation, of which out-party disdain is a core component, has been linked to increased voter dissatisfaction with democracy. Political out-group hostility makes losing elections to opponents worse, leading to a cascade of dissatisfaction with democratic institutions. As Iyengar argues, "those who impugn the motives and character of political opponents are less likely to treat as legitimate the decisions and policies enacted when the opponents control government, and may also be less satisfied with institutions that respond to popular will" (Iyengar, Sood and Lelkes, 2012). Comparative research has shown that higher levels of affective polarisation are indeed linked with higher levels of voter dissatisfaction with democracy (Wagner, 2021). Higher partisan out-group dislike is also associated with lower levels of individual-level trust in political institutions and politicians (Torcal and Carty, 2022).

In a similar vein, negative partisanship has been linked to lower satisfaction with the functioning of democracy (Ridge, 2020; Samuels and Zucco, 2018), particularly for individuals who strongly dislike a major party. Negative partisans have stronger reactions to experiences of electoral victory and defeat than positive partisans: negative partisans are more satisfied with democracy when they win and more dissatisfied with democracy when they lose to the party they dislike (Ridge, 2020). However, other research has argued that this is not necessarily the case everywhere: in Latin American countries where positive partisanship is particularly low, negative partisanship can provide an important structuring device for political behaviour and preferences, and thereby also a gateway to

greater democratic satisfaction (Haime and Cantú, 2022). Individuals who at least hold negative partisanship, rather than no partisanship at all, are relatively more supportive of democracy (Haime and Cantú, 2022). In line with the theoretical framework used in this thesis, this finding highlights that a nuanced partisanship typology is important for our understanding of the link between negative partisanship and democracy. It also shows that the reference group we compare negative partisans to matters: while positive partisanship may be better for democracy, negative partisanship is still preferable to complete apathy towards political parties.

Beyond voter satisfaction with democracy, affective polarisation and negative partisanship have also been argued to have implications for institutional stability. One of the rare positive side-effects mentioned in the literature is the stabilising effect of negative partisanship on voting behaviour, and therefore political institutions. In two-party systems like the US, negative partisanship, even among Independents, increases party loyalty and straight-ticket voting (Abramowitz and Webster, 2016). This might be beneficial for democratic functioning because it leads to more stable electoral outcomes. However, in multi-party systems, the systemic implications of negative partisanship may look different. Early research on negative partisanship in European post-Communist countries argued that negative partisanship increases volatility and vote-switching between parties, thereby destabilising political institutions (Rose and Mishler, 1998). However, my thesis also speaks to the idea that this volatility may be time-limited: negative identities may also transform into positive ones in the long run. Paper 1 shows that negative partisanship activation prompts negative partisans (individuals who hold negative partisanship and no positive partisanship) to feel closer to their preferred party. If negative partisanship can be channelled into more positive attachments towards a political party, volatility should go down again.

Negative partisanship and affective polarisation have also been linked to reductions in the quality of democracy. For example, heightened political out-group dislike is associated with a greater willingness to engage in political violence (Mason and Kalmoe, 2022), out-group dehumanisation (Martherus et al., 2021) and reduced support for democratic norms (Kingzette et al., 2021). Comparative research finds that affective polarisation is positively correlated with democratic backsliding – countries with increases in affective polarisation are more likely to regress in terms of their democratic quality, with reductions in government accountability, individual civil rights and opportunities for democratic deliberation (Orhan, 2022). This finding builds on research arguing that polarisation leads to democratic erosion, reduced respect for counterarguments and reduced consultation with other political actors (McCoy, Rahman and Somer, 2018). However, more recent experimental research has called into question whether the link between affective polarisation and democratic backsliding is spurious. Experimental research finds

that affective polarisation does not lead to reduced support for democratic norms or bipartisanship (Broockman, Kalla and Westwood, 2022).

This thesis contributes to the literature on democracy, negative partisanship and affective polarisation by showing what effects the activation of negative identities has on political participation. While negative identities may have harmful effects on many features of democratic institutions, this thesis shows that they also bolster political participation – a key component of democratic functioning. Paper 2 shows that activating negative political identities has powerful mobilising effects on party supporters. This thesis therefore contributes to a nuanced understanding of what implications negative political identities have for democracy.

Moreover, affective polarisation and out-party hostility also affect social cohesion and inter-group contact, important prerequisites for democratic functioning. For example, affective polarisation leads to individuals feeling more uncomfortable about having out party friends and neighbours, more uncomfortable with their children marrying someone from the out party, and less willing to talk about politics with an out partisan (Broockman, Kalla and Westwood, 2022). This is deeply problematic because partisans both in the US and in many other Western democracies are already increasingly sorted along socio-demographic, geographic and religious lines (Mason, 2018; Brown and Enos, 2021; Harteveld, 2021*b*).

This overlap between other social characteristics and partisanship is problematic because it reduces cross-cutting identities that could alleviate the pull of out-group hostility. The meshing of social and political divisions also means that political out-group hostility more easily finds its way into attitudes and behaviours outside the political realm. For example, job applicants with minority political views face discrimination in the recruitment process (Gift and Gift, 2015). Other research has shown that partisan divisions even affect dating decisions, where individuals on average see potential romantic partners from the out-party as less attractive and dateable (Easton, Holbein and Batten, 2021).

Scholars have proposed a range of solutions to tackle the inter-personal distance and ensuing problems resulting from out-group dislike. One proposed approach is increased inter-group contact, with the aims of increasing empathy, reducing hostility and correcting beliefs about who the modal member of the out-party is. This is important because partisans tend to think of overly engaged and ideological voters when they picture a characteristic out-party member (Druckman et al., 2022). Correcting these beliefs lowers inter-personal affective polarisation (Druckman et al., 2022).

Another potential remedy for affective polarisation is emphasising cross-cutting identities. Focusing on shared identities that cut across partisan lines should lower the relative salience of partisan identities and therefore reduce the negative implications of partisan divisions. National identity has been put forward as one example of a cross-cutting identity that voters from across the aisle could identify with. Experimental research conducted in 2015 shows that priming American identity among US partisans indeed reduces out-party dislike (Levendusky, 2018). However, national identity can also become tied to partisan divisions in which case it would no longer be considered cross-cutting. After American identity has become more politicised during Trump's presidency, it is questionable whether feeling American still constitutes a bridge across the political divide. Indeed, finding common identities that reduce partisan hostility is tricky once sorting has happened and partisan identities are strong. For example, invoking common identities like gender can also backfire and actually decrease trust between groups of opposing partisans (Klar, 2018).

Yet, new cross-cutting identities may not just be ineffective at lowering partisan divisions, they may also create new societal divisions and patterns of hostile group conflict. The dark side of political identity is by no means limited to partisanship: other political identities can also generate visceral political conflict and entail strong negative out-group feelings (Hobolt, Leeper and Tilley, 2020). This thesis demonstrates that activating these feelings of dislike towards an issue out-group is a powerful mobilising tool and can be more effective than rallying supporters around a disliked partisan out-group when issue identities are salient (Paper 2). It also speaks to the idea that out-group hostility is a key part of issue identities (Paper 2). Rather than lowering political group conflict, new cross-cutting political identities may just move group conflict into a new domain. The emergence of new cross-cutting political identities is therefore not necessarily good news for democracy.

5.2 The role of political parties in the relationship between out-group disdain and democracy

While mass-level out-party disdain is linked to democratic functioning, political parties also play a key role in the relationship between negative partisanship, affective polarisation and democracy. Political parties and politicians' actions and words can either lower or raise the temperature of political conflict. For example, elite polarisation has been linked to mass polarisation (Banda and Cluverius, 2018). There is also evidence that negative campaigning in political ads increases affective polarisation (Lau et al., 2017), and that political campaigns inflate the gap in affect between voter's least and

most preferred party, and therefore lead to increases in affective polarisation (Hansen and Kosiara-Pedersen, 2015).

How political parties and politicians behave towards their opponents also matters for voters' trust in political institutions and politicians. Early studies on the relationship between negative campaigning and democracy argued that uncivil attacks against the opponent decrease voters' political trust (Mutz and Reeves, 2005), while others did not find that uncivil attacks have an effect on political trust or efficacy (Brooks and Geer, 2007). More recent research shows that voters' perceptions of "dirty campaigning", politicians using frequent uncivil attacks against the other side, are related to a decrease in their political trust (Reiter and Matthes, 2021). Experimental research confirms that exposure to politicians' uncivil communication – insults and interruptions of opponents - lowers voters' political trust (Goovaerts and Marien, 2020). Incivility between politicians during televised political debates has also been linked to lower political trust, as well as respect for opposing viewpoints (Mutz, 2015).

However, political parties can also play an important role in reducing out-party hostility and its potentially harmful implications for democracy. For example, political cooperation between parties has been linked to a decrease in mass-level affective polarisation. Experimental evidence from Israel shows when political parties signal that they are willing to enter a unity coalition government, this significantly reduces levels of out-party dislike in the electorate (Bassan-Nygate and Weiss, 2021). Other research has shown that smaller cooperative behaviours also already affect voters' perceptions of political parties (Adams, Weschle and Wlezien, 2021). More cooperative relationships between political parties influence voters' perceptions of the ideological distance between two parties (Adams, Weschle and Wlezien, 2021).

Yet, cooperation between political parties is not an automatic guarantor of lasting reductions in political out-group hostility. For example, observational research shows that countries where coalition governments are more common do not necessarily have lower levels of negative partisanship (Anderson, McGregor and Stephenson, 2021). Relationships between former coalition partners can change and governing together does not preclude the possibility of future hostility between parties. In line with this, Paper 2 shows that invoking negative feelings towards political out-groups, including towards a former coalition partner, rallies up party supporters, compared to invoking positive feelings towards political in-groups.

However, Paper 2 is also set in the specific institutional context of the UK where coalition governments are the exception. Research from Austria, a multi-party system with frequent coalition governments, shows that former coalition partners indeed engage in less virulent attacks against each other (Haselmayer and Jenny, 2018). In a setting

where political parties want to potentially re-enter coalition governments, parties have incentives not to alienate coalition partners and their supporters (Haselmayer and Jenny, 2018). Future research should examine how long-lasting the effects of political cooperation on affective polarisation are, and if invoking out-group hostility towards a former coalition partner is indeed ineffective in countries where governing together is the norm.

Finally, political parties can also contribute to lowering affective polarisation among their supporters. Recent research from the US shows that outreach activities that involve in-depth, two-way conversations with voters can lead to reductions in affective polarisation among party supporters (Kalla and Broockman, 2022). It is thus clearly possible for campaigns to design activities and messages in a way that fulfil strategic goals and simultaneously reduce out-group hostility among supporters. Political parties have an obligation to consider what their internal messages and activities mean for the levels of out-group hostility among their support base. While this thesis shows that invoking out-group dislike may spur supporters into taking political action momentarily, it also shows that this strategy contains serious trade-offs for political parties (Paper 2). Political parties should consider the long-term implications of their messages and activities, and ideally design them in a way that aligns their objectives with reductions in out-group disdain.

Lastly, political parties may also try to defuse strong negative feelings towards them. This thesis investigated this at the example of radical right-wing parties trying to make some of their core views more societally acceptable (Paper 3). I find that normative repackaging, a strategy which couples an unacceptable view with a liberal democratic norm, normalises the unacceptable view. When an anti-immigration statement is presented as a defence of women's rights and gender equality, voters think that saying something bad about immigrants is more acceptable. Normative repackaging allows extreme political parties to shift the boundaries of acceptable political discourse to their advantage with little to no cost to themselves.

This finding has worrisome implications for democratic institutions. It means that important societal norms are perhaps more fragile than previously thought, and that extreme political actors can actively chip away at them without having to adapt or moderate their positions. In the process, radical right-wing parties do not just hollow out the target norm (the anti-prejudice norm in this case) but also the liberal democratic norm that they use to defend it (that of gender equality). If liberal democratic norms can just be used to defend any position, no matter how incongruent, this also risks diluting the meaning and power of the liberal democratic norm itself. In other words, if anyone can claim to be a defender of gender equality, no matter what their actual substantive policies or track record on this topic are, this problematically widens the

boundaries around what gender equality means and how it can be used. Normative repackaging thus has potentially wide-ranging and deeply problematic implications for liberal democratic values that future research should investigate.

5.3 Generalisability and avenues for further research

Having discussed the implications of negative partisanship for democracy, I discuss avenues for further research in this section. I also discuss to what extent I expect my research to generalise to other settings and questions, and how future studies could expand on the research presented in this thesis.

This thesis investigated what the activation of negative partisanship does to voter evaluations of political parties. One of the key findings of Paper 1 was that negative partisanship activation lead negative partisans to feel closer to their most-preferred party. I would expect these findings to travel to other multi-party systems. Canada has average levels of affective polarisation compared to other Western democracies, so these findings should also hold in other Western countries with similar levels of out-party dislike. Unfortunately, one of the limitations of Paper 1 is the suboptimal measurement of pre-treatment partisan type based on thermometer scores. Future studies could improve on this by fielding a two-wave study where pre-treatment partisan type is measured using both identity and affect batteries in Wave 1, while the actual experiment takes place in Wave 2. A slightly larger sample would also allow for interesting sub-group comparisons, examining e.g., whether negative partisanship activation has different effects when the disliked party is a mainstream party compared to an extreme party, or when the disliked party is in the voter's own ideological camp (left or right) compared to the other side.

This thesis demonstrated that when political parties activate negative political identities in their campaign messages, this has mobilising effects on their supporters. The field experiment presented in Paper 2 took place in the UK during the 2019 general election campaign. I would expect that the overall finding that negative identity cues are more impactful than positive ones should travel to many two- and multi-party systems where affective polarisation is relatively high, as was the case in the UK. The finding that negative issue-based identity cues outperform negative party identity cues should travel to other political contexts where an issue identity is highly salient, e.g., during a ballot measure in the US or a referendum campaign in a European country. For example, ballot measures on abortion were up for a public vote in states like Louisiana in 2020 in the US – as abortion identities were likely salient during this vote, they might have outperformed traditional partisan identities during this specific election context.

Generally, I would expect that voters with congruent political identities react strongest to their activation. In other words, supporters who hold negative partisanship probably reacted more positively to the negative partisanship activation email in Paper 2 than supporters who were positive partisans. However, it is unfortunately hard to test this without any pre-treatment survey data about the sample. Future research could try to collaborate with political organisations to field a pre-treatment survey to supporters to elicit their political identity types or could also run further online survey experiments on different populations to test how individual-level pre-treatment characteristics (e.g., the identity salience, direction and intensity) shape responses to identity activation.

In regard to the target and sender of negative identity campaigning, I would generally expect negative identity campaigning to work between most party pairs. Research on negative partisanship has shown that it is not just limited to extreme parties – negative partisanship towards mainstream parties is also common (Anderson, McGregor and Stephenson, 2021). Negative party identity activation should thus generally work between mainstream parties, between extreme parties and between mainstream and extreme parties. Having said that, negative party identity activation should be particularly effective for mobilising supporters when dislike towards the disliked party is high. For example, negative identity activation used by a mainstream party against a highly disliked party, like the radical right, should be more impactful than when used against a less disliked party, such as a former coalition partner.

This thesis also examined how political parties may try to reduce negative feelings towards them. In particular, it showed that normative repackaging can be used to make previously unacceptable views seem more acceptable (Paper 3). I tested this argument at the example of gender equality and women's rights being used to defend anti-immigration views. Expanding the scope of the current research, I would expect that normative repackaging would also apply to other liberal democratic values (e.g., freedom of assembly, freedom of speech, individual autonomy see e.g., Ferrin Kriesi, 2016) and other unacceptable views (e.g., conspiracy beliefs, homophobia). While these examples are about how extreme right-wing beliefs might become more acceptable through normative repackaging, it could also be applied to understand how left-wing political views that are considered unacceptable by some voters, such as pro-abortion views or drug legalisation, are repackaged to become more acceptable to these voters.

However, I also expect that there are limits to the impact of normative repackaging. I expect that normative repackaging only works when the liberal democratic norm used evokes broad societal consensus and is not yet associated with a particular political party. When a political party consistently uses a liberal democratic norm and starts being associated with this, normative repackaging no longer works. For example, as

right-wing political actors have increasingly used freedom of speech in conjunction with concerns over "cancel culture" to justify the expression of nativist views, freedom of speech might start being associated with right-wing discourse, and therefore no longer function as a normative repackaging message.

Experimental research shows that framing a rally of an extremist group as permissible due to concerns over cancel culture indeed does not increase partisans' support for the rally (Fahey, Roberts and Utych, 2022). This is in contrast to the foundational framing survey experiments that show invoking freedom of speech increases support for permitting a rally of an extremist group (Nelson, Clawson and Oxley, 1997), which has been replicated in 2015 in the US (Mullinix et al., 2015). I would argue that this discrepancy is due to the fact that "cancel culture" as a reframed version of freedom of speech concerns is clearly partisan, and therefore does not work as a normative repackaging strategy. In other words, when normative repackaging becomes obvious – that a political party is trying to claim a universal liberal democratic norm to its advantage – then it should cease to be effective. Future research should test both if normative repackaging works applied to other issues and liberal democratic values, and if its impact is indeed limited by partisan cueing.

5.4 Policy implications

This thesis has shown that strong negative feelings towards political parties and political out-groups matter for attitudes and behaviour. Beyond the empirical and theoretical contributes to academic research on these topics, this thesis also has important implications for practitioners.

This thesis can help mainstream parties understand how extreme parties try to shake off accusations of racism and normalise previously unacceptable views. Paper 3 shows that political parties can use normative repackaging to make anti-immigration views more acceptable. When an anti-immigration statement is presented as a defence of women's rights and gender equality, voters think that saying something bad about immigrants is more acceptable. Normative repackaging allows extreme political parties to shift the boundaries of acceptable political discourse to their advantage.

To mainstream parties, these findings should be concerning: once a norm erodes, it is unclear if and how it can be strengthened again. An expansion to the extreme of what views and policies are seen as fair game in politics is also likely to hurt mainstream parties electorally. Mainstream parties now have to compete on this expanded issue dimension, and simply copying the extreme position is not a promising electoral strategy

(Krause, Cohen and Abou-Chadi, 2022). Mainstream parties should therefore try to curb the impact of normative repackaging. A good starting point would be not to copy and repeat narratives that tap into normative repackaging. Yet, there is evidence that mainstream German media outlets and politicians e.g., used narratives around the presence of refugees compromising women's rights and safety (Holzberg, Kolbe and Zaborowski, 2018; Wigger, Yendell and Herbert, 2021).

Rather than fuelling the legitimacy of these frames, mainstream political parties and media sources should be careful not to replicate normative repackaging. Beyond this, mainstream political parties could also try to actively challenge normative repackaging. I discussed in the previous section that I would expect normative repackaging only to work when the norm invoked reads as a universal, rather than a partisan cue to voters. When a liberal democratic norm, like freedom of speech and cancel culture concerns, becomes clearly associated with the rhetoric of one political party, normative repackaging may lose its bite.

Instead of waiting for this to happen, mainstream political parties could also directly call out extreme parties for using this strategy. For example, Macron tried to actively challenge Marine Le Pen's attempts to link her headscarf ban to gender equality and women's rights during the latest French Presidential election campaign. When Marine Le Pen quipped "You have not read my [proposed] law," Macron replied: "No, but I have read the French Constitution" (quoted in Kirby 2022a). Macron went on to argue that the headscarf ban goes against constitutional values and that the proposed policy would "create a civil war in certain areas of our cities" (Macron quoted in Kirby 2022a). By making it obvious to voters that the liberal democratic norm is not invoked in a genuine way, that it is incongruent with the liberal democratic norms, or with the wider aims and positions of the sender, mainstream parties could potentially stop the impact of normative repackaging in its tracks. Applied to gender-immigration messages, mainstream parties could e.g., clarify what gender equality entails, show voters that other radical right positions are not compatible with gender equality, and call this out as an opportunistic rhetorical strategy.

This thesis also has wider implications for the voter outreach efforts of political parties. Paper 1 shows that among negative partisans (voters without positive party attachment but who really like a political party), priming negative partisanship can also reduce their felt distance to other political parties. This suggests that feelings of closeness and proximity to other political parties may also, in the long run, grow out strong antipathy towards another political party. In other words, negative identities may also lead to positive ones in the future (Zhong et al., 2008). While it is not a given that these positive identities would not contain any out-group hostility, the possibility that voters

may feel connected to a party rather than just hostile to a political party seems like a relative improvement.

Political parties should explore how they can engage with negative partisans, and if positive feelings of attachment could supersede negative partisan affect. A key goal of this process should be to transform a negative identity, into a positive, forward-looking identity focused on common goals and solutions, rather than just adding a positive component to an otherwise dominant negative identity. While this type of identity building may be slow and costly, political parties have incentives to engage in this: positive party identification is waning, with obvious electoral costs to parties. Positive community building focused on retaining and attracting supporters should therefore already be a strategic priority for party elites.

Finally, the findings from this thesis also demonstrate that political parties and politicians need to be mindful of the consequences of activating out-group dislike. Paper 2 shows that political parties can use negative identity messaging to mobilise their supporters: rallying supporters around a disliked out-group generates a higher number of donations. However, Paper 2 also shows that this is not necessarily a winning strategy for political parties. Negative identity campaigning does not generate more money for a political campaign compared to not sending out an email at all, and is thus not an effective fundraising strategy. Exploratory analyses show that the party supporters who donated most in the past do not respond to the activation of negative feelings towards a political out-group. Going negative against the other side thus fails to mobilise the most committed party supporters – who have the highest likelihood of making a substantial donation to the campaign. All in all, negative identity campaigning may widen the donor pool but also entails costs for political parties. Considering the potentially harmful long-term implications of entrenching political out-group hostility, political parties may want to re-focus their resources on trying out alternative campaign strategies and analysing how different types of supporters respond to them.

To be clear, I am not arguing that political parties need to banish negative comments about other political parties from their campaigns – after all, critical engagement and debate is part and parcel of a functioning democracy. It is more about the type of disagreement: one can strongly criticise another political party or politician without resorting to uncivil attacks, or portraying this disagreement as an unsolvable, existential group conflict between "us" and "them". While some voters and party supporters are motivated by their desire to not see a rival party elected, political parties can also channel these negative motivations in ways that are less hostile, and less focused on group conflict.

Finally, if political parties choose to go negative to activate supporters, they should also consider how they integrate this into a wider strategy of positive outreach and campaigning. While discontent over a rival party's actions may jolt passive supporters into action once in a while, parties need to consider how to consistently engage these supporters and make them feel they are fighting for something worthwhile, rather than just against something. A shift towards positive identity building and campaigning may not just help parties build a pool of committed, engaged supporters, but could also alleviate pressing problems around democratic functioning and societal cohesion in an age of polarisation.

Appendix A

Appendix Paper 1

A.0.1 Balance checks

	Control	Negative prime	Positive prime
Variable	Mean	Mean	Mean
Age	48(16.72)	47(16.34)	48(17.28)
Women	0.52	0.51	0.50
Employed	0.53	0.59	0.57
Bachelor	0.23	0.24	0.28
Quebec	0.29	0.28	0.27
N	540	508	505

TABLE A.1: Pre-treatment means on key covariates, including age, gender (% of women), employment (% in full-time employment), education (% with an undergraduate degree as highest education), province (% living in Quebec).

A.0.2 Descriptives

	All groups			
Variable	Mean	SD	Min	Max
Left-right self-placement	5.12	2.18	0	10
Distance to most disliked party	3.89	2.71	0	10
Distance to most liked party	1.26	1.47	0	9

TABLE A.2: Means and standard deviations for key outcome variables pooled across all experimental conditions

A.0.3   Treatment uptake

Overall, 49% of respondents in the negative partisanship condition were able to name a party they dislike straight away (Q. Is there any party that you dislike more than all the others?). This was very similar in the positive partisanship condition, where 48% of respondents responded positively to the partisanship feeder question (Q. Is there any party that you like more than all the others?). If participants responded No/Don't know to this first question, they were encouraged one more time to name a party. 35.6% of respondents in the negative partisanship condition named a party then, and 38% of respondents in the positive partisanship condition. This means that overall, 86% of the individuals in the positive partisanship condition were given the main treatment (prompted to write a text about the party they like). The most liked parties were the Liberals, the Conservatives and the NDP in this order (see Figure A.4). Overall, 84.6% of participants in the negative partisanship condition were given the actual treatment (prompted to write a text about the party they dislike). The most commonly disliked parties were the Conservatives, the Liberals and the Bloc Quebecois in this order (see Figure A.5).

The main treatment manipulation consisted of writing a short text in response to the question: Q3. What do you like/dislike about this party? [textbox]. Respondents engaged well with the writing task. Of those respondents exposed to the writing prompt, the majority (71%) wrote something in response to the prompt. This means that of all those assigned to the positive or negative treatment, 60.1% wrote something.

The content written by respondents also suggests that the engagement with the treatment manipulation went in the intended direction. On average, individuals used around 3.2 unique words (tokens) to describe why they liked or disliked a party. As this metric excludes stopwords and punctuation, this means respondents mostly wrote a short sentence in response to the treatment. Respondents wrote slightly more in the negative partisanship condition than in the positive condition [2.9 in positive/3.5 in negative condition].

Looking at the actual content of the sentences written by respondents, we can see that these match onto the negative vs positive partisanship prompts. Negative words such as bad, no, don't, arrogant, racist, dishonest are among the most frequently used words in the negative partisanship condition. On the other hand, individuals in the positive partisanship condition used more positive words such as better, strong, great, best, values, open.
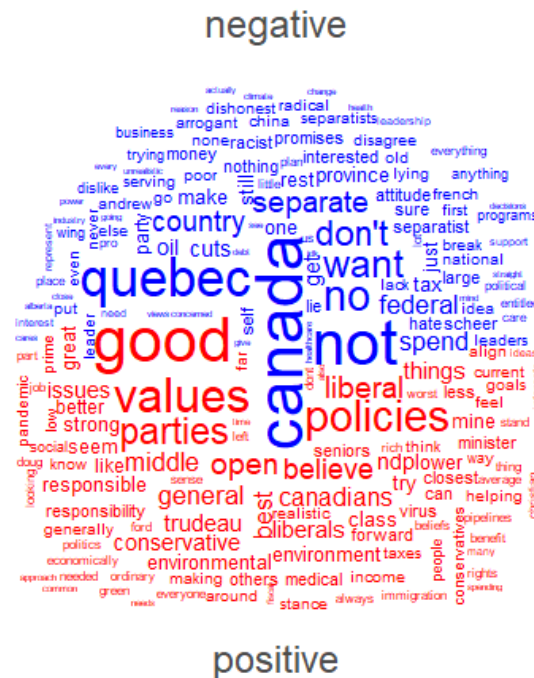
FIGURE A.1: Most common words in the negative and positive partisanship treatment condition

Example negative prime responses:

"Justin Trudeau (The Prime Mistake of Canada) and the rest of his peanut gallery" "Arrogant and self serving" "I do not believe they are honest with people" "Disagree with policies such as legalizing marijuana"

Example positive prime responses:

"Justin Trudeau has been doing good and proving himself" "seem more for the people not wealthy rich snobs" "I've always been conservative" "more rational around climate, environment, pipelines and other things"

A.0.4 Main results: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Control	1.25***	1.31***	1.17***	1.12***	1.13***	0.99***
	(0.08)	(0.10)	(0.15)	(0.15)	(0.18)	(0.21)
Treatment effects						
T1: Negative prime	0.10	0.11	0.42*	0.10	0.10	0.40*
	(0.11)	(0.11)	(0.22)	(0.12)	(0.11)	(0.22)
T2: Positive prime	−0.08	−0.10	−0.00	−0.08	−0.10	−0.02
	(0.11)	(0.11)	(0.18)	(0.11)	(0.11)	(0.18)
Pre-T covariates						
Closed partisan		−0.10	0.12		−0.08	0.12
		(0.11)	(0.20)		(0.11)	(0.20)
Non-partisan		−0.29**	−0.20		−0.29**	−0.22
		(0.13)	(0.21)		(0.13)	(0.21)
Negative partisan		−0.12	0.15		−0.10	0.16
		(0.15)	(0.23)		(0.15)	(0.23)
Interaction effects						
NegativePrime*Closedpartisan			−0.37			−0.36
			(0.28)			(0.28)
PositivePrime*Closedpartisan			−0.27			−0.25
			(0.26)			(0.25)
NegativePrime*Non-partisan			−0.37			−0.33
			(0.33)			(0.33)
PositivePrime*Non-partisan			0.09			0.13
			(0.28)			(0.29)
NegativePrime*Negativepartisan			−0.84***			−0.83***
			(0.32)			(0.32)
PositivePrime*Negativepartisan			0.12			0.12
			(0.40)			(0.40)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.01	0.02
Adj. R^2	0.00	0.00	0.01	0.00	0.00	0.01
Num. obs.	1086	1012	1012	1086	1012	1012
RMSE	1.47	1.41	1.41	1.47	1.42	1.41

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.3: Treatment to Control ITT, DV: Distance to most liked party
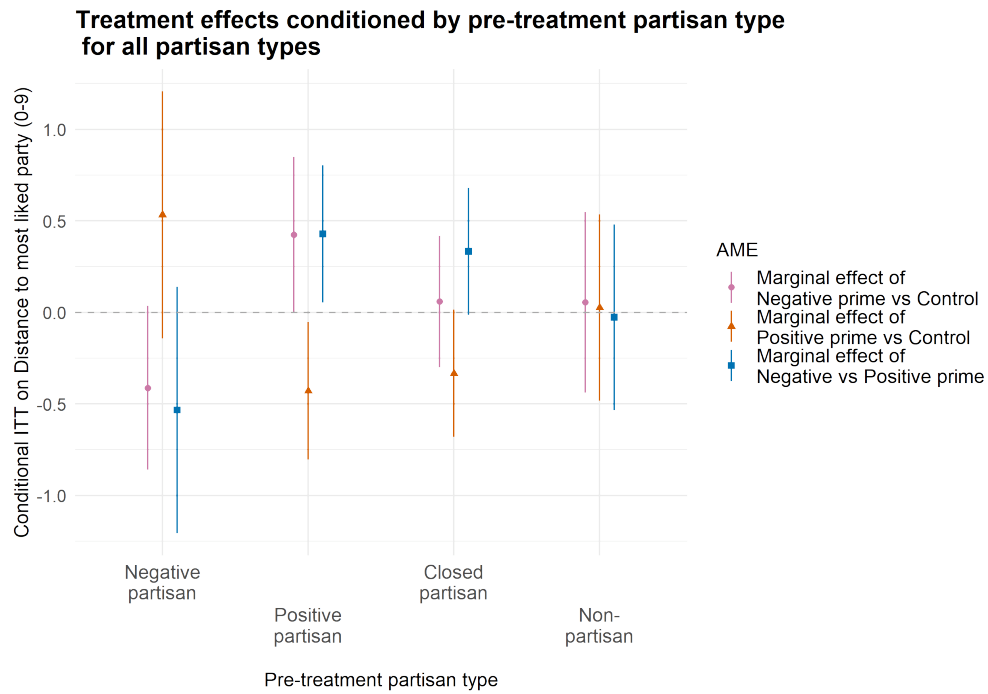
FIGURE A.2: Overview of treatment effects conditioned by partisan type for all partisans

