They Just Don't Look Like Me:
How Social Sorting Increases Affective Polarization

By

DAVID BRACKEN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

POLITICAL SCIENCE

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Bradford Jones, Chair

James Adams

Cheryl Boudreau

Amber Boydstun

Committee in Charge

2023

Abstract

Affective polarization - the growing dislike and distrust between supporters of opposing political parties - has become a growing concern in the United States and other Western democracies. While a number of potential explanations for this phenomena have been offered, such as elite ideological polarization or income inequality, in this dissertation I focus on the role of social identity and its relationship to polarization. I explore whether the growing inter-party hatred we witness is due less to what our opponents think, and more about *who* they represent. I build on social sorting theory to argue that individuals who perceive opposing parties as looking *less like them* will form more polarized views of the parties. In this dissertation, I offer three tests of this theory, and find several instances in which the alignment between an individual's social identities — whether that be racial, gender, religious, age or other core identities — and those they attach to a political party, influences their feelings towards those parties.

The first paper develops a theory of identity alignment, explaining how and why we might expect individuals to form views of the parties that are based on a comparison between an individual's view of themselves and their views of those who support each party. I then provide three experimental tests of this theory, two studies using undergraduate subjects from a large public university, and a third replication study using a national sample, and find tentative evidence that supports the idea that the groups we associate with a political party influence our affective evaluations of the party, perceived ideological placement of the party, and responsiveness to party cues.

The second paper studies the rise of social sorting in the United States and its relationship with affective polarization at the aggregate and individual level. I provide a new way to measure social sorting using the tools of machine learning, and validate this method by also applying it to the more well established case of ideological sorting. I then demonstrate that social sorting has risen at roughly the same rate as ideological sorting within the US, and that both phenomena are independently associated with increasing levels of affective polarization. I show that individuals do indeed dislike their out-party because of what they believe, but they *also* form increasingly negative views because the two parties now represent wholly different social groups, comprised of few overlapping identities. This chapter provides strong evidence that the US parties

now represent two groups who find it increasingly difficult to see themselves in the other side.

The final chapter provides a comparative perspective by studying the relationship between social sorting and one core social identity — gender. I explore whether the gender gap between two parties — the difference in the proportion of female supporters within each party — impacts the affective evaluations that partisans make of out-parties. I argue that the gender composition of one's own party represents an important identity marker, an indication of the identity composition of the party, and a yard stick by which individuals judge other parties. I argue that the more a party differs in its gender composition from an individual's own party, the less warmly the individual will feel towards the party. I test this, along with competing theories regarding the gender composition of party supporters, on a dataset containing evaluations of 159 parties across 25 countries between 1996-2020, and find that an increased gender gap between parties is indeed associated with increasingly polarized evaluations of the out-party. I also find tentative evidence that individuals, particularly women, prefer parties with more women supporters, but that this relationship is attenuated by the effects of the gender gap.

Taken together, these three papers present evidence that the increased partisan animosity that we see in many contemporary polities is driven, at least in part, by the social divisions between parties. As parties (particularly those in the United States) come to represent wholly different social groups who no longer share large numbers of cross-cutting identities, individuals find it increasingly difficult to find those among their political opponents who look like them, and so become increasingly negative, distrustful, and unresponsive to out-parties. Unlike other potential explanations for affective polarization such as elite ideological polarization, income inequality, or majoritarian electoral systems, this polarization driven by social sorting may be a much more difficult problem to solve. Without incentives for political parties to court supporters of social groups that are currently associated with their political opponents, it is hard to see how and why parties would take the necessary steps to decrease social-based polarization. Indeed, it is more likely that the results here are part of a vicious cycle — as groups become sorted with one party or another, this feeds into the stereotypes people have about the parties, forcing them to update their own identification and retrench into the party associated with their group, thus increasing levels of social sorting, and driving further polarization.

Acknowledgements

A dissertation is in many ways a lonely journey, but were it not for the unending support of my family, friends, committee, and others, this work would not exist.

To my incredible wife Charlotte, without your love, support, enthusiasm, and encouragement, this dissertation would never have happened, or I would have gone insane writing it. I came to Davis with a potential blind date lined up, and leave with a family. Wherever this journey leads us, I will always be boundlessly happy that you're the one taking it with me.

To my tiny co-author, I apologise in advance that your dad will now introduce himself as Doctor wherever he goes, and for the endless Doctor Who jokes you'll have to put up with when you're older. But I am glad you were always there to cheer me up when writing was not going well.

To my parents, and particularly my mum, the greatest proof reader in history, words will never be enough to thank you for all the opportunities you have given me. I promise I'll never ask you to read any of these chapters again. And to my brother, whose nocturnal schedule meant he was often the only Brit available in some of my most difficult moments, thank you for always being there with something to lift my spirits or distract my brain.

To my committee, without your help and guidance, this dissertation would not have ever come together, I would have undoubtedly folded without the continuous support along the way. Amber, Brad, Cheryl, and Jim, I count myself lucky to have had such fantastically distinguished advisors who also happened to be phenomenally nice people. Thank you for everything.

To the UC Davis Political Science Department, when I arrived I was genuinely conflicted as to whether I'd made the right choice. As I leave, I now mourn leaving somewhere I truly consider to be home. I'll miss every single individual here, it is rare to get to work and spend time in a place full of intelligent, kind, curious people, trying to ask the right questions about important things. And I'll miss the bake off too.

Finally, to my friends. I count myself incredibly lucky to have friends who have supported me on every step of this journey, from near and far, through all the highs and lows. I cannot possibly name every person who accompanied me on this journey, but I also cannot leave without thanking a few individually. To the all

the players of the Greatest Game of Dungeons and Dragon Ever Played, I thank you for the weekly breaks from the madness of graduate school to indulge in the insanity of D&D. To Brad, Craig, and especially Lou, thank you for cheering me on from across the pond. To Mary, knowing I had a friend waiting when I arrived made coming here easier, knowing you were always a text away made the whole journey better, and I don't think I'll I ever be able to properly thank a friend who not only has been supporting me for over a decade, but even found me a wife. And to Hannah, Davis would not have been Davis without you, from game theory to The Bachelor, my experience of graduate school was constantly improved by your help, your wit, and your company.

To anyone and everyone I met along the way, I thank you. I wouldn't be here without you, and there's nowhere I'd rather be.

Contents

Chapter 1

The Effect of Social Sorting on Affect, Identity, Proximity and Cues

Abstract

To what extent does the increase in social sorting explain changes in party ID, affective polarization, ideological perceptions, and responsiveness to party cues? I explore this question by experimentally manipulating the groups associated with political parties and organizations, to induce a change in how well a respondent is socially sorted with their in or out-group. Among undergraduates, I find that increased social sorting has modest effects, but decreased social sorting significantly reduces identity strength, affect, ideological proximity, and responsiveness to party cues. However, these effects do not appear to hold among the wider US population, posing questions about the generalizability of the group theory of party identification and social sorting.

Introduction

There appears general agreement among scholars that Americans have become increasingly polarized in their affective evaluations of the two major political parties over recent decades (Hetherington et al. 2016, Iyengar et al. 2019, Kalmoe and Mason 2019, Matherus et al. 2021). Affective polarization, generally defined as increasing antipathy towards an out-party and possibly, though not necessarily, accompanied by increasingly warm feelings towards an in-party, has been increasing in the USA at an alarming rate (Iyengar et al. 2019, Adams et al 2020). Given evidence that affective polarization is tied to preferential treatment of co-partisans

(Lelkes and Westwood 2017), discrimination against out-partisans in economic transactions (McConnell et al. 2018; Carlin and Love 2018), reductions in trust in government (Hetherington and Rudolph 2015), and even increasing dehumanization (Matherus et al. 2021) and political violence (Kalmoe and Mason 2019), scholars have turned their focus to the potential causes of this phenomenon. Elite polarization (Abramwotiz and Webster 2016, Rogowski and Sutherland 2015, Adams, Gidron, & Horne 2020, Reiljan 2019), the electoral system (Adams, Gidron, & Horne 2018), internet usage (Lelkes, Sood, & Iyengar 2017), and income inequality (Gunderson 2022) among other subjects, have all been linked to increasing levels of partisan animosity.

One potential explanation has been the increased levels of social sorting occurring within the electorate - as partisan identities have shifted along socio-demographic lines, the parties have become more internally homogeneous, and more distinct from one another in terms of the social groups they represent (Mason 2016, 2018). Social sorting theory posits that this decrease in the number of shared identities between the parties, and therefore, between partisans, and the increasingly divergent groups associated with each party, leads individuals to form more polarized views of the parties (Mason 2016, Mason & Wronski 2018).

In this paper, I test the theory that social sorting may contribute to mass polarization, by exploring the consequences of social sorting in four key areas. I test whether the alignment between one's social and partisan identities affects strength of party identification, responsiveness to party cues, ideological perceptions, and affective evaluations. Among undergraduate subjects, I find tentative evidence that increased levels of social-partisan identity alignment result in increased identification with a party and warmer feelings towards them, with stronger evidence of an effect in decreasing identity alignment. I find that as individuals view a given party or group as "looking less like them" in terms of their demographic identities, they become less responsive to cues from the group, perceive the group as less ideologically aligned with their own beliefs, and feel less warmly towards them. Taken together, my findings indicate that social sorting at the individual level may be associated with increases in negative partisanship - as individuals perceive their out-party as being "less like them", they move away from that party without necessarily forming stronger attachments towards their in-party. However, these findings do not appear to replicate beyond college students, posing questions as to whether these effects are felt more strongly among younger voters, or whether these effects exist at all. If these findings can be validated, the relationship between social sorting and mass polarization in the United States would appear to be one driven by the increasingly homogeneous nature of the two parties resulting in two groups of partisans who find themselves "out-sorted" from their out-party, and reacting by forming more negative views of the party, its policy positions, and its ideological outlook.

Social Sorting and its Consequences

Work on social sorting has grown significantly in the last decade within political science. This work builds on traditional social identity theory [SIT] (Tajfel & Turner 1979, Huddy 2001, Kalin & Sambanis 2018) to develop a theory that explicitly links social identities to a core foundational object within American politics: party identity. Even before the seminal The American Voter (Campbell, Converse, Miller & Stokes 1960), scholars within American politics understood that party identity is a central object which shapes other political values, judgments, and identities, something further supported by recent scholarship (Iyengar and Westwood 2015, Achen & Bartels 2016). More recent theorizing regarding party identification has focused on understanding party identity as a social identity like many others. Green, Palmquist, and Schickler (2002) compared party identity to a religious or ethnic identity, and argue that it should be understood as an identity formed in response to the groups which an individual associates with the party. They state that a party "exists as a [stable] stereotype in the minds of voters, who in turn harbor a sense of attachment toward this group image" (p. 26). They argue that, while changes in party identification can be caused by changes in individual circumstances and national events, these changes often revert back to the mean over time. Instead, major changes in party identification are due to changes in this 'stereotype' associated with the parties in the minds of voters - when the answer to "What is a Democrat?" changes, individuals change their party identification to accord with the new definition of "a Democrat". Ahler (2018) describes this as the "group theory" of party identification, namely individuals form their party identifications by evaluating the groups associated with a particular party, comparing this "group image" of the party to their own group membership, and choosing to identify with a party (or strengthen their identification with that party) based on the similarity or difference in this comparison. Ahler documents a range of evidence in favor of this social group-based notion of party identification, arguing that "citizens see parties as collectives of more fundamental groups in society and evaluate them according to how well the parties reflect their own identities" (p. 6). This evaluation is based on the psychological concept of "comparative fit", (Turner 1999, Hogg et al. 2004) - the idea that individuals compare their identity and characteristics with those of some prototype or group image, and identify with the image which is most similar to their own.