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	1.17***	1.26***	1.16***	1.15***	1.28***	1.16***
	(0.07)	(0.10)	(0.10)	(0.18)	(0.22)	(0.23)
T1: Negative prime	0.19*	0.22**	0.43**	0.18	0.23**	0.44**
	(0.11)	(0.11)	(0.19)	(0.11)	(0.11)	(0.19)
Closed partisan		−0.20	−0.15		−0.20	−0.15
		(0.13)	(0.16)		(0.13)	(0.17)
Non-partisan		−0.32**	−0.11		−0.30*	−0.09
		(0.16)	(0.19)		(0.16)	(0.20)
Negative partisan		−0.26	0.27		−0.26	0.27
		(0.19)	(0.33)		(0.19)	(0.34)
NegativePrime*Closedpartisan			−0.10			−0.11
			(0.26)			(0.26)
NegativePrime*Non-partisan			−0.46			−0.47
			(0.32)			(0.33)
NegativePrime*Negativepartisan			−0.96**			−0.96**
			(0.39)			(0.40)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.02	0.03
Adj. R^2	0.00	0.01	0.01	−0.00	0.00	0.01
Num. obs.	715	664	664	715	664	664
RMSE	1.46	1.40	1.40	1.46	1.40	1.40

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.4: Treatment to Treatment comparison ITT, DV: Distance to most liked party

A.0.5 Main results: Distance to disliked party

	M1	M2	M3	M4	M5	M6
Ref:Control	3.89***	3.00***	3.20***	4.60***	3.27***	3.48***
	(0.15)	(0.18)	(0.27)	(0.31)	(0.33)	(0.37)
Treatment effects						
T1: Negative prime	−0.02	0.09	−0.35	−0.03	0.07	−0.36
	(0.21)	(0.20)	(0.35)	(0.21)	(0.19)	(0.35)
T2: Positive prime	0.02	0.25	0.11	−0.02	0.21	0.10
	(0.20)	(0.19)	(0.33)	(0.20)	(0.19)	(0.33)
Pre-T covariates						
Closed partisan		2.08***	1.84***		2.05***	1.85***
		(0.19)	(0.36)		(0.19)	(0.36)
Non-partisan		−0.70***	−1.24***		−0.69***	−1.28***
		(0.21)	(0.34)		(0.21)	(0.34)
Negative partisan		0.59**	0.52		0.53**	0.53
		(0.26)	(0.45)		(0.26)	(0.46)
Interaction effects						
NegativePrime*Closedpartisan			0.55			0.51
			(0.49)			(0.48)
PositivePrime*Closedpartisan			0.14			0.04
			(0.47)			(0.47)
NegativePrime*Non-partisan			1.23**			1.30**
			(0.54)			(0.54)
PositivePrime*Non-partisan			0.47			0.57
			(0.48)			(0.47)
NegativePrime*Negativepartisan			0.24			0.17
			(0.61)			(0.61)
PositivePrime*Negativepartisan			−0.15			−0.28
			(0.70)			(0.68)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.17	0.17	0.03	0.19	0.19
Adj. R^2	−0.00	0.16	0.16	0.02	0.18	0.18
Num. obs.	1075	1000	1000	1075	1000	1000
RMSE	2.71	2.46	2.46	2.69	2.44	2.44

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.5: Treatment to Control ITT, DV: Distance to disliked party

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	3.91***	3.17***	3.32***	4.81***	3.49***	3.70***
	(0.14)	(0.17)	(0.20)	(0.36)	(0.37)	(0.39)
T1: Negative prime	−0.03	−0.15	−0.46	0.01	−0.12	−0.47
	(0.20)	(0.19)	(0.30)	(0.20)	(0.19)	(0.30)
Closed partisan		2.17***	1.97***		2.11***	1.87***
		(0.22)	(0.30)		(0.22)	(0.30)
Non-partisan		−0.43	−0.77**		−0.41	−0.75**
		(0.26)	(0.33)		(0.26)	(0.33)
Negative partisan		0.56*	0.37		0.46	0.24
		(0.32)	(0.53)		(0.32)	(0.51)
NegativePrime*Closedpartisan			0.41			0.48
			(0.44)			(0.45)
NegativePrime*Non-partisan			0.76			0.75
			(0.53)			(0.53)
NegativePrime*Negativepartisan			0.39			0.45
			(0.67)			(0.65)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.17	0.18	0.03	0.19	0.19
Adj. R^2	−0.00	0.17	0.17	0.02	0.18	0.18
Num. obs.	700	650	650	700	650	650
RMSE	2.64	2.37	2.38	2.60	2.36	2.36

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.6: Treatment to Treatment comparison ITT, DV: Distance to disliked party

A.0.6   Validity checks: Pre-treatment partisanship measure

This section presents validity checks for my pre-treatment partisanship measures based on party like-dislike ratings. I show that my partisanship measures are correlated with standard partisanship questions, as well as partisanship items included in the experimental survey. This bolsters my claim that these measures are picking up on partisan affect, and are measuring what they are set out to do.

The online panel, in which this experiment was embedded, asked respondents one standard positive party identification question ("In federal politics, do you usually think of yourself as a: Liberal/Conservative/NDP/Bloc/Green/Other/DK/None") pre-treatment. I recode this into a binary variable (1=any party, 0= no party/DK) and test how this is associated with my own binary positive partisan affect measure (1=positive partisan or closed partisan, 0=non-partisan or negative partisan). I collapse positive and closed partisans into one category here as both of these groups of partisans display strong positive affect, and should thus feature among the individuals who also answer positively to the standard party identification question. I find that my own positive partisan affect measure and the standard positive party identification measure are positively correlated (correlation coefficient: 0.44, p<0.001, also see Table A.7).
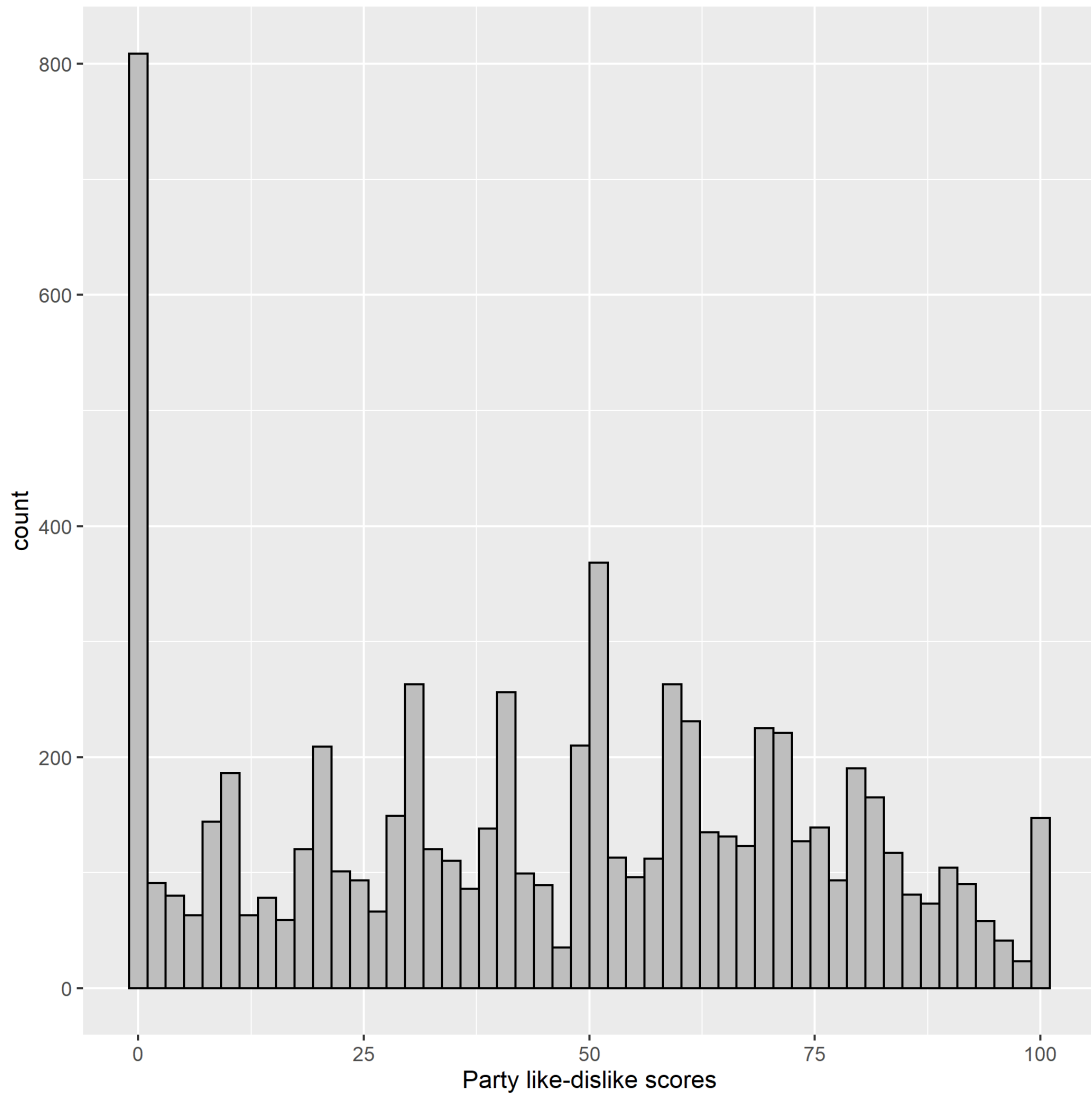
FIGURE A.3: Histogram of party like-dislike scores in the full experimental sample
(N=1553)

	No positive partisan (thermometer measure)	Positive partisan (thermometer measure)
No positive partisan (standard measure)	211	79
Positive partisan (standard measure)	349	889

TABLE A.7: Frequency table: positive partisanship standard vs thermometer measure

Respondents in the treatment groups were asked which party they either really like or dislike, and are also asked two additional questions about their partisan identity. If the pre-treatment measures of partisanship measure positive and negative partisan affect, then we would expect the political party in the pre-treatment measure to be congruent with the political party in the treatment partisanship feeder question. In other words, if a

respondent was classified as a negative partisan against the Conservative Party based on pre-treatment thermometer scores, we would expect that this respondent would answer "Conservative Party" to the question "Which party do you dislike more than all the others?". The same should be true for positive partisans. If a respondent was classified as a positive partisan for the Liberal Party pre-treatment, we would expect that this respondent would answer "Liberal Party" to the question "Which party do you like more than all the others?". Figure A.5 and Figure A.4 show which political parties were the most liked and disliked according to the pre-treatment measure of partisanship, and the in-treatment measure of partisanship. Both figures show that the pre-treatment and in-treatment distributions of which parties are most liked and disliked are similar.
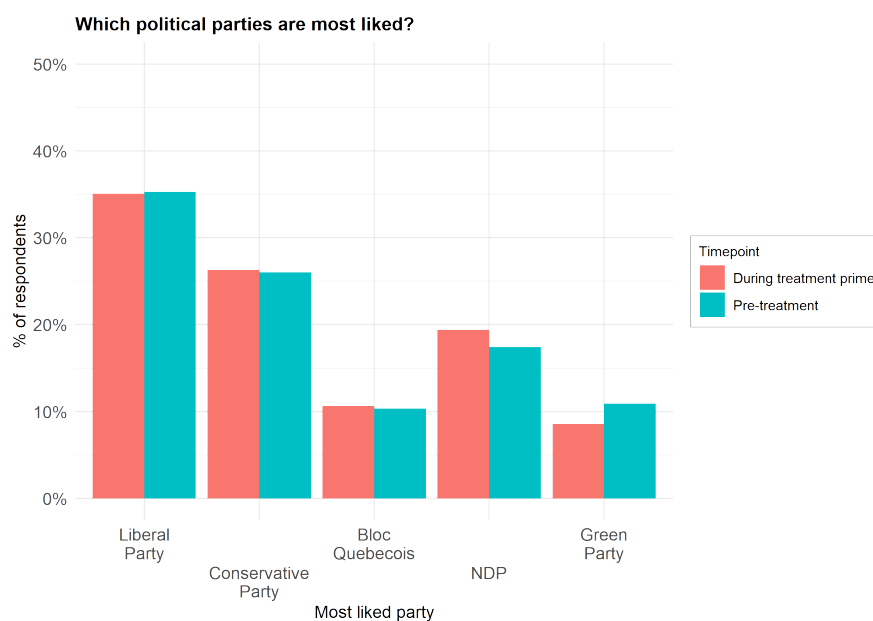
FIGURE A.4: Most liked party before treatment (in the whole sample) and during treatment (among respondents in the positive prime condition)

Among respondents who were assigned to the negative partisanship prime and named a party they dislike, 74% (or 320 out of 430 respondents) said they dislike the same political party that they gave the lowest thermometer score pre-treatment. Among respondents who were assigned to the negative partisanship prime, were classified as negative or closed partisans pre-treatment and named a party they dislike, 79% (or 195 out of 247 respondents) said they dislike the same political party that they were classified as having negative partisan affect towards pre-treatment. Among respondents who were assigned to the positive partisanship prime and named a party they like, 87% (or 384 out of 442 respondents) said they like the same political party that they gave the highest thermometer score pre-treatment. Among respondents who were classified as pre-treatment positive partisans, assigned to the positive partisanship prime and named a party they like, 90% (or 291 out of 323 respondents) said they like the same
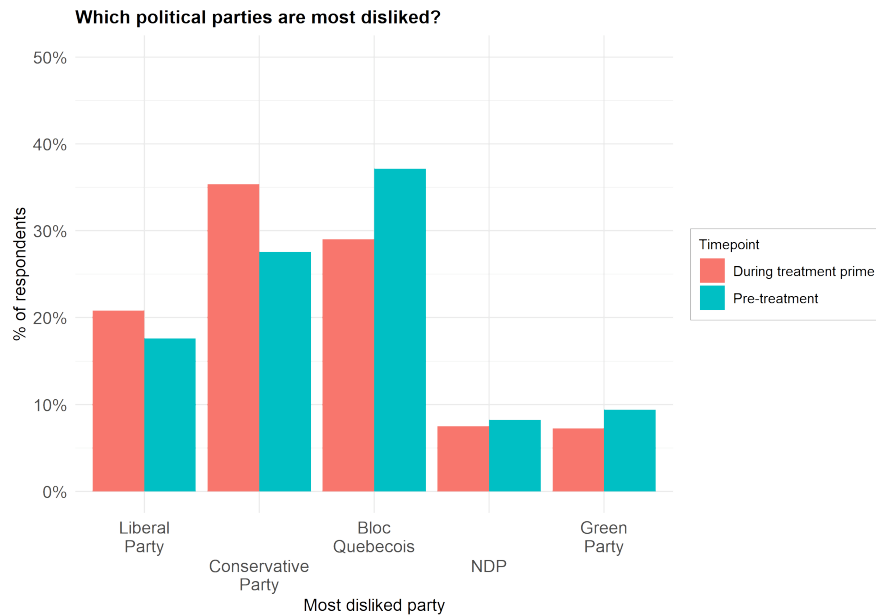
FIGURE A.5: Most disliked party before treatment (in the whole sample) and during treatment (among respondents in the negative prime condition)

political party that they were classified as having positive partisanship towards pre-treatment. This shows that a vast majority of respondents gave answers to the partisan affect questions that are congruent with their pre-treatment partisanship categorisation. This is encouraging in terms of the partisanship categorisation based on thermometer scores. Respondents' self-reported partisan affect before the treatment primes matches the pre-treatment categorisation well.

A.0.7   Alternative outcome measure

It was pre-registered that distance to the most liked party would be the absolute distance on the left-right scale to the party with the least subjective left-right distance. Distance to the most disliked party was pre-registered as the distance to the political party which the respondent placed furthest away from themselves. However, this measurement of distance risks underestimating potential changes in distance because it assumes that the most liked party at baseline is already the party that the respondent places closest to themselves ideologically. Yet, the treatment may alter the ordering of ideological proximity of political parties. Determining the most liked and disliked party through pre-treatment thermometer scores is more appropriate because it is not mechanically related to the left-right distance outcome. Thermometer scores are also used for determining the pre-treatment partisan type, and this outcome measure is therefore more consistent with the other measures used in the analysis. Still, the results for the original pre-registered outcomes are presented in this section. Table A.8 and Table A.9 show the results for the

distance to most disliked party, while Table A.10 and Table A.11 show the results for the distance to the most disliked party. The results using this alternative outcome measure are substantively similar to the main results. The negative prime (compared to the Control) has a different effect on negative, compared to positive partisans, and makes negative partisans feel closer to their most-liked party, even when using the alternative outcome measure (see Table A.10). The coefficient on the effect of the negative vs positive prime (see Table A.11) is not significant but substantively similar to the one reported in the main results.

	M1	M2	M3	M4	M5	M6
Ref:Control	4.96***	4.25***	4.29***	5.82***	4.78***	4.82***
	(0.13)	(0.16)	(0.24)	(0.26)	(0.29)	(0.33)
Treatment effects						
T1: Negative prime	−0.14	−0.04	−0.15	−0.16	−0.06	−0.18
	(0.18)	(0.17)	(0.32)	(0.18)	(0.17)	(0.32)
T2: Positive prime	−0.24	−0.08	−0.08	−0.30*	−0.14	−0.10
	(0.18)	(0.17)	(0.30)	(0.18)	(0.17)	(0.30)
Pre-T covariates						
Closed partisan		1.79***	1.79***		1.72***	1.77***
		(0.17)	(0.31)		(0.17)	(0.31)
Non-partisan		−0.89***	−1.14***		−0.90***	−1.22***
		(0.20)	(0.38)		(0.20)	(0.37)
Negative partisan		0.41*	0.45		0.34	0.40
		(0.23)	(0.38)		(0.23)	(0.38)
Interaction effects						
NegativePrime*Closedpartisan			0.08			0.03
			(0.42)			(0.42)
PositivePrime*Closedpartisan			−0.08			−0.20
			(0.41)			(0.41)
NegativePrime*Non-partisan			0.68			0.80
			(0.54)			(0.53)
PositivePrime*Non-partisan			0.15			0.24
			(0.48)			(0.48)
NegativePrime*Negativepartisan			−0.17			−0.13
			(0.55)			(0.55)
PositivePrime*Negativepartisan			0.07			−0.03
			(0.60)			(0.58)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.18	0.18	0.03	0.19	0.20
Adj. R^2	−0.00	0.17	0.17	0.03	0.19	0.18
Num. obs.	1118	1036	1036	1118	1036	1036
RMSE	2.46	2.21	2.22	2.42	2.20	2.20

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.8: Treatment to Control ITT, DV: Distance to disliked party (0-10) [alternative measure]

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	4.71***	4.15***	4.21***	5.68***	4.69***	4.77***
	(0.12)	(0.15)	(0.18)	(0.29)	(0.31)	(0.33)
T1: Negative prime	0.10	0.05	−0.07	0.14	0.07	−0.08
	(0.18)	(0.17)	(0.28)	(0.17)	(0.17)	(0.28)
Closed partisan		1.78***	1.70***		1.69***	1.58***
		(0.19)	(0.27)		(0.20)	(0.28)
Non-partisan		−0.75***	−0.99***		−0.74***	−1.00***
		(0.24)	(0.29)		(0.24)	(0.29)
Negative partisan		0.38	0.52		0.28	0.38
		(0.30)	(0.46)		(0.30)	(0.45)
NegativePrime*Closedpartisan			0.16			0.22
			(0.39)			(0.39)
NegativePrime*Non-partisan			0.53			0.56
			(0.49)			(0.48)
NegativePrime*Negativepartisan			−0.24			−0.16
			(0.61)			(0.60)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.17	0.17	0.04	0.19	0.19
Adj. R^2	−0.00	0.16	0.16	0.03	0.18	0.17
Num. obs.	736	680	680	736	680	680
RMSE	2.40	2.16	2.16	2.36	2.14	2.15

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.9: Treatment to Treatment comparison ITT, DV: Distance to disliked party (0-10) [alternative measure]

	M1	M2	M3	M4	M5	M6
Ref:Control	0.55***	0.44***	0.37***	0.74***	0.60***	0.54***
	(0.05)	(0.05)	(0.06)	(0.11)	(0.10)	(0.10)
Treatment effects						
T1: Negative prime	0.07	0.09	0.18*	0.07	0.10	0.18*
	(0.07)	(0.06)	(0.10)	(0.07)	(0.06)	(0.10)
T2: Positive prime	−0.07	−0.04	0.05	−0.07	−0.05	0.05
	(0.06)	(0.05)	(0.08)	(0.06)	(0.05)	(0.08)
Pre-T covariates						
Closed partisan		0.15***	0.21**		0.13**	0.21**
		(0.06)	(0.09)		(0.06)	(0.09)
Non-partisan		−0.09	−0.02		−0.09	−0.03
		(0.07)	(0.10)		(0.07)	(0.10)
Negative partisan		0.14	0.32*		0.13	0.31*
		(0.09)	(0.16)		(0.09)	(0.16)
Interaction effects						
NegativePrime*Closedpartisan			−0.06			−0.07
			(0.14)			(0.14)
PositivePrime*Closedpartisan			−0.11			−0.13
			(0.13)			(0.13)
NegativePrime*Non-partisan			−0.02			−0.01
			(0.20)			(0.19)
PositivePrime*Non-partisan			−0.16			−0.15
			(0.13)			(0.13)
NegativePrime*Negativepartisan			−0.41*			−0.40*
			(0.21)			(0.21)
PositivePrime*Negativepartisan			−0.14			−0.15
			(0.21)			(0.21)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.02	0.03	0.01	0.03	0.03
Adj. R^2	0.00	0.01	0.02	0.00	0.02	0.02
Num. obs.	1118	1036	1036	1118	1036	1036
RMSE	0.88	0.77	0.77	0.88	0.77	0.77

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.10: Treatment to Control ITT, DV: Distance to most liked party [alternative measure]

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	0.48***	0.42***	0.42***	0.69***	0.64***	0.65***
	(0.04)	(0.05)	(0.06)	(0.12)	(0.12)	(0.13)
T1: Negative prime	0.14**	0.14**	0.13	0.14**	0.15**	0.13
	(0.06)	(0.06)	(0.10)	(0.06)	(0.06)	(0.10)
Closed partisan		0.13*	0.10		0.11	0.07
		(0.07)	(0.09)		(0.07)	(0.09)
Non-partisan		−0.12	−0.18**		−0.12	−0.18**
		(0.09)	(0.08)		(0.09)	(0.08)
Negative partisan		0.03	0.18		0.01	0.16
		(0.09)	(0.14)		(0.09)	(0.14)
NegativePrime*Closedpartisan			0.05			0.06
			(0.14)			(0.14)
NegativePrime*Non-partisan			0.14			0.14
			(0.19)			(0.19)
NegativePrime*Negativepartisan			−0.27			−0.27
			(0.19)			(0.19)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.01	0.02	0.03	0.02	0.03	0.03
Adj. R^2	0.01	0.01	0.02	0.01	0.02	0.02
Num. obs.	736	680	680	736	680	680
RMSE	0.84	0.77	0.77	0.84	0.77	0.77

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.11: Treatment to Treatment comparison ITT, DV: Distance to most liked party [alternative measure]

A.0.8 Additional results: Distance to second-liked party

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	2.14***	1.84***	1.86***	3.52***	3.12	3.16***
	(0.11)	(0.14)	(0.17)	(0.84)		(0.71)
T1: Negative prime	0.06	−0.04	−0.09	0.12	−0.04	−0.07
	(0.16)	(0.16)	(0.24)	(0.16)		(0.24)
Closed partisan		0.96***	0.87***		0.85	0.76***
		(0.19)	(0.27)			(0.27)
Non-partisan		−0.24	−0.39		−0.21	−0.38
		(0.19)	(0.25)			(0.27)
Negative partisan		0.19	0.64		0.17	0.73
		(0.27)	(0.44)			(0.45)
NegativePrime*Closedpartisan			0.20			0.20
			(0.39)			(0.37)
NegativePrime*Non-partisan			0.35			0.37
			(0.38)			(0.42)
NegativePrime*Negativepartisan			−0.78			−0.96*
			(0.55)			(0.57)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.06	0.06	0.16	0.20	0.20
Adj. R^2	−0.00	0.05	0.05	0.10	0.12	0.12
Num. obs.	703	654	654	703	654	654
RMSE	2.10	2.01	2.01	1.99	1.94	1.93

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.12: Treatment to Treatment comparison ITT, DV: Distance to second most liked party

	M1	M2	M3	M4	M5	M6
Ref:Control	2.13***	1.73***	1.76***	3.70***	3.23***	3.29***
	(0.11)	(0.15)	(0.21)	(0.91)	(0.69)	(0.67)
Treatment effects						
T1: Negative prime	0.07	0.07	0.01	0.09	0.07	−0.00
	(0.16)	(0.16)	(0.27)	(0.16)	(0.16)	(0.27)
T2: Positive prime	0.01	0.11	0.09	−0.03	0.08	0.10
	(0.16)	(0.16)	(0.27)	(0.16)	(0.16)	(0.27)
Pre-T covariates						
Closed partisan		0.93***	0.87***		0.82***	0.83***
		(0.16)	(0.29)		(0.16)	(0.28)
Non-partisan		−0.23	−0.22		−0.23	−0.27
		(0.16)	(0.32)		(0.17)	(0.33)
Negative partisan		0.17	0.14		0.13	0.08
		(0.20)	(0.32)		(0.21)	(0.31)
Interaction effects						
NegativePrime*Closedpartisan			0.20			0.12
			(0.40)			(0.39)
PositivePrime*Closedpartisan			−0.00			−0.15
			(0.40)			(0.40)
NegativePrime*Non-partisan			0.19			0.32
			(0.43)			(0.44)
PositivePrime*Non-partisan			−0.17			−0.15
			(0.40)			(0.42)
NegativePrime*Negativepartisan			−0.27			−0.23
			(0.46)			(0.47)
PositivePrime*Negativepartisan			0.51			0.56
			(0.54)			(0.55)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.05	0.06	0.12	0.16	0.16
Adj. R^2	−0.00	0.05	0.05	0.08	0.11	0.10
Num. obs.	1072	1000	1000	1072	1000	1000
RMSE	2.12	2.04	2.04	2.04	1.98	1.98

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.13: Treatment to Control ITT, DV: Distance to second most liked party

A.0.9 Additional results: Political participation

In this section, I present results for additional outcomes related to political participation. The treatment primes did not have a significant effect on political participation. Results for turnout are reported in Tables A.14 and A.15. Turnout is coded as a continuous measure, ranging from 1 to 5, with higher values denoting a greater likelihood to turn out to vote. (Q: If a federal election was held tomorrow, how likely is it that you would vote? (1=highly likely to 5=very unlikely)). Results for political participation are reported in Tables A.16 and A.17. Political participation is coded as an index ranging from 0-5, with higher values denoting a higher intention to participate in politics (contacting a representative, donating to a party, donating to a charity, sharing political content on social media, signing a petition).

	M1	M2	M3	M4	M5	M6
Ref:Control	4.46***	4.39***	4.42***	4.19***	4.38***	4.27***
	(0.05)	(0.07)	(0.10)	(0.76)	(0.72)	(0.76)
Treatment effects						
T1: Negative prime	−0.05	−0.06	−0.06	−0.04	−0.05	−0.04
	(0.07)	(0.07)	(0.13)	(0.06)	(0.06)	(0.12)
T2: Positive prime	−0.00	0.00	−0.07	−0.02	−0.03	−0.03
	(0.06)	(0.06)	(0.13)	(0.06)	(0.06)	(0.11)
Pre-T covariates						
Closed partisan		0.33***	0.25**		0.20***	0.17
		(0.06)	(0.12)		(0.06)	(0.11)
Non-partisan		−0.11	−0.22		−0.13*	−0.20
		(0.09)	(0.15)		(0.08)	(0.14)
Negative partisan		−0.01	0.14		−0.14	0.03
		(0.10)	(0.14)		(0.10)	(0.13)
Interaction effects						
NegativePrime*Closedpartisan			0.03			0.02
			(0.16)			(0.15)
PositivePrime*Closedpartisan			0.19			0.06
			(0.15)			(0.14)
NegativePrime*Non-partisan			0.08			0.00
			(0.23)			(0.20)
PositivePrime*Non-partisan			0.25			0.20
			(0.20)			(0.18)
NegativePrime*Negativepartisan			−0.19			−0.15
			(0.22)			(0.21)
PositivePrime*Negativepartisan			−0.36			−0.49**
			(0.26)			(0.24)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.03	0.04	0.19	0.23	0.24
Adj. R^2	−0.00	0.03	0.03	0.16	0.20	0.20
Num. obs.	1490	1338	1338	1490	1338	1338
RMSE	1.04	0.97	0.97	0.95	0.89	0.88

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.14: Treatment to Control ITT, DV: Turnout

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	4.45***	4.38***	4.35***	4.18***	4.29***	4.23***
	(0.05)	(0.07)	(0.08)	(0.71)	(0.70)	(0.73)
T1: Negative prime	−0.05	−0.06	0.02	−0.02	−0.02	−0.02
	(0.07)	(0.07)	(0.12)	(0.06)	(0.06)	(0.10)
Closed partisan		0.36***	0.45***		0.21***	0.23**
		(0.07)	(0.09)		(0.07)	(0.09)
Non-partisan		−0.05	0.03		−0.09	−0.01
		(0.11)	(0.14)		(0.09)	(0.12)
Negative partisan		−0.11	−0.22		−0.28**	−0.49**
		(0.14)	(0.22)		(0.14)	(0.20)
NegativePrime*Closedpartisan			−0.17			−0.04
			(0.14)			(0.13)
NegativePrime*Non-partisan			−0.17			−0.18
			(0.21)			(0.19)
NegativePrime*Negativepartisan			0.16			0.35
			(0.28)			(0.27)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.04	0.04	0.21	0.25	0.25
Adj. R^2	−0.00	0.03	0.03	0.16	0.20	0.20
Num. obs.	976	875	875	976	875	875
RMSE	1.04	0.98	0.98	0.95	0.89	0.89

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.15: Treatment to Treatment comparison ITT, DV: Turnout

	M1	M2	M3	M4	M5	M6
Ref:Control	3.64***	3.47***	3.60***	4.27***	4.08***	4.18***
	(0.04)	(0.06)	(0.09)	(0.54)	(0.50)	(0.46)
Treatment effects						
T1: Negative prime	0.02	0.00	−0.09	−0.00	−0.03	−0.05
	(0.06)	(0.07)	(0.12)	(0.06)	(0.07)	(0.12)
T2: Positive prime	−0.01	0.04	−0.22*	−0.04	0.00	−0.18
	(0.06)	(0.07)	(0.13)	(0.06)	(0.07)	(0.12)
Pre-T covariates						
Closed partisan		0.07	−0.14		0.01	−0.12
		(0.07)	(0.12)		(0.07)	(0.12)
Non-partisan		0.28***	0.16		0.15*	0.09
		(0.08)	(0.14)		(0.08)	(0.13)
Negative partisan		0.47***	0.29**		0.36***	0.26*
		(0.09)	(0.14)		(0.09)	(0.15)
Interaction effects						
NegativePrime*Closedpartisan			0.15			0.06
			(0.17)			(0.17)
PositivePrime*Closedpartisan			0.43***			0.30*
			(0.17)			(0.17)
NegativePrime*Non-partisan			0.04			−0.04
			(0.19)			(0.19)
PositivePrime*Non-partisan			0.27			0.21
			(0.19)			(0.19)
NegativePrime*Negativepartisan			0.18			0.06
			(0.20)			(0.20)
PositivePrime*Negativepartisan			0.36			0.24
			(0.22)			(0.23)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.03	0.04	0.10	0.12	0.12
Adj. R^2	−0.00	0.02	0.03	0.06	0.07	0.07
Num. obs.	1348	1207	1207	1348	1207	1207
RMSE	0.95	0.94	0.94	0.92	0.92	0.92

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.16: Treatment to Control ITT, DV: Political participation

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	3.64***	3.46***	3.38***	4.05	3.89	3.81
	(0.05)	(0.07)	(0.09)			
T1: Negative prime	0.03	−0.04	0.13	0.02	−0.05	0.10
	(0.06)	(0.07)	(0.12)			
Closed partisan		0.16*	0.30**		0.07	0.19
		(0.08)	(0.12)			
Non-partisan		0.32***	0.43***		0.16	0.26
		(0.10)	(0.14)			
Negative partisan		0.56***	0.65***		0.41	0.51
		(0.11)	(0.17)			
NegativePrime*Closedpartisan			−0.28*			−0.25
			(0.17)			
NegativePrime*Non-partisan			−0.23			−0.23
			(0.20)			
NegativePrime*Negativepartisan			−0.18			−0.20
			(0.22)			
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.04	0.04	0.13	0.15	0.15
Adj. R^2	−0.00	0.03	0.03	0.07	0.08	0.08
Num. obs.	887	794	794	887	794	794
RMSE	0.96	0.95	0.95	0.93	0.93	0.93

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.17: Treatment to Treatment comparison ITT, DV: Political participation

A.0.10   Additional results: Voting for most-liked party

	M1	M2	M3	M4	M5	M6
Ref:Control	0.60***	0.64***	0.67***	0.11	0.30**	0.30**
	(0.02)	(0.03)	(0.04)	(0.11)	(0.12)	(0.12)
Treatment effects						
T1: Negative prime	0.02	0.01	−0.01	0.01	−0.00	−0.02
	(0.03)	(0.03)	(0.06)	(0.03)	(0.03)	(0.06)
T2: Positive prime	0.01	−0.00	−0.07	0.01	−0.01	−0.06
	(0.03)	(0.03)	(0.06)	(0.03)	(0.03)	(0.06)
Pre-T covariates						
Closed partisan		0.16***	0.11**		0.13***	0.08
		(0.03)	(0.05)		(0.03)	(0.05)
Non-partisan		−0.24***	−0.28***		−0.24***	−0.28***
		(0.04)	(0.06)		(0.04)	(0.07)
Negative partisan		−0.17***	−0.19***		−0.20***	−0.21***
		(0.04)	(0.07)		(0.04)	(0.07)
Interaction effects						
NegativePrime*Closedpartisan			0.06			0.06
			(0.07)			(0.07)
PositivePrime*Closedpartisan			0.09			0.07
			(0.07)			(0.07)
NegativePrime*Non-partisan			−0.01			−0.02
			(0.09)			(0.09)
PositivePrime*Non-partisan			0.11			0.11
			(0.09)			(0.09)
NegativePrime*Negativepartisan			−0.02			−0.01
			(0.10)			(0.10)
PositivePrime*Negativepartisan			0.08			0.04
			(0.11)			(0.11)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.10	0.11	0.06	0.15	0.15
Adj. R^2	−0.00	0.10	0.10	0.03	0.12	0.11
Num. obs.	1553	1392	1392	1553	1392	1392
RMSE	0.49	0.46	0.46	0.48	0.45	0.45

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.18: Treatment to Control ITT, DV: Voting for preferred party

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	0.61***	0.63***	0.60***	0.10	0.24	0.19
	(0.02)	(0.03)	(0.04)	(0.14)	(0.15)	(0.15)
T1: Negative prime	0.01	0.01	0.05	0.01	0.01	0.06
	(0.03)	(0.03)	(0.06)	(0.03)	(0.03)	(0.06)
Closed partisan		0.19***	0.20***		0.16***	0.17***
		(0.04)	(0.05)		(0.04)	(0.05)
Non-partisan		−0.22***	−0.17**		−0.21***	−0.15**
		(0.05)	(0.07)		(0.05)	(0.07)
Negative partisan		−0.17***	−0.11		−0.18***	−0.15*
		(0.05)	(0.08)		(0.06)	(0.09)
NegativePrime*Closedpartisan			−0.03			−0.04
			(0.07)			(0.07)
NegativePrime*Non-partisan			−0.12			−0.14
			(0.09)			(0.10)
NegativePrime*Negativepartisan			−0.09			−0.07
			(0.11)			(0.11)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.11	0.11	0.09	0.17	0.17
Adj. R^2	−0.00	0.10	0.10	0.04	0.12	0.12
Num. obs.	1013	907	907	1013	907	907
RMSE	0.49	0.46	0.46	0.48	0.45	0.45

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.19: Treatment to Treatment comparison ITT, DV: Voting for preferred party
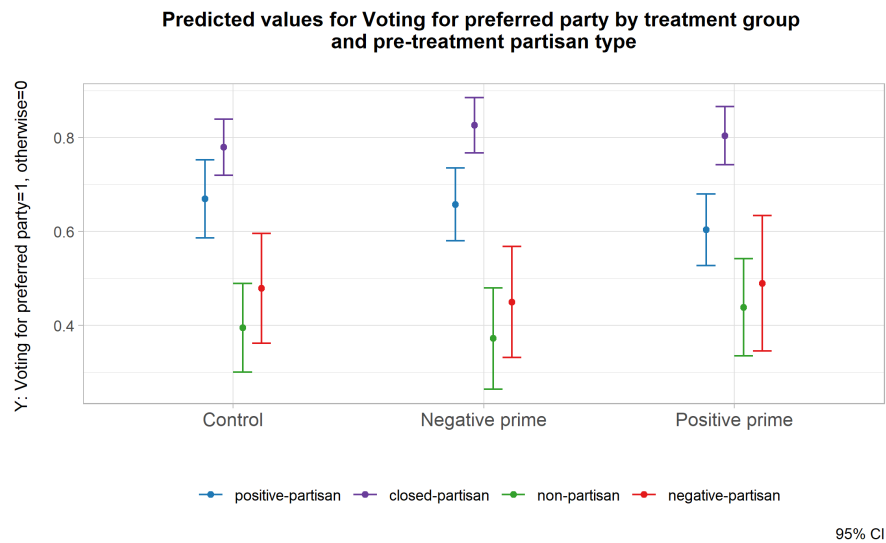
FIGURE A.6: Voting for most-liked party by treatment condition and pre-treatment partisan type

A.0.11 Additional results: Left-right self-placement

	M1	M2	M3	M4	M5	M6
Ref:Control	5.03***	5.40***	5.27***	5.46***	5.81***	5.71***
	(0.11)	(0.15)	(0.21)	(0.23)	(0.27)	(0.31)
Treatment effects						
T1: Negative prime	0.19	0.09	0.28	0.22	0.12	0.24
	(0.15)	(0.16)	(0.29)	(0.15)	(0.16)	(0.29)
T2: Positive prime	0.07	0.01	0.19	0.09	0.03	0.22
	(0.15)	(0.16)	(0.29)	(0.15)	(0.16)	(0.29)
Pre-T covariates						
Closed partisan		−0.46***	−0.35		−0.47***	−0.38
		(0.17)	(0.29)		(0.17)	(0.29)
Non-partisan		−0.43***	0.02		−0.44***	−0.01
		(0.16)	(0.27)		(0.16)	(0.28)
Negative partisan		−0.40**	−0.36		−0.36**	−0.35
		(0.18)	(0.32)		(0.18)	(0.31)
Interaction effects						
NegativePrime*Closedpartisan			−0.12			−0.03
			(0.42)			(0.41)
PositivePrime*Closedpartisan			−0.18			−0.18
			(0.42)			(0.41)
NegativePrime*Non-partisan			−0.73*			−0.63
			(0.39)			(0.40)
PositivePrime*Non-partisan			−0.62*			−0.67*
			(0.37)			(0.38)
NegativePrime*Negativepartisan			−0.10			−0.00
			(0.42)			(0.42)
PositivePrime*Negativepartisan			0.04			0.04
			(0.47)			(0.44)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.01	0.04	0.05	0.05
Adj. R^2	−0.00	0.01	0.00	0.03	0.04	0.04
Num. obs.	1210	1111	1111	1210	1111	1111
RMSE	2.18	2.17	2.17	2.14	2.13	2.14

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.20: Treatment to Control ITT, DV: Left-right self-placement (0-10)

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	5.10***	5.46***	5.45***	5.47***	5.79***	5.82***
	(0.11)	(0.16)	(0.20)	(0.27)	(0.32)	(0.35)
T1: Negative prime	0.12	0.08	0.09	0.13	0.08	0.03
	(0.15)	(0.16)	(0.28)	(0.15)	(0.16)	(0.29)
Closed partisan		−0.50**	−0.53*		−0.49**	−0.56*
		(0.21)	(0.29)		(0.21)	(0.29)
Non-partisan		−0.66***	−0.61**		−0.66***	−0.67***
		(0.19)	(0.25)		(0.19)	(0.26)
Negative partisan		−0.40*	−0.32		−0.33	−0.30
		(0.22)	(0.34)		(0.21)	(0.32)
NegativePrime*Closedpartisan			0.06			0.15
			(0.42)			(0.42)
NegativePrime*Non-partisan			−0.11			0.02
			(0.38)			(0.39)
NegativePrime*Negativepartisan			−0.14			−0.04
			(0.44)			(0.42)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.01	0.05	0.06	0.06
Adj. R^2	−0.00	0.01	0.01	0.04	0.05	0.04
Num. obs.	791	725	725	791	725	725
RMSE	2.18	2.16	2.16	2.13	2.12	2.12

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.21: Treatment to Treatment comparison ITT, DV: Left-right self-placement (0-10)

A.0.12 Robustness checks: Alternative partisanship cut-offs

	M1	M2	M3	M4	M5	M6
Ref:Control	1.25***	1.43***	1.41***	1.12***	1.24***	1.22***
	(0.08)	(0.12)	(0.19)	(0.15)	(0.18)	(0.23)
Treatment effects						
T1: Negative prime	0.10	0.10	0.24	0.10	0.09	0.22
	(0.11)	(0.11)	(0.26)	(0.12)	(0.11)	(0.26)
T2: Positive prime	−0.08	−0.11	−0.19	−0.08	−0.11	−0.20
	(0.11)	(0.11)	(0.23)	(0.11)	(0.11)	(0.23)
Pre-T covariates						
Closed partisan		−0.33**	−0.24		−0.32**	−0.22
		(0.16)	(0.31)		(0.16)	(0.31)
Non-partisan		−0.30***	−0.29		−0.29**	−0.30
		(0.11)	(0.21)		(0.11)	(0.21)
Negative partisan		−0.23	−0.21		−0.21	−0.18
		(0.15)	(0.25)		(0.15)	(0.25)
Interaction effects						
NegativePrime*Closedpartisan			−0.23			−0.25
			(0.40)			(0.40)
PositivePrime*Closedpartisan			−0.05			−0.07
			(0.40)			(0.40)
NegativePrime*Non-partisan			−0.19			−0.18
			(0.29)			(0.30)
PositivePrime*Non-partisan			0.17			0.18
			(0.27)			(0.27)
NegativePrime*Negativepartisan			−0.15			−0.17
			(0.40)			(0.40)
PositivePrime*Negativepartisan			0.10			0.07
			(0.33)			(0.34)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.01	0.01	0.02	0.02
Adj. R^2	0.00	0.01	0.00	0.00	0.01	0.00
Num. obs.	1086	1000	1000	1086	1000	1000
RMSE	1.47	1.41	1.41	1.47	1.41	1.41

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on the 10th and 90th percentile cut-offs of pre-treatment thermometer scores

TABLE A.22: 90-10 Cut-off Treatment to Control ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	1.17***	1.33***	1.22***	1.15***	1.28***	1.17***
	(0.07)	(0.12)	(0.14)	(0.18)	(0.21)	(0.21)
T1: Negative prime	0.19*	0.21*	0.43*	0.18	0.21*	0.43*
	(0.11)	(0.11)	(0.23)	(0.11)	(0.11)	(0.23)
Closed partisan		−0.37**	−0.29		−0.37**	−0.28
		(0.18)	(0.25)		(0.18)	(0.25)
Non-partisan		−0.30**	−0.12		−0.28**	−0.11
		(0.13)	(0.17)		(0.14)	(0.18)
Negative partisan		−0.23	−0.11		−0.23	−0.11
		(0.20)	(0.22)		(0.20)	(0.22)
NegativePrime*Closedpartisan			−0.18			−0.20
			(0.36)			(0.36)
NegativePrime*Non-partisan			−0.37			−0.37
			(0.27)			(0.28)
NegativePrime*Negativepartisan			−0.25			−0.25
			(0.38)			(0.39)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.02	0.02	0.01	0.02	0.02
Adj. R^2	0.00	0.01	0.01	−0.00	0.00	0.00
Num. obs.	715	660	660	715	660	660
RMSE	1.46	1.39	1.40	1.46	1.40	1.40

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on the 10th and 90th percentile cut-offs of pre-treatment thermometer scores

TABLE A.23: 90-10 Cut-off Treatment to Treatment comparison ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Control	1.25***	1.33***	1.30***	1.12***	1.12***	1.10***
	(0.08)	(0.11)	(0.16)	(0.15)	(0.18)	(0.21)
Treatment effects						
T1: Negative prime	0.10	0.08	0.28	0.10	0.07	0.25
	(0.11)	(0.11)	(0.23)	(0.12)	(0.12)	(0.23)
T2: Positive prime	−0.08	−0.12	−0.21	−0.08	−0.12	−0.23
	(0.11)	(0.11)	(0.19)	(0.11)	(0.11)	(0.19)
Pre-T covariates						
Closed partisan		−0.12	−0.04		−0.11	−0.03
		(0.12)	(0.23)		(0.12)	(0.23)
Non-partisan		−0.18	−0.19		−0.18	−0.22
		(0.11)	(0.20)		(0.11)	(0.20)
Negative partisan		−0.12	−0.03		−0.11	−0.02
		(0.15)	(0.25)		(0.15)	(0.25)
Interaction effects						
NegativePrime*Closedpartisan			−0.34			−0.33
			(0.31)			(0.31)
PositivePrime*Closedpartisan			0.06			0.08
			(0.30)			(0.30)
NegativePrime*Non-partisan			−0.18			−0.13
			(0.30)			(0.30)
PositivePrime*Non-partisan			0.19			0.25
			(0.26)			(0.26)
NegativePrime*Negativepartisan			−0.45			−0.44
			(0.35)			(0.35)
PositivePrime*Negativepartisan			0.27			0.25
			(0.38)			(0.39)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.01	0.01	0.01	0.02
Adj. R^2	0.00	0.00	−0.00	0.00	−0.00	−0.00
Num. obs.	1086	1007	1007	1086	1007	1007
RMSE	1.47	1.42	1.42	1.47	1.42	1.42

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on the 15th and 85th percentile cut-offs of pre-treatment thermometer scores

TABLE A.24: 85-15 Cut-off Treatment to Control ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	1.17***	1.23***	1.09***	1.15***	1.18***	1.04***
	(0.07)	(0.09)	(0.10)	(0.18)	(0.21)	(0.21)
T1: Negative prime	0.19*	0.20*	0.49***	0.18	0.20*	0.50**
	(0.11)	(0.11)	(0.19)	(0.11)	(0.11)	(0.19)
Closed partisan		−0.17	0.02		−0.16	0.04
		(0.14)	(0.19)		(0.15)	(0.20)
Non-partisan		−0.17	0.00		−0.15	0.03
		(0.14)	(0.16)		(0.14)	(0.17)
Negative partisan		−0.17	0.24		−0.17	0.22
		(0.18)	(0.29)		(0.19)	(0.30)
NegativePrime*Closedpartisan			−0.40			−0.42
			(0.29)			(0.29)
NegativePrime*Non-partisan			−0.37			−0.39
			(0.28)			(0.28)
NegativePrime*Negativepartisan			−0.72*			−0.69*
			(0.38)			(0.39)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.01	0.02
Adj. R^2	0.00	0.00	0.00	−0.00	−0.00	−0.00
Num. obs.	715	665	665	715	665	665
RMSE	1.46	1.40	1.40	1.46	1.40	1.40

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on the 15th and 85th percentile cut-offs of pre-treatment thermometer scores

TABLE A.25: 85-15 Cut-off Treatment to Treatment comparison ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Control	1.25***	1.28***	1.15***	1.12***	1.06***	0.94***
	(0.08)	(0.10)	(0.15)	(0.15)	(0.18)	(0.21)
Treatment effects						
T1: Negative prime	0.10	0.13	0.39*	0.10	0.12	0.36*
	(0.11)	(0.11)	(0.21)	(0.12)	(0.11)	(0.22)
T2: Positive prime	−0.08	−0.07	0.03	−0.08	−0.07	0.01
	(0.11)	(0.11)	(0.18)	(0.11)	(0.11)	(0.19)
Pre-T covariates						
Closed partisan		−0.06	0.07		−0.04	0.08
		(0.10)	(0.18)		(0.10)	(0.18)
Non-partisan		−0.37**	−0.22		−0.37**	−0.26
		(0.15)	(0.26)		(0.15)	(0.26)
Negative partisan		−0.11	0.25		−0.09	0.27
		(0.16)	(0.26)		(0.16)	(0.26)
Interaction effects						
NegativePrime*Closedpartisan			−0.24			−0.22
			(0.27)			(0.27)
PositivePrime*Closedpartisan			−0.14			−0.11
			(0.24)			(0.24)
NegativePrime*Non-partisan			−0.46			−0.39
			(0.39)			(0.39)
PositivePrime*Non-partisan			0.04			0.09
			(0.36)			(0.35)
NegativePrime*Negativepartisan			−0.88**			−0.85**
			(0.35)			(0.35)
PositivePrime*Negativepartisan			−0.21			−0.22
			(0.41)			(0.40)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.01	0.01	0.01	0.02
Adj. R^2	0.00	0.00	0.00	0.00	0.00	0.00
Num. obs.	1086	1026	1026	1086	1026	1026
RMSE	1.47	1.42	1.42	1.47	1.42	1.42

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on the 25th and 75th percentile cut-offs of pre-treatment thermometer scores

TABLE A.26: 75-25 Cut-off Treatment to Control ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	1.17***	1.26***	1.18***	1.15***	1.23***	1.15***
	(0.07)	(0.10)	(0.11)	(0.18)	(0.22)	(0.23)
T1: Negative prime	0.19*	0.20*	0.36*	0.18	0.20*	0.36*
	(0.11)	(0.11)	(0.19)	(0.11)	(0.11)	(0.19)
Closed partisan		−0.11	−0.07		−0.11	−0.06
		(0.12)	(0.16)		(0.13)	(0.16)
Non-partisan		−0.42**	−0.18		−0.39**	−0.15
		(0.19)	(0.25)		(0.19)	(0.25)
Negative partisan		−0.31	0.05		−0.31	0.03
		(0.19)	(0.31)		(0.19)	(0.31)
NegativePrime*Closedpartisan			−0.10			−0.10
			(0.25)			(0.25)
NegativePrime*Non-partisan			−0.50			−0.51
			(0.39)			(0.39)
NegativePrime*Negativepartisan			−0.67*			−0.64*
			(0.39)			(0.39)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.02	0.02
Adj. R^2	0.00	0.01	0.01	−0.00	0.00	0.00
Num. obs.	715	677	677	715	677	677
RMSE	1.46	1.42	1.42	1.46	1.42	1.42

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on the 25th and 75th percentile cut-offs of pre-treatment thermometer scores

TABLE A.27: 75-25 Cut-off Treatment to Treatment comparison ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Control	1.25***	1.27***	1.23***	1.12***	1.05***	1.00***
	(0.08)	(0.11)	(0.17)	(0.15)	(0.19)	(0.22)
Treatment effects						
T1: Negative prime	0.10	0.14	0.25	0.10	0.13	0.23
	(0.11)	(0.11)	(0.24)	(0.12)	(0.11)	(0.24)
T2: Positive prime	−0.08	−0.08	−0.07	−0.08	−0.08	−0.09
	(0.11)	(0.11)	(0.21)	(0.11)	(0.11)	(0.21)
Pre-T covariates						
Closed partisan		−0.05	−0.05		−0.02	−0.04
		(0.10)	(0.19)		(0.11)	(0.19)
Non-partisan		−0.49**	−0.46		−0.50**	−0.51
		(0.21)	(0.33)		(0.21)	(0.32)
Negative partisan		−0.11	0.35		−0.08	0.37
		(0.17)	(0.31)		(0.18)	(0.31)
Interaction effects						
NegativePrime*Closedpartisan			−0.02			−0.00
			(0.28)			(0.28)
PositivePrime*Closedpartisan			0.04			0.07
			(0.25)			(0.25)
NegativePrime*Non-partisan			−0.23			−0.18
			(0.53)			(0.53)
PositivePrime*Non-partisan			0.16			0.22
			(0.51)			(0.51)
NegativePrime*Negativepartisan			−0.95**			−0.93**
			(0.39)			(0.39)
PositivePrime*Negativepartisan			−0.43			−0.44
			(0.46)			(0.46)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.01	0.02
Adj. R^2	0.00	0.00	0.00	0.00	0.00	0.00
Num. obs.	1086	1031	1031	1086	1031	1031
RMSE	1.47	1.42	1.42	1.47	1.42	1.42

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on the 30th and 70th percentile cut-offs of pre-treatment thermometer scores

TABLE A.28: 70-30 Cut-off Treatment to Control ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	1.17***	1.20***	1.16***	1.15***	1.17***	1.13***
	(0.07)	(0.10)	(0.12)	(0.18)	(0.22)	(0.24)
T1: Negative prime	0.19*	0.22**	0.32	0.18	0.22**	0.32
	(0.11)	(0.11)	(0.20)	(0.11)	(0.11)	(0.20)
Closed partisan		−0.03	−0.01		−0.03	0.00
		(0.12)	(0.15)		(0.13)	(0.16)
Non-partisan		−0.49*	−0.30		−0.49*	−0.29
		(0.28)	(0.39)		(0.28)	(0.39)
Negative partisan		−0.36*	−0.08		−0.36*	−0.10
		(0.20)	(0.34)		(0.21)	(0.35)
NegativePrime*Closedpartisan			−0.06			−0.07
			(0.25)			(0.25)
NegativePrime*Non-partisan			−0.39			−0.43
			(0.57)			(0.57)
NegativePrime*Negativepartisan			−0.52			−0.49
			(0.42)			(0.42)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.02	0.02
Adj. R^2	0.00	0.01	0.01	−0.00	0.00	0.00
Num. obs.	715	680	680	715	680	680
RMSE	1.46	1.42	1.42	1.46	1.42	1.42

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on the 30th and 70th percentile cut-offs of pre-treatment thermometer scores

TABLE A.29: 70-30 Cut-off Treatment to Treatment comparison ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Control	1.25***	1.46***	1.56***	1.12***	1.24***	1.33***
	(0.08)	(0.15)	(0.25)	(0.15)	(0.19)	(0.29)
Treatment effects						
T1: Negative prime	0.10	0.09	0.24	0.10	0.08	0.24
	(0.11)	(0.11)	(0.34)	(0.12)	(0.11)	(0.34)
T2: Positive prime	−0.08	−0.12	−0.52	−0.08	−0.12	−0.53*
	(0.11)	(0.11)	(0.32)	(0.11)	(0.11)	(0.32)
Pre-T covariates						
Closed partisan		−0.39*	−0.34		−0.37*	−0.30
		(0.20)	(0.38)		(0.20)	(0.38)
Non-partisan		−0.27*	−0.43		−0.27*	−0.43
		(0.14)	(0.27)		(0.14)	(0.27)
Negative partisan		−0.29*	−0.38		−0.27	−0.34
		(0.17)	(0.30)		(0.17)	(0.30)
Interaction effects						
NegativePrime*Closedpartisan			−0.25			−0.31
			(0.50)			(0.50)
PositivePrime*Closedpartisan			0.07			0.05
			(0.49)			(0.49)
NegativePrime*Non-partisan			−0.15			−0.17
			(0.37)			(0.37)
PositivePrime*Non-partisan			0.59*			0.60*
			(0.34)			(0.34)
NegativePrime*Negativepartisan			−0.18			−0.22
			(0.44)			(0.44)
PositivePrime*Negativepartisan			0.44			0.40
			(0.39)			(0.39)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.02	0.02
Adj. R^2	0.00	0.01	0.01	0.00	0.00	0.01
Num. obs.	1086	990	990	1086	990	990
RMSE	1.47	1.41	1.41	1.47	1.41	1.41

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.30: 95-5 Cut-off Treatment to Control ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	1.17***	1.30***	1.03***	1.15***	1.21***	0.92***
	(0.07)	(0.16)	(0.20)	(0.18)	(0.22)	(0.25)
T1: Negative prime	0.19*	0.21*	0.76**	0.18	0.21*	0.78**
	(0.11)	(0.11)	(0.30)	(0.11)	(0.11)	(0.30)
Closed partisan		−0.41*	−0.27		−0.42*	−0.25
		(0.22)	(0.31)		(0.22)	(0.31)
Non-partisan		−0.19	0.16		−0.19	0.17
		(0.17)	(0.22)		(0.17)	(0.22)
Negative partisan		−0.24	0.06		−0.24	0.06
		(0.21)	(0.25)		(0.21)	(0.25)
NegativePrime*Closedpartisan			−0.33			−0.37
			(0.44)			(0.44)
NegativePrime*Non-partisan			−0.74**			−0.76**
			(0.33)			(0.33)
NegativePrime*Negativepartisan			−0.62			−0.62
			(0.41)			(0.42)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.02	0.03
Adj. R^2	0.00	0.01	0.01	−0.00	0.00	0.01
Num. obs.	715	652	652	715	652	652
RMSE	1.46	1.40	1.40	1.46	1.40	1.40

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.