Importantly, a growing body of scholarship in American politics has demonstrated the increasing linkages between demographic identities, such as race (Mason 2015; Jackson 2011; Juddy & Carey 2009; Huddy et al. 2016; Bowler & Segura 2011) age (Shively 1979; Franklin & Jackson 1983), gender (Box-Steffensmeier, De Boef, & Lin 2004; Kittilson 2016), and religiosity (Conover 1988; Patrikios 2008) to party identification. Perhaps the clearest example of the importance of demographics comes from Rothschild et al. (2019) who use open-ended survey questions to explore what kinds of characteristics are most associated with the group

images of the Democratic and Republican parties. They find that trait-based characteristics are among the most common stereotypes of the parties, with stereotypes related to race, age, religiosity, and gender forming key components of the group images of both parties - indeed, words related to race, age, and religiosity were among the ten most common terms to describe both parties.

Social sorting is the process through which individuals' identities shift over time in order to form a better "alignment" of their social identities with their partisan identity. More specifically, it involves the alignment of an individual's social identities with the identities that they attach to their partisan in-group. It is the process through which individuals shape their partisan identities to conform to the group theory of party identification as described by Green Palmquist, & Schickler (2002) and Ahler (2018). For example, an individual may slowly change their partisan identity in order to form a better match between their own racial identity and the racial identity they attach to the political parties. At the aggregate level, this results in particular groups gradually showing greater levels of support for one particular party, as that group's identity comes into alignment with the party identity (Mason 2016, Mason and Wronski 2018). While the theory does not necessarily dictate which identities should be expected to change in order to come into alignment, many of the identities generally studied by social sorting theorists are "immutable" identities such as race or gender - those same demographic characteristics previously mentioned - from which it is very challenging for individuals to divest themselves. Therefore, the most common direction of the theory that is studied explores whether individuals bring their partisan identity into line with their more immutable or deep seated social and demographic identities, rather than changing their other social identities to suit their partisan identity (although see Campbell et al. 2018, Egan 2019). For the purposes of this paper, I focus on this concept of a latent measure of "identity alignment" - specifically the unconscious measure of alignment between an individual's social identities and their partisan identity - and the consequences of changes in this alignment.

Strength of Partisan Attachments

Much of the research regarding the consequences of social sorting looks specifically at identity strength, and whether those whose identities are most well sorted along some specified social fault line (i.e. those with the highest degree of alignment between social and partisan identities), also tend to be the strongest partisans. Theoretically, social sorting theory predicts that, as the alignment between an individual's social or demographic identities and their party identity increases, so too should the strength of their attachment to their party. This has been a consistent focus of scholars from Green, Palmquist, & Schickler (2002) onwards. More recently, Mason and Wronski (2018) demonstrate that those whose social identities most closely align with

their partisan identity (in their view, White, Christians, Conservatives and the Republican party, and Black, Secular, Liberals and the Democratic party) also show the strongest identification with their party. In this paper, I experimentally test this theory of increasing alignment between social identity and partisan identity resulting in increased strength of partisan attachments. Given the extent to which partisan attachment shape other political views and opinions, evidence that social sorting affects party identification would have significant ramifications for our understanding of the causes of affective polarization. This provides my first hypothesis:

> ***H1: Party Identification Hypothesis***: Greater (weaker) identity alignment between individual and party results in stronger (weaker) identification with the party

Party Cues

A second potential consequence of changes in levels of identity alignment is in the way individuals respond to cues from political parties. Much research has demonstrated that party labels provide a valuable heuristic as to the extent to which a particular policy is in the interests of a given individual (Lupia 1994, Cohen 2003), or in the case of out-party cues, the extent to which a policy goes against their interests (Nicholson 2011). This is because cues activate "group motivational processes" which encourage the individual to view the in-party as a representative of the individual's own group interests, and thus to support the party's position (Petersen et al. 2012).

However, the efficacy of this party cue may rest on the extent to which the party itself is representative of the individual's group interests. If it is the stimulation of "group motivational processes" that drive the effects of the cue, then any deviation that may affect the activation of this process should also change the effect of the cue. That is, if the party is perceived as being a good representative of the individual and their social groups (i.e. identity alignment is high), this group motivational process is more likely to be activated, leading the individual to respond to the party cue by becoming more supportive of their party. But if the party is perceived as either not representing their group well, or of representing the individual's out-group (i.e. identity alignment is low), the cue may either not activate these group motivational processes, thus the cue will have no effect, or the cue may activate this process but against the party providing the cue, a boomerang' effect, in which support for the party's position may decrease (as evidenced by work from Nicholson 2012). If proven, this theory provides valuable context for when party cues may be effective, or when individuals will instead rely on other information, such as policy information (Boudreau and MacKenzie

2014).

Mason and Wronski (2018) provide some evidence that social sorting (through identity alignment) may affect the efficacy of party cues. They argue that individuals for whom social and partisan identities align can use the trends in aggregate partisan sorting within the electorate (along racial, religious, and ideological lines) as "cues" which enable them "to perceive their party identity as a singular social identity" (p. 261), thus enabling them to simplify the cognitively strenuous tasks of forming political opinions, by relying on the single identity for cues and information. If true, we should expect those who perceive themselves as being well sorted to respond more strongly to party cues, and those who view themselves as less similar to the cue-giving party to respond either weakly, not at all, or possibly even negatively. Similarly, Mason (2016) demonstrates that those with larger numbers of cross-cutting identities (i.e. less well sorted partisans) respond less negatively to threats towards their parties, and argues that, as social sorting has increased, resulting in partisans with fewer cross-cutting identities, this dampening effect has decreased, resulting in greater levels of hostility in response to the political environment. It is in this reduction of the potential for cross-cutting identities to moderate party cues, that enables party cues to increase mass political polarization. That is, if party cues are only effective sometimes, for some of the population, we would expect a less polarized public. However, if the forces which reduce the effects of cues are eliminated, we would expect to see more polarized views among the public. In this paper, I directly test whether changes in identity alignment alter the efficacy of party cues.

> ***H2 Cue hypothesis:*** Greater (weaker) identity alignment between individual and party results in stronger (weaker) responses to party cues.

Affective Evaluations:

There is now substantial evidence that affective polarization has markedly increased within the US (Iyengar, Sood, & Lelkes 2012, Iyengar et al. 2019), and some evidence that this increase is not unique to the US but reflects a wider trend (Reiljan 2019, Adams et al. 2020). These authors have demonstrated a growing number of influences on affective polarization, such as income inequality (Gunderson 2019), the electoral system (Adams, Gidron, & Horne 2018), internet usage (Lelkes, Sood, & Iyengar 2017), and most consistently, ideological polarization (Abramwotiz and Webster 2016, Rogowski and Sutherland 2015, Adams, Gidron, & Horne 2020, Reiljan 2019). At the same time, within the US there is now ample evidence that social sorting has increased, in concert with partisan polarization. Theoretically, we would expect this kind of sorting to increase affective polarization, by compounding the natural tendency to favor one's in-group and denigrate one's out-group. As parties become more sorted, they no longer only represent a divide in terms of

the ideological in or out-group of the respondent. Instead, more demographically sorted parties become the embodiment of a respondent's ideological and demographic in or out-groups, essentially activating further inter-group bias, leading to greater affective polarization.

Some scholars have demonstrated that individuals with certain specific identities which are aligned with their partisan identity are more affectively polarized, such as racial identity (Oliver 2018, Westwood and Peterson 2019) or religious identity (Davis 2018). Ahler and Sood (2018) demonstrate that those who perceive their out-party as highly sorted (i.e. as being comprised mostly of disliked groups) also hold highly negative views of that party, and Rothschild et al. (2019) demonstrate that correcting these misperceptions about party group images or stereotypes can reduce levels of affective polarization. I build on this scholarship by exploring whether the extent to which a political party (or in the case of this study, a fictional group serving as a proxy for the party) reflects the identities of an individual (i.e. the extent to which the individual shares high levels of identity alignment with a party) affects their affective evaluations of the group.

> ***H3 Affective Evaluation Hypothesis:*** Greater (weaker) identity alignment between individual and political group results in more positive (negative) evaluations of the group

Ideological Perceptions

One final consequence to be explored here is the possibility that social sorting changes the way individuals perceive parties ideologically. Brady & Sniderman (1985) and others have demonstrated that individuals are able to infer party positions based on what they know and feel about particular groups associated with the party, indicating that individuals do use non-policy based information to infer ideological placement. Applying social sorting theory here, we would expect that individuals who have high levels of identity alignment with a party would be likely to infer that this group shares their policy preferences also, and thus place the party close to them on the ideological spectrum. Conversely, parties which share low levels of identity alignment with an individual are likely to be viewed as less proximate, and possibly more extreme than the individual. Some evidence for this already exists. Ahler and Sood (2018) demonstrate that the extent to which individuals perceive (or in fact, misperceive) parties as being composed of "party-stereotypical groups" affects individuals' feelings about, and judgments of, particular parties, demonstrating that individuals form beliefs about the parties based on the groups they attach to the party, a key component of the theory of social sorting. Rothschild et al. (2019) demonstrate that when individuals hold incorrect party group images (or stereotypes in their terms), then correcting these misperceptions can reduce perceived ideological differences, demonstrating that when individuals view the group image of the opposing party as more similar to their

own identity, they become less distrustful of the party and view them as less extreme.

> ***H4 Ideological Perceptions Hypothesis:*** Greater (weaker) identity alignment between individual and political group results in closer (more distant) perceived ideological placement

Methodology

There remain some key challenges with using experimental methods to study the consequences of social sorting. As Ahler and Sood (2018) state: "We still largely lack direct evidence that the parties' social composition drives partisanship" (2018, p. 964), and thus, lack direct causal evidence that social sorting is driving affective polarization. Ahler (2018) argues that "Experimentally manipulating attitudes toward social groups is quite hard and manipulating peoples' identities nearly impossible" (p. 8, 2018), which has thus made it difficult to experimentally test the "group theory" of party identification, and by extension, to experimentally test the effects of social sorting.

In this paper I build on previous scholarship regarding the consequences of social sorting by devising an experimental method to directly test the effects of changes in identity alignment. Through a series of treatment articles, I manipulate the extent to which the respondent's own demographic in and out-groups are associated with their own political party (study one and three), their out-party (study three) or with three fictional political organizations (serving as proxies for parties, in study two). In each study, I use the respondent's own demographic information on race, gender, age, and religiosity to determine in and out-social groups for each category. I then manipulate the extent to which a given party's or political group's supporters contain large numbers of the respondent's in or out-social groups. In the congruent treatment, a group or party will be portrayed as containing high numbers of the respondent's social in-groups, with the aim of increasing their latent sense of identity alignment. In the incongruent treatment, a group is portrayed as featuring high numbers of the respondent's social out-groups, in order to reduce latent identity alignment. Theoretically, a respondent confronted with a congruent in-party and incongruent out-party reflects a situation in which social sorting within the electorate is high. This approach provides two key benefits that distinguish it from previous research.

Firstly, almost all of these studies rely on what Mason and Wronski (2018) term as "objective" measures of social sorting - that is, they use existing literature to determine which identities should line up with a particular partisan identity, based on the extent to which certain existing groups (for example, evangelical Christians) report high levels of support for one particular party (in this case, the Republican party). These groups may be racial, religious, gender, or age based groups, or in some studies, groups based on sexual

orientation or financial means (Ahler & Sood 2019). In any case, in each study the authors make a judgment call about which is the "correct" party for each identity to be linked to, and perform their analysis based on these assumptions. The key issue here is that we assume individuals accurately perceive that these groups are linked to certain parties, when research suggests party group perceptions are highly inaccurate and contain wide variation (Ahler & Sood 2019). It also assumes that each individual should base their partisan identities on the same set of identities, yet we lack research that conclusively supports this. Furthermore, treating these identities as singularly "correctly" or "incorrectly" sorted masks significant sub-group heterogeneity, and thus means scholars either treat identities as in conflict (see Mason's 2016 discussion of cross-cutting identities) or ignores this heterogeneity entirely. Yet individuals may in fact perceive that one overall group is sorted to one particular party, yet view their specific sub-group as associated with another party. For example, an older, white, evangelical woman may accurately perceive that women broadly favor the Democratic party, but may see older, white, evangelical women as favoring the Republican party, without feeling any associated cross-pressures of being a woman, as they see themselves as belonging to a coherent, sorted sub-group. This work attempts to remedy this by manipulating subjective identity alignment based on the identities of the individual in question, presenting certain parties and political groups as being more or less similar to the individual, rather than as being more or less well represented by specific predetermined groups.