TABLE A.31: 95-5 Cut-off Treatment to Treatment comparison ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Control	1.25***	1.40***	1.51***	1.12***	1.17***	1.30***
	(0.08)	(0.17)	(0.29)	(0.15)	(0.22)	(0.33)
Treatment effects						
T1: Negative prime	0.10	0.11	0.41	0.10	0.10	0.39
	(0.11)	(0.12)	(0.43)	(0.12)	(0.12)	(0.42)
T2: Positive prime	−0.08	−0.11	−0.68**	−0.08	−0.11	−0.72**
	(0.11)	(0.11)	(0.34)	(0.11)	(0.11)	(0.34)
Pre-T covariates						
Closed partisan		−0.17	−0.16		−0.16	−0.16
		(0.20)	(0.39)		(0.20)	(0.38)
Non-partisan		−0.24	−0.51*		−0.24	−0.53*
		(0.17)	(0.31)		(0.16)	(0.30)
Negative partisan		−0.15	−0.16		−0.13	−0.15
		(0.18)	(0.32)		(0.18)	(0.32)
Interaction effects						
NegativePrime*Closedpartisan			−0.46			−0.46
			(0.52)			(0.52)
PositivePrime*Closedpartisan			0.34			0.38
			(0.49)			(0.49)
NegativePrime*Non-partisan			−0.14			−0.12
			(0.46)			(0.45)
PositivePrime*Non-partisan			0.86**			0.91**
			(0.37)			(0.36)
NegativePrime*Negativepartisan			−0.55			−0.55
			(0.47)			(0.47)
PositivePrime*Negativepartisan			0.51			0.53
			(0.40)			(0.40)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.01	0.03
Adj. R^2	0.00	0.00	0.01	0.00	0.00	0.01
Num. obs.	1086	964	964	1086	964	964
RMSE	1.47	1.43	1.42	1.47	1.43	1.42

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on absolute cut-offs of pre-treatment thermometer scores

TABLE A.32: Absolute Cut-off Treatment to Control ITT, DV: Distance to most liked party

	M1	M2	M3	M4	M5	M6
Ref:Positive prime	1.17***	1.23***	0.83***	1.15***	1.18***	0.75***
	(0.07)	(0.18)	(0.17)	(0.18)	(0.26)	(0.27)
T1: Negative prime	0.19*	0.23**	1.09***	0.18	0.23**	1.11***
	(0.11)	(0.11)	(0.36)	(0.11)	(0.12)	(0.36)
Closed partisan		−0.19	0.19		−0.19	0.21
		(0.23)	(0.29)		(0.23)	(0.30)
Non-partisan		−0.12	0.35*		−0.11	0.37*
		(0.19)	(0.20)		(0.19)	(0.20)
Negative partisan		−0.16	0.35		−0.15	0.36
		(0.21)	(0.24)		(0.21)	(0.24)
NegativePrime*Closedpartisan			−0.81*			−0.84*
			(0.46)			(0.46)
NegativePrime*Non-partisan			−1.00**			−1.02***
			(0.39)			(0.39)
NegativePrime*Negativepartisan			−1.06**			−1.06**
			(0.42)			(0.42)
Covariates?	No	No	No	Yes	Yes	Yes
R^2	0.00	0.01	0.02	0.01	0.01	0.03
Adj. R^2	0.00	0.00	0.01	−0.00	−0.00	0.01
Num. obs.	715	636	636	715	636	636
RMSE	1.46	1.42	1.41	1.46	1.42	1.41

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Partisan ref. category: Positive partisan.
Partisanship categories based on absolute cut-offs of pre-treatment thermometer scores

TABLE A.33: Absolute Cut-off Treatment to Treatment comparison ITT, DV: Distance to most liked party

Appendix B

Appendix Paper 2

B.0.1 Design

TABLE B.1: Assignment of experimental subjects to treatment condition

Group	Positivity	Identity	Ask	Assignment
Control	n/a	n/a	n/a	20%
T1	Positive	In-party	donation	10%
T2	Negative	Out-party	donation	10%
T3	Positive	Pro remain in EU	donation	10%
T4	Negative	Anti Brexit	donation	10%
T5	Positive	In-party	donation volunteer	5%
T6	Negative	Out-party	donation volunteer	5%
T7	Positive	Pro remain in EU	donation volunteer	5%
T8	Negative	Anti-Brexit	donation volunteer	5%
T9	Positive	In-party	volunteer donation	5%
T10	Negative	Out-party	volunteer donation	5%
T11	Positive	Pro remain in EU	volunteer donation	5%
T12	Negative	Anti Brexit	volunteer donation	5%

B.0.2 Balance checks

B.0.3 Main results: Additional graphs

TABLE B.2: Balance on key pre-treatment covariates

Variable	Control	Neg-Issue	Neg-Party	Pos-Issue	Pos-Party
Member	0.88 (0.32)	0.88 (0.32)	0.88 (0.33)	0.89 (0.32)	0.88 (0.32)
South East	0.2 (0.4)	0.2 (0.4)	0.2 (0.4)	0.2 (0.4)	0.2 (0.4)
Fundraiser	0.03 (0.17)	0.03 (0.17)	0.03 (0.17)	0.03 (0.16)	0.03 (0.17)
Volunteer	0.08 (0.28)	0.08 (0.27)	0.08 (0.27)	0.08 (0.27)	0.08 (0.27)
Donor	0.21 (0.41)	0.21 (0.41)	0.21 (0.41)	0.21 (0.41)	0.21 (0.41)
Membership	28.06 (33.78)	27.98 (34.43)	27.7 (33.94)	28.02 (35.49)	28.02 (34.55)

Mean and (SD) for key covariates.
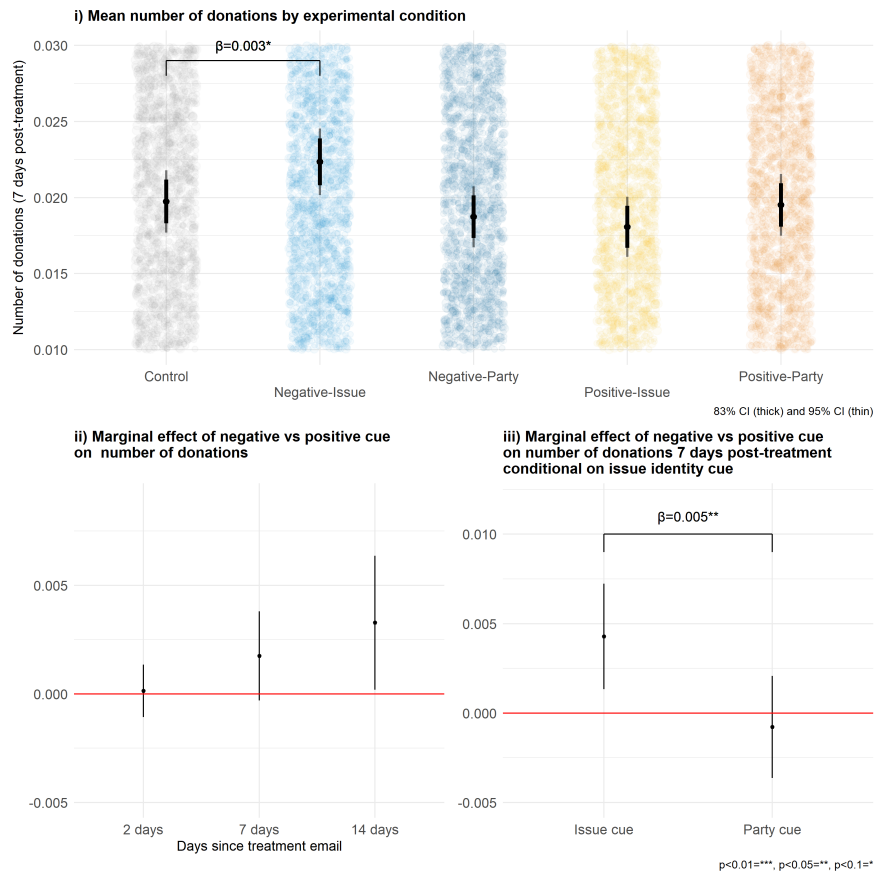
FIGURE B.1: Effect of negative vs positive cue on number of donations

i) Mean donation amount (logged) by experimental condition

83% CI (thick) and 95% CI (thin)

ii) Marginal effect of negative vs positive cue
on logged donation amount

iii) Marginal effect of negative vs positive cue
on logged donation amount 7 days post-treatment
conditional on issue identity cue

β=0.019***

p<0.01=***, p<0.05=**, p<0.1=*
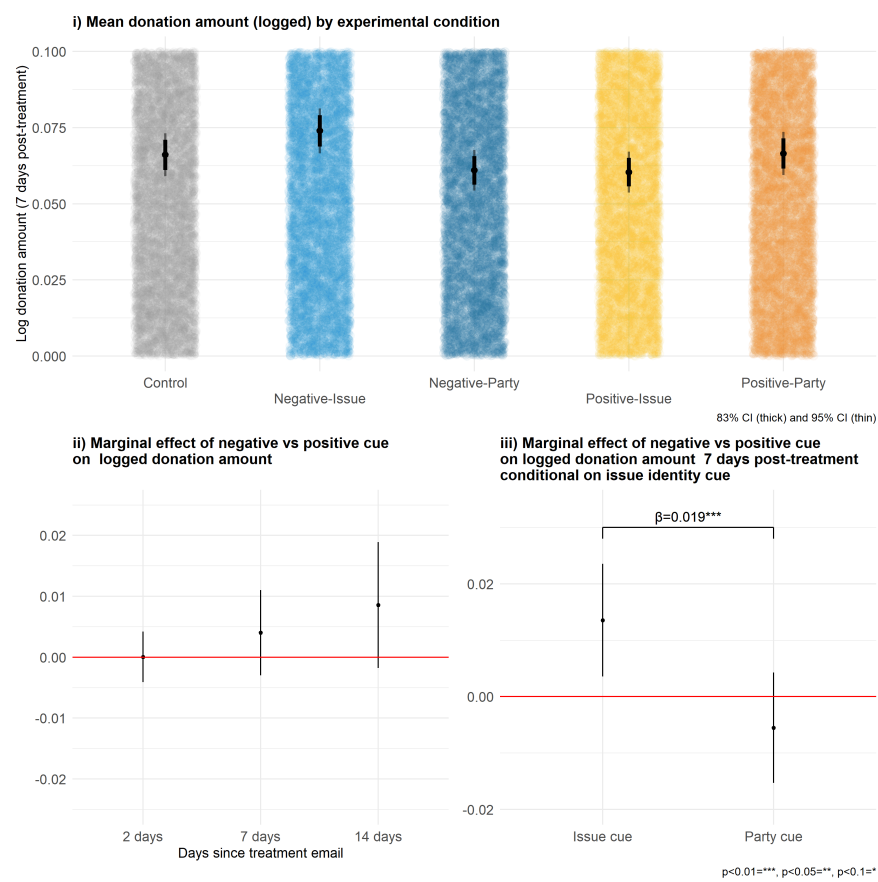
FIGURE B.2: Effect of negative vs positive cue on donation amount (log)

B.0.4 Main results: Regression tables

	M1	M2	M3	M4
Mean(Positive party cue)	0.018***	0.019***	0.001	0.002
	(0.001)	(0.001)	(0.003)	(0.003)
Negative cue	0.002*	−0.001	0.002*	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Issue cue	0.001	−0.002	0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Negative*Issue		0.005**		0.005**
		(0.002)		(0.002)
Covariates?	No	No	Yes	Yes
R^2	0.000	0.000	0.031	0.031
Adj. R^2	0.000	0.000	0.030	0.031
Num. obs.	71952	71952	71952	71952
RMSE	0.138	0.138	0.136	0.136

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
Covariates include region, membership status and amount, past volunteering, fundraising and donations behaviour

TABLE B.3: ITT effects of campaign emails on DV: Donated (=1), did not donate (=0), 7 days post-treatment

	M1	M2	M3	M4
Mean(Positive party cue)	0.018***	0.020***	0.000	0.002
	(0.001)	(0.001)	(0.003)	(0.003)
Negative cue	0.002*	−0.001	0.002*	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Issue cue	0.001	−0.001	0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Negative*Issue		0.005**		0.005**
		(0.002)		(0.002)
Covariates?	No	No	Yes	Yes
R^2	0.000	0.000	0.031	0.031
Adj. R^2	0.000	0.000	0.030	0.030
Num. obs.	71952	71952	71952	71952
RMSE	0.140	0.140	0.138	0.138

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
Covariates include region, membership status and amount, past volunteering, fundraising and donations behaviour

TABLE B.4: ITT effects of campaign emails on DV: How many times did an individual donate? 7 days post-treatment

	M1	M2	M3	M4
Mean(Positive party cue)	0.062***	0.066***	−0.001	0.004
	(0.003)	(0.004)	(0.009)	(0.009)
Negative cue	0.004	−0.006	0.004	−0.005
	(0.004)	(0.005)	(0.004)	(0.005)
Issue cue	0.003	−0.006	0.004	−0.006
	(0.004)	(0.005)	(0.004)	(0.005)
Negative*Issue		0.019***		0.018***
		(0.007)		(0.007)
Covariates?	No	No	Yes	Yes
R^2	0.000	0.000	0.031	0.031
Adj. R^2	0.000	0.000	0.031	0.031
Num. obs.	71952	71952	71952	71952
RMSE	0.478	0.478	0.471	0.471

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Covariates include region, membership status and amount, past volunteering, fundraising and donations behaviour

TABLE B.5: ITT effects of campaign emails on DV: Logged DV: How much did an individual donate? (in GBP) 7 days post-treatment

B.0.5 Robustness checks

	M1	M2	M3	M4
Mean(Positive party cue)	−3.993***	−3.923***	−5.197***	−5.129***
	(0.048)	(0.054)	(0.170)	(0.172)
Negative cue	0.090*	−0.048	0.099*	−0.041
	(0.054)	(0.077)	(0.055)	(0.079)
Issue cue	0.052	−0.088	0.060	−0.083
	(0.054)	(0.078)	(0.055)	(0.079)
Negative*Issue		0.270**		0.276**
		(0.108)		(0.110)
Covariates?	No	No	Yes	Yes
AIC	13805.511	13801.267	12155.892	12151.610
BIC	13833.063	13838.002	12357.934	12362.837
Log Likelihood	−6899.756	−6896.634	−6055.946	−6052.805
Deviance	13799.511	13793.267	12111.892	12105.610
Num. obs.	71952	71952	71952	71952

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Covariates include region, membership status and amount, past volunteering, fundraising and donations behaviour

TABLE B.6: Logit model: ITT effects of campaign emails on DV: Donated (=1), did not donate (=0), 7 days post-treatment. Coefficients represent Log odds.

	M1	M2	M3	M4
Mean(Positive party cue)	0.013***	0.013***	0.000	0.001
	(0.001)	(0.001)	(0.002)	(0.002)
Negative cue	0.001*	−0.001	0.001*	−0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Issue cue	0.001	−0.001	0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Negative*Issue		0.004**		0.003**
		(0.001)		(0.001)
Covariates?	No	No	Yes	Yes
R^2	0.000	0.000	0.031	0.031
Adj. R^2	0.000	0.000	0.031	0.031
Num. obs.	71952	71952	71952	71952
RMSE	0.097	0.097	0.095	0.095

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Covariates include region, membership status and amount, past volunteering, fundraising and donations behaviour

TABLE B.7: ITT effects of campaign emails on DV: Logged: How many times did an individual donate? 7 days post-treatment

	M1	M2	M3	M4
Mean(Positive party cue)	0.805***	0.870***	0.013	0.073
	(0.068)	(0.078)	(0.225)	(0.243)
Negative cue	−0.011	−0.140	−0.005	−0.127
	(0.082)	(0.110)	(0.082)	(0.110)
Issue cue	0.036	−0.093	0.038	−0.084
	(0.082)	(0.114)	(0.082)	(0.114)
Negative*Issue		0.258		0.244
		(0.164)		(0.164)
Covariates?	No	No	Yes	Yes
R^2	0.000	0.000	0.013	0.013
Adj. R^2	−0.000	−0.000	0.012	0.012
Num. obs.	71952	71952	71952	71952
RMSE	10.969	10.969	10.901	10.901

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Covariates include region, membership status and amount, past volunteering, fundraising and donations behaviour

TABLE B.8: ITT effects of campaign emails on DV: How much did an individual donate? (in GBP) 7 days post-treatment

B.0.6 Main results at different post-treatment time period cut-offs: regression tables

B.0.7 Main results at different post-treatment time period cut-offs: graphs

	2 days	2 days	7 days	7 days	14 days	14 days
Mean(Positive party cue)	0.007***	0.007***	0.018***	0.019***	0.039***	0.040***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Negative cue	0.000	−0.001	0.002*	−0.001	0.003**	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Issue cue	0.000	−0.001	0.001	−0.002	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Negative*Issue		0.002		0.005**		0.002
		(0.001)		(0.002)		(0.003)
Covariates?	No	No	No	No	No	No
R²	0.000	0.000	0.000	0.000	0.000	0.000
Adj. R²	−0.000	−0.000	0.000	0.000	0.000	0.000
Num. obs.	71952	71952	71952	71952	71952	71952
RMSE	0.083	0.083	0.138	0.138	0.201	0.201

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
The 14 days model shows the results 14 days post-treatment.
The 7 days model shows the results 7 days post-treatment.
The 2 days model shows the results 2 days post-treatment.

TABLE B.9: ITT effects of campaign emails on DV: Donated (=1), did not donate (=0)

	2 days	2 days	7 days	7 days	14 days	14 days
Mean(Positive party cue)	0.007***	0.007***	0.018***	0.020***	0.041***	0.041***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Negative cue	0.000	−0.001	0.002*	−0.001	0.003**	0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Issue cue	0.000	−0.001	0.001	−0.001	0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Negative*Issue		0.002		0.005**		0.001
		(0.001)		(0.002)		(0.003)
Covariates?	No	No	No	No	No	No
R²	0.000	0.000	0.000	0.000	0.000	0.000
Adj. R²	−0.000	−0.000	0.000	0.000	0.000	0.000
Num. obs.	71952	71952	71952	71952	71952	71952
RMSE	0.083	0.083	0.140	0.140	0.211	0.211

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
The 14 days model shows the results 14 days post-treatment.
The 7 days model shows the results 7 days post-treatment.
The 2 days model shows the results 2 days post-treatment.

TABLE B.10: ITT effects of campaign emails on DV: How many times did an individual donate?

	2 days	2 days	7 days	7 days	14 days	14 days
Mean(Positive party cue)	0.022***	0.024***	0.062***	0.066***	0.136***	0.139***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.005)
Negative cue	0.000	−0.003	0.004	−0.006	0.009	0.004
	(0.002)	(0.003)	(0.004)	(0.005)	(0.005)	(0.007)
Issue cue	0.001	−0.003	0.003	−0.006	0.006	0.002
	(0.002)	(0.003)	(0.004)	(0.005)	(0.005)	(0.007)
Negative*Issue		0.007*		0.019***		0.009
		(0.004)		(0.007)		(0.011)
Covariates?	No	No	No	No	No	No
R^2	0.000	0.000	0.000	0.000	0.000	0.000
Adj. R^2	−0.000	−0.000	0.000	0.000	0.000	0.000
Num. obs.	71952	71952	71952	71952	71952	71952
RMSE	0.283	0.283	0.478	0.478	0.707	0.707

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
The 14 days model shows the results 14 days post-treatment.
The 7 days model shows the results 7 days post-treatment.
The 2 days model shows the results 2 days post-treatment.

TABLE B.11: ITT effects of campaign emails on DV: Logged DV: How much did an individual donate? (in GBP)
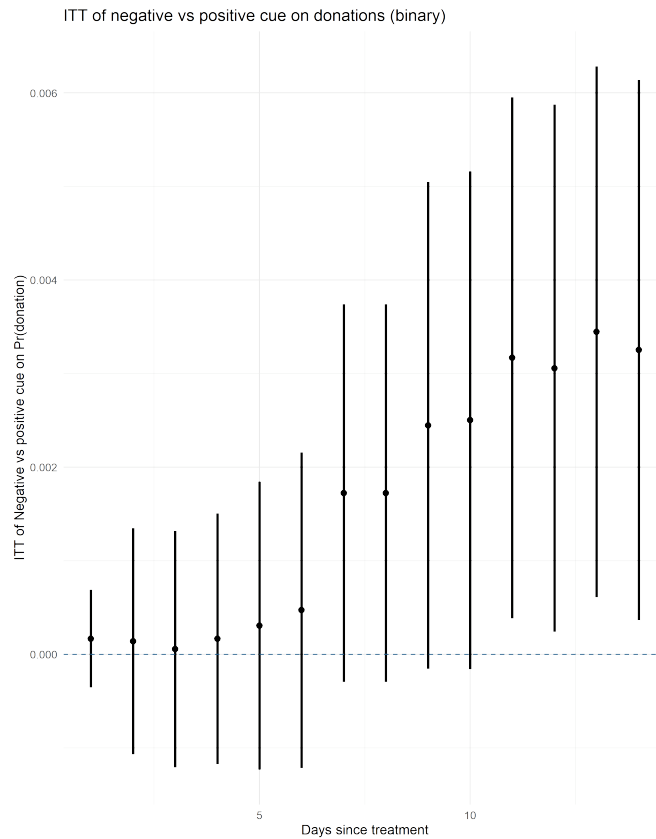
FIGURE B.3: ITT of negative vs positive cue on Pr(donation) at different post-treatment time periods
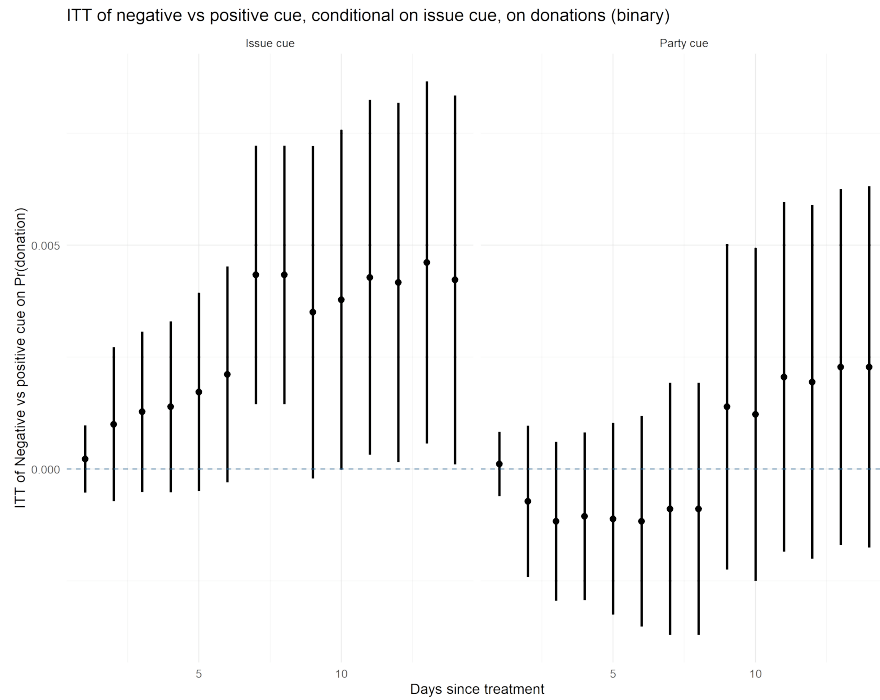
FIGURE B.4: ITT of negative vs party cue, conditional on issue cue, on Pr(donation) at different post-treatment time periods

B.0.8   Additional results: Treatment-Control comparisons
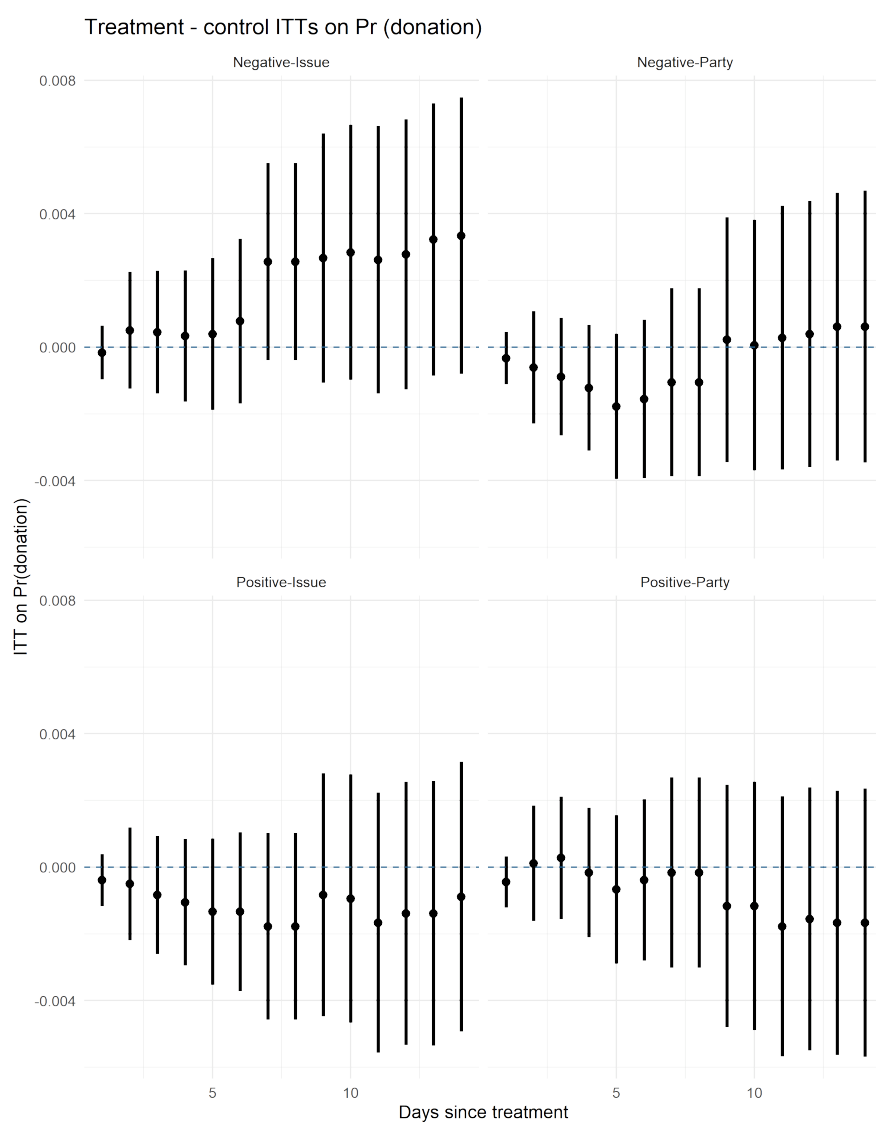
FIGURE B.5: Treatment-Control ITT on Pr(donation) from 1 to 14 days post-treatment
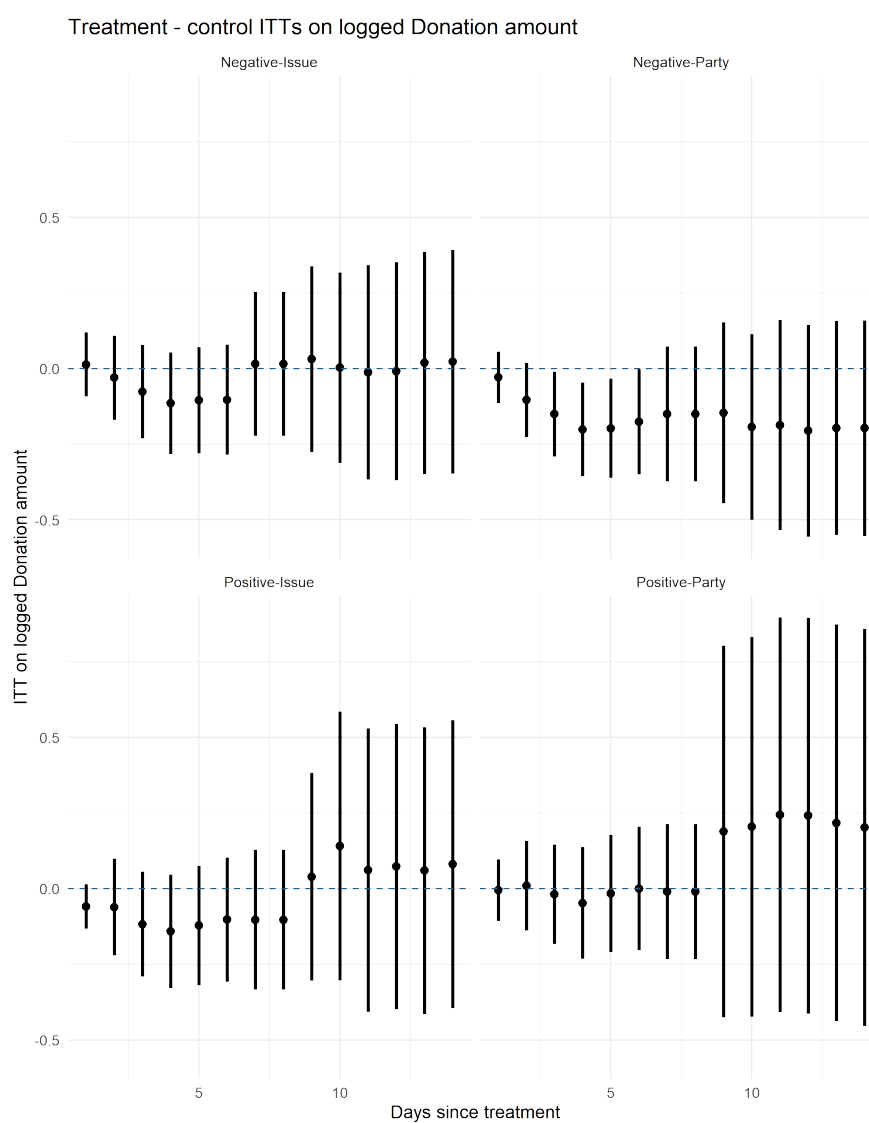
Figure B.6: Treatment-Control ITT on Logged donation amount from 1 to 14 days post-treatment

B.0.9   Additional results: Interaction with pre-treatment donor type

Figures/citt_donortype_donationpr_forallcues.png

FIGURE B.7: CITT of treatment cues on Pr(donation), by pre-treatment donor type

Figures/citt_donortype_donationamount_forallcues.png

FIGURE B.8: CITT of treatment cues on donation amount, by pre-treatment donor type

Appendix C

Appendix Paper 3

C.0.1   Treatment texts

Study 1 - Norwegian treatment texts

Treatment 1: Gender+immigration

Norwegian: Vennligst les den følgende hypotetiske politiske uttalelsen: Høy innvandring over flere tiår har fått store konsekvenser for samfunnsutviklingen. Vi trenger å holde innvandringsnivået lavt. Mange av de som kommer respekterer ikke de verdiene det norske samfunnet er bygget på. Likestilling mellom kjønnene er en norsk kjerneverdi. Vi mener at undertrykkende holdninger overfor kvinner ikke hører hjemme i samfunnet vårt. Vi mener at alle mennesker er likeverdige og vi aksepterer ikke diskriminering av kvinner. Dette er grunnen til at vi ønsker en lavere innvandringsrate og en streng innvandringspolitikk. Vi trenger en innvandringspolitikk som tydelig krever at innvandrere innretter seg etter norske lover, regler og verdier.

English: Please read the following hypothetical political statement: High immigration over several decades has had major consequences for the development of society. We need to keep immigration levels low. Many of those who come do not respect the values on which Norwegian society is founded. Gender equality is a Norwegian core value. We believe that oppressive attitudes towards women do not belong in our society. We believe that all people are equal and we do not accept discrimination against women. This is why we want a lower immigration rate and a strict immigration policy. We need an immigration policy that clearly requires immigrants to comply with Norwegian laws, rules and values.

Treatment 2: Immigration

Norwegian: Vennligst les den følgende hypotetiske politiske uttalelsen: Høy innvandring over flere tiår har fått store konsekvenser for samfunnsutviklingen. Vi trenger å holde innvandringsnivået lavt. Mange av de som kommer respekterer ikke de verdiene det norske samfunnet er bygget på. Dette er grunnen til at vi ønsker en lavere innvandringsrate og en streng innvandringspolitikk. Vi trenger en innvandringspolitikk som tydelig krever at innvandrere innretter seg etter norske lover, regler og verdier.

English: Please read the following hypothetical political statement: High immigration over several decades has had major consequences for the development of society. We need to keep immigration levels low. Many of those who come do not respect the values on which Norwegian society is founded. This is why we want a lower immigration rate and a strict immigration policy. We need an immigration policy that clearly requires immigrants to comply with Norwegian laws, rules and values.