Secondly, many of the previous studies regarding this topic have focused on the linkages between party identity and just one other identity, such as race (Oliver 2018, Westwood & Peterson 2019) or religion (Davis 2018), or small group of identities (Mason & Wronski 2018). This limits our ability to make generalizations about the extent to which party identity is group based in general, or based on just a few specific groups. It also masks the effects of "compound identities" (Westwood & Peterson 2019, see also Mason & Davis 2015, Mason & Wronski 2018), in which the effects of multiple well-sorted identities are greater than the effects of each individually sorted identity combined. That is, while being well sorted on gender, age, racial, or religious or ideological grounds may strengthen partisan attachments or affective polarization slightly, being well sorted on all of those grounds strengthens identification and polarization more than the sum of the individual parts. This work seeks to address this issue by using not one, but four identities in experimental manipulations. This presents something of a "blunt force" approach, in which all four identities are manipulated simultaneously, forcing the individual to perceive themselves as more or less well aligned. Therefore, if there is a causal relationship between overall levels of identity alignment and polarization, it should be evidenced here. By also manipulating four identities simultaneously, the groups or parties should better represent the particular subgroup that an individual belongs to (at least in terms of the interaction between the four identities used here), thus again directly affecting the individual's level of identity alignment.

Study One: Party Identity and Party Cues

Subjects were recruited using a convenience sample of political science students at a large public university. The study took place over four days, and subjects were awarded extra class credit in exchange for their participation. Analysis is limited from the original 500 students who participated to those who identified themselves as Democrats (N = 346) or Republicans (N = 55), with Independents and leaners excluded as the theory posits no clear effects for these groups. This criteria gave a total N = 401 adults (272 women, 128 men, 1 who chose not to identify), aged between 18-50.

The respondents were asked a short battery of demographic and political knowledge questions, before being randomly sorted into one of three treatment groups, and asked to read a short article about a fictional study of the supporters of a particular party (always the respondent's self-identified in-party). Within this study, four demographic groups associated with the party's supporters (age, race, gender, and religiosity) were manipulated, to be either more similar or more different to the subject's own demographic characteristics, using the information they provided at the start of the experiment. These characteristics were chosen as they have been shown over numerous studies (see Mason and Wronski 2018, Rothschild et al. 2019) to be strongly associated with perceptions of a party's group image, and therefore could be significant in impacting an individual's perceived level of identity alignment.

In the control treatment group, the demographic characteristics in question were mentioned but with no detail. In the congruent treatment group, these characteristics were manipulated such that supporters of the party are portrayed as being more like the subject than authors of the study previously expected, while in the incongruent treatment group, these characteristics were manipulated such that supporters of the party are shown to be more unlike the subject than previously expected. Absolute information about the relative prevalence of demographic groups among the party's supporters is not provided, instead the treatment states that these characteristics are more or less prevalent than the fictional researchers anticipated. A sample sentence from the treatment reads "For example, women comprise a larger share of the party's supporters than the researchers had anticipated". The intention of this manipulation was to move the party's image away from whatever baseline the subject holds without requiring knowledge of the baseline party image for each subject, in order to change the level of identity alignment between the respondent's own identities and the party's supporter's demographic identities.

After reading the treatment article, the respondents were asked a series of questions regarding, first, the strength of their partisan attachments, and secondly, their opinions on particular policy issues. To test my party identity hypothesis (H1), I analyze whether the initial congruent/control/incongruent treatment affects the extent to which individuals identify with their in-party, measured using the strength of partisan-

ship scale developed by Huddy, Mason, and Aarøe's (2015). To test my party cue hypothesis (H2), after responding to the strength of partisan identification questions, the subjects were further randomly divided into a total of six treatment groups. Subjects were asked to provide their opinions on five policy questions (healthcare, affirmative action, the minimum wage, income tax, and right-to-work laws). When responding to these questions, each original group was split into two sub-groups, with one sub-group provided a party cue treatment, and the second sub-group provided no party cue (for example, on a given policy they were told "The Republican Party opposes this policy, while the Democratic Party supports this policy"). The treatment groups thus combine the original identity alignment treatment with a party-cue treatment, resulting in six treatment groups.

Table 1.1: Number of subjects per treatment condition

	Congruent	Control	Incongruent
Party Cue	70	71	67
No Cue	62	64	67

Cell entries are number of subjects within each treatment condition.

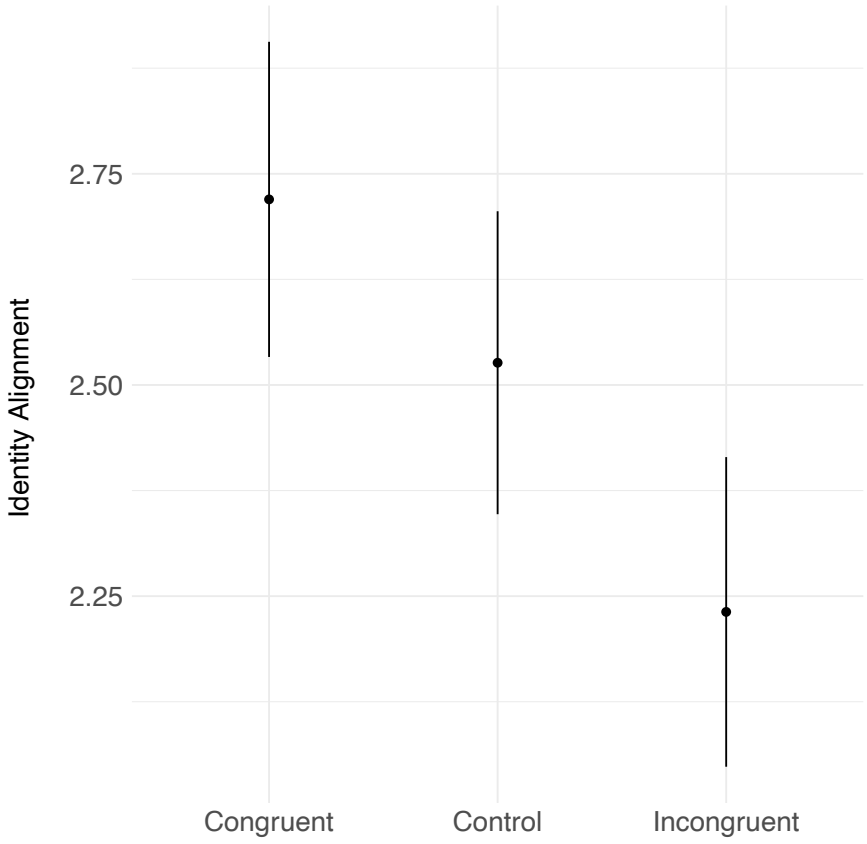
To test the second hypothesis, I use these policy positions to test whether individuals respond more or less positively to party cues, as measured by becoming stronger or weaker supporters of their party's policy positions, when their party is presented as being more or less like the subject. The primary dependent variable is a scale of the five policy questions combined, coded such that higher values equate to greater agreement with the respondent's self-identified party's position.

Manipulation Check

As a preliminary manipulation check, I examined whether the treatment affected perceptions of the party, particularly the extent to which individuals felt that their party's supporters shared their identities. After the treatments offered in parts one and two of this study, respondents were asked a series of questions regarding their perceptions of Democrats and Republicans. For example, "thinking about (Democrats / Republicans), do they tend to be...? Male / Female". Using responses to these questions and the demographic information provided by the subjects, I construct a measure of identity alignment, which runs from 0-4. This measures the extent to which individuals believe their own characteristics match their typical co-partisan, using only the demographic categories that were manipulated in the treatment. A 0 on this scale indicates that respondents believed their typical co-partisan shared none of their own demographic characteristics, while a 4 indicates the respondent believes their typical co-partisan matches them in terms of their race, gender, age group, and religiosity. This gives a measure of how well aligned individuals perceived their own identities to be to those

associated with their party. Figure 1.1 displays the mean value of identity alignment by treatment group.

Figure 1.1: Mean of Party Identity Scale by treatment, with 95% confidence intervals



I find that the treatment appears to have the expected effect - those in the congruent condition show the highest level of identity alignment while those in the incongruent condition show the lowest level. Moving from the control to the congruent condition results in roughly a 5% increase in identity alignment (difference of 0.2, p=0.07 in a one-tailed test), while moving from the control to the incongruent condition results in a decrease in identity alignment of roughly 7.4% (a difference of 0.3, p=0.01 in a one-tailed test). This provides important evidence that the results to follow are a direct result of changes in underlying levels of alignment between the individual's demographic identities and their party's associated group identities.

Results: Party Identity

The primary dependent variable for the first part of this study is the strength of partisan identification of subjects in the post-treatment survey. The variable is measured using an 8-item scale, based on Huddy, Mason, and Aarøe's (2015) expressive partisanship measure. Full text of this measure is available in Appendix **??**. This strength of party identification variable runs from 0 (very weak) to 8 (very strong), has a mean of 5.08,

a standard deviation of 1.26, and a reliability coefficient of = 0.81. One-tailed tests are used to determine whether differences between the treatment groups are statistically significant in the expected direction. If the party identification hypothesis (H1) is correct, we would expect to see that partisan attachments would be stronger among those in the congruent condition, and weaker among those in the incongruent condition, than those in the control group.

Figure 1.2: Mean of party identity scale by treatment, with 95% confidence intervals

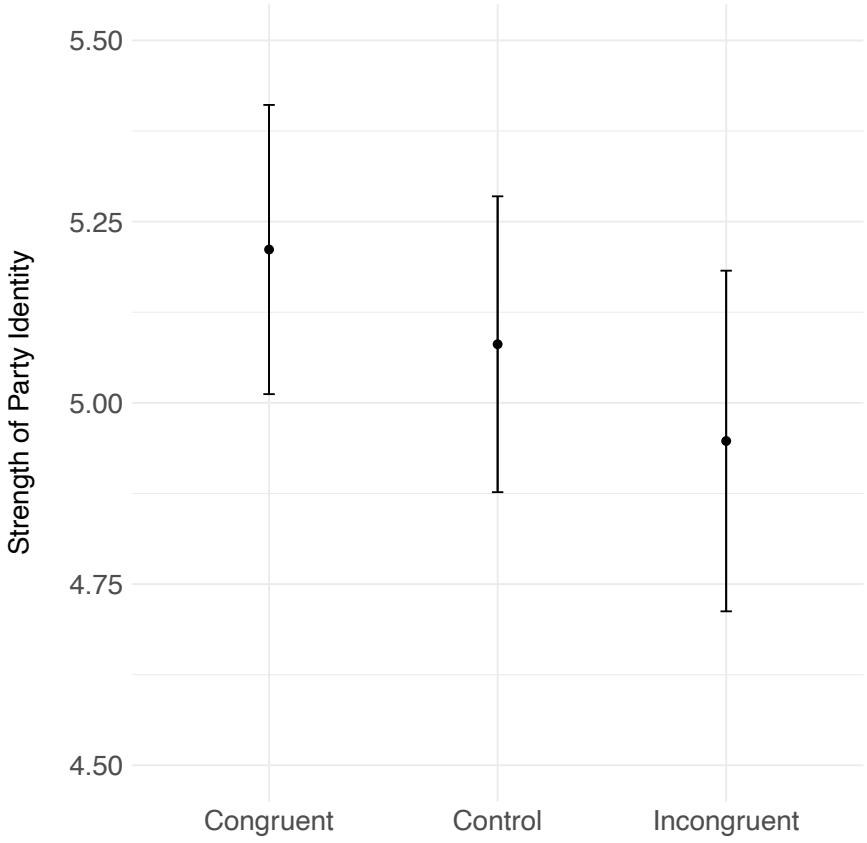


Figure 1.2 displays the mean of my strength of party identity scale for each of the treatment groups. While the differences between the treatment and control groups are not statistically significant, the difference between the congruent and incongruent groups is significant (a difference of 0.26, p = 0.05 in a one-tailed test), and the direction of the effect conforms to the expectations of the theory - those in the congruent condition appear to be the strongest identifiers, while those in the incongruent condition appear the weakest identifiers. These findings are encouraging, though not conclusive. There appears to be a linear pattern, and given a larger sample size it's possible the differences between each treatment and control group could reach conventional levels of significance. That a treatment as simple as the one offered in this study, one short fictional article, could have had detectable effects on the strength of partisan attachments - a phenomenon