TABLE C.1: Overview of treatment texts- Study 1 (Norway)

Study 1 - Norwegian treatment texts

Treatment 3: Gender

Norwegian: Vennligst les den følgende hypotetiske politiske uttalelsen: Likestilling mellom kjønnene er en norsk kjerneverdi. Vi mener at undertrykkende holdninger overfor kvinner ikke hører hjemme i samfunnet vårt. Vi mener at alle mennesker er likeverdige og vi aksepterer ikke diskriminering av kvinner.

English: Please read the following hypothetical political statement: Gender equality is a Norwegian core value. We believe that oppressive attitudes towards women do not belong in our society. We believe that all people are equal and we do not accept discrimination against women.

Treatment 4: Mediator+T2

Norwegian: Innvandringsdebatten har fått mye oppmerksomhet de siste årene. I Institutt for samfunnsforsknings spørreundersøkelse «Integreringsbarometeret» fra 2018 sa 41 % av nordmenn at vi bør ta imot betydelig færre eller noen færre innvandrere. Dette viser at folk er komfortable med å dele sine bekymringer for innvandrere i spørreundersøkelser og å utrykke varierte meninger om innvandring.
Vennligst les den følgende hypotetiske politiske uttalelsen: Høy innvandring over flere tiår har fått store konsekvenser for samfunnsutviklingen. Vi trenger å holde innvandringsnivået lavt. Mange av de som kommer respekterer ikke de verdiene det norske samfunnet er bygget på. Dette er grunnen til at vi ønsker en lavere innvandringsrate og en streng innvandringspolitikk. Vi trenger en innvandringspolitikk som tydelig krever at innvandrere innretter seg etter norske lover, regler og verdier.

English: The immigration debate has received a lot of attention in recent years. In the Institute for Social Research's survey "The Integration Barometer" of 2018, 41% of Norwegians said that we should accept significantly fewer or some fewer immigrants. This shows that people are comfortable sharing their concerns about immigrants in surveys and expressing varied opinions on immigration.
Please read the following hypothetical policy statement: High immigration over several decades has had major consequences for the development of society. We need to keep immigration levels low. Many of those who come do not respect the values on which Norwegian society is founded. This is why we want a lower immigration rate and a strict immigration policy. We need an immigration policy that clearly requires immigrants to comply with Norwegian laws, rules and values.

Table C.2: Overview of treatment texts - Study 1 (Norway)

Study 2 - German treatment texts

Treatment 1: Gender+immigration
EN: Mass immigration poses a grave threat to our future. The consequences of unchecked immigration are dramatic. Women and girls are increasingly becoming victims of violence in Germany. We cannot be indifferent to this development: Women's shelters and counselling centres are becoming more and more popular. We stand up for women's freedom and equality at all political levels - because women's rights are part of our fundamental values and are non-negotiable. In order to protect our basic values and population, mass immigration must be stopped. We demand a policy that puts the values and interests of our own country and the indigenous population back at the centre. Germany must stop illegal entries at its own borders and end unregulated mass immigration.
DE: Die Massenzuwanderung stellt eine schwere Gefährdung unserer Zukunft dar. Die Folgen ungebremster Zuwanderung sind dramatisch. Zunehmend werden Frauen und Mädchen Opfer von Gewalt in Deutschland. Diese Entwicklung darf uns nicht gleichgültig sein: Frauenhäuser und Beratungsstellen haben immer mehr Zulauf. Wir setzen uns für die Freiheit und Gleichheit von Frauen auf allen politischen Ebenen ein - denn Frauenrechte gehören zu unseren Grundwerten und sind nicht verhandelbar. Um unsere Grundwerte und Bevölkerung zu schützen, muss die Masseneinwanderung gestoppt werden. Wir fordern eine Politik, die die Werte und Interessen unseres eigenen Landes und der einheimischen Bevölkerung wieder ins Zentrum rückt. Deutschland muss illegale Einreisen an den eigenen Grenzen unterbinden und die ungeregelte Massenzuwanderung beenden.

Treatment 2: Immigration
EN: Mass immigration poses a grave threat to our future. The consequences of unchecked immigration are dramatic. In order to protect our fundamental values and population, mass immigration must be stopped. We demand a policy that puts the values and interests of our own country and the native population back at the centre. Germany must stop illegal entries at its own borders and end unregulated mass immigration.
DE: Die Massenzuwanderung stellt eine schwere Gefährdung unserer Zukunft dar. Die Folgen ungebremster Zuwanderung sind dramatisch. Um unsere Grundwerte und Bevölkerung zu schützen, muss die Masseneinwanderung gestoppt werden. Wir fordern eine Politik, die die Werte und Interessen unseres eigenen Landes und der einheimischen Bevölkerung wieder ins Zentrum rückt. Deutschland muss illegale Einreisen an den eigenen Grenzen unterbinden und die ungeregelte Massenzuwanderung beenden.

TABLE C.3: Overview of treatment texts - Study 2 (Germany 2021)

Study 2 - German treatment texts

Treatment 3: Gender
EN: Women and girls are increasingly becoming victims of violence in Germany. We must not be indifferent to this development: Women's shelters and counselling centres are becoming more and more popular. We stand up for the freedom and equality of women at all political levels - because women's rights belong to our basic values and are not negotiable.
DE: Zunehmend werden Frauen und Mädchen Opfer von Gewalt in Deutschland. Diese Entwicklung darf uns nicht gleichgültig sein: Frauenhäuser und Beratungsstellen haben immer mehr Zulauf. Wir setzen uns für die Freiheit und Gleichheit von Frauen auf allen politischen Ebenen ein - denn Frauenrechte gehören zu unseren Grundwerten und sind nicht verhandelbar.

TABLE C.4: Overview of treatment texts- Study 2 (Germany 2021)

C.0.2   Balance checks

The treatment groups were balanced along key covariates, including age, education and gender, see Tables C.5 and C.6.

TABLE C.5: Balance table for key demographic variables in Study 1 (Norway).

		Control (N=578)		T1 (N=588)		T2 (N=507)		T3 (N=573)		T4 (N=574)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Left-right		5.9	2.3	5.8	2.3	6.1	2.3	5.7	2.3	5.9	2.3
		N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.
Citizenship	Citizen of another country	23	4.0	32	5.4	29	5.7	30	5.2	37	6.4
	Norwegian citizen	545	94.3	545	92.7	466	91.9	534	93.2	526	91.6
Born	1959 or earlier	262	45.3	249	42.3	239	47.1	274	47.8	273	47.6
	1960-1989	272	47.1	293	49.8	229	45.2	254	44.3	264	46.0
	1990 or later	44	7.6	46	7.8	39	7.7	45	7.9	37	6.4
Gender	Male	311	53.8	287	48.8	232	45.8	299	52.2	273	47.6
	Female	267	46.2	301	51.2	275	54.2	274	47.8	301	52.4
Education	No education/elementary school	32	5.5	27	4.6	21	4.1	25	4.4	33	5.7
	Upper secondary education	170	29.4	188	32.0	163	32.1	173	30.2	163	28.4
	University/University college	363	62.8	359	61.1	311	61.3	357	62.3	370	64.5
	Other	578	100.0	588	100.0	507	100.0	573	100.0	574	100.0

Left-right: left-right self-placement (0=Left, 10-Right).

TABLE C.6: Balance table for key demographic variables in Study 2 (Germany).

		Control (N=752)		T1 (N=747)		T2 (N=753)		T3 (N=750)	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Age		47.3	16.3	48.2	16.1	48.8	16.5	48.4	16.5
		N	Pct.	N	Pct.	N	Pct.	N	Pct.
Migration background	No	626	83.2	627	83.9	621	82.5	638	85.1
	Yes	126	16.8	120	16.1	132	17.5	112	14.9
Gender	Female	394	52.4	355	47.5	386	51.3	368	49.1
	Male	357	47.5	391	52.3	367	48.7	381	50.8
	NA	1	0.1	1	0.1	0	0.0	1	0.1
Education	University	149	19.8	173	23.2	152	20.2	166	22.1
	College	52	6.9	50	6.7	55	7.3	59	7.9
	A-level	98	13.0	109	14.6	105	13.9	122	16.3
	Traineeship	256	34.0	235	31.5	265	35.2	223	29.7
	GCSE	185	24.6	165	22.1	160	21.2	171	22.8
	Other	12	1.6	15	2.0	16	2.1	9	1.2
Employment	Full-time employed	315	41.9	332	44.4	312	41.4	328	43.7
	Other	437	58.1	415	55.6	441	58.6	422	56.3

Age: in years. Migration background: Yes=at least one parent born outside Germany. Education: highest level of education.

C.0.3   Regression tables: Main results

	M1	M2	M3	M4
Ref:Control	4.195***	4.768***	4.991***	5.230***
	(0.107)	(0.158)	(0.200)	(0.222)
Immigration	0.007	−0.317	0.006	−0.260
	(0.153)	(0.225)	(0.151)	(0.225)
GenderEquality	−0.064	−0.404*	−0.077	−0.357
	(0.152)	(0.221)	(0.150)	(0.220)
Immigration*GenderEquality	0.239	0.455	−0.532***	0.368
	(0.217)	(0.311)	(0.111)	(0.310)
Female		−1.080***		−0.999***
		(0.212)		(0.214)
Immigration*Female		0.595*		0.507*
		(0.304)		(0.303)
GenderEquality*Female		0.613**		0.538*
		(0.302)		(0.301)
Immigration*GenderEquality*Female		−0.390		−0.227
		(0.434)		(0.433)
Covariates?	No	No	Yes	Yes
R^2	0.001	0.013	0.021	0.023
Adj. R^2	−0.000	0.011	0.018	0.019
Num. obs.	2995	2992	2992	2992
RMSE	2.974	2.959	2.948	2.946

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Covariates include age, education, employment, region and migration background

TABLE C.7: Study 2 (Germany) - Treatment interaction model. DV: Is it acceptable to say something negative about immigrants? 0-10, 10=Very acceptable

	M1	M2	M3	M4
Ref:Control	4.195***	4.768***	4.991***	5.230***
	(0.107)	(0.158)	(0.200)	(0.222)
Gender+Immigration(T1)	0.182	−0.266	0.164	−0.249
	(0.154)	(0.216)	(0.153)	(0.216)
Immigration(T2)	0.007	−0.317	0.006	−0.260
	(0.153)	(0.225)	(0.151)	(0.225)
Gender(T3)	−0.064	−0.404*	−0.077	−0.357
	(0.152)	(0.221)	(0.150)	(0.220)
Female		−1.080***		−0.999***
		(0.212)		(0.214)
Female*Gender+Immigration(T1)		0.818***		0.819***
		(0.308)		(0.307)
Female*Immigration(T2)		0.595*		0.507*
		(0.304)		(0.303)
Female*Gender(T3)		0.613**		0.538*
		(0.302)		(0.301)
Covariates?	No	No	Yes	Yes
R^2	0.001	0.013	0.021	0.023
Adj. R^2	−0.000	0.011	0.018	0.019
Num. obs.	2995	2992	2992	2992
RMSE	2.974	2.959	2.948	2.946

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Covariates include age, education, employment, region and migration background

TABLE C.8: Study 2 (Germany) - Treatment to Control comparison. DV: Is it acceptable to say something negative about immigrants? 0-10, 10=Very acceptable

	Model 1	Model 2	Model 3	Model 4
Ref:Control	3.456***	4.029***	4.682***	4.571***
	(0.121)	(0.176)	(0.496)	(0.505)
Immigration	0.056	0.105	0.120	0.169
	(0.177)	(0.262)	(0.176)	(0.263)
GenderEquality	−0.283*	−0.304	−0.270	−0.220
	(0.166)	(0.240)	(0.164)	(0.242)
Immigration*GenderEquality	0.688***	0.980***	0.655***	0.858**
	(0.242)	(0.357)	(0.240)	(0.362)
Female		−1.239***		−1.112***
		(0.235)		(0.241)
Immigration*Female		0.090		−0.125
		(0.347)		(0.354)
GenderEquality*Female		0.077		−0.113
		(0.321)		(0.325)
Immigration*GenderEquality*Female		−0.665		−0.380
		(0.472)		(0.481)
Covariates?	No	No	Yes	Yes
R^2	0.009	0.064	0.083	0.084
Adj. R^2	0.008	0.061	0.076	0.076
Num. obs.	2234	2234	2142	2142
RMSE	2.860	2.781	2.758	2.758

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.
M1= Main effects, M2= Interaction with respondent gender, M3= Main effects (cov.adjusted), M4= Interaction with respondent gender (cov.adjusted)
Covariates include age, income, education, and citizenship

TABLE C.9: Study 1 (Norway) - Treatment interaction model. DV: Is it acceptable to say something negative about immigrants? 0-10, 10=Very acceptable

	Model 1	Model 2	Model 3	Model 4
Ref:Control	3.456***	4.029***	4.686***	4.623***
	(0.121)	(0.176)	(0.445)	(0.457)
Immigration(T2)	0.056	0.105	0.113	0.173
	(0.177)	(0.262)	(0.175)	(0.263)
Gender+Immigration(T1)	0.460***	0.781***	0.494***	0.800***
	(0.172)	(0.252)	(0.168)	(0.251)
GenderEquality(T3)	−0.283*	−0.304	−0.272*	−0.222
	(0.166)	(0.240)	(0.164)	(0.242)
Acceptability+Immigration(T4)	0.411**	0.268	0.440***	0.241
	(0.170)	(0.252)	(0.169)	(0.250)
Female		−1.239***		−1.109***
		(0.235)		(0.240)
Female*Immigration(T2)		0.090		−0.126
		(0.347)		(0.353)
Female*Gender+Immigration(T1)		−0.498		−0.612*
		(0.331)		(0.335)
Female*GenderEquality(T3)		0.077		−0.109
		(0.321)		(0.326)
Female*Acceptability+Immigration(T4)		0.417		0.373
		(0.333)		(0.338)
Covariates?	No	No	Yes	Yes
T1 - T4 = 0 (Chi-square)	0.084		0.105	
Pr(>Chisq)	0.772		0.746	
Female*T1 - Female*T4 = 0 (Chi-square)		7.603		7.603
Pr(>Chisq)		0.006		0.006
R^2	0.009	0.058	0.073	0.076
Adj. R^2	0.008	0.055	0.067	0.068
Num. obs.	2806	2806	2694	2694
RMSE	2.855	2.787	2.769	2.766

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
M1= Main effects, M2= Interaction with respondent gender, M3= Main effects (cov.adjusted), M4= Interaction with respondent gender (cov.adjusted)
Covariates include age, income, education, and citizenship

TABLE C.10: Study 1 (Norway) - Treatment to Control comparison. DV: Is it acceptable to say something negative about immigrants? 0-10, 10=Very acceptable

C.0.4 Figures: Additional results
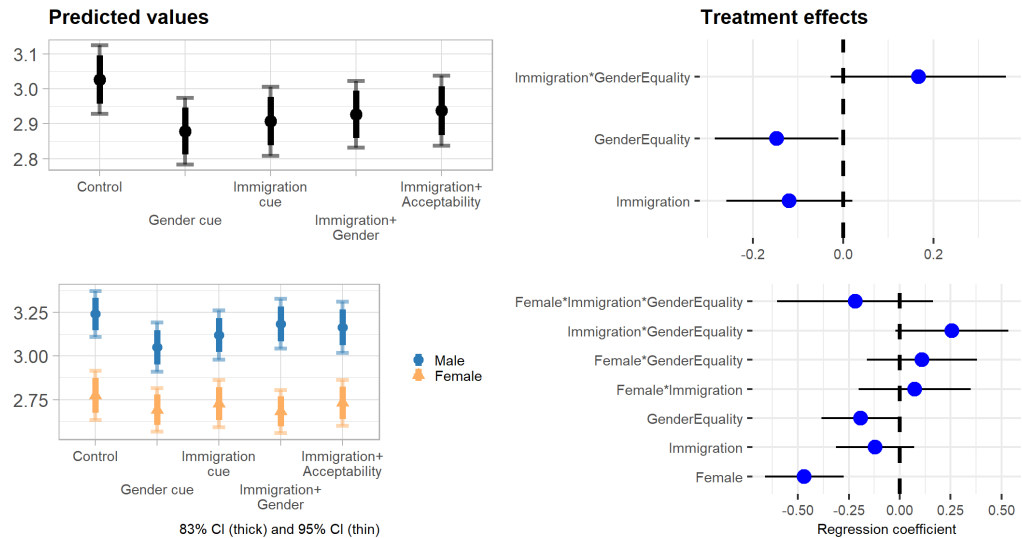
FIGURE C.1: Immigration attitudes (Study 1 - Norway)

FIGURE C.2: Immigration attitudes (Study 2 - Germany 2021)

Results for radical right vote (Norway)
Y: Would vote for the radical right (FrP) in a GE (=1), would vote for another party (=0)
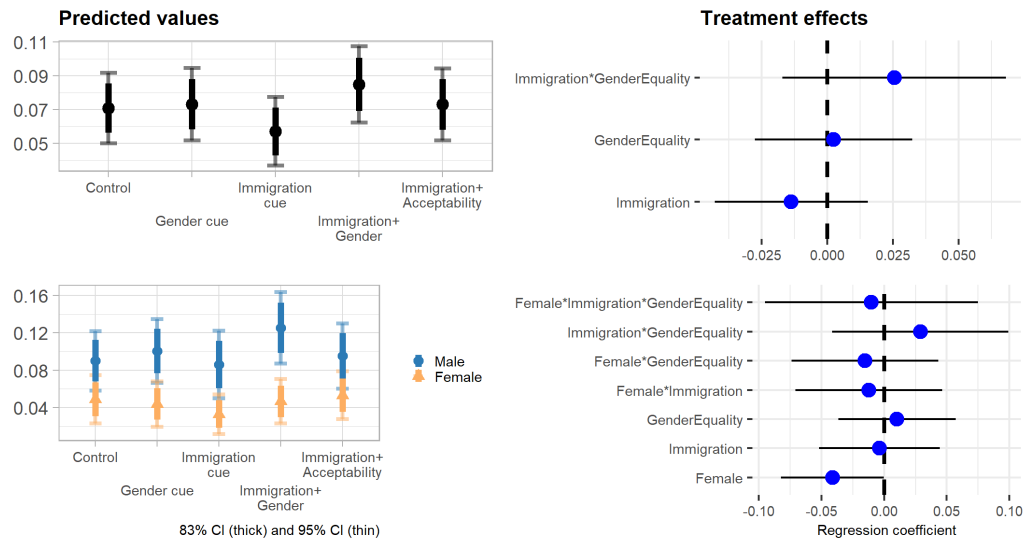
FIGURE C.3: Radical right vote (Study 1 - Norway)

Results for radical right vote (Germany 2021)
Y: Would vote for the radical right in a GE (1), otherwise (0)
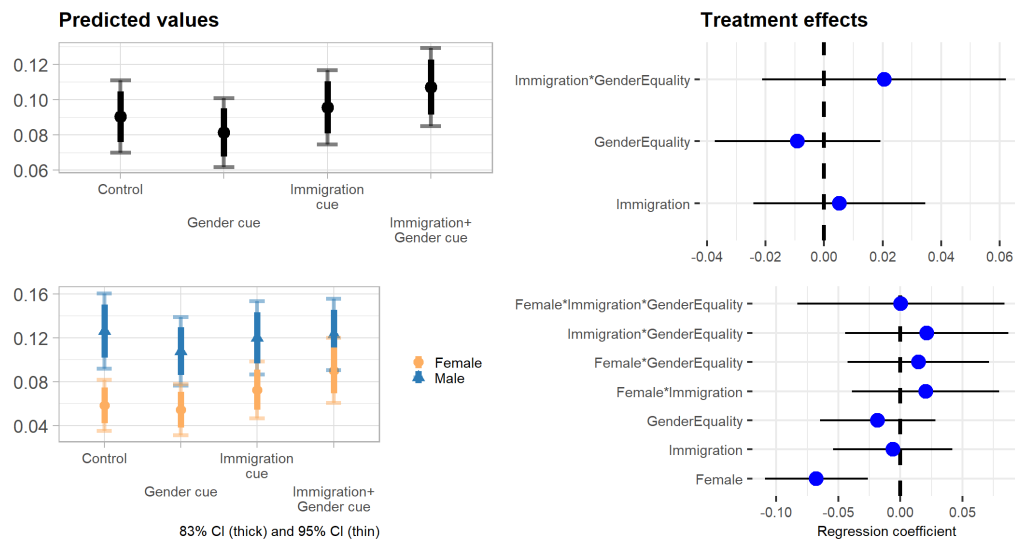
FIGURE C.4: Radical right vote (Study 2 - Germany 2021)

C.0.5 Regression tables: Additional results

	M1	M2	M3	M4
Ref:Control	4.832***	4.739***	5.161***	5.091***
	(0.100)	(0.149)	(0.186)	(0.206)
Immigration	0.174	0.332	0.200	0.327
	(0.144)	(0.209)	(0.141)	(0.207)
GenderEquality	0.082	0.226	0.087	0.178
	(0.144)	(0.206)	(0.140)	(0.202)
Immigration*GenderEquality	-0.281	-0.511^*	0.051	-0.459
	(0.206)	(0.290)	(0.104)	(0.285)
Female		0.164		0.197
		(0.200)		(0.198)
Immigration*Female		-0.290		-0.244
		(0.289)		(0.283)
GenderEquality*Female		-0.274		-0.176
		(0.288)		(0.281)
Immigration*GenderEquality*Female		0.433		0.257
		(0.413)		(0.403)
Covariates?	No	No	Yes	Yes
R^2	0.001	0.001	0.057	0.057
Adj. R^2	-0.000	-0.001	0.054	0.054
Num. obs.	2994	2991	2991	2991
RMSE	2.820	2.820	2.741	2.742

$^{***}p < 0.01$; $^{**}p < 0.05$; $^*p < 0.1$.
Covariates include age, education, employment, region and migration background

TABLE C.11: Study 2 (Germany) - Treatment interaction model. DV: Does immigration undermine or enrich cultural life?, 0=Under, 10=Enrich

	M1	M2	M3	M4
Ref:Control	3.702***	3.790***	4.412***	4.351***
	(0.094)	(0.140)	(0.174)	(0.192)
Immigration	−0.042	−0.009	−0.007	0.004
	(0.136)	(0.197)	(0.131)	(0.195)
GenderEquality	0.089	0.165	0.087	0.127
	(0.133)	(0.193)	(0.128)	(0.188)
Immigration*GenderEquality	−0.142	−0.105	−0.270***	−0.076
	(0.190)	(0.271)	(0.095)	(0.265)
Female		−0.183		−0.161
		(0.189)		(0.185)
Immigration*Female		−0.052		−0.020
		(0.271)		(0.265)
GenderEquality*Female		−0.161		−0.074
		(0.266)		(0.257)
Immigration*GenderEquality*Female		−0.100		−0.248
		(0.380)		(0.369)
Covariates?	No	No	Yes	Yes
R^2	0.001	0.005	0.070	0.071
Adj. R^2	−0.000	0.002	0.067	0.067
Num. obs.	2994	2991	2991	2991
RMSE	2.601	2.595	2.510	2.510

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Covariates include age, education, employment, region and migration background

TABLE C.12: Study 2 (Germany) - Treatment interaction model. DV: Should Germany welcome more or fewer immigrants than today? 0=Fewer, 10=More

	Model 1	Model 2	Model 3	Model 4
Ref:Control	3.026***	3.242***	3.503***	3.492***
	(0.050)	(0.068)	(0.214)	(0.216)
Immigration	−0.119*	−0.121	−0.096	−0.074
	(0.071)	(0.098)	(0.070)	(0.099)
GenderEquality	−0.147**	−0.191*	−0.152**	−0.161
	(0.070)	(0.099)	(0.069)	(0.099)
Immigration*GenderEquality	0.167*	0.256*	0.163*	0.181
	(0.099)	(0.142)	(0.097)	(0.142)
Female		−0.468***		−0.353***
		(0.099)		(0.102)
Immigration*Female		0.074		−0.045
		(0.140)		(0.141)
GenderEquality*Female		0.109		0.016
		(0.138)		(0.138)
Immigration*GenderEquality*Female		−0.219		−0.039
		(0.195)		(0.194)
Covariates?	No	No	Yes	Yes
R^2	0.002	0.037	0.087	0.087
Adj. R^2	0.001	0.034	0.080	0.079
Num. obs.	2237	2237	2145	2145
RMSE	1.174	1.155	1.121	1.121

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
M1= Main effects, M2= Interaction with respondent gender, M3= Main effects (cov.adjusted), M4= Interaction with respondent gender (cov.adjusted)
Covariates include age, income, education, and citizenship

TABLE C.13: Study 1 (Norway) - Treatment interaction model. DV: Should Norway welcome more or fewer immigrants than today?, 1-5, 5=Sig. fewer immigrants

	M1	M2	M3	M4
Ref:Control	0.090***	0.126***	0.130***	0.138***
	(0.010)	(0.018)	(0.020)	(0.023)
Immigration	0.005	−0.006	0.004	−0.004
	(0.015)	(0.024)	(0.015)	(0.024)
GenderEquality	−0.009	−0.018	−0.010	−0.015
	(0.014)	(0.024)	(0.014)	(0.024)
Immigration*GenderEquality	0.021	0.021	−0.047***	0.016
	(0.021)	(0.034)	(0.011)	(0.033)
Female		−0.068***		−0.061***
		(0.021)		(0.021)
Immigration*Female		0.020		0.015
		(0.030)		(0.030)
GenderEquality*Female		0.014		0.008
		(0.029)		(0.029)
Immigration*GenderEquality*Female		0.000		0.012
		(0.043)		(0.042)
Covariates?	No	No	Yes	Yes
R^2	0.001	0.009	0.018	0.019
Adj. R^2	0.000	0.007	0.015	0.015
Num. obs.	3002	2999	2999	2999
RMSE	0.291	0.290	0.289	0.289

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
Covariates include age, education, employment, region and migration background

TABLE C.14: Study 2 (Germany) - Treatment interaction model. DV: Would vote for the radical right in a GE (1), otherwise (0)

	Model 1	Model 2	Model 3	Model 4
Ref:Control	0.071***	0.090***	0.100*	0.092*
	(0.011)	(0.016)	(0.053)	(0.052)
Immigration	−0.014	−0.004	−0.009	0.002
	(0.015)	(0.025)	(0.015)	(0.025)
GenderEquality	0.002	0.010	0.004	0.018
	(0.015)	(0.024)	(0.015)	(0.024)
Immigration*GenderEquality	0.025	0.029	0.021	0.014
	(0.022)	(0.036)	(0.022)	(0.036)
Female		−0.041**		−0.026
		(0.021)		(0.022)
Immigration*Female		−0.012		−0.024
		(0.030)		(0.031)
GenderEquality*Female		−0.015		−0.031
		(0.030)		(0.030)
Immigration*GenderEquality*Female		−0.010		0.016
		(0.043)		(0.044)
Covariates?	No	No	Yes	Yes
R^2	0.001	0.015	0.034	0.035
Adj. R^2	0.000	0.011	0.027	0.026
Num. obs.	2246	2246	2152	2152
RMSE	0.259	0.257	0.254	0.254

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

M1= Main effects, M2= Interaction with respondent gender, M3= Main effects (cov.adjusted), M4= Interaction with respondent gender (cov.adjusted)

Covariates include age, income, education, and citizenship

TABLE C.15: Study 1 (Norway) - Treatment interaction model. DV: Would vote for the radical right (FrP) in a GE (=1), would vote for another party (=0)

C.0.6 Manipulation checks

Did respondents actually engage with the treatment text? Encouragingly, the majority of respondents correctly remembered the topic of the treatment text they were assigned to. Participants exposed to an immigration text were significantly more likely to recall the text they read was about immigration (see Tables C.19 and C.22), while respondents assigned to a gender equality text were able to recall that the text was about gender equality (see Tables C.18 and C.21). In Study 1 (Norway), 77% of respondents in the gender+immigration (T1) condition correctly recalled that the text was about immigration, and of these, 46% of respondents condition remembered that the text was about gender equality and immigration. 69% of respondents in the immigration only condition (T2) remembered that the text was only about immigration. In the gender text condition (T3), 51% of respondents correctly remembered that the text was only about gender equality. In T4 (acceptability + immigration), 73% of participants correctly recalled that the text was only about immigration. In Study 2 (Germany), 68% of respondents remembered that the gender+immigration text (T1) was about immigration, and of these, around 25% recalled that the text was about immigration and gender equality. 76% of respondents in T2 (immigration only text) correctly recalled that the text was only about immigration. 73% of respondents in T3 (gender only text) correctly remembered that the text was only about gender equality. The proportion of respondents who said that the gender+immigration text was about both gender equality and immigration may have been smaller because respondents who did not read the question instructions carefully might have not understood that multiple responses were possible. All other multiple choice questions in the survey required a single response. Women were slightly better than men at remembering that the gender equality plus immigration text was about gender equality (Tables C.18 and C.21). Respondents also correctly identified the immigration texts, including the gendered immigration message, as coming from the radical right-wing party (see Tables C.20 and C.23). Overall, these results are encouraging and suggest that the treatment manipulation was successful - respondents paid attention to the treatment texts and were able to remember the topic of the text they read.

The following manipulation check question was asked:

Q7. When you think back to the hypothetical political statement you read earlier, do you remember what topic it was about? Please tick all options that apply. [up to three choices are allowed, order of first four response items randomised] Immigration/ Healthcare/ Gender equality/ Foreign policy/ Don't know

Respondents were also asked to guess which political party the statement they read might have come from:

Q8. Which party or parties do you think the hypothetical political statement you read earlier could have come from? [Multiple response options allowed, list of political parties as answer options]

TABLE C.16: Manipulation check in Study 1 (Norway).

		Control (N=578)		T1 (N=588)		T2 (N=507)		T3 (N=573)		T4 (N=574)	
		N	Pct.	N	Pct.	N	Pct.	N	Pct.	N	Pct.
Topic	Gender equality	0	0.0	14	2.4	4	0.8	295	51.5	3	0.5
	Gender equality & Other	0	0.0	12	2.0	11	2.2	22	3.8	9	1.6
	Immigration	0	0.0	189	32.1	351	69.2	72	12.6	418	72.8
	Immigration & Other	0	0.0	16	2.7	53	10.5	27	4.7	48	8.4
	Immigration & Gender Equality	0	0.0	268	45.6	14	2.8	64	11.2	21	3.7
	Immigration & Gender Equality & Other	0	0.0	68	11.6	47	9.3	57	9.9	48	8.4
	Other	0	0.0	12	2.0	11	2.2	17	3.0	14	2.4
	Dont know	0	0.0	9	1.5	16	3.2	19	3.3	13	2.3
	Not asked	578	100.0	0	0.0	0	0.0	0	0.0	0	0.0

Q: When you think back to the hypothetical political statement you read earlier, do you remember what topic it was about?
Please tick all options that apply. [up to three choices allowed, order randomised]:
Immigration/Healthcare/Gender equality/Foreign policy/Dont know.

TABLE C.17: Manipulation check in Study 2 (Germany).

Topic	Control (N=752)		T1 (N=747)		T2 (N=753)		T3 (N=750)	
	N	Pct.	N	Pct.	N	Pct.	N	Pct.
Gender equality	0	0.0	81	10.8	5	0.7	551	73.5
Gender equality & Other	0	0.0	4	0.5	5	0.7	17	2.3
Immigration	0	0.0	316	42.3	571	75.8	49	6.5
Immigration & Other	0	0.0	45	6.0	82	10.9	17	2.3
Immigration & Gender Equality	0	0.0	190	25.4	3	0.4	27	3.6
Immigration & Gender Equality & Other	0	0.0	39	5.2	14	1.9	13	1.7
Other	0	0.0	28	3.7	31	4.1	33	4.4
Dont know	0	0.0	44	5.9	42	5.6	43	5.7
Not asked	752	100.0	0	0.0	0	0.0	0	0.0

Q: When you think back to the hypothetical political statement you read earlier, do you remember what topic it was about?
Please tick all options that apply. [up to three choices allowed, order randomised]:
Immigration/Healthcare/Gender equality/Foreign policy/Dont know.

	Model 1	Model 2
Ref:Immigration Only (T2)	0.150***	0.147***
	(0.016)	(0.023)
Gender+Immigration(T1)	0.466***	0.407***
	(0.026)	(0.037)
GenderEquality(T3)	0.614***	0.623***
	(0.024)	(0.034)
Acceptability+Immigration(T4)	−0.009	0.015
	(0.022)	(0.032)
Female		0.006
		(0.032)
Female*Gender+Immigration(T1)		0.114**
		(0.051)
Female*GenderEquality(T3)		−0.016
		(0.048)
Female*Acceptability+Immigration(T4)		−0.044
		(0.043)
R^2	0.314	0.318
Adj. R^2	0.313	0.316
Num. obs.	2242	2242
RMSE	0.410	0.409

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE C.18: Study 1 (Norway). Manipulation check. DV: Manipulation check: text was about gender equality (=1), Other topic (=0)

	Model 1	Model 2
Ref:Immigration Only (T2)	0.917***	0.922***
	(0.012)	(0.018)
Gender+Immigration(T1)	0.003	−0.003
	(0.017)	(0.024)
GenderEquality(T3)	−0.533***	−0.524***
	(0.024)	(0.033)
Acceptability+Immigration(T4)	0.015	−0.007
	(0.016)	(0.024)
Female		−0.010
		(0.025)
Female*Gender+Immigration(T1)		0.010
		(0.033)
Female*GenderEquality(T3)		−0.020
		(0.048)
Female*Acceptability+Immigration(T4)		0.041
		(0.032)
R^2	0.329	0.329
Adj. R^2	0.328	0.327
Num. obs.	2242	2242
RMSE	0.337	0.337

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE C.19: Study 1 (Norway). Manipulation check. DV: Manipulation check: text was about Immigration (=1), Other topic (=0)

	Model 1	Model 2
Ref:Immigration Only (T2)	0.801***	0.810***
	(0.018)	(0.026)
Gender+Immigration(T1)	0.007	−0.037
	(0.024)	(0.036)
GenderEquality(T3)	−0.611***	−0.593***
	(0.024)	(0.035)
Acceptability+Immigration(T4)	−0.027	−0.045
	(0.025)	(0.036)
Female		−0.018
		(0.036)
Female*Gender+Immigration(T1)		0.085*
		(0.048)
Female*GenderEquality(T3)		−0.039
		(0.048)
Female*Acceptability+Immigration(T4)		0.033
		(0.050)
R^2	0.302	0.304
Adj. R^2	0.301	0.302
Num. obs.	2242	2242
RMSE	0.402	0.401

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

TABLE C.20: Study 1 (Norway). Manipulation check. DV: Text came from which party? FrP (=1), Other party (=0)

	Model 1	Model 2
Ref:Immigration Only (T2)	0.036***	0.046***
	(0.007)	(0.011)
Gender+Immigration(T1)	0.384***	0.312***
	(0.019)	(0.027)
GenderEquality(T3)	0.775***	0.765***
	(0.016)	(0.023)
Female		−0.020
		(0.014)
Female*Gender+Immigration(T1)		0.152***
		(0.038)
Female*GenderEquality(T3)		0.019
		(0.032)
R^2	0.411	0.417
Adj. R^2	0.411	0.416
Num. obs.	2250	2248
RMSE	0.379	0.378

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

TABLE C.21: Study 2 (Germany). Manipulation check. DV: Manipulation check: text was about gender (=1), Other topic (=0)

	Model 1	Model 2
Ref:Immigration Only (T2)	0.890***	0.880***
	(0.011)	(0.017)
Gender+Immigration(T1)	−0.100***	−0.103***
	(0.019)	(0.027)
GenderEquality(T3)	−0.748***	−0.731***
	(0.017)	(0.025)
Female		0.019
		(0.023)
Female*Gender+Immigration(T1)		0.006
		(0.038)
Female*GenderEquality(T3)		−0.035
		(0.034)
R^2	0.462	0.462
Adj. R^2	0.461	0.461
Num. obs.	2250	2248
RMSE	0.359	0.359

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

TABLE C.22: Study 2 (Germany). Manipulation check. DV: Manipulation check: text was about immigration (=1), Other topic (=0)

	Model 1	Model 2
Ref:Immigration Only (T2)	0.495***	0.482***
	(0.018)	(0.026)
Gender+Immigration(T1)	−0.056**	−0.040
	(0.026)	(0.036)
GenderEquality(T3)	−0.453***	−0.435***
	(0.020)	(0.028)
Female		0.025
		(0.036)
Female*Gender+Immigration(T1)		−0.031
		(0.052)
Female*GenderEquality(T3)		−0.035
		(0.039)
R^2	0.185	0.185
Adj. R^2	0.184	0.183
Num. obs.	2250	2248
RMSE	0.423	0.424

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

TABLE C.23: Study 2 (Germany). Manipulation check. DV: Text came from which party? radical right (=1), other party (=0)

C.0.7   Robustness checks: Study 1 survey groups

The survey experiment was fielded as a part of the Norwegian Citizen Panel to three distinct survey groups which received different sets of questions before and after this survey experiment (survey groups 2,6 and 7). Two of these survey groups (groups 6 and 7) received a set of political questions and primes directly related to the topic of and outcomes of the experiment just before being shown the survey experiment (see Table C.24). One of the groups participated in an immigration policy and a welfare policy experiment, and the other group answered a number of questions about politics, as well as party like/dislike questions, which are similar to the vote intention outcomes used in this survey experiment. The figures and tables presented in the main results are thus restricted to survey group 6 which received the least priming prior to engaging in the survey experiment. However, when including the other survey groups in the analysis, the results still hold (see Table C.25). As expected, the treatment effects are smaller in size when the pre-primed survey groups are included, but are still statistically significant and follow the same substantive direction as the treatment effects uncovered in the group which only received one unrelated survey question prior to engaging with the treatment texts.

Participants in survey group 2 received a longer survey with immigration policy primes, candidate experiments and propensity to vote questions that are very similar to the outcomes used in my own experiment. Participants in group 7 answered a number of general questions about politics, as well as party like/dislike questions, which are similar to the vote intention outcomes used in my own survey experiment.

Participants in group 6 received the least priming by other question batteries before engaging with my survey experiment.

When looking at the predicted values for each of the outcomes by survey group, there seem to be differences in how respondents in each of the survey groups responded to the treatments. Respondents in Group 7 (blue squares in the Figures C.8 to Figure C.10 below) are those respondents who are participating in the survey for the first time, and who filled in political and party evaluation questions just before my survey experiment. These respondents seem to have a stronger reaction to the simple anti-immigration frame (T2) compared to the other treatments, and to respondents in the other survey groups.

Respondents in Group 2 (red circles in the Figures C.8 to Figure C.10 below) are those respondents who were exposed to several candidate experiments before my treatment, and who also engaged with an immigration policy experiment prior to participating in my survey experiment. What is striking in this respondent group is their higher baseline (Control) support for the radical right. While around 8 to 9% of respondents

Question group	Survey group 2	Survey group 6	Survey group 7
1	Climate change	Microplastics	Trust, life satisfaction
2	Voted which party in 2017	**Survey experiment**	Most important issue
3	Self-placement L-r scale	COVID attitudes	Perception: economic situation in Norway
4	Attitudes: Police reform, meat production, wealth inequality	COVID attitudes	Satisfaction: democracy in Norway, current government
5	Positive/negative partisanship	COVID attitudes	Like/dislike political parties
6	Candidate experiment	COVID attitudes	Self-placement l-r scale
7	Experiment: welfare policy and PTV scores (1-10)	COVID attitudes	Confidence in parliament
8	Experiment: immigration policy and PTV scores (1-10)	COVID attitudes	Confidence in politicians
9	Experiment: share pictures on Facebook	COVID attitudes	Confidence in news
10	**Survey experiment**	COVID attitudes	**Survey experiment**
Recruitment	Participated in previous panel	Participated in previous panel	Newly recruited
Included?	Excluded from analysis	Included in analysis	Excluded from analysis
N	1255	2820	2041

TABLE C.24: Overview of survey flow by survey group

report voting for the Progress Party if there was a general election in the groups 6 and 7, around 17% of respondents report that they would vote for the Progress Party in Group 2 (see Figure C.9).

This suggests that some of the effects or priming from the previous questions are still present in my survey experiment. The treatment in groups 2 and 7 is thus more like a compound treatment of how respondents read my own treatment texts, *in combination with* the previous questions and stimuli, rather than a "clean" test of just my own experimental treatments. Reducing the sample to participants in Group 6 provides the best test of my hypotheses because this is the group which received the least priming before engaging with my experiment.

But why would this matter? Wouldn't treatment assignment in the previous experiments be independent of treatment assignment in my own experiment - therefore this should not be a problem? A chi-square test was carried out to assess whether treatment assignment in the immigration policy experiment (that participants were engaging with just before my treatment) was associated with treatment assignment in my experiment. Treatment assignment in the "pre-treatment" experiments is independent of assignment in my own survey experiment (p= 0.414). However, it is still plausible that the assignment in the previous experiment (e.g. to a treatment text about immigration policy) affects how individuals understand and respond to my treatments. This is a problem when individuals do not react symmetrically (e.g. people assigned to the treatment group in the preceding experiment react more strongly to my treatment, and differently than people who were in the control group in the preceding experiment). In that scenario, treatment effects do not cancel each other out, and the average treatment effect recovered would not resemble the ATE we would obtain from a "clean" sample that was not pre-exposed to an immigration policy/candidate experiment.

What happened in Group 2? To further justify why it might make sense to exclude Group 2 participants from my sample, I look at how assignment in the previous experiments/survey questions affects responses to my own experimental outcomes. In the welfare/immigration policy experiment that participants engaged with before my treatment, participants were assigned to one of four treatment texts: a text about an increase in refugees for which the Progress Party is responsible ("Immigration_FrP"), a text about an increase in refugees for which the Left Party is responsible ("Immigration_Left"), a text about an increase in unemployment for which the Labour Party is responsible ("Welfare_Labour"), and a text about an increase in unemployment for which the Conservative Party is responsible ("Welfare_Con"). The treatment text in the welfare condition reads: "We will now introduce you to an imaginary political situation. We will then ask you some questions about your reactions to this situation. Imagine

that a general election is to take place in the autumn. *The Norwegian Labour Party wins the election and their leader becomes prime minister in a red-green majority government consisting of the Norwegian Labour Party, the Centre Party and the Socialist Left Party/The Conservative Party of Norway wins the election their leader becomes prime minister in a conservative majority government consisting of the Conservative Party of Norway, the Progress Party and the Liberal Party of Norway.* After the election, oil prices fall dramatically and unemployment rises sharply. In the face of this economic crisis, government chooses to cut welfare funding. This includes cuts to unemployment benefits and pensions, as well as more expensive kindergarten fees." The immigration treatment text reads: "We will now introduce you to an imaginary political situation. We will then ask you some questions about your reactions to this situation. Imagine that a general election is to take place in the autumn. *The Progress Party does very well in the election and becomes responsible for immigration in a majority government consisting of the Conservative Party of Norway and the Progress Party/The Socialist Left Party does very well in the election and becomes responsible for immigration in a majority government consisting of the Norwegian Labour Party and the Socialist Left Party.* After the election, the conflict in Syria flares up again and Norway experiences a dramatic increase in the number of asylum seekers. In the face of this refugee crisis, the government chooses to soften its immigration policy so that very many asylum seekers are allowed to stay in Norway.Think about the government's decision to soften immigration policy in the face of the refugee crisis. Then, imagine that a general election is to take place tomorrow."

In Figures C.5 to C.7, I look at how the responses to my outcomes vary between these pre-treatment groups. If the effects from the welfare/immigration experiment carry over to my experiment, I would expect participants exposed to the immigration treatments to mainly drive the surprising effects. As the sample is now restricted to respondents in survey group 2, there are only around 50 - 60 respondents in each subgroup, making it difficult to draw conclusive inferences. Tentatively looking at Figures C.5 to C.7, it does look like the diverging treatment effects in Group 2 could be driven by the Immigration Left Party treatment and the Immigration Progress Party treatment. Participants might react to my immigration treatments by decreasing their support for the Progress Party because these are at odds with what they have just been exposed to. While the pre-treatment immigration policy texts convey that the government is softening immigration policy, my immigration treatment is a political statement arguing in favour of restricting immigration. This dissonance between the information conveyed in the previous treatment and my treatment text might have angered or confused participants, and therefore altered how they engage with my treatment texts.

**Predicted values for immigration attitudes
by pre-treatment experimental group**

Y: Should Norway welcome more or fewer immigrants than today?, 1-5, 5=Sig. fewer immigrants
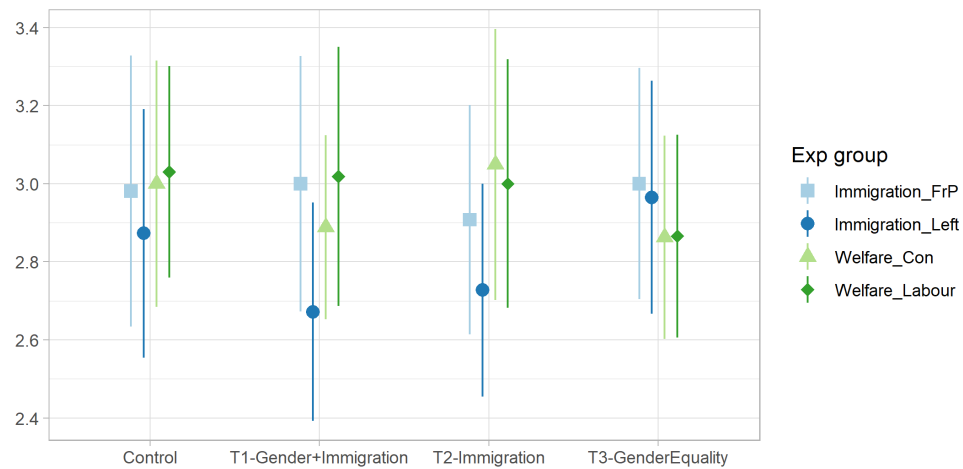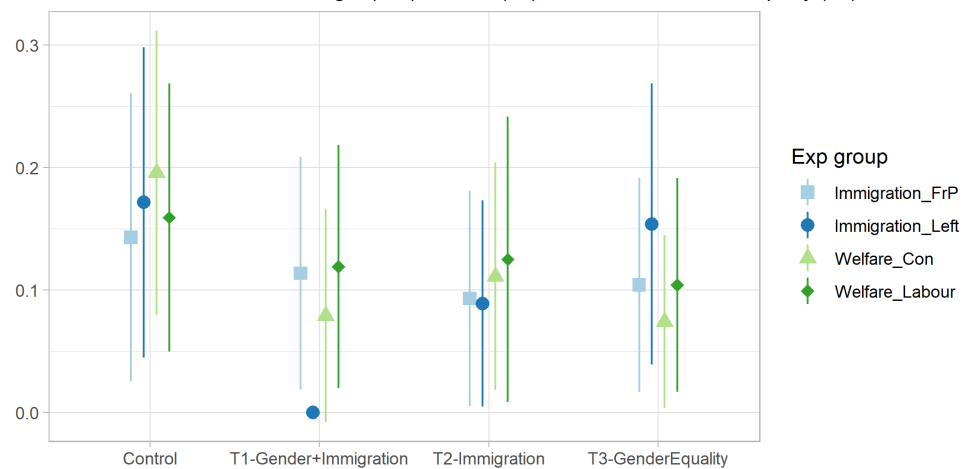

95% CI, values calculated from the regression model: Y=Survey group*GenderEquality*Immigration

FIGURE C.5: Immigration attitudes by pre-treatment experimental group

**Predicted values for radical right vote
by pre-treatment experimental group**

Y: Would vote for the radical right (FrP) in a GE (=1), would vote for another party (=0)

95% CI, values calculated from the regression model: Y=Survey group*GenderEquality*Immigration

FIGURE C.6: Radical right vote by pre-treatment experimental group

Predicted values for acceptability of anti-immigration positions by pre-treatment experimental group

Y: Is it acceptable to say something negative about immigrants? 0-10, 10=Very acceptable
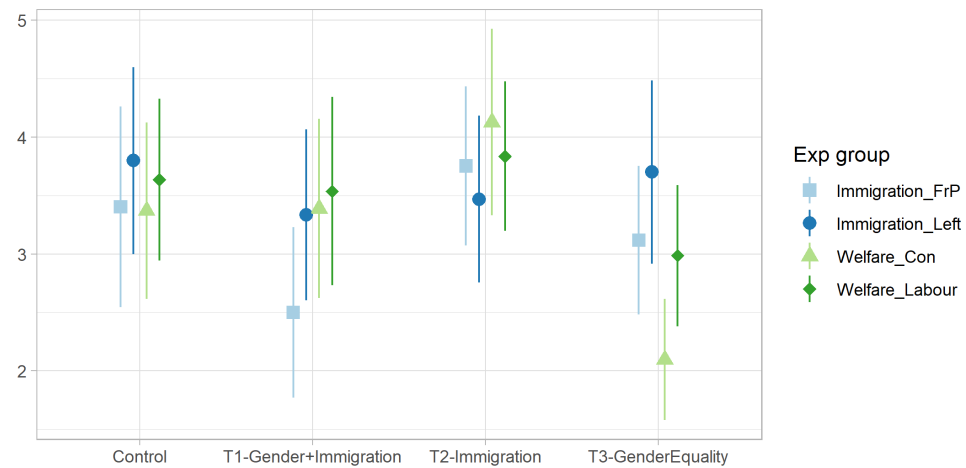
95% CI, values calculated from the regression model: Y=Survey group*GenderEquality*Immigration

FIGURE C.7: Acceptability by pre-treatment experimental group

FIGURE C.8: Immigration attitudes by survey group

FIGURE C.9: Radical right vote by survey group

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Ref:Control	3.456***	4.029***	3.340***	3.983***	3.380***	3.982***
	(0.121)	(0.176)	(0.090)	(0.133)	(0.082)	(0.120)
Immigration	0.056	0.105	0.149	0.225	0.172	0.250
	(0.177)	(0.262)	(0.130)	(0.195)	(0.117)	(0.174)
GenderEquality	−0.283*	−0.304	−0.235*	−0.328*	−0.313***	−0.437***
	(0.166)	(0.240)	(0.126)	(0.183)	(0.112)	(0.164)
Immigration*GenderEquality	0.688***	0.980***	0.390**	0.494*	0.318*	0.459*
	(0.242)	(0.357)	(0.181)	(0.270)	(0.163)	(0.242)
Female		−1.239***		−1.330***		−1.261***
		(0.235)		(0.175)		(0.159)
Immigration*Female		0.090		−0.028		−0.056
		(0.347)		(0.254)		(0.228)
GenderEquality*Female		0.077		0.214		0.286
		(0.321)		(0.244)		(0.219)
Immigration*GenderEquality*Female		−0.665		−0.274		−0.316
		(0.472)		(0.354)		(0.318)
Survey groups included	Group 6	Group 6	6 and 7	6 and 7	2, 6, 7	2, 6, 7
R^2	0.009	0.064	0.005	0.059	0.005	0.052
Adj. R^2	0.008	0.061	0.004	0.057	0.004	0.051
Num. obs.	2234	2234	3874	3874	4837	4837
RMSE	2.860	2.781	2.823	2.747	2.829	2.762

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.
M1 and M2 are the models presented in the main results.

TABLE C.25: Full linear regression model (comparing all Ts and Control). DV: Is it acceptable to say something negative about immigrants? 0-10, 10=Very acceptable

Predicted values for acceptability of anti-immigration positions by survey group

Y: Is it acceptable to say something negative about immigrants? 0-10, 10=Very acceptable

95% CI, values calculated from the regression model: Y=Survey group*GenderEquality*Immigration
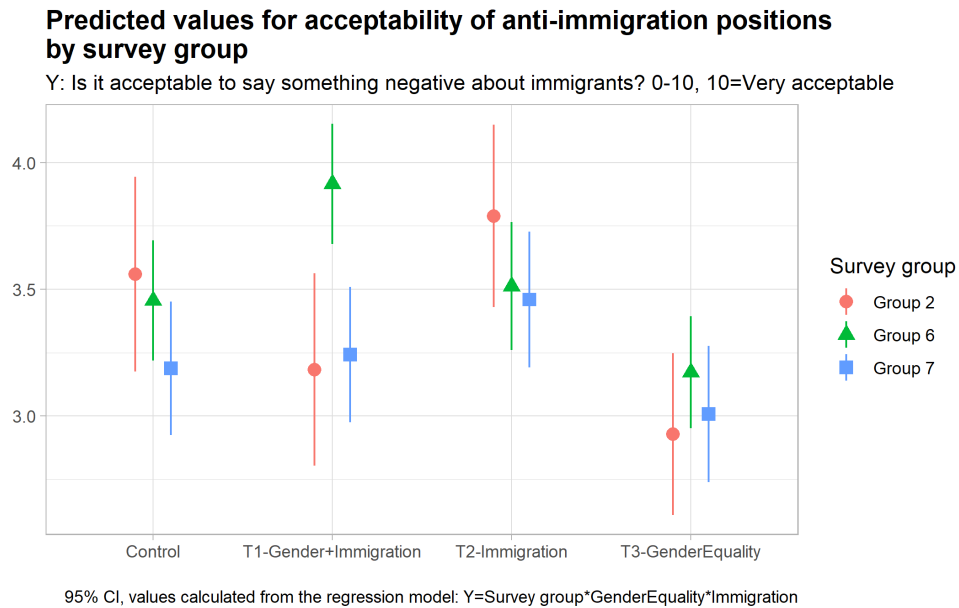
FIGURE C.10: Acceptability by survey group

C.0.8 Additional results - Pilot (Germany 2018)

C.0.8.1 Pilot design

The pilot consisted of an online survey experiment fielded in Germany in 2018. The survey was administered to an online convenience sample (n=385) from April 18-19 2018 via Qualtrics and distributed through the crowdsourcing platform Clickworker.com. Clickworker.com, similar to Amazon MTurk in the U.S., allows individuals and companies to distribute Human Intelligence Microtasks to a crowd of individual workers. In contrast to Amazon MTurk, the majority of individuals registered on Clickworker.com (an estimated 150,000) are located in Germany. Participants were recruited informing them that they could earn 1 Euro for completing a 5-7 minute survey on political attitudes and behaviour. The survey length and accessibility was tested in a pilot.

In Study 1 (Germany), the survey experiment consisted of a between-subjects design with three experimental groups to which participants were randomly assigned. As the main treatment manipulation, participants either read an anti-immigration text that explicitly referred to women (Treatment), or an anti-immigration text that didn't refer to women (Control). Participants from all three treatment groups also read a distractor text about climate change before reading the main treatment text about immigration. This was meant to make it more difficult for participants to infer the purpose of the treatment manipulation. Besides the Treatment and the Control group, the experiment also included a third condition. Respondents in the Placebo group read a text about climate change that included an explicit gender cue. The placebo was meant to test

whether the effect of the gender-immigration prime is really due to the unique combination of gender and immigration messages, or whether a gender prime connected to another topic would have also elicited this reaction. A summary of the three experimental groups can be found in Table C.26. Climate change was chosen as the topic for the distractor text because it provided a credible, ecologically valid gender prime on a salient issue unrelated to immigration policy. Other experimental research has successfully used environmental primes as placebos for ethnic identity and immigration-related considerations (Valenzuela and Michelson, 2016). In the pilot, the treatment manipulation itself only involved a two word change - replacing "people" with "women and girls". The statement on immigration was taken from the AfD's 2016 regional election manifesto [1]. The statement on climate change was taken from the Greens' 2017 General Election manifesto [2].

TABLE C.26: Overview of experimental groups in Study 1 (Germany)

Experimental group	Content	N
Control	Climate change text (no gender) Immigration text (no gender)	119
Treatment	Climate change text (no gender) Immigration text (with gender)	118
Placebo	Climate change text (with gender) Immigration text (no gender)	117
Sample size		354

C.0.8.2 Pilot treatment texts

C.0.8.3 Outcomes and Estimation

To measure immigration attitudes in the pilot, two standard questions on immigration attitudes were fielded. The first immigration question tapped into immigration as economic threat (Would you say that it is generally good or bad for Germany's economy

---

[1] Accessible online at `http://www.weiterdenken.de/sites/default/files/uploads/2016/06/scholten_einwanderungs-_flucht-_und_asylpolitik_der_alternative_fuer_deutschland.pdf`

[2] Accessible online at: `https://www.gruene.de/fileadmin/user_upload/Dokumente/BUENDNIS_90_DIE_GRUENEN_Bundestagswahlprogramm_2017.pdf`

Treatment (Immigration+gender)
"Climate change is one of the biggest challenges for our country. If we do not act now, temperatures will rise up to 4 degrees until the end of the century. Globally, it is the poorest, [**omitted**] who are most affected by environmental destruction- even though they contribute the least to it."
"Mass migration is a big threat to our future. The **women and girls** in Germany are the biggest victims of our open borders policy. We demand that the interests of our own country and our native population are put centre-stage again."

Control (Immigration)
"Climate change is one of the biggest challenges for our country. If we do not act now, temperatures will rise up to 4 degrees until the end of the century. Globally, it is the poorest, [**omitted**] who are most affected by environmental destruction- even though they contribute the least to it."
"Mass migration is a big threat to our future. The **people** in Germany are the biggest victims of our open borders policy. We demand that the interests of our own country and our native population are put centre-stage again."

Placebo (Climate change+gender)
"Climate change is one of the biggest challenges for our country. If we do not act now, temperatures will rise up to 4 degrees until the end of the century. Globally, it is the poorest, **women and girls** who are most affected by environmental destruction- even though they contribute the least to it."
"Mass migration is a big threat to our future. The **people** in Germany are the biggest victims of our open borders policy. We demand that the interests of our own country and our native population are put centre-stage again."

Note: Treatment manipulations in bold.

TABLE C.27: Overview of treatment texts - Pilot (Germany)

that people come to live here from other countries?). The second question tapped into immigration as cultural threat (Would you say Germany's cultural life is generally undermined or enriched by people coming to live here from other countries?). For both questions, participants answered on a scale from 0 (Bad for the economy/Cultural life undermined) to 10 (Good for the economy/Cultural life enriched). Higher scores on the immigration scales thus denote more pro-immigration attitudes, while lower scores denote anti-immigration attitudes. For the German survey experiment, Equation C.1 and Equation C.2 were estimated. The main estimator of interest is the β_1 coefficient on the Treatment variable, providing the effect of the gendered anti-immigration message compared to the simple anti-immigration message. Treatment is a binary indicator, coded 1= treatment text contains the climate change text, followed by the anti-immigration text with a gender prime, 0=otherwise. Placebo is a binary indicator, coded 1= treatment text contains the climate change text with a gender prime, followed by the anti-immigration text, 0=otherwise. The reference group is the Control group

which received the climate change text, followed by the anti-immigration text, without any gender primes. I use Equation C.2 to test Hypotheses 1A, 2A and 3A, exploring whether the treatment had a different effect on women, compared to men. The main estimator of interest is the β_4 coefficient on the Treatment and Gender interaction, providing the differential effect of the gendered anti-immigration message for women, compared to men.

$$Y_i = \alpha + \beta_1 * Treatment_i + \beta_2 * Placebo_i + \epsilon_i \qquad (C.1)$$

$$Y_i = \alpha + \beta_1 * Treatment_i + \beta_2 * Placebo_i + \beta_3 * FemaleRespondent_i +$$
$$\beta_4 * Treatment_i * FemaleRespondent_i \; + \beta_5 * Placebo_i * FemaleRespondent_i + \epsilon_i$$
$$(C.2)$$

C.0.8.4 Results (Pilot)

In the pilot, exposure to the gendered immigration message did not have significant effects on aggregate immigration attitudes. However, men and women responded very differently to the gendered anti-immigration text (Figure C.11).
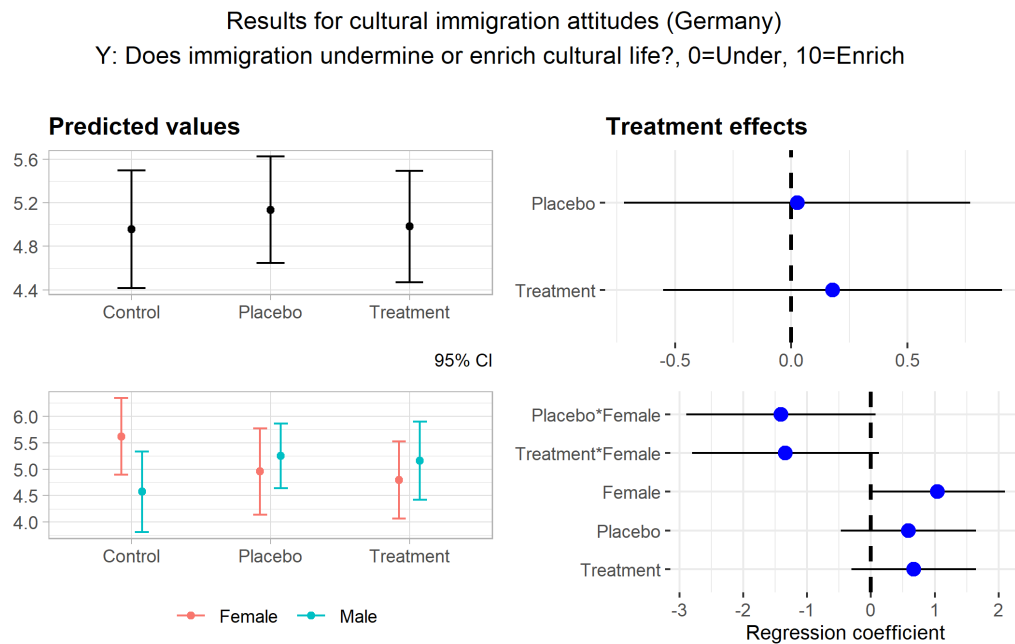
FIGURE C.11: Immigration attitudes (Pilot - Germany 2018)

In the pilot, receiving the gendered immigration message (rather than the immigration only message) made women more likely to state that cultural life in Germany is being undermined by immigration. This 1.4 effect on the 0-10 scale constitutes a change of more than half a Standard Deviation in immigration attitudes and thus presents a sizeable effect. There were some significant differences between the placebo and the control group in cultural immigration attitudes but these were not robust to the inclusion of covariates. Respondents in the German pilot experiment were also asked if they thought immigration was good or bad for the economy. There were no significant treatment effects on economic immigration attitudes. Additionally, respondents were asked a general question about immigration (Should immigration be reduced or increased?). There were significant heterogeneous treatment effects by gender, with women expressing more anti-immigrant views in response to the gender-immigration message than men.