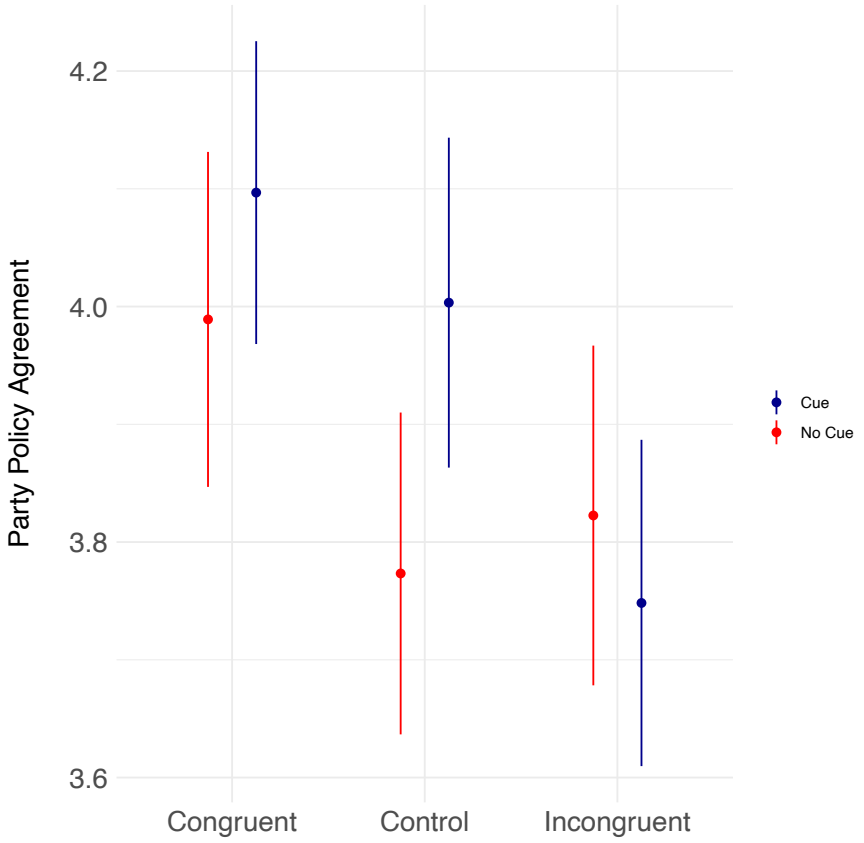
once described as the "unmoved mover" (Johnston 2006) - is worth noting. The magnitude of the differences is not large (a difference in strength of identity of roughly 5% between the congruent and incongruent conditions), but nor would we expect it to be, given that partisan identities change slowly over time (Green, Palmquist, and Schickler 2002). That this treatment may show signs of any effect provides some evidence that changes in the party group image - particularly changes regarding the demographic characteristics of party support - resulting in changes to the alignment between the identities of the individual and the party, do exert an effect on the partisan attachments of individuals.

Results: Party Cues

The dependent variable for the test of the party cues hypothesis (H2) measures levels of party policy agreement - support for particular policy positions espoused by the respondent's party. I recoded individual responses to these policy questions according to the policy positions of the respondent's party (for example, Republicans who strongly opposed Obamacare score highly on party policy agreement, and Democrats who strongly support Obamacare also score highly in party policy agreement). I first present a summary measure of these policies, by combining all five into a scale, which runs from 1 (low party policy agreement) to 5 (perfect party policy agreement), and has a mean of 3.91 and a standard deviation of 0.58.

Figure 1.3 presents the results of this analysis. Here we find some supporting evidence for the party cue hypothesis, particularly regarding the effects of the incongruent treatment. We see that in the control condition, the cue has the positive effect we would expect based on previous literature surrounding party cues, and this difference is significant at the .05 level (a difference of 0.23, $p = 0.02$ in a one-tailed test). In the congruent condition, we see again a positive effect of the cue, however, this effect is of a smaller magnitude than the effect of the cue in the control condition, and is insignificant, counter to expectations (a difference of 0.11, $p = 0.15$). This is likely due to the somewhat unexpected effect of the congruent treatment in the no-cue condition, which may reflect that some individuals are responding to the congruent treatment even when the cue is not provided (possibly due to already having knowledge of their party's positions on these issues), by becoming stronger supporters of their party. Nonetheless, the fact that the cue continues to have a positive effect in the congruent condition does somewhat support the theory proposed here, as well as the overall level of support in the congruent cue condition being greater than the control cue condition (though this difference is not significant). The most interesting finding is the effect of the cue in the incongruent condition. Here, we see no statistically significant difference (a difference of -0.07, $p = 0.98$ in a two- tailed test). The level of support shown for the subject's party's policies appears exactly the same in the control - no cue condition and the incongruent condition whether the cue is provided or not - indicating that the incongruent treatment effectively eliminated any positive effects from the cue. When disaggregating these

Figure 1.3: Mean of party policy agreement scale by treatment and cue, with 95% confidence intervals



results by policy area, I find that, while the strength of the cue effect in the control and congruent condition does vary by policy, the effect of the incongruent condition in negating the effects of the cue is consistent across all policy areas. Given the extensive literature on party cues, this is a consequential finding - party cues regarding policy appear to be ineffective when that party's group image is less aligned with the identities of the individual receiving the cue. This provides strong support for the theory that social sorting, through identity alignment, moderates the effects of party cues, particularly when the party's image is shown to be less like the individual, significantly reducing (or in this case eliminating) the effects of cues.

Combining the findings from study one, I find tentative effects of two consequences of social sorting, providing evidence that social sorting may impact mass polarization in two distinct ways. The analysis of H1 indicates individuals may base the strength of their partisan attachments on the extent to which their own identities align with a given party, though the results are statistically insignificant. If these results can be confirmed, we would expect as the parties become more homogeneous, certain individuals are likely to form stronger or weaker attachments to each party, potentially resulting in two parties comprised of partisans with strong attachments to their own party, and very weak attachments to the opposing party. Given the

effects demonstrated by previous research of the impact of partisan strength on political views and opinions, particularly regarding affective evaluations of the party, this provides evidence of one causal path through which social sorting affects mass polarization. The results from H2 provide another consequence of social sorting: if individuals become less well sorted within their party, they become less responsive to party cues. This would seem at first a positive result for potentially decreasing polarization: if citizens are less responsive to party cues, they may become less polarized in their issue positions (or express less polarizing views on the issues, whatever their true issue positions may be). However, given that sorting in the electorate is increasing rather than decreasing, instead these results indicate that, while sorting remains high, responsiveness to cues will remain high, and thus as elites polarize and provide more polarizing cues, we should expect well sorted individuals to form more polarized policy views.

Study 2: Affective Evaluations and Ideological Perceptions

I now turn to study two, in which I use the same experimental approach from study one, but move away from manipulating the social groups associated with parties, and instead manipulate the social groups associated with fictional campus organizations to explore two other dependent variables: affective polarization, and ideological perceptions. As the quantities of interest here do not require the direct use of parties in the treatments, proxy groups are used to represent political parties instead. This provides a number of benefits. Firstly, individuals have no strong preconceived attachments to the group, and thus their views should be more malleable. Secondly, they also have no strong notions of the demographic makeup of the group's supporters, thus making it less necessary to be concerned about moving the group image away from any concrete baselines. Instead, individuals will form completely novel views of the group based only on the information provided in the treatment. Third, using this method I am able to explore reactions to both in and out-groups (both in terms of social identity and ideology) with the same respondents. Finally, while with parties it may seem implausible to argue that certain groups are more prevalent than expected among certain party's support, here fictional groups could easily be associated with any particular demographic group, therefore making the treatment less susceptible to being dismissed by more educated respondents.

In order to test hypotheses three and four, I undertook an experiment using a second convenience sample of 467 undergraduate students at a large public university. Subjects were informed they were being asked for their opinions on a randomly selected subset of a number of international organizations who were considering opening up chapters on campus. These students were first asked the same series of demographic questions as study one, along with questions about their political activities, their partisanship and their political knowledge. The subjects were then provided with short descriptions of three fictional organizations, with

questions following the description of each group. The names and descriptions of each group included an ideological signal, either left-leaning, right-leaning, or non-ideological. The organization "Union Leaders of Tomorrow", which works to recruit future labor union members and leaders, was created to send a left-leaning ideological signal, while the organization "Movement for Consumer Freedom", which works to campaign for deregulation and less government involvement in the economy, was created to send a right-leaning signal. A third group was included with no explicit political leaning. Other than their political persuasions, the organizations were described as offering training and support to students, including financial support to attend international events.

The description of each organization also included demographic information about the supporters of the group. As in study one, within these descriptions, four demographic groups associated with the organization's supporters (groups related to age, race, gender, and religiosity) were randomly manipulated, to be either more similar or more different to the subject's own demographic characteristics, using the information they provided at the start of the experiment. In the control treatment group, no demographic characteristics of supporters were mentioned. In the congruent treatment group, these characteristics were manipulated such that supporters of the organization are portrayed as being similar to the subject, while in the incongruent treatment group, these characteristics were manipulated such that supporters of the organization are shown to be dissimilar to the subject. For example, a young, male, religious, white respondent reading the incongruent treatment about a group, would be told the group's supporters tended to be older, female, non-religious, minorities. This provided six different treatment groups, in which the demographic treatments were assigned to each of the three fictional groups in turn, such that a single subject saw all descriptions of both political organizations and the third non-political organization, and told one group was similar to them demographically, one was different, and one provided no demographic information, with each demographic signal randomly assigned to one group. The sample was then restricted to those who described themselves as Liberal (N=317) or Conservative (N=44), as I have no clear expectation regarding moderates. Table 1.2 displays the number of subjects in each condition.

Table 1.2: Number of subjects per treatment condition

	Congruent	Control	Incongruent
In-group	116	125	115
Out-group	120	115	121

Cell entries are number of subjects within each treatment condition.

After reading each description, subjects were asked to provide evaluations of the group in question. In order to test hypothesis three, regarding affective polarization, I utilized a modified form of Ahler and Sood's

(2018) social distance battery, with changes intended to better reflect the direct circumstances of student experience. In this case, the subjects were asked a series of questions about their willingness to tolerate or support the group's presence on campus. These included questions as to whether they would support the group opening a chapter on campus, how likely they would be to join the group, how likely they would be to donate to the group, and how happy they would be for the university to provide support and funding to the group. Each question was asked on a 5 or 7 point scale. These responses were combined into a scale of group evaluations for both respondents' ideological in-group and ideological out-group, which runs from 0 (most negative) to 1 (most positive), Alpha = 0.81. I used the respondent's own ideological placements, and their self-described ideology (Liberal, Conservative, Moderate, or Other) to determine which group represents their ideological in and out-groups.

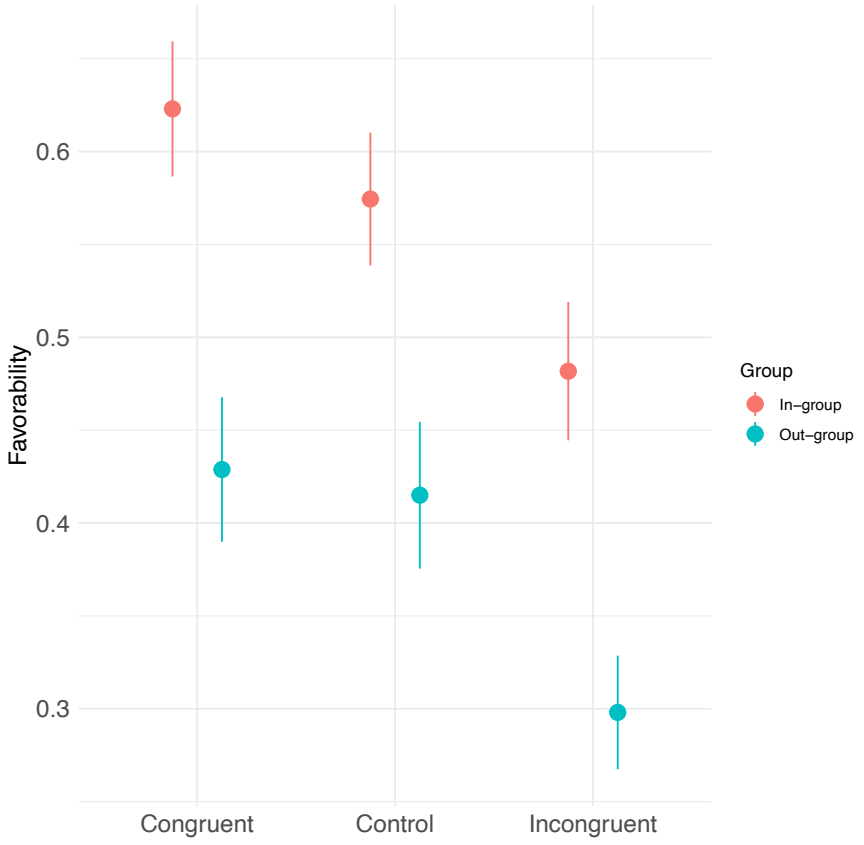
My fourth hypothesis (H4) explores whether identity alignment affects the perceived ideology of political groups. To test this, subjects were asked to place themselves, the major political parties, and the fictional campus groups on a 10-point ideological scale. I used this information to create a measure of the absolute difference between the subject's ideological placements of themselves, and their placements of each group, to form a measure of ideological distance. Finally, subjects were asked a series of screener questions to measure attention to the experiment, in order to eliminate the effects of non-compliance.

Results: Affective Evaluations

Figure 1.4 summarizes the results regarding their feelings towards the groups. Firstly, we see that, absent any demographic information (i.e. in the control condition) subjects formed significantly more favorable views towards their ideological in-group than their ideological out-group, as would be expected. More importantly, we see that when told that a group is less similar to them on demographic grounds, they formed more negative views about the group, regardless of whether that group shared their ideological disposition, and these differences are statistically significant for both in and out-groups. However, when told the group is similar to them on demographic grounds, we see weak effects on the ideological in-group (a difference of 0.05, p=0.03), and no effect whatsoever on the ideological out-group. That is, discovering a group is unlike you in terms of demographic identity matters, regardless of the group's ideology, but discovering they are similar to you demographically matters very slightly if they share your ideology, but does not change your views at all if they oppose your ideology.

To put this into the context of two-highly sorted parties within the American electorate, (using these fictional groups as proxies for political parties), we can compare the effects on feelings towards political groups of only ideological disagreement (the control condition), compared to the effects of ideological disagreement and social sorting (comparing the congruent ideological in-group to the incongruent ideological out-group).

Figure 1.4: Mean of affective evaluations scale of groups by treatment, with 95% confidence intervals



Here, we see that social sorting essentially doubles the gap between respondents' feelings. When individuals are aligned with a group in terms of only ideology, respondents are roughly 16% more favorable to their in-group than their out-group. When individuals are aligned with a group in terms of both ideology and demographic identity, respondents are then roughly 32% more favorable to their in-group than their out-group.

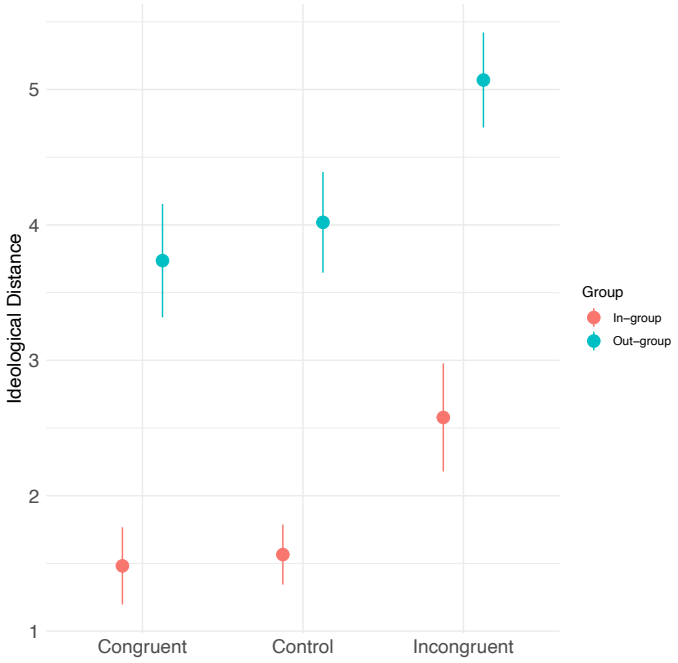
These results provide support for H3. Subjects do indeed take demographic information into account when forming their views about ideological groups, in somewhat narrow circumstances. Information that a group is similar to yourself has limited effects when the group is also similar to you ideologically, and shows no effect when the group is your ideological out-group. However, information that a group is demographically dissimilar to yourself is an important factor in forming opinions about groups, regardless of the ideological predisposition of the group. This provides evidence that information about demographic differences between political groups, such as the two main political parties, may contribute towards more negative feelings towards opposing groups, even if it leads to only very slightly more positive feelings towards the in-group. And more importantly, it shows that alignment with a group in terms of not only ideology but also demographic

identity, matters when forming evaluations of the group. These results offer support for the theory that demographic and social sorting may lead to increased affective polarization, as out-parties cease to represent only ideological opponents, but also social identity out-groups as well.

Results: Ideological Distance

Hypothesis four posits that individuals use this demographic information to infer the political preferences of the group. That is, if respondents assume that these demographic groups have defined group interests, and view those interests as distinct from their own interests, they may use the demographic information on show to infer something about the ideological disposition of the group. I test this possibility by exploring whether respondents place groups more or less closely to themselves on an ideological scale, based on the demographic information provided about the group. The results are shown in Figure 1.5.

Figure 1.5: Mean of ideological distance by treatment, with 95% confidence intervals



We see that information that a group is more aligned in terms of demographic identity with the respondent does not alter their ideological placement of the group at all, regardless of the ideological disposition of the group. However, the results show that information that a group is unaligned with the subject in terms of demographic identity causes them to place the group roughly one-point further away from themselves on the 1-10 ideological scale. That is, when an individual finds themselves unaligned with a group in terms of their social identities, they perceive that group as being less ideologically aligned with them as well. This provides

evidence for hypothesis four, and suggests that individuals do indeed interpret this information regarding the demographic identity of group supporters as an ideological signal. Again we see the effects of two sorted parties by comparing the difference between the in and out-groups in the control condition, and the congruent in-group and ideological out-group . We see the difference between these two ideological distances increases by roughly 25% when these groups represent both ideological and demographic differences, and that this result is driven almost entirely by the effect of lack of identity alignment with the ideological out-group.

Taken together, the results from study two provide strong experimental evidence of the consequences of social sorting, particularly regarding increases in partisan polarization in the US at the mass level. Individuals who view themselves as less aligned with a political group in terms of social identity view that group less warmly, and view them as more ideologically distant. Whether the increased ideological distance causes the decrease in warm feelings, I am not yet able to state conclusively. But it is clear that, if these results generalize to the major parties, then social sorting does have a profound effect on polarization in this context.

However, it is interesting that these effects are contingent on an interaction between identity alignment and ideological alignment. Ideological in-groups do not appear to receive strong affective or proximity "bonuses" from increased identity alignment, whereas ideological out-groups appear only to be punished by decreased identity alignment, but not rewarded for increased alignment. It is possible that this is because individuals assume some level of increased identity alignment based on the ideological label of the group - that is, they may believe that "I agree with this group ideologically, therefore I presume they look like me", and thus providing information that the group is similar to them in terms of their social identities is simply "old news". It may also result from the previously established findings that negative 'frames' are more impactful in altering opinion than positive frames (Boystun et al 2017, Ledgerwood & Boystun 2014). If we are to think how the information protrayed here as a way of framing the groups in terms of their identity composition, and of the incongruent frame as being a 'negative' frame of the group, then it may make sense that incongruent information has a greater impact on perceptions of the groups. The asymmetric nature of this relationship deserves further study.

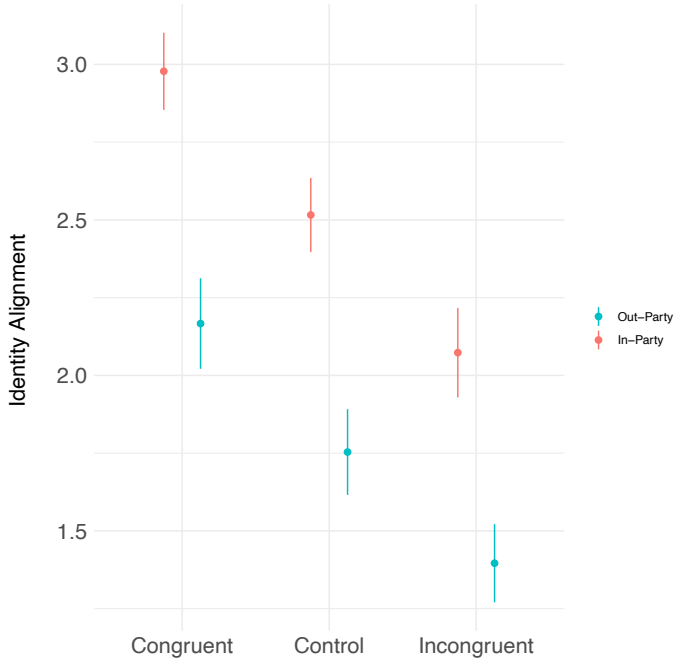
Study Three: Failure to Replicate

Study three attempted to replicate and extend the results of study one with a nationally representative sample. 1,800 Democrats and Republicans (including leaners) were initially recruited from Amazon's Cloud Research Prime Panels, using a census matching template, to represent the United States population based on age, race, and gender. 102 failed initial attention checks, and so were replaced by a further 102 respondents. Of those 1,800, a further 149 failed basic quality control measures (such as the exclusion of "straight line"

answers), and were excluded from analysis, resulting in a sample of 1,651 partisans, 866 Democrats and 785 Republicans. These respondents were subjected to the same experimental design used as study one, with some minor modifications. Firstly, the subjects were assigned to read an article about either their in- or out-party, extending the analysis from study one to include effects of identity alignment with the out-party. Secondly the policy questions were refined to match the time period, and expanded to include a fictional policy regarding local policy on cycling, in an attempt to explore the effects of cues in the absence of previous policy preferences. Finally, the study included further measures of social identity strength, in order to explore whether the effects of the identity alignment treatments varied by how strongly an individual held the identities within the treatment. No further modifications were made to the treatments or procedures.