In the pilot, gender-immigration message did not have any aggregate effects on support for the radical right, but significantly increased women's intention for vote for the radical right (Figure C.12).
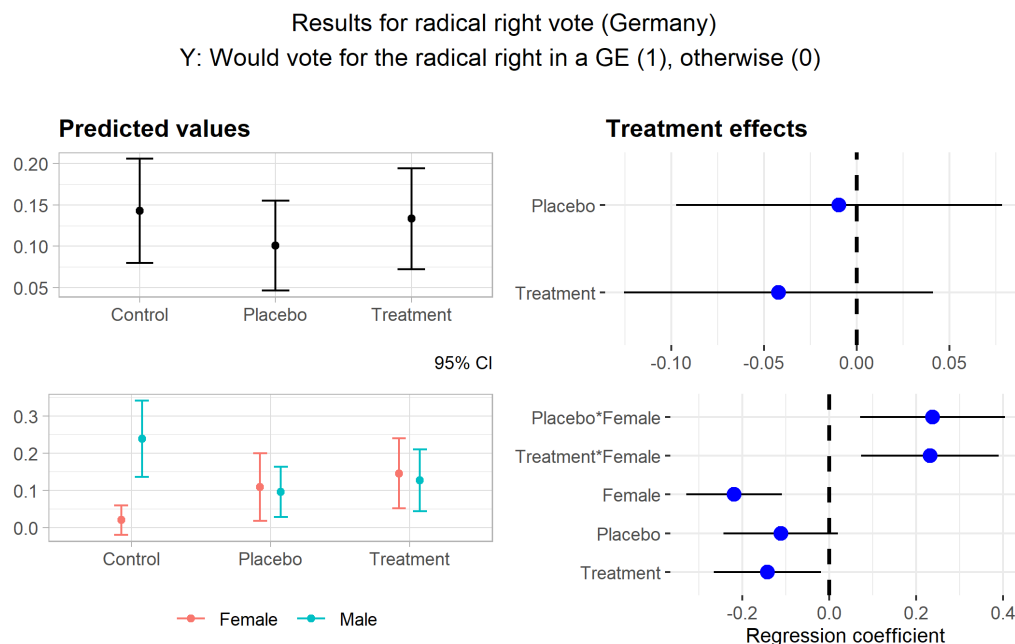
FIGURE C.12: Radical right vote (Pilot - Germany 2018)

In the pilot, for women, receiving the gendered immigration text quadrupled their likelihood to vote for the AfD. While women who read the simple immigration text had an estimated 4% likelihood of voting for the AfD, women in the treatment group had a 17% likelihood of voting for the party. Men had the opposite reaction to the gendered immigration message: reading the gendered immigration statement cut their likelihood to support the AfD in half. Male participants responded to the treatment with a drop in

their likelihood to vote for the AfD, going from an estimated 30% likelihood of support-
ing the party in the control condition, to an estimated 16% likelihood of supporting the
AfD in the treatment condition (Figure C.12). The gender gap in probability of voting
for the AfD is thus around 28 percentage points larger in the control condition than in
the treatment group where the gender gap becomes small and statistically insignificant.
These heterogeneous treatment effects by gender represent sizeable and substantively
important effects. Being exposed to the placebo condition (the gendered climate change
text and the immigration text) also significantly increased women's vote intention for the
radical right. One potential explanation for this is that they gender prime in the climate
change text was so strong that it still affected how respondents read and interpreted the
immigration text.

In the pilot, socio-demographic characteristics were balanced across treatment groups
and similar to a nationally representative sample of the German population, see Table
C.28.

	Gender	Age	Education	Education	Employment	Income	Immigration background	Religiosity
Reference category	Male		A-Level (Abitur)	Apprenticeship (Abgeschlossene Ausbildung)	Full- time job (30 hours or more/week)	2000 to below 2500 Euro/month*	No immigrant background	Never attends church
	%	Mean	%	%	%	%	%	%
Treatment	53.2	37 (SD=12.6)	21.1	19.5	45.4	12.6	79.7	63.0
Control	55.9	35.9 (SD=13.2)	16.4	20.3	44.1	10.2	73.4	58.5
Placebo	58.1	35.1 (SD=12.1)	14.7	19.4	41.2	17.6	80.6	55.5
Average	55.8	36 (SD=12.6)	17.4	19.7	43.5	13.5	77.9	59
Population sample	50.8	50.6(SD=19.4)	18.46	32.4	40.6	14.28	75.9	44
Test	Pearson chi2(2) = 0.6386 Pr = 0.727	ANOVA SS=213.89232 df=2 MS=106.94616 F=0.67 Pr=0.5132	Pearson chi2(16) = 19.6971 Pr = 0.234		Pearson chi2(14) = 13.9562 Pr = 0.453	Pearson chi2(26) = 30.4247 Pr = 0.250	Pearson chi2(2) = 2.2741 Pr = 0.321	Pearson chi2(8) = 7.9644 Pr = 0.437

Note: Income: most common category was "prefer not to say". *Sample averages calculated from the German National Election Study 2017(N=4279).

TABLE C.28: Balance among key covariates in the Pilot (Germany 2018)

Bibliography

Abou-Chadi, Tarik, Magdalena Breyer and Theresa Gessler. 2021. "The (Re)politicisation of gender in western europe." *European Journal of Politics and Gender* 4(2).

Abou-Chadi, Tarik and Werner Krause. 2020. "The Causal Effect of Radical Right Success on Mainstream Parties' Policy Positions: A Regression Discontinuity Approach." *British Journal of Political Science* 50(3):829–847.
URL: *https://www.cambridge.org/core/journals/british-journal-of-political-science/article/abs/causal-effect-of-radical-right-success-on-mainstream-parties-policy-positions-a-regression-discontinuity-approach/6C78B1EF4B39361A9A2B38DF86B24A90*

Abramowitz, A. and S. Webster. 2017. "'Negative Partisanship' Explains Everything.".
URL: *https://www.politico.com/magazine/story/2017/09/05/negative-partisanship-explains-everything-215534/*

Abramowitz, Alan I. and Steven Webster. 2016. "The rise of negative partisanship and the nationalization of U.S. elections in the 21st century." *Electoral Studies* 41.

Abrams, Dominic, Fanny Lalot and Michal A Hogg. 2021. "Intergroup and intragroup dimensions of COVID-19: A social identity perspective on social fragmentation and unity." *Group Processes and Intergroup Relations* 24(2):201–209.

Achen, C and L. Bartels. 2016. *Democracy for Realists*. Princeton University Press.
URL: *https://press.princeton.edu/books/hardcover/9780691169446/democracy-for-realists*

Adams, James, Michael Clark, Lawrence Ezrow and Garrett Glasgow. 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.
URL: *http://doi.wiley.com/10.1111/j.1540-5907.2006.00199.x*

Adams, James, Simon Weschle and Christopher Wlezien. 2021. "Elite Interactions and Voters' Perceptions of Parties' Policy Positions." *American Journal of Political Science* 65:101–114.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/ajps.12510 https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12510 https://onlinelibrary.wiley.com/doi/10.1111/ajps.12510*

Adler, Katya. 2022. "French election: A battle of bad reputations for Le Pen and Macron.".
URL: *https://www.bbc.co.uk/news/world-europe-61175730*

AfD. 2020. "AfD Press release: "Violence against women starts with the hijab".".
URL: *https://afd-brandenburg.de/2020/11/25/birgit-bessin-beim-kopftuch-beginnt-bereits-die-gewalt-gegen-frauen/*

Akkerman, Tjitske. 2015. "Gender and the radical right in Western Europe: a comparative analysis of policy agendas." *Patterns of Prejudice* 49(1-2):37–60.
URL: *http://www.tandfonline.com/doi/full/10.1080/0031322X.2015.1023655*

Akkerman, Tjitske and Sarah L. de Lange. 2012. "Radical Right Parties in Office: Incumbency Records and the Electoral Cost of Governing." *Government and Opposition* 47(04):574–596.
URL: *http://www.journals.cambridge.org/abstract_S0017257X00000737*

Akkerman, Tjitske, Sarah L. de Lange and Matthijs Rooduijn. 2016. *Radical right-wing populist parties in Western Europe : into the mainstream?* Routledge.
URL: *https://www.routledge.com/Radical-Right-Wing-Populist-Parties-in-Western-Europe-Into-the-Mainstream/Akkerman-de-Lange-Rooduijn/p/book/9781138914988*

Allport, G.W. 1954. *The nature of prejudice.* Addison-Wesley.
URL: *https://psycnet.apa.org/record/1954-07324-000*

Anderson, Cameron D., Michael McGregor and Laura B. Stephenson. 2021. "Us versus them: Do the rules of the game encourage negative partisanship?" *European Journal of Political Research* .

Ansolabehere, Stephen, Shanto Iyengar, Adam Simon and Nicholas Valentino. 1994. "Does Attack Advertising Demobilize the Electorate?" *American Political Science Review* 88:829–838.
URL: *https://www.cambridge.org/core/journals/american-political-science-review/article/abs/does-attack-advertising-demobilize-the-electorate/1C63CF214B59466F29FA08E4ABB69292*

APSA. 2020. "Principles and Guidance for Human Subjects Research.".
URL: *https://connect.apsanet.org/hsr/principles-and-guidance/*

Arceneaux, Kevin and David W. Nickerson. 2010. "Comparing Negative and Positive Campaign Messages." *American Politics Research* 38:54–83.
URL: *http://journals.sagepub.com/doi/10.1177/1532673X09331613*

Art, David. 2018. "The AfD and the End of Containment in Germany?" *German Politics and Society* 36:76–86.
URL: *https://www.berghahnjournals.com/view/journals/gps/36/2/gps360205.xml*

Arzheimer, Kai and Carl C. Berning. 2019. "How the Alternative for Germany (AfD) and their voters veered to the radical right, 2013–2017." *Electoral Studies* 60:102040.

Avdagic, Sabina and Lee Savage. 2021. "Negativity Bias: The Impact of Framing of Immigration on Welfare State Support in Germany, Sweden and the UK." *British Journal of Political Science* 51:624–645.
URL: *https://www.cambridge.org/core/journals/british-journal-of-political-science/article/abs/negativity-bias-the-impact-of-framing-of-immigration-on-welfare-state-support-in-germany-sweden-and-the-uk/58A3F1DEA8CF17C57AB077F1B27B0856*

Bafumi, Joseph and Robert Y. Shapiro. 2009. "A new partisan voter." *Journal of Politics* 71:1–24.
URL: *https://www.journals.uchicago.edu/doi/full/10.1017/S0022381608090014*

Bakker, Bert N., Gijs Schumacher and Matthijs Rooduijn. 2021. "Hot Politics? Affective Responses to Political Rhetoric." *American Political Science Review* 115:150–164.
URL: *https://doi.org/10.1017/S0003055420000519*

Bale, Tim and Cristobal Rovira Kaltwasser. 2021. "The Netherlands: How the Mainstream Right Normalized the Silent Counter-Revolution.".
URL: *https://www.cambridge.org/core/books/riding-the-populist-wave/netherlands-how-the-mainstream-right-normalized-the-silent-counterrevolution/7D9CFDBEB09DA2235C0C56BC7FA25318*

Banda, Kevin K. 2021. "Reexamining the Effects of Electoral Competition on Negative Advertising." *American Politics Research* 0(0).
URL: *https://doi.org/10.1177/1532673X211042001*

Banda, Kevin K. and John Cluverius. 2018. "Elite polarization, party extremity, and affective polarization." *Electoral Studies* 56:90–101.

Bankert, Alexa. 2020. "Negative and Positive Partisanship in the 2016 U.S. Presidential Elections." *Political Behavior* pp. 1–19.
URL: *https://doi.org/10.1007/s11109-020-09599-1*

Bankert, Alexa. 2021. "Reflections on the Past and Present of Research on Partisan Identity." *Forum (Germany)* 19:459–480.
URL: *https://www-degruyter-com.gate3.library.lse.ac.uk/document/doi/10.1515/for-2021-2024/html*

Bankert, Alexa, Leonie Huddy and Martin Rosema. 2017. "Measuring Partisanship as a Social Identity in Multi-Party Systems." *Political Behavior* .

Bartels, Larry M. 2002. "Beyond the running tally: Partisan Bias in Political Perceptions." *Political Behavior* 24:117–150.
URL: *http://ezproxy-prd.bodleian.ox.ac.uk:2084/stable/1558352*

Barton, Jared, Marco Castillo and Ragan Petrie. 2016. "Negative campaigning, fundraising, and voter turnout: A field experiment." *Journal of Economic Behavior and Organization* 121:99–113.

Bassan-Nygate, Lotem and Chagai M. Weiss. 2021. "Party Competition and Cooperation Shape Affective Polarization: Evidence from Natural and Survey Experiments in Israel:." *https://doi.org/10.1177/00104140211024283* 55:287–318.
URL: *https://journals.sagepub.com/doi/full/10.1177/00104140211024283?casa_token = 40d7gPRY3sAAAAAA%3AAYFi8lGtt12YlJOVvUtuYiO_a1dz_aP9BLlC6xF0Ga − uWm75It7OL1BcEsnanhZRhQKu7_yJz79KKQ*

Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer and Kathleen D. Vohs. 2001. "Bad Is Stronger Than Good." *Review of General Psychology* 5.

BBC News. 2022. "France election: 'Many voted to build a barricade against the far-right', says Macron.".
URL: *https://www.bbc.co.uk/news/av/world-europe-61211790*

Belmont report. 1979. "The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research.".
URL: *https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html*

Ben-Shitrit, Lihi, Julia Elad-Strenger and Sivan Hirsch-Hoefler. 2022. "'Pinkwashing' the radical-right: Gender and the mainstreaming of radical-right policies and actions." *European Journal of Political Research* 61:86–110.

Berenson, Tessa. 2016. "Donald Trump Proposes 'Extreme Vetting' for Immigrants.".
  URL: *https://time.com/4452970/donald-trump-immigration-isis-terrorism/*

Bischof, Daniel and Markus Wagner. 2019. "Do Voters Polarize When Radical Parties
  Enter Parliament?" *American Journal of Political Science* 63(4):888–904.
  URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/ajps.12449*

Blinder, Scott, Robert Ford and Elisabeth Ivarsflaten. 2013. "The Better Angels of Our
  Nature: How the Antiprejudice Norm Affects Policy and Party Preferences in Great
  Britain and Germany." *American Journal of Political Science* .

Bliuc, Ana Maria, Craig McGarty, Katherine Reynolds and Daniela Muntele. 2007.
  "Opinion-based group membership as a predictor of commitment to political action."
  European Journal of Social Psychology 37:7–23.

Bolin, Niklas, Stefan Dahlberg and Sofie Blombäck. 2022. "The stigmatisa-
  tion effect of the radical right on voters' assessment of political proposals."
  https://doi.org/10.1080/01402382.2021.2019977 pp. 1–22.
  URL: *https://www.tandfonline.com/doi/abs/10.1080/01402382.2021.2019977*

Borbáth, Endre and Theresa Gessler. 2021. "How Do Populist Radical Right Parties
  Differentiate their Appeal? Evidence from the Media Strategy of the Hungarian
  Jobbik Party." *Government and Opposition* pp. 1–22.
  URL: *https://www.cambridge.org/core/journals/government-and-
  opposition/article/how-do-populist-radical-right-parties-differentiate-
  their-appeal-evidence-from-the-media-strategy-of-the-hungarian-jobbik-
  party/0EF74F56137DF7FD8AF0FF22A171347A*

Bos, Linda, Sophie Lecheler, Moniek Mewafi and Rens Vliegenthart. 2016. "It's the frame
  that matters: Immigrant integration and media framing effects in the Netherlands."
  International Journal of Intercultural Relations 55:97–108.

Bosson, Jennifer K., Amber B. Johnson, Kate Niederhoffer and William B. Swann. 2006.
  "Interpersonal chemistry through negativity: Bonding by sharing negative attitudes
  about others.".

Boxell, Levi, Matthew Gentzkow and Jesse M. Shapiro. 2022. "Cross Country Trends
  in Affective Polarization." *The Review of Economics and Statistics* pp. 1–60.

Brader, Ted, Nicholas A. Valentino and Elizabeth Suhay. 2008. "What Triggers Public
  Opposition to Immigration? Anxiety, Group Cues, and Immigration Threat." *Amer-
  ican Journal of Political Science* 52(4):959–978.
  URL: *http://doi.wiley.com/10.1111/j.1540-5907.2008.00353.x*

Brandão, Amélia and Paolo Popoli. 2022. ""I'm hatin' it"! Negative consumer–brand relationships in online anti-brand communities." *European Journal of Marketing* 56:622–650.

Brewer, Marilynn B. 1999. "The Psychology of Prejudice: Ingroup Love and Outgroup Hate?" *Journal of Social Issues* 55:429–444.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/0022-4537.00126 https://onlinelibrary.wiley.com/doi/abs/10.1111/0022-4537.00126 https://spssi.onlinelibrary.wiley.com/doi/10.1111/0022-4537.00126*

Broockman, David, Joshua Kalla and Sean Westwood. 2022. "Does Affective Polarization Undermine Democratic Norms or Accountability? Maybe Not." *AJPS* .
URL: *https://osf.io/9btsq/*

Brooks, Deborah Jordan and John G. Geer. 2007. "Beyond negativity: The effects of incivility on the electorate." *American Journal of Political Science* 51:1–16.

Brown, Jacob R. and Ryan D. Enos. 2021. "The measurement of partisan sorting for 180 million voters." *Nature Human Behaviour 2021 5:8* 5:998–1008.
URL: *https://www.nature.com/articles/s41562-021-01066-z*

Bursztyn, Leonardo and Robert Jensen. 2017. "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure." *Annual Review of Economics* 9(1):131–153.
URL: *http://www.annualreviews.org/doi/10.1146/annurev-economics-063016-103625*

Bush, Sarah Sunn and Pär Zetterberg. 2021. "Gender Quotas and International Reputation." *American Journal of Political Science* 65(2):326–341.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/ajps.12557 https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12557 https://onlinelibrary.wiley.com/doi/10.1111/ajps.12557*

Butler, David and Donald Stokes. 1974. *Political Change in Britain*. Palgrave Macmillan.

Campbell, Angus, Converse Philip E., Miller Waren E. and Donald E. Stokes. 1960. *The American Voter*.

Campbell, Rosie and Silvia Erzeel. 2018. "Exploring Gender Differences in Support for Rightist Parties: The Role of Party and Gender Ideology.".

Caruana, Nicholas J., R. Michael McGregor and Laura B. Stephenson. 2015. "The Power of the Dark Side: Negative Partisanship and Political Behaviour in Canada." *Canadian Journal of Political Science* .

Charness, Gary and Yan Chen. 2020. "Social Identity, Group Behavior, and Teams." *https://doi.org/10.1146/annurev-economics-091619-032800* 12:691–713.
   URL: *https://www.annualreviews.org/doi/abs/10.1146/annurev-economics-091619-032800*

Choi, Donghyun Danny, Mathias Poertner and Nicholas Sambanis. 2021. "The Hijab Penalty: Feminist Backlash to Muslim Immigrants." *American Journal of Political Science* .

Chong, Dennis and James N. Druckman. 2007. "Framing Theory." *Annual Review of Political Science* 10(1):103–126.
   URL: *http://www.annualreviews.org/doi/10.1146/annurev.polisci.10.072805.103054*

Coffé, Hilde. 2018. *Gender and the Radical Right.* Vol. 1 Oxford University Press.
   URL: *http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780190274559.001.0001/oxfordhb-9780190274559-e-10*

Crawford, Brandon L, Kristen N. Jozkowski, Ronna C Turner and Wen-Juo Lo. 2021. "Examining the Relationship Between *Roe v. Wade* Knowledge and Sentiment Across Political Party and Abortion Identity." *Sexuality Research and Social Policy* Online First.

Cutts, David, Matthew Goodwin, Oliver Heath and Paula Surridge. 2020. "Brexit, the 2019 General Election and the Realignment of British Politics." *The Political Quarterly* 91(1):7–23.

Cyphers, Karen H., Hans J.G. Hassell and Kai Ou. 2022. "Racial and Partisan Identities Prompt Campaign Giving: Evidence from a Field Experiment." *Working paper* .

Dancygier, Rafaela. 2020. "Another Progressive's Dilemma: Immigration, the Radical Right Threats to Gender Equality — American Academy of Arts and Sciences.".
   URL: *https://www.amacad.org/publication/another-progressives-dilemma-immigration-radical-right-threats-gender-equalityfootnote39_rkwl1me*

De Lange, Sarah L. and Liza M. Mügge. 2015. "Gender and right-wing populism in the Low Countries: Ideological variations across parties and time." *Patterns of Prejudice* .

de Vries, Catherine and Sara Hobolt. 2020. *Political entrepreneurs: the rise of challenger parties in Europe.* Princeton University Press.

Dekeyser, Elizabeth and Michael Freedman. 2021. "Elections, Party Rhetoric, and Public Attitudes Toward Immigration in Europe." *Political Behavior 2021* pp. 1–13.
   URL: *https://link.springer.com/article/10.1007/s11109-021-09695-w*

Dias, Nicholas and Yphtach Lelkes. 2021. "The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identification." *forthcoming in AJPS* .
URL: *https://osf.io/6dq52/*

Dinas, Elias, Erin Hartman and Joost van Spanje. 2016. "Dead Man Walking: The Affective Roots of Issue Proximity Between Voters and Parties." *Political Behavior* 38:659–687.

Dinas, Elias and Ksenia Northmore-Ball. 2019. "The Ideological Shadow of Authoritarianism." *Comparative Political Studies* .

Dovidio, John F., Ana Validzic and Samuel L. Gaertner. 1998. "Intergroup Bias: Status, Differentiation, and a Common In-Group Identity." *Journal of Personality and Social Psychology* .

Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky and John Barry Ryan. 2020*a*. "Affective polarization, local contexts and public opinion in America." *Nature Human Behaviour 2020 5:1* 5:28–38.
URL: *https://www.nature.com/articles/s41562-020-01012-5*

Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky and John Barry Ryan. 2020*b*. "Affective polarization, local contexts and public opinion in America." *Nature Human Behaviour 2020 5:1* 5:28–38.
URL: *https://www.nature.com/articles/s41562-020-01012-5*

Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky and John Barry Ryan. 2022. "(Mis)estimating Affective Polarization." *Journal of Politics* .
URL: *https://www.journals.uchicago.edu/doi/full/10.1086/715603*

Easton, Matthew J, John B Holbein and Frank Batten. 2021. "The Democracy of Dating: How Political Affiliations Shape Relationship Formation." *Journal of Experimental Political Science* 8:260–272.
URL: *https://www.cambridge.org/core/journals/journal-of-experimental-political-science/article/democracy-of-dating-how-political-affiliations-shape-relationship-formation/79CD8CFC3E8862E586B2661E6C85AD89*

Eberl, Jakob-Moritz, Christine E. Meltzer, Tobias Heidenreich, Beatrice Herrero, Nora Theorin, Fabienne Lind, Rosa Berganza, Hajo G. Boomgaarden, Christian Schemer and Jesper Strömbäck. 2018. "The European media discourse on immigration and its effects: a literature review." *https://doi.org/10.1080/23808985.2018.1497452* 42:207–223.
URL: *https://www.tandfonline.com/doi/abs/10.1080/23808985.2018.1497452*

EGAP. 2011. "EGAP Research Principles.".
URL: *https://egap.org/wp-content/uploads/2020/05/egap-research-principles.pdf*

Elsbach, Kimberly D and C B Bhattacharya. 2001. "Defining Who You Are by What You're Not: Organizational Disidentification and the National Rifle Association." *Organization Science* 12(4):393–413.

Fahey, James, Damon Roberts and Stephen Utych. 2022. "Principled or Partisan? The Effect of Cancel Culture Framings on Support for Free Speech." *Working paper* .
URL: *https://www.researchgate.net/publication/354046920principled$_o$rpartisan$_T$he$_E$ffect$_o$f$_C$ance*

Farris, Sara R. 2017. *In the name of women's rights: the rise of femo-nationalism.* Durham: Duke University Press.

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

Ford, Robert, Tim Bale, Will Jennings and Paula Surridge. 2021. Get Brexit Done: The National Campaign. In *The British General Election of 2019*, ed. Robert Ford, Tim Bale, Will Jennings and Paula Surridge. Basingstoke: Palgarve MacMillan.

Fowler, Anthony. 2020. "Partisan Intoxication or Policy Voting?" *Quarterly Journal of Political Science* 15.

FrP. 2020. "Immigration and Integration policy.".
URL: *https://www.frp.no/stories/innvandring-og-integrering*

Galasso, Vincenzo, Tommaso Nannicini and Salvatore Nunnari. 2020. "Positive Spillovers from Negative Campaigning." *Working paper* .
URL: *https://papers.ssrn.com/sol3/papers.cfm?abstract$_i$d = 3526036*

Gamson, W. A. 1992. "The social psychology of collective action.".

Gamson, W. A. and A. Modigliani. 1987. The changing culture of affirmative action. In *Research in political sociology.* pp. 137–177.

Garzia, Diego and Frederico Ferreira da Silva. 2022. "The Electoral Consequences of Affective Polarization? Negative Voting in the 2020 US Presidential Election:." *https://doi.org/10.1177/1532673X221074633* 2022:303–311.
URL: *https://journals.sagepub.com/doi/full/10.1177/1532673X221074633*

Geer, John G. 2006. *In Defense of Negativity: Attack Ads in Presidential Campaigns.* University of Chicago Press.

Gerber, A.S, D.P Green and M Green. 2003. "Partisan mail and voter turnout: results from randomized field experiments." *Electoral Studies* 22(4):563–579.
URL: *https://www.sciencedirect.com/science/article/pii/S026137940200029X*

Gidron, Noam, James Adams and Will Horne. 2019. "Toward a Comparative Research Agenda on Affective Polarization in Mass Publics." *APSA Comparative Politics Newsletter* 29:30–36.

Gidron, Noam, James Adams and Will Horne. 2020. "American Affective Polarization in Comparative Perspective." *Elements in American Politics* .
URL: *https://www.cambridge.org/core/elements/american-affective-polarization-in-comparative-perspective/1E3584B482D51DB25FFFB37A8044F204*

Gift, Karen and Thomas Gift. 2015. "Does Politics Influence Hiring? Evidence from a Randomized Experiment." *Political Behavior* .

Goovaerts, Ine and Sofie Marien. 2020. "Uncivil Communication and Simplistic Argumentation: Decreasing Political Trust, Increasing Persuasive Power?" *Political Communication* 37:768–788.
URL: *https://www.tandfonline.com/doi/abs/10.1080/10584609.2020.1753868*

Gopffarth, Julian. 2018. "The far-right's online party — Chatham House.".
URL: *https://www.chathamhouse.org/publications/twt/far-right-s-online-party*

Granberg, Donald and Edward Brent. 1980. "Perceptions of Issue Positions of Presidential Candidates: Candidates are often perceived by their supporters as holding positions on the issues that are closer to the supporters' views than they really are." *American Scientist* 68:617–625.
URL: *http://www.jstor.org/stable/27850130*

Grand, Peter and Guido Tiemann. 2013. "Projection effects and specification bias in spatial models of European Parliament elections:." *http://dx.doi.org/10.1177/1465116513490238* 14:497–521.
URL: *https://journals.sagepub.com/doi/full/10.1177/1465116513490238*

Green, Donald P, Bradley Palmquist and Eric Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.

Grimmer, Justin. 2012. *Representational Style in Congress: What Legislators Say and Why It Matters.* Cambridge University Press.
URL: *https://books.google.co.uk/books?id=O552AgAAQBAJdq=Grimmer,+J.+2012.+Representatio*

Haime, Agustina and Francisco Cantú. 2022. "Negative Partisanship in Latin America." *Latin American Politics and Society* 64:72–92.
URL: *https://www.cambridge.org/core/journals/latin-american-politics-and-society/article/negative-partisanship-in-latin-america/F8818D32454ADFF1BC85E5BE329FECA7*

Han, Hahrie. 2016. "The Organizational Roots of Political Activism: Field Experiments on Creating a Relational Context." *American Political Science Review* 110(2):296–307.

Hansen, Kasper M. and Karina Kosiara-Pedersen. 2015. "How campaigns polarize the electorate: Political polarization as an effect of the minimal effect theory within a multi-party system." *http://dx.doi.org/10.1177/1354068815593453* 23:181–192.
URL: *https://journals.sagepub.com/doi/full/10.1177/1354068815593453*

Harteveld, Eelco. 2016. "Winning the 'losers' but losing the 'winners'? The electoral consequences of the radical right moving to the economic left." *Electoral Studies* 44:225–234.
URL: *http://ezproxy-prd.bodleian.ox.ac.uk:2054/science/article/pii/S0261379416303390*

Harteveld, Eelco. 2021*a*. "Fragmented foes: Affective polarization in the multiparty context of the Netherlands." *Electoral Studies* 71:102332.