Figure 1.6 presents the results of the manipulation check for the identity alignment treatment, the same check used in study one, here applied to whichever party the treatment was applied to for that respondent (either their in- or out-party, known as the "treatment party" from here on). The identity alignment measure represents the extent to which an individual feels supporters of the treatment party match their own individual identities.

Figure 1.6: Mean of Identity Alignment scale by treatment party, with 95% confidence intervals



The figure demonstrates that the treatment successfully altered levels of identity alignment between the individual and the treatment party, no matter whether the treatment was applied to the respondent's in- or out-party. The results are also interesting in themselves, as we see in the control condition a baseline

level of identity alignment that individuals feel towards their in- or out-parties. The results indicate that, on average, individuals feel their in-party shares roughly 62.5% of their identities in terms their race, gender, age, or religiosity. In contrast, on average respondents believe their out-party shares only 43.8% of these same identities, indicating a clear identity gap, that supports the arguments of the group theory of party identity and social sorting theory. Further study is needed here — while previous research has indicated that individuals are able to determine which political party certain groups (such as high-income individuals or gays and lesbians) are most associated with (Ahler & Sood 2018), there is insufficient research studying the extent to which individuals perceive the parties as being representative of their own core social identities. The findings here offer some initial evidence that a clear gap exists in the minds of voters about the two parties, even on such core identities as race, gender, and age.

However, the remaining analysis of this study provided no further support for these theories. Across a range of domains including strength of party identification, responses to party cues, ideological proximity and affective evaluations, the treatments administered had no statistically significant effects. The results of study one did not replicate with the larger representative sample, beyond the effects of the initial identity alignment treatment on levels of identity alignment with the treated party. While some results were directionally consistent with the theories proposed here, none reached significance, and the results across different domains were sufficiently inconsistent that it would be unwise to draw many conclusions from the results.

There are a number of potential explanations for these null results, with the precise cause unclear. However, the contrast with the consistently supportive findings of the original studies with the null results here make it hard to draw inferences as to how problematic the results of study three are for the theories proposed here. I will lay out the potential causes for these null findings and discuss their potential implications.

Firstly, it may be that levels of identity alignment form an important part of political decision making among students and younger Americans than among those who are no longer of college age. It is a well established fact that college students are generally unrepresentative of the population as a whole (Henrich et al 2010), and differ on many important characteristics. However, they also differ in terms of the level of development of their political beliefs and convictions, with college traditionally being seen as a time in which political beliefs, even partisan identities, are more subject to change, a finding that dates back to Newcomb's 1943 study (Newcomb 1943, 1968, Milem 1998). The contradiction between the results of the studies conducted on undergraduates and the study conducted on a larger representative sample, may simply imply that those who are still forming and developing their political identities and beliefs are more susceptible to changes in the levels of identity alignment between themselves and the parties. Older Americans who may have more entrenched political identities and beliefs may still perceive a change in identity alignment caused by the treatment, but whose party identity and other political beliefs are simply much less susceptible to

change. Further study is needed here to determine if this is the case.

A second, related cause, may be that younger voters, and particularly college students, are more affected by "identity politics" than older voters. Whether this is a generational difference or simply a factor of individual's interests and preferences with regard to politics changing as they age, it may be that younger voters and older voters base their partisan identities and political opinions on different considerations, with identity figuring as a much more prominent consideration among younger voters. If this is the case, it may be that the treatment offered here was effective in altering the perceptions of identity alignment, but this only has further downstream effects among younger voters. Unfortunately, the limited number of younger participants in study three makes further exploration of this impossible with the existing data.

A third possible explanation is simple performance effects - this result was undertaken by individuals on Amazon's Cloud Research Prime Panels, who have taken multiple online surveys in the past, and thus may be more likely to be able to identify experimental manipulations as they are administered. If so, the effects on identity alignment may not be true effects, but instead may simply be individuals providing the researcher with the answers they believe I want, or simply reporting the information they remember from the treatment article without having actually been affected by it. This may point to a weakness in the treatment if it can be so easily detected, and likely requires further refinement of the treatment and exploration of a different platform for the research.

A final potential explanation involves some of the peculiarities of the online sample collected here. Despite several attention checks during the survey experiment resulting in expulsion from the survey, there are potential issues with the data that pose questions about the validity of the results here. Firstly, only two-thirds of respondents were able to identify which party they read about in the fictional survey only minutes before, indicating that the attention checks may have been ineffective in entirely ejecting inattentive survey takers. Secondly, while the data was representative on key demographic lines, it appears unrepresentative with regards to attention paid to politics, with subjects indicating very high levels of political interest and knowledge, which does not generally accord with most findings about the US public. The unrepresentative nature of the sample may mean these findings do not mirror the true effects of social sorting that occur among the general population.

Overall, the null results here certainly give cause for pause in drawing concrete conclusions from the initial studies conducted with undergraduates. It is possible that the relationship between an individual's perceptions of the extent to which a party matches their core social identities, and their partisan identity and political preferences, exists and responds in the ways indicated in the initial studies. It's also possible this relationship declines as voters age and form more concrete political identities and opinions, or begin to base these on other considerations. Alternatively, it may be that the initial results are simply a quirk

of working with college students as experimental subjects, and that the findings will not hold with future replications. Further work is clearly needed.

Conclusion: Social Sorting, Polarization, and Negative Partisanship

Combining the individual findings from the first two studies conducted on college students paints an intriguing picture of the effects of social sorting. Rather than increased social sorting resulting in strengthening the tie between an individual and their in-party, and thus resulting in greater polarization, instead it appears that decreased identity alignment (i.e. socially sorting away from a given party) has a more significant effect. Decreased identity alignment affects the extent to which individuals respond to cues from their in-party, and the extent to which they feel warmly towards or ideologically closer to either in or out-parties. This may be because individuals come with a pre-existing view that they are already well socially sorted with their in-party / ideological group, and therefore congruent information is perceived as "nothing new", whereas incongruent information forces them to adjust their perceptions. However, it is telling that even with fictional groups about whom they will likely have much weaker preconceived ideas about their social identities, it is incongruent information that reflects a lower level of identity alignment that has by far the greater effect.

This points to a rather different view of how social sorting is related to polarization than many traditional theories may posit. As individuals increasingly align their party identity with their social identities, causing the parties to become increasingly homogeneous, rather than forming stronger attachments towards their own party, they instead now perceive their out-party as not only an ideological out-group, but a social out-group as well, and thus form more negative views towards the out-group.

As previously discussed, the inability to replicate these findings among a more representative sample poses a number of questions. However, I do not believe it marks the end of this line of inquiry. The cause of this lack of replication deserves further study, in order to determine if the findings of these first two studies were simply a fluke, or whether the effects demonstrated here are real, but concentrated only among certain sections of the population, particularly younger voters. If the latter finding is accurate (which I believe is both plausible and probable), it may spur further research on the role that identity plays in political socialization, how, if, and when voters come to change the considerations that inform their political identities and opinions, and the importance of identity alignment for different groups within the US.

Previous research has shown that increased affective polarization is primarily driven by increasingly negative feelings towards the out-party (Adams et al. 2020), while the negative partisanship literature has

shown that individuals often position themselves against their out-party, rather than in relation to their in-party (Abramowitz & Webster, 2016), and if the initial findings from studies one and two are validated, then it seems social sorting may operate in a similar way, increasing polarization due to the increased identity gap between parties, rather than because of increased alignment between individuals and their in-party. Those concerned about the effects of social sorting should perhaps focus more efforts on the effects of being un-sorted with a political out-group, rather than the effects of increasing levels of sorting within one's own party.

Chapter 2

A Sorted Electorate is a Divided Electorate: The Rise of Social Sorting in the US

Abstract

Scholars have argued that increased social "sorting" — the process through which individuals' social identities come into increasing alignment with their partisan identity — is a significant factor in explaining rising levels of partisan animus, dubbed "affective polarization". However, significant challenges have plagued the measurement of sorting, and therefore questions remain. Has social sorting significantly increased in the US? Has it increased at a similar rate to ideological sorting? Are social and ideological sorting equally responsible for increasing affective polarization? Here, I introduce new measures of sorting derived from machine learning. I find that social sorting has increased at a similar rate to ideological sorting, and that both phenomena are important contributors to affective polarization within the American electorate.

Introduction

While previous works have extensively documented the rise in ideological sorting in the United States over time, questions remain about the extent to which social sorting — the process through which partisan identities become increasingly aligned social identities — has kept pace. Previous work has demonstrated significant increases in sorting among particular groups, and increasing linkages between certain identities.

However, the extent to which the American electorate as a whole has become more "socially sorted" over time remains in question.

Many scholars have also explored the consequences of both ideological and social sorting, with a particular focus on affective polarization. Affective polarization has been one of the defining characteristics of the most recent period of American political behavior (Mason 2015, Iyengar & Westwood 2015, Iyengar et al, 2019). Large numbers of Republicans and Democrats see the opposing party in an increasingly negative light, and are less willing to work with, cooperate with, or have their children marry, their partisan opponents (Iyengar et al, 2019). While many of pointed to both ideological and social sorting as a potential cause of increased affective polarization, limitations in measurement have limited our ability to compare the effects of different kinds of sorting on interparty dislike. This work attempts to remedy these two issues.

In this paper, I use the American National Elections Surveys (ANES) from 1984-2020 and the tools of machine learning to develop a new measure of sorting. I employ an ensemble machine learning method to generate "sorting scores", defined as the predicted probability that an individual is a member of a given political party, based on either their ideological beliefs or their social identities. Using these scores, I show that the aggregate levels of of both social sorting and ideological sorting have increased substantially since 1984, coinciding with the rise of affective polarization. I then employ linear regression to show that, at the individual level, even when controlling for traditional factors that we believe increase affective polarization, both social and ideological sorting are independently associated with heightened levels of affective polarization.

This paper aims to make several important contributions to the literature regarding sorting and affective polarization. Firstly, I propose and validate a measure which solves a number of the challenges of the measurement of sorting, particularly social sorting, that have plagued previous research. Secondly, this new measure enables us to compare the rise and change in aggregate levels of ideological and social sorting on the same scale, something not previously possible. Thirdly, this provides an approach for the measurement of sorting that can be easily replicated, for sorting in different contexts, and for examining other phenomena over time. Finally, and most importantly, this work provides evidence that both social and ideological sorting independently affect levels of affective polarization, and that both phenomena present challenges for the health of American democracy. The results of this paper portray an American electorate that is more divided now than any time within this dataset, both in terms of ideology and identity. Put simply, Americans can no longer see their reflection in their political opponents. Instead, they find a group that represents everything they are not, not only in terms of policy views, but even in terms of their racial, religious, class, and other core social identities. And the results of this increasingly divided electorate, is severely heightened levels of animosity between partisans. And most worryingly, the process shows no sign

of slowing down — indeed, as parties continue to target particular social groups, and retrench further to their ideological extremes, this situation is likely to get worse, not better.

Ideological sorting: increasing over time

That ideological sorting within the US has increased over time now appears a fairly settled debate. Numerous studies show that the political parties at the elite level have become sorted along ideological lines (Hetherington 2009, Binder 2016; McCarty, Poole, & Rosenthal 2006, McCarty 2019; Sinclair 2014), and consistently show that Members of Congress with more liberal policy views have become aligned with the Democratic party as conservatives have become aligned with the Republican party. At the mass level, the traditional view of ideological sorting in the electorate argues that individuals have, over time, changed their party identification to suit their existing policy preferences (Abramowitz & Saunders 1998, Fiorina, Abrams, & Pope 2008). Other scholars argue instead that individuals have shifted their policy views to conform to the dogma now adopted by their party (Johnson 2006, Lenz 2009, Levendusky 2009). While they may differ on the mechanism through which sorting occurs, both theories portray an electorate that is sorted in terms of operational ideology, with the two parties representing supporters who sincerely hold a collection of divergent policy preferences.

This consensus within the political science literature provides a useful basis to validate new measures of sorting. If my methods for measuring sorting reflect the consensus that ideological sorting has increased over time, this provides support for this new approach to measure different forms of sorting within the electorate. Therefore, my first hypothesis acts as a necessary condition to verify my methods for measuring sorting.

Ideological Sorting Hypothesis$_1$: Ideological sorting has increased over time

Social sorting: increasing over time?

Work on social sorting has grown significantly in the last decade within political science. This work builds on traditional social identity theory [SIT] (Tajfel & Turner 1979, Huddy 2001, Kalin & Sambanis 2018) to develop a theory that explicitly links social identities to party identity. Understanding party identity as intrinsically linked to other social identities dates back at least to *The American Voter* (Campbell, Converse, Miller & Stokes 1960), with Green, Palmquist, and Schickler (2002) crafting one of the most effective explanations for this link. These scholars compared party identity to a religious or ethnic identity, and argue that it should be understood as an identity formed in response to the groups which an individual

associates with it. They state that a party exists as a [stable] stereotype in the minds of voters, who in turn harbor a sense of attachment toward this group image (p. 26). They argue that major changes in party identification are due to changes in the stereotype associated with the parties in the minds of voters — when the answer to "What is a Democrat?" changes, individuals change their party identification to accord with the new definition of "a Democrat". Ahler (2018) describes this as the group theory of party identification, namely individuals form their party identifications by evaluating the groups associated with a particular party, comparing this "group image" of the party to their own group membership, and choosing to identify with a party (or strengthen their identification with that party) based on the similarity or difference in this comparison.

Social sorting is the process through which individuals' identities shift over time in order to form a better "alignment" between their social identities and their partisan identity, in order to conform to the "group image" of the party (Mason 2016, 2018). More specifically, it involves the alignment of an individual's social identities with those that they attach to their partisan in-group. For example, an individual may slowly change their partisan identity in order to form a better match between their own racial identity and the racial identities they attach to the political parties. At the aggregate level, it involves certain groups gradually showing greater levels of support for one particular party, as that group's identity comes into alignment with the party identity. Many works explore different groups in the aggregate, such as racial or religious groups, finding significant evidence that particular groups have become more sorted over time (Mason 2016, 2018, Claassen et al. 2019).

The consequences of social sorting

Social sorting is important, due to the consequences we find in increased sorting. Mason and Wronski (2018) demonstrate that those whose social identities most closely align with their partisan identity (in their view, White, Christians, Conservatives and the Republican party, and Black, Secular, Liberals and the Democratic party) also show the strongest identification with their party. They argue that individuals for whom these identities align can use the trends in aggregate partisan sorting within the electorate (along racial, religious, and ideological lines) as cues which enable them "to perceive their party identity as a singular social identity" (p. 261), thus enabling them to simplify the cognitively strenuous tasks of forming political opinions, by relying on the single identity for cues and information. Mason (2016) has shown that those with fewer cross cutting identities react more strongly to partisan messages, and that a larger number of identities which traverse this partisan divide can reduce anger towards the out-party. DellaPosta (2020) demonstrates that mass polarization within the US is driven primarily by the collapse in cross-cutting

social and cultural alignments, leading to ideological polarization through belief consolidation. Ahler and Sood (2018) demonstrate that the extent to which individuals perceive (or misperceive) of parties as being composed of "party-stereotypical groups" affects individuals' feelings about, and judgments of, particular parties. In sum, previous research supports the notion that when multiple political and social identities overlap with each other in a manner which decreases cross-cutting clevages, political conflict takes on a harsher, more uncompromising tenor.

However, key questions remain about social sorting, particularly, exactly how widespread this kind of sorting is beyond certain groups. What percentage of the American electorate is well sorted along lines of social identity? How has this changed over time? These issues remain due to the challenges of measuring social sorting.

The challenges of measuring sorting

Many of the previous studies regarding social sorting have focussed on the between party identity and just one other identity, such as race (Oliver 2018, Westwood & Peterson 2019) or religion (Davis 2018), or small group of identities (Mason & Wronski 2018). This limits our ability to make generalizations about the extent to which party identity is group-based in general, or based on just a few specific groups. It also masks the effects of compound identities (Westwood & Peterson 2019, see also Mason & Davis 2015, Mason & Wronski 2018), in which the effects of multiple well-sorted identities are greater than the effects of each individually sorted identity combined. Without a method that can account for a large variety of identities, and potential interactions between them, it is impossible to make a confident assessment of how widespread social sorting may be, or the effects of it in the aggregate.

Secondly, almost all studies of both ideological and social sorting rely on what Mason and Wronski (2018) term as objective measures of sorting - that is, they use existing literature to determine which identities or beliefs *should* accord with each partisan identity. In the social sorting literature, these objective measures tend to be based on the extent to which certain existing groups (for example, evangelical Christians) report high levels of support for one particular party (the Republican party). In each study the authors make a judgment call about which is the correct party for each identity or belief to be linked to, and perform their analysis based on these assumptions.

The issue with using these "objective" measures of social sorting is that we assume individuals accurately perceive that these groups are linked to certain parties, when research suggests party group perceptions are highly inaccurate, and contain wide variation (Lupton et al. 2015, Ahler & Sood 2019). It assumes that each individual should base their partisan identities on the same set of identities, yet we lack research that

conclusively supports this. Furthermore, treating these identities as singularly correct or incorrectly sorted masks significant sub-group heterogeneity, and thus means scholars either treat identities as being in conflict (see Masons 2016 discussion of cross-cutting identities) or ignores this heterogeneity entirely. Particularly, smaller sub-groups within a wider identity group may be sorted differently from the whole - something that has been frequently discussed in relation to Hispanic Americans. At the individual level, citizens may accurately perceive that on large identity group is sorted to one particular party, yet view their specific sub-group as associated with another party. For example, an older, married, educated, woman may accurately perceive that women broadly favor the Democratic party, but may see older, married, educated, women as favoring the Republican party, without feeling any associated cross-pressures, as they see themselves as belonging to a coherent, sorted group.

This issue also exists with ideological sorting, as we know that most citizens have complex mixtures of ideological beliefs, and tend to rarely fall squarely in line with the Liberal or Conservative elite consensus (Claasen, Tucker & Smith 2014). It is possible that many, if not most, partisans view their own heterogeneous preferences as being perfectly in-line with their party, as again we tend to assume that citizens are accurately able to perceive the beliefs of both party elites and other partisans, yet we know that is often not the case (Ahn 2021). Therefore, measuring ideological sorting with objective measures of ideological beliefs may ignore that those with heterogeneous preferences may still be well sorted, if large groups share similar kinds of heterogeneous beliefs and sort in similar ways.

In this paper, I expand on previous research on the sorting of particular groups or identities, to explore whether the electorate as a whole has followed the levels of sorting found in specific groups, and is now more well-sorted in terms in terms of social identity. Given previous findings that certain groups have become more socially sorted over time, and that individuals whose identities better align with one party display stronger signs of partisanship, I expect these trends should reflect an aggregate increase in the overall level of social sorting within the electorate.

Social Sorting Hypothesis$_1$: Social sorting has increased over time

Sorting and Affective Polarization

There is now substantial evidence that affective polarization has markedly increased within the US (Iyengar, Sood, & Lelkes 2012, Iyengar et al. 2019), and some evidence that this increase is not unique to the US but reflects a wider trend (Reiljan 2019). These authors have demonstrated a growing number of influences on affective polarization, such as income inequality (Gunderson 2022), the electoral system (Adams, Gidron, &

Horne 2022), internet usage (Lelkes, Sood, & Iyengar 2017), and most consistently, ideological polarization (Abramwotiz & Webster 2016, Rogowski & Sutherland 2015, Adams, et al. 2022, Reiljan 2019). This increase in affective polarization is happening at the same time as the hypothesized increases in ideological and social sorting.

Research already exists which indicates that ideological sorting is associated with affective polarization. Theoretically, a well sorted individual in terms of ideology sees their party and co-partisans not only as a political or social in-group, but as an ideological in-group that shares their particular policy beliefs. Rogowski & Sutherland (2018) find precisely this: those who are well ideologically sorted within their party form more polarized views of the parties and key figures within them. They also find that this is particularly true amongst those who have stronger levels of ideological commitments and greater political knowledge. Webster & Abramwotiz (2017) and Bougher (2017) find similar supporting evidence that those who hold more party-consistent ideological beliefs also tend to be more affectively polarized. I again intend to validate my measure of sorting by replicating these findings regarding ideological sorting.

Another line of research suggests ideological sorting may impact affective polarizations through the adoption of ideological *identity* by partisans (Devine 2014, Mason 2018). That is, individuals have adopted ideological labels and identities that are linked with their party, or with other groups associated with their party, without necessarily changing their policy views (Ellis & Stinson 2012, Claasen, Tucker & Smith 2014). This line of thinking would result in an electorate well sorted in terms of identity, but not necessarily policy views. Both Mason (2018) and Devine (2014) find compelling evidence that social ideological identity (ISI) affects political behavior independent from operational ideology, and increases affective polarization. I intend to extend these findings by testing whether operational ideological sorting at the individual level increases affective polarization, even when controlling for strength of ideological social identity, to determine whether both operational and identity-based ideological sorting increase affective polarization independently of each other, and to explore the relative weight of each component.

> ***Ideological Hypothesis***$_2$: Ideological sorting is positively associated with affective polarization.

There are several mechanisms through which social sorting could increase affective polarization. Many have been tested with particular groups, but not at the aggregate level. One key mechanism through which we would expect social sorting to impact affective polarization is through changes in the way an individual seeks to maintain positive self-esteem. SIT states that individuals seek to maintain a positive self-image

through positively distinguishing their in-groups from their out-groups. This may be through either forming more positive views of their in-group (and behaving more positively towards members of said in-group), or through forming more negative views (and behaving more negatively) towards one or many out-groups (Tajfel & Turner 1989). The larger number of identities shared by an individual and a group, the greater the incentives to view this group positively (and any opposing groups negatively), as the group becomes more central to one's own levels of self-esteem (Roccas & Brewer 2002).

Social sorting increases polarization by increasing the number of identities associated with political parties as in or out-groups, and thus increasing the incentives to more clearly distinguish one's own party from one's out-party through affective evaluations (Mason 2016, Mason & Wronski 2018). A socially well sorted individual views their party as not only sharing their political leanings, but also their social identities. These well sorted individuals then have a stronger incentive to positively distinguish the party through positive evaluations, as the party represents not just one, but many of their in-groups. More socially and ideologically sorted parties become the embodiment of a respondent's ideological and social in or out-groups, essentially activating further inter-group bias, leading to greater affective polarization. Put simply, as one's conception of self becomes more intertwined with a political party, the incentive to maintain a positive self-image becomes increasingly reliant on maintaining a positive view of that party, and distinguishing it from any out-party.

A further mechanism through which social sorting may increase affective polarization is through increasing the impact of elite cues. Previous research has demonstrated that stronger partisans are likely to have stronger responses to elite cues from their party (Malka & Lelkes 2010) and that affective polarization is associated with the increasingly negative information provided by political campaigns (Sood & Iyengaar 2016). Given that we expect well sorted individuals to feel stronger attachments to their party, we would then expect them to form stronger responses to elite cues. As elites have polarized in the United States, as has campaign rhetoric has increased in negativity (Geer 2012) we would then expect those who follow elite cues to equally form more polarized views of the political parties.