Harteveld, Eelco. 2021*b*. "Ticking all the boxes? A comparative study of social sorting and affective polarization." *Electoral Studies* 72:102337.

Harteveld, Eelco and Elisabeth Ivarsflaten. 2016. "Why Women Avoid the Radical Right: Internalized Norms and Party Reputations." *British Journal of Political Science* pp. 1–16.
URL: *http://www.journals.cambridge.org/abstract_S0007123415000745*

Harteveld, Eelco, Stefan Dahlberg, Andrej Kokkonen and Wouter Van Der Brug. 2019. "Gender Differences in Vote Choice: Social Cues and Social Harmony as Heuristics." *British Journal of Political Science* 49.

Harteveld, Eelco, Stefan Dahlberg, Andrej Kokkonen and Wouter Van Der Brug. 2019*a*. "Gender Differences in Vote Choice: Social Cues and Social Harmony as Heuristics." *British Journal of Political Science* 49(3).

Harteveld, Eelco, Stefan Dahlberg, Andrej Kokkonen and Wouter van der Brug. 2019*b*. "Social Stigma and Support for the Populist Radical Right: An Experimental Study." *Scandinavian Political Studies* 42:296–307.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/1467-9477.12153*

Harteveld, Eelco, Wouter Van Der Brug, Stefan Dahlberg and Andrej Kokkonen. 2015. "The gender gap in populist radical-right voting: examining the demand side in Western and Eastern Europe." *Patterns of Prejudice* 49(1-2):103–134.
URL: *http://www.tandfonline.com/doi/full/10.1080/0031322X.2015.1024399*

Haselmayer, Martin and Marcelo Jenny. 2018. "Friendly fire? Negative campaigning among coalition partners:." *https://doi.org/10.1177/2053168018796911* 5.
URL: *https://journals.sagepub.com/doi/10.1177/2053168018796911*

Hassell, Hans JG and J Quin Monson. 2014. "Campaign targets and messages in direct mail fundraising." *Political Behavior* 36(2):359–376.

Heider, Fritz. 1958. *The psychology of interpersonal relations*. Wiley.

Hernández, Enrique, Eva Anduiza and Guillem Rico. 2021. "Affective polarization and the salience of elections." *Electoral Studies* 69:102203.

Hersch, Eitan. 2020. *Politics is for Power. How to Move Beyond Political Hobbyism, Take Action, and Make Real Change*. New York: Scribner.

Hobolt, Sara B., Thomas J. Leeper and James Tilley. 2020. "Divided by the Vote: Affective Polarization in the Wake of the Brexit Referendum." *British Journal of Political Science* .

Hobolt, Sara B and Toni Rodon. 2020. "Cross-cutting issues and electoral choice. EU issue voting in the aftermath of the Brexit referendum." *Journal of European Public Policy* 27(2):227–245.

Holbrook, Allyson L., Jon A. Krosnick, Penny S. Visser, Wendi L. Gardner and John T. Cacioppo. 2001. "Attitudes toward Presidential Candidates and Political Parties: Initial Optimism, Inertial First Impressions, and a Focus on Flaws." *American Journal of Political Science* 45:930.
URL: *https://www.jstor.org/stable/2669333?origin=crossref*

Holzberg, Billy, Kristina Kolbe and Rafal Zaborowski. 2018. "Figures of Crisis: The Delineation of (Un)Deserving Refugees in the German Media:." *https://doi.org/10.1177/0038038518759460* 52:534–550.
URL: *https://journals.sagepub.com/doi/full/10.1177/0038038518759460?casa_token = FnBvIsX_PyoAAAAA%3AtXCmUC2Ad9lXjSu7adQeNVReGR64MSGb3q3UlQ13NsxsItQ6nzli rgm_uU5M*

Huddy, Leonie and Alexa Bankert. 2017. "Political Partisanship as a Social Identity." *Oxford Research Encyclopedia of Politics* .
URL: *https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-250*

Huddy, Leonie, Alexa Bankert and Caitlin Davies. 2018. "Expressive Versus Instrumental Partisanship in Multiparty European Systems." *Political Psychology* 39:173–199.
URL: *http://doi.wiley.com/10.1111/pops.12482*

Huddy, Leonie, Lilliana Mason and Lene Aaroe. 2015. "Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity." *American Political Science Review* 109(1):1–17.

Hutter, Swen and Hanspeter Kriesi. 2021. "Politicising immigration in times of crisis." *Journal of Ethnic and Migration Studies* .
URL: *https://www.tandfonline.com/doi/abs/10.1080/1369183X.2020.1853902*

Inglehart, Ronald and Hans Dieter Klingemann. 1976. "Party Identification, Ideological Preference and the Left-Right Dimension among Western Mass Publics.".

Ivaldi, Gilles. 2016. "A new course for the french radical right? the front national and 'de-demonisation'.".

Ivarsflaten, Elisabeth. 2007. "What Unites Right-Wing Populists in Western Europe?: Re-Examining Grievance Mobilization Models in Seven Successful Cases." *http://dx.doi.org/10.1177/0010414006294168* 41:3–23.
URL: *https://journals.sagepub.com/doi/10.1177/0010414006294168?icid=int.sj-abstract.similar-articles.1*

Iyengar, Shanto, Gaurav Sood and Yphtach Lelkes. 2012. "Affect, Not Ideology." *Public Opinion Quarterly* 76:405–431.
URL: *https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfs038*

Iyengar, Shanto and Masha Krupenkin. 2018. "The Strengthening of Partisan Affect." *Political Psychology* .

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22:annurev–polisci–051117–073034.
URL: *https://www.annualreviews.org/doi/10.1146/annurev-polisci-051117-073034*

Johns, Rob. 2021. "As You Were: The Scottish Parliament Election of 2021." *The Political Quarterly* 92:493–499.

Kalla, Joshua L. and David E. Broockman. 2022. "Voter Outreach Campaigns Can Reduce Affective Polarization among Implementing Political Activists: Evidence from Inside Three Campaigns." *American Political Science Review* pp. 1–7.
URL: *https://www.cambridge.org/core/journals/american-political-science-review/article/voter-outreach-campaigns-can-reduce-affective-polarization-among-implementing-political-activists-evidence-from-inside-three-campaigns/866E1938791042F642F1A8E472B41EDB*

Khan-Ruf, Safya. 2022. "Le Pen is bad, but many French Muslims like me don't want to vote for Macron either.".
URL: *https://www.theguardian.com/commentisfree/2022/apr/22/macron-or-le-pen-many-french-muslims-may-say-neither*

Killian, Mitchell and Clyde Wilcox. 2021. "Do Abortion Attitudes Lead to Party Switching?" *Political Research Quarterly* 61(4):561–573.

Kingzette, Jon, James N Druckman, Samara Klar, Yanna Krupnikov, Matthew Levendusky and John Barry Ryan. 2021. "How Affective Polarization Undermines Support for Democratic Norms." *Public Opinion Quarterly* 85:663–677.
URL: *https://academic.oup.com/poq/article/85/2/663/6373858*

Kirby, Paul. 2022*a*. "French election: Macron and Le Pen clash in TV presidential debate.".
URL: *https://www.bbc.co.uk/news/world-europe-61166601*

Kirby, Paul. 2022*b*. "French election: Macron and Le Pen trade taunts as campaigning ends.".
URL: *https://www.bbc.co.uk/news/world-europe-61194641*

Klandermans, B, J M Subecedo, M Rodriguez and M de Weerd. 2002. "Identity processes in collective action participation: Farmers' identity and farmers' protest in the Netherlands and Spain." *Political Psychology* 23:235–251.

Klar, Samara. 2014. "Partisanship in a Social Setting." *American Journal of Political Science* 58:687–704.
URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/ajps.12087*

Klar, Samara. 2018. "When Common Identities Decrease Trust: An Experimental Study of Partisan Women." *American Journal of Political Science* 62:610–622.
URL: *http://doi.wiley.com/10.1111/ajps.12366*

Krause, Werner, Denis Cohen and Tarik Abou-Chadi. 2022. "Does accommodation work? Mainstream party strategies and the success of radical right parties." *Political Science Research and Methods* pp. 1–8.
URL: *https://www.cambridge.org/core/journals/political-science-research-and-methods/article/does-accommodation-work-mainstream-party-strategies-and-the-success-of-radical-right-parties/5C3476FCD26B188C7399ADD920D71770*

Kustov, Alexander, Dillon Laaker and Cassidy Reller. 2021. "The stability of immigration attitudes: Evidence and implications." *Journal of Politics* 83:1478–1494.
URL: *https://www.journals.uchicago.edu/doi/abs/10.1086/715061*

Lancaster, Caroline Marie. 2019. "Not So Radical After All: Ideological Diversity Among Radical Right Supporters and Its Implications:." *https://doi.org/10.1177/0032321719870468* 68:600–616.
URL: *https://journals.sagepub.com/eprint/KS339UQ2HFFNQ66KRKNI/full*

Lancaster, Caroline Marie. 2022. "Value Shift: Immigration Attitudes and the Socio-cultural Divide." *British Journal of Political Science* 52:1–20.
  URL: *https://www.cambridge.org/core/journals/british-journal-of-political-science/article/value-shift-immigration-attitudes-and-the-sociocultural-divide/5F893991E485BEB4CF33727F55DDA58B*

Lau, Richard R., David J. Andersen, Tessa M. Ditonto, Mona S. Kleinberg and David P. Redlawsk. 2017. "Effect of Media Environment Diversity and Advertising Tone on Information Search, Selective Exposure, and Affective Polarization." *Political Behavior* 39:231–255.
  URL: *https://link.springer.com/article/10.1007/s11109-016-9354-8*

Lau, Richard R., Lee Sigelman and Ivy Brown Rovner. 2007. "The Effects of Negative Political Campaigns: A Meta-Analytic Reassessment." *The Journal of Politics* 69:1176–1209.
  URL: *https://www.journals.uchicago.edu/doi/10.1111/j.1468-2508.2007.00618.x*

Lavine, Howard G., Christopher D. Johnston and Marco R. Steenbergen. 2012. *The Ambivalent Partisan: How Critical Loyalty Promotes Democracy.* Oxford University Press.

Le Pen, Marine. 2016. "Marine Le Pen : Un référendum pour sortir de la crise migratoire.".
  URL: *https://www.lopinion.fr/politique/marine-le-pen-un-referendum-pour-sortir-de-la-crise-migratoire*

Lelkes, Yphtach. 2021. "What Do We Mean by Negative Partisanship?" *Forum (Germany)* 19.

Lelkes, Yphtach and Sean J. Westwood. 2017. "The Limits of Partisan Prejudice." *The Journal of Politics* 79:485–501.
  URL: *https://www.journals.uchicago.edu/doi/10.1086/688223*

Lenz, Gabriel S. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science* 53:821–837.
  URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-5907.2009.00403.x https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5907.2009.00403.x https://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2009.00403.x*

Lenz, Gabriel S. 2012. *Follow the leader? : how voters respond to politicians' policies and performance.*

Levendusky, Matthew S. 2018. "Americans, not partisans: Can priming American national identity reduce affective polarization?" *Journal of Politics* 80:59–70.
URL: *https://www.journals.uchicago.edu/doi/full/10.1086/693987*

MacDonald, Stuart Elaine, George Rabinowitz and Ola Listhaug. 2007. "Simulating Models of Issue Voting." *Political Analysis* 15:406–427.
URL: *https://www.cambridge.org/core/journals/political-analysis/article/simulating-models-of-issue-voting/73AA6CB322CE4730530F727406394FE9*

MacKuen, Michael B., Robert S. Erikson and James A. Stimson. 1989. "Macropartisanship." *American Political Science Review* 83.

Markus, Gregory B. and Philip E. Converse. 1979. "A Dynamic Simultaneous Equation Model of Electoral Choice." *American Political Science Review* 73:1055–1070.
URL: *https://www.cambridge.org/core/journals/american-political-science-review/article/abs/dynamic-simultaneous-equation-model-of-electoral-choice/5682182CA9C8AFBA27A416C2B3F533DB*

Martherus, James L., Andres G. Martinez, Paul K. Piff and Alexander G. Theodoridis. 2021. "Party Animals? Extreme Partisan Polarization and Dehumanization." *Political Behavior* 43.

Mason, Liliana and Nathan Kalmoe. 2022. *Radical American Partisanship Mapping Violent Hostility, Its Causes, and the Consequences for Democracy.* Chicago Studies in American Politics.

Mason, Lilliana. 2018. *Uncivil agreement : how politics became our identity.* University of Chicago Press.

Mayer, Nonna. 2013. "From jean-marie to marine le pen: Electoral change on the far right." *Parliamentary Affairs* 66(1):160–178.

Mayer, Nonna. 2015. "The closing of the radical right gender gap in France?" *French Politics* 13(4):391–414.
URL: *http://link.springer.com/10.1057/fp.2015.18*

Mayer, Sabrina Jasmin. 2017. "How negative partisanship affects voting behavior in Europe: Evidence from an analysis of 17 European multi-party systems with proportional voting." *Research Politics* 4:205316801668663.
URL: *http://journals.sagepub.com/doi/10.1177/2053168016686636*

McCoy, Jennifer, Tahmina Rahman and Murat Somer. 2018. "Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities:." *https://doi.org/10.1177/0002764218759576* 62:16–42.
URL: *https://journals.sagepub.com/doi/10.1177/0002764218759576*

McDonnell, Duncan and Annika Werner. 2018. "Respectable radicals: why some radical right parties in the European Parliament forsake policy congruence." *Journal of European Public Policy* 25:747–763.
URL: *https://www.tandfonline.com/doi/abs/10.1080/13501763.2017.1298659*

McGarty, Craig, Ana Maria Bliuc, Emma F Thomas and Renata Bongiorno. 2009. "Collective Action as the Material Expression of Opinion-Based Group Membership." *Journal of Social Issues* 65(4):839–857.

McGregor, R. Michael, Nicholas J. Caruana and Laura B. Stephenson. 2015. "Negative Partisanship in a Multi-party System: The Case of Canada." *Journal of Elections, Public Opinion and Parties* .

Medeiros, Mike and Alain Noël. 2014. "The Forgotten Side of Partisanship: Negative Party Identification in Four Anglo-American Democracies." *Comparative Political Studies* .

Meléndez, Carlos and Cristóbal Rovira Kaltwasser. 2017. "Political identities: The Missing Link in the Study of Populism." *Party Politics* .

Meléndez, Carlos and Cristóbal Rovira Kaltwasser. 2021. "Negative partisanship towards the populist radical right and democratic resilience in Western Europe." *https://doi.org/10.1080/13510347.2021.1883002* 28:949–969.
URL: *https://www.tandfonline.com/doi/abs/10.1080/13510347.2021.1883002*

Mendes, Mariana S. and James Dennison. 2020. "Explaining the emergence of the radical right in Spain and Portugal: salience, stigma and supply." *https://doi.org/10.1080/01402382.2020.1777504* 44:752–775.
URL: *https://www.tandfonline.com/doi/abs/10.1080/01402382.2020.1777504*

Merolla, Jennifer, Karthick Ramakrishnan and Chris Haynes. 2013. ""Illegal," "Undocumented," or "Unauthorized": Equivalency Frames, Issue Frames, and Public Opinion on Immigration." *Perspectives on Politics* 11:789–807.
URL: *https://www.cambridge.org/core/journals/perspectives-on-politics/article/illegal-undocumented-or-unauthorized-equivalency-frames-issue-frames-and-public-opinion-on-immigration/2DCBA83F6C25DB76C07EED45022E81AB*

Merrill, Samuel, Bernard Grofman and James Adams. 2001. "Assimilation and contrast effects in voter projections of party locations: Evidence from Norway, France, and the USA." *European Journal of Political Research 2001 40:2* 40:199–223.
URL: *https://link.springer.com/article/10.1023/A:1012975221087*

Miller, Joanne M. and Jon A. Krosnick. 2004. "Threat as a motivator of political activism: A field experiment." *Political Psychology* 25:507–523.

Mitchell, James and Ailsa Henderson. 2020. "Tribes and Turbulence: The 2019 UK General Election in Scotland." *Parliamentary Affairs* 73:142–156.

Motta, Matt, Timothy Callaghan, Steven Sylvester and Kristin Lunz-Trujillo. 2021. "Identifying the prevalence, correlates, and policy consequences of anti-vaccine social identity." *Politics, Groups and Identities* Online First.

Mudde, Cas. 2007. *Populist radical right parties in Europe.*

Mudde, Cas. 2019. *The Far Right Today.* Polity.

Mudde, Cas and Cristóbal Rovira Kaltwasser. 2015. "Vox populi or vox masculini? Populism and gender in Northern Europe and South America." *Patterns of Prejudice* 49(1-2):16–36.
URL: *http://www.tandfonline.com/doi/full/10.1080/0031322X.2015.1014197*

Mullinix, Kevin J, Thomas J Leeper, James N Druckman, Jeremy Freese, Adam Berinsky, Emily Cochran Bech, Scott Clifford, Adrienne Hosek, Cindy Kam, Lasse Laustsen, Diana Mutz, Helene Helboe Pedersen, Richard Shafranek, Flori So, Rune Slothuus, Rune Stubager and Magdalena Wojcieszak. 2015. "The Generalizability of Survey Experiments*." *Journal of Experimental Political Science* 2:109–138.
URL: *https://www.cambridge.org/core/journals/journal-of-experimental-political-science/article/generalizability-of-survey-experiments/72D4E3DB90569AD7F2D469E9DF3A94CB*

Mutz, Diana C. 2015. *In-Your-Face Politics.* Princeton University Press.

Mutz, Diana C. and Byron Reeves. 2005. "The New Videomalaise: Effects of Televised Incivility on Political Trust." *American Political Science Review* 99:1–15.
URL: *https://www.cambridge.org/core/journals/american-political-science-review/article/new-videomalaise-effects-of-televised-incivility-on-political-trust/093762E57EF0CFA2E4A0328572DE0009*

Nai, Alessandro. 2020. "Going Negative, Worldwide: Towards a General Understanding of Determinants and Targets of Negative Campaigning." *Government and Opposition* 55.

Nelson, Thomas E., Rosalee A. Clawson and Zoe M. Oxley. 1997. "Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance." *American Political Science Review* 91.

Neundorf, Anja, Daniel Stegmueller and Thomas J. Scotto. 2011. "The individual-level dynamics of bounded partisanship." *Public Opinion Quarterly* .

Niven, David. 2006. "A Field Experiment on the Effects of Negative Campaign Mail on Voter Turnout in a Municipal Election." *Political Research Quarterly* 59:203–210.
  URL: *http://journals.sagepub.com/doi/10.1177/106591290605900203*

Norocel, Ov Cristian. 2016. "Populist radical right protectors of the folkhem: Welfare chauvinism in Sweden." *Critical Social Policy* 36(3):371–390.
  URL: *http://journals.sagepub.com/doi/10.1177/0261018315621991*

Orhan, Yunus. 2022. "The Relationship between Affective Polarization and Democratic Backsliding: Comparative Evidence.".
  URL: *https://osf.io/ym2xs/*

Oshri, Odelia, Liran Harsgor, Reut Itzkovitch-Malka and Or Tuttnauer. 2022. "Risk Aversion and the Gender Gap in the Vote for Populist Radical Right Parties." *American Journal of Political Science* .
  URL: *https://onlinelibrary.wiley.com/doi/full/10.1111/ajps.12696*
  https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12696
  https://onlinelibrary.wiley.com/doi/10.1111/ajps.12696

Ostrom, E. 2000. "Collective action and the evolution of social norms." *The Journal of Economic Perspectives* 14(3):137–158.
  URL: *http://www.jstor.org/stable/2646923%5Cnpapers3://publication/doi/10.2307/2646923*

Perez-Truglia, Ricardo and Guillermo Cruces. 2017. "Partisan interactions: Evidence from a field experiment in the united states." *Journal of Political Economy* 125(4):1208–1243.

Pew. 2019. "Partisan Antipathy: More Intense, More Personal — Pew Research Center.".
  URL: *https://www.pewresearch.org/politics/2019/10/10/partisan-antipathy-more-intense-more-personal/*

Phillips, Trisha. 2021. "Ethics of Field Experiments." *https://doi.org/10.1146/annurev-polisci-041719-101956* 24:277–300.
  URL: *https://www.annualreviews.org/doi/abs/10.1146/annurev-polisci-041719-101956*

Popp, Bastian, Claas Christian Germelmann and Benjamin Jung. 2016. "We love to hate them! Social media-based anti-brand communities in professional football." *International Journal of Sports Marketing and Sponsorship* 17:349–367.

Prosser, Christopher. 2021. "The end of the EU affair: the UK general election of 2019." *West European Politics* 44(2):450–461.

Reiljan, Andres. 2020. "'Fear and loathing across party lines' (also) in Europe: Affective polarisation in European party systems." *European Journal of Political Research* 59:376–396.

Reiter, Franz and Jörg Matthes. 2021. ""The Good, the Bad, and the Ugly": A Panel Study on the Reciprocal Effects of Negative, Dirty, and Positive Campaigning on Political Distrust." *Mass Communication and Society* .
URL: *https://www.tandfonline.com/doi/abs/10.1080/15205436.2021.1934702*

Ridge, Hannah M. 2020. "Enemy Mine: Negative Partisanship and Satisfaction with Democracy." *Political Behavior* .

Rooduijn, M., S. Van Kessel, C. Froio, A. Pirro, P. De Lange, S., Halikiopoulou, D., Lewis, C. Mudde and P. Taggart. 2019. "The PopuList: An Overview of Populist, Far Right, Far Left and Eurosceptic Parties in Europe.".
URL: *http://www.popu-list.org*

Rose, Richard and William Mishler. 1998. "Negative and positive party identification in post-Communist countries." *Electoral Studies* 17:217–234.

Rydgren, Jens. 2005. "Is extreme right-wing populism contagious? Explaining the emergence of a new party family." *European Journal of Political Research* 44:413–437.

Samuels, David J. and Cesar Zucco. 2018. *Partisans, Antipartisans, and Nonpartisans.* Cambridge University Press.

Schonfeld, Bryan and Sam Winter-Levy. 2021. "Policy or partisanship in the united kingdom? Quasi-experimental evidence from brexit." *Journal of Politics* 83:1450–1461.
URL: *https://doi.org/10.1086/712143*

Scrinzi, Francesca. 2017*a*. A 'New' National Front? Gender, Religion, Secularism and the French Populist Radical Right. In *Gender and Far Right Politics in Europe.*

Scrinzi, Francesca. 2017*b*. "Gender and women in the Front National discourse and policy: from 'mothers of the nation' to 'working mothers'?" *New Formations* .

Sherif, Muzafer and Carl I Hovland. 1961. *Social judgment: Assimilation and contrast effects in communication and attitude change.* Yale Univer. Press.

Simonsen, Kristina B. and Bart Bonikowski. 2022. "Moralizing Immigration: Political Framing, Moral Conviction, and Polarization in the United States and Denmark:." *https://doi.org/10.1177/00104140211060284* 2022:1–34.
URL: *https://journals.sagepub.com/doi/full/10.1177/00104140211060284?casa$_t$oken =*
$OIfFhwYejIsAAAAA\%3A3cJlrl5rD1jThoU1pkuTms4QnPevFbLJ77AGbkpMEOtANFlmkm8$
$0WBt - e_O$

Sobolewska, Maria and Robert Ford. 2020. *Brexitland.* Cambridge University Press.

Sorace, Miriam and Sara B. Hobolt. 2020. "A tale of two peoples: Motivated reasoning in the aftermath of the Brexit Vote." *Political Science Research and Method* .

Soroka, Stuart, Patrick Fournier and Lilach Nir. 2019. "Cross-national evidence of a negativity bias in psychophysiological reactions to news." *Proceedings of the National Academy of Sciences of the United States of America* 116:18888–18892.
URL: *www.pnas.org/cgi/doi/10.1073/pnas.1908369116*

Spanje, Joost Van and Rachid Azrout. 2019. "Tainted Love: How Stigmatization of a Political Party in News Media Reduces Its Electoral Support." *International Journal of Public Opinion Research* 31:283–308.
URL: *https://academic.oup.com/ijpor/article/31/2/283/4953818*

Spierings, Niels. 2021. "Homonationalism and Voting for the Populist Radical Right: Addressing Unanswered Questions by Zooming in on the Dutch Case." *International Journal of Public Opinion Research* 33:171–182.
URL: *https://academic.oup.com/ijpor/article/33/1/171/5811110*

Spierings, Niels, Marcel Lubbers and Andrej Zaslove. 2017. "'Sexually modern nativist voters': do they exist and do they vote for the populist radical right?" *https://doi.org/10.1080/09540253.2016.1274383* 29:216–237.
URL: *https://www.tandfonline.com/doi/abs/10.1080/09540253.2016.1274383*

Tajfel, Henri. 1974. "Social identity and intergroup behaviour." *Social Science Information* 13:65–93.

Tajfel, Henri and John C Turner. 1979. "An Integrative Theory of Intergroup Conflict.".

Torcal, Mariano and Emily Carty. 2022. "Partisan Sentiments and Political Trust: A Longitudinal Study of Spain." *https://doi.org/10.1080/13608746.2022.2047555* pp. 1–26.
URL: *https://www.tandfonline.com/doi/abs/10.1080/13608746.2022.2047555*

Trumm, Siim, Caitlin Milazzo and Joshua Townsley. 2020. "The 2016 EU Referendum: Explaining Support for Brexit Among Would-Be British MPs." *Political Studies* 68(4):819–836.

Turnbull-Dugarte, Stuart J, Joshua Townsley, Florian Foos and Denise Baron. 2022. "Mobilising supporters when the (Brexit) stakes are high: A large-scale experiment on mass email lobbying." *European Journal of Political Research* 61(2):601–619.

Tversky, Amos and Daniel Kahneman. 1992. "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and Uncertainty* .

Valentim, Vicente. 2021. "Parliamentary Representation and the Normalization of Radical Right Support." *Comparative Political Studies* 54.

Valentino, Nicholas A., Stuart N. Soroka, Shanto Iyengar, Toril Aalberg, Raymond Duch, Marta Fraile, Kyu S. Hahn, Kasper M. Hansen, Allison Harell, Marc Helbling, Simon D. Jackman and Tetsuro Kobayashi. 2019. "Economic and Cultural Drivers of Immigrant Support Worldwide." *British Journal of Political Science* 49:1201–1226.
URL: *https://www.cambridge.org/core/journals/british-journal-of-political-science/article/economic-and-cultural-drivers-of-immigrant-support-worldwide/02BBCF09B063FCD0C252B6D78E748DE8*

Valenzuela, Ali A. and Melissa R. Michelson. 2016. "Turnout, status, and identity: Mobilizing Latinos to vote with group appeals." *American Political Science Review* .

van Heerden, Sjoerdje Charlotte and Wouter van der Brug. 2017. "Demonisation and electoral support for populist radical right parties: A temporary effect." *Electoral Studies* 47:36–45.

van Stekelenburg, Jacquelien and Bert Klandermans. 2010. "Individuals in Movements: A Social Psychology of Contention.".

Vesely, Stepan, Torsten Masson, Parissa Chokrai, Anna M. Becker, Immo Fritsche, Christian A. Klöckner, Lorenza Tiberio, Giuseppe Carrus and Angelo Panno. 2021. "Climate change action as a project of identity: Eight meta-analyses." *Global Environmental Change* 70:102322.

Vries, Catherine E. De and Sara B. Hobolt. 2012. "When dimensions collide: The electoral success of issue entrepreneurs." *European Union Politics* 13:246–268.
URL: *http://journals.sagepub.com/doi/10.1177/1465116511434788*

Wagner, Markus. 2012. "When do parties emphasise extreme positions? How strategic incentives for policy differentiation influence issue importance." *European Journal of*

Political Research 51(1):64–88.
URL: *http://doi.wiley.com/10.1111/j.1475-6765.2011.01989.x*

Wagner, Markus. 2021. "Affective polarization in multiparty systems." *Electoral Studies* 69:102199.

Wagner, Markus and Jakob-Moritz Eberl. 2022. "Divided by the Jab: On the Nature, Origins, and Consequences of COVID-19 Vaccination Identities." *OSF Pre-Print* .

Walter, Annemarie S. and Cees van der Eijk. 2019. "Unintended consequences of negative campaigning: Backlash and second-preference boost effects in a multi-party context." *British Journal of Politics and International Relations* 21.

Wattenberg, Martin P. and Craig Leonard Brians. 1999. "Negative Campaign Advertising: Demobilizer or Mobilizer?" *American Political Science Review* 93:891–899.
URL: *https://www.cambridge.org/core/journals/american-political-science-review/article/abs/negative-campaign-advertising-demobilizer-or-mobilizer/8E7DECAAF139E01E829B0F4FFA564071*

Weeks, Ana, Hilde Coffe, Bonnie Meguid and Miki Kittilson. 2022. "When do Männerparteien Elect Women? Radical Right Populist Parties and Strategic Descriptive Representation." *American Political Science Review* .

Wigger, Iris, Alexander Yendell and David Herbert. 2021. "The end of 'Welcome Culture'? How the Cologne assaults reframed Germany's immigration discourse:." *https://doi.org/10.1177/02673231211012173* 37:21–47.
URL: *https://journals.sagepub.com/doi/full/10.1177/02673231211012173*

Wildman, Sarah. 2017. "Almost everyone in French politics is working to stop Le Pen.".
URL: *https://www.vox.com/world/2017/4/24/15408742/marine-le-pen-firewall-politicians-republican-front-stop-far-right*

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zhong, Chen Bo, Katherine W. Phillips, Geoffrey J. Leonardelli and Adam D. Galinsky. 2008. "Negational categorization and intergroup behavior." *Personality and Social Psychology Bulletin* .

Zingher, Joshua N. and Michael E. Flynn. 2018. "From on High: The Effect of Elite Polarization on Mass Attitudes and Behaviors, 1972–2012." *British Journal of Political Science* 48:23–45.
URL: *https://www.cambridge.org/core/journals/british-journal-of-political-science/article/from-on-high-the-effect-of-elite-polarization-on-mass-attitudes-and-behaviors-19722012/B2F5A39BB79D7A26848BEA4AD73A42A5*