Studies that have explored this relationship have found promising evidence. Mason (2016) demonstrates that cross-cutting identities (i.e. less well sorted partisans) respond less negatively to threats towards their parties, and argues that as social sorting has increased, resulting in partisans with fewer cross-cutting identities, this dampening effect has decreased, resulting in greater levels of hostility in response to the political environment. Other scholars have demonstrated that individuals with certain specific identities which are aligned with their partisan identity are more affectively polarized, such as racial identity (Oliver 2018, Westwood & Peterson 2019) or religious identity (Davis 2018). Ahler and Sood (2018) demonstrate that those who perceive their out-party as highly sorted (i.e. as being comprised mostly of disliked groups) also hold

highly negative views of that party, and Rothschild et al. (2019) demonstrate that when individuals hold incorrect party group images or stereotypes, then correcting these misperceptions can reduce levels of affective polarization and perceived ideological differences. Through this study, I intend to demonstrate that social sorting does impact affective polarization, even when controlling for traditional factors that influence affective polarization such as ideological sorting, ideological identity, partisan strength, and political interest.

Social Hypothesis$_2$: Social sorting is positively associated with affective polarization

Methods

To test my hypotheses, I require measures of social and ideological sorting at the aggregate and individual level. My approach is based on a simple corollary of social sorting theory: if individuals are becoming increasingly sorted in partisan camps according to their social identities or ideological beliefs, then it stands to reason that the social identities or ideological beliefs they hold *should become increasingly predictive of party identification over time*. As parties become more homogeneous in their social identities and ideological preferences, so too should those identities and preferences increase in predictive power when attempting to estimate whether an individual is a Democrat or Republican. Therefore, my measure of sorting at the aggregate level is the overall accuracy with which partisan divides in the electorate can be predicted, based on either social identity or ideological identity. My measure of individual-level sorting is the predicted probability an individual is a member of their stated in-party, based on either their social identities or ideological beliefs. Turning the task of measuring sorting into one of prediction calls for the us of predictive tools, and so I employ a machine learning approach, and use the American National Elections Survey (ANES) data, from 1984-2020.

I employ an ensemble classification method to classify each individual as either a Democrat or Republican, based solely on, in the first instance, their ideological beliefs, and in the second instance, their social identities. An ensemble method is valuable here for a number of reasons (see Speiss (2017), Schapire 1990; Laan, Polley, & Hubbard 2007 for some of the benefits of utilizing an ensemble method), and has already shown great promise in political science. Grimmer, Messing, and Westwood demonstrate the utility of weighting different methods to increase the accuracy in estimating heterogeneous treatment effects, Hare and Kutsuris (2022) employ an ensemble approach to estimate the likelihood of any individual being a 'swing' voter, and Broniecki et al (2022) show that ensemble modelling can increase the accuracy of multilevel regression with poststratification (MrP). In general, the key benefit of ensemble modeling is that it employs a "wisdom of the crowds" approach of finding the best fit for the data by utilising a wide range of models that include

different functional forms and assumptions.

This method provides a large number of benefits. Firstly, it removes the requirement that the researcher use prior existing literature to make a priori assumptions about which identities or beliefs *should* be sorted within each party. Instead, the ML algorithm makes its own determination about which identities or beliefs (and interactions between different identities or beliefs) is associated with each party directly from the data, allowing for a more "objective" measure of sorting. Secondly, it allows for large numbers of interactions between identities in the one model, and beliefs in the second, in order to account for sub-group social sorting that other authors may have missed, and to account for the potential for individuals with heterogeneous beliefs to still sort in similar ways. [1]. This enables the algorithm to detect sorting at the sub-group level, and among those with heterogeneous beliefs, a key weakness in previous methods. Thirdly, it enables me to generate a parsimonious measure of sorting that is based on more than one, or a small group, of identities or beliefs, enabling me to measure the overall level to which an individual may be in conflict with their party. Fourth, it produces useful measures of overall aggregate sorting, something that previous research has lacked, as it has been limited by the focus on specific groups, identities, or beliefs (though see Selway 2011 for one approach). Finally, it allows for ideological and social sorting to be placed on the same scale, enabling easy comparisons between both individual levels of sorting and levels of aggregate sorting over time.

To develop the classification algorithms for both ideological and social sorting, each survey year is divided into a training set containing 70% of the observations from each survey year ($N = 14,523$), and a testing set, containing the remaining observations from each year ($N = 6,213$). [2]. Within the training set, training the model slowly improves its performance at translating the identities or ideological beliefs of the individual into predictions regarding the party identification of the individual, until no more improvement can be found. In essence, the model is attempting to map the social and ideological cleavages within the electorate that translate into differences between the parties. The final ensemble includes a selection vector machine, a random forest, a gradient boosting machine, classification and regression trees (CART), k-nearest neighbors, a neural network, and a generalized linear model, thus covering a wide variety of methods, each with its own strengths and weaknesses. For example, tree-based methods naturally fit interactions as a byproduct of their structure, while selection vector machines and neural nets are flexible enough to fit nonlinear relationships in high-dimensional parameter space. The ensemble approach thus benefits from the diversity of the component models in fitting different kinds of patterns manifest in the data. The model is trained using the CARET

[1] However, to reduce overfitting within the training set, the number of interactions is limited to a maximum of 8-way interactions

[2] As a robustness check, to ensures that the findings were not driven by a bias caused by one survey year having a more significant impact on the overall results, the analysis has also been conducted on a training set with the an equal number of respondents from each survey year ($N = 500$), rather than a percentage of each survey year. The overall results can be seen in Appendix **??**, and the findings remain unchanged.

package, using five-fold cross-validation, repeated three-times. Full specifications of the model are available in Appendex **??** but are removed here for brevity.

Due to my focus on the effects of sorting on affective polarization (and due to the limitations in using ensemble methods to predict multi-category outcomes), I limit the analysis to only Democrats and Republicans, as I have no predictions about how social or ideological sorting effects the partisan evaluations of independent voters.

This method provides the two key measures needed to evaluate my hypotheses. To determine if sorting is increasing overall, I require an aggregate measure of the extent to which social identities or idealogical beliefs predict party identification. This is measured using the area under the Receiver Operating Characteristic (ROC) curve in a given survey year. This value (between 0-1) provides a measure of how well the model is able to distinguish between the two potential response categories — in this case, Democrats and Republicans, at different levels of sensitivity. Because the ROC is based on the model's ability to distinguish predictive power based on the area accounted for under the curve, from here on I refer to this fit measure as the "area under the curve" or AUC. This is a measure of the model's *discriminatory* power—how well it can distinguish between Democrats and Republicans. An AUC value of 0.5 indicates the model is no better than random guessing, while a value of 1 indicates the model is perfectly able to distinguish between partisans. These AUC values can also be interprested as the probability that the model will correctly distinguish between a randomly chosen "positive" (for example, a Democrat) observation and a randomly chosen "negative" observation, (in this case, a Republican).

If, as I argue, a more sorted electorate is one in which either ideological beliefs or social identities are more predictive of partisanship, then the machine learning model should be able to better categorize respondents within a more sorted electorate than a less sorted one. That is, if the model was able to perfectly determine whether each individual within the dataset was a Republican or Democrat based solely on their social identities or ideological beliefs, indicating that the individuals were perfectly sorted along their social identities or ideological beliefs into the two parties, we would expect the AUC to be 1. An AUC of 0.5 indicates that the algorithm is able to correctly determine whether any given individual is a Republican or Democrat, based solely on their identities or beliefs, 50% of the time, indicating that social identities or ideological beliefs were not significantly predictive of party identification (indeed, they would be no better than random chance in this two-party system). Given that the precise AUC of the model at any given time is vulnerable to changes in specification of the algorithm; what we are interested in here is *change* in the level of accuracy over time, holding the algorithm itself constant. In order to evaluate the over time hypotheses (HI1 and HS1), I compare the AUC under the same specifications of the machine learning model at different survey years within the test set.

To test the second set of hypotheses regarding sorting and affective polarization, I require a measure of how well sorted an individual is with their political party. For each observation, the model uses the individual's identities or beliefs to generate a predicted probability that that individual is a member of party X (in this case, the Democratic party). As this is a two-party system and Independents are excluded from the model, finding the probability they are a Republican is simply the inverse of the probability that they are a Democrat. This predicted value can be interpreted as a measure of how well that individual is socially or ideologically sorted into that party, according to the alignment determined by the algorithm, a "sorting score". That is, if the value is 1, the model has determined this set of identities or beliefs perfectly sorts into the Democratic party. I recode these values such that each individual has two predicted probabilities of being a member of their self-reported political party, the first based on their ideological beliefs, the second based on their social identities. As an example, individual j is a white, 68-year old, retired, protestant woman, from a union household in the South, with a weak racial identity, and identifies with both unions and big business (indicating a weak class identity). j is strongly in favor of increasing government spending in health, social security, and on government services, is strongly in favor of a government health insurance option, and believes abortion should be legal only in the cases of rape, incest, and risk to the health of the mother. She identifies as a Democrat. The model fits her ideological beliefs relatively well, with the algorithm generating an ideological sorting score of 0.84. However, the model does not fit her social identities as effectively, giving her a social sorting score of just 0.29.

To use this method to study sorting over time requires consistent measures of the demographic and social identities and ideological beliefs of the individuals within the electorate. Measures of ideological preferences are fairly consistent throughout the period, with 12 questions that can be categorized as ideological in nature appearing in each survey. Questions include topics such as federal spending, education spending, and the government's role in providing guaranteed jobs and aid to black Americans. These questions generally indicate a preference for either Liberal or Conservative approaches to government, and thus a combination of them should be expected to provide a fairly accurate measure of ideological sorting. See Table B1 in Appendeix **??** for a full list of variable used to develop the ideological sorting measure. By first confirming that the method effectively detects ideological sorting, I am able to more confidently state that increases in accuracy represent increases in social or demographic sorting.

However, the ANES data is somewhat limited in the extent to which individuals were questioned about their social identities. Until recently, high quality questions about identity strength were not common. Therefore, I use two groups of measures of social identity: the first, self-reported membership of particular demographic groups (race, gender, age, education, religious denomination, union membership, marital status, region, and native born status). Reporting membership of a group is a prerequisite for any level of

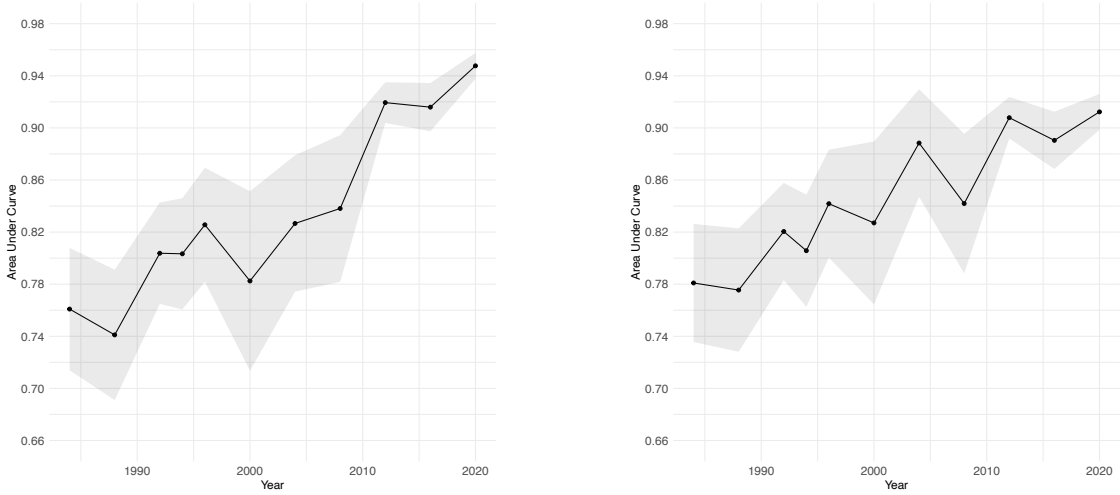
identification with the group, and thus while these measures can not perfectly capture whether an individual holds a certain identity, they do provide a baseline. The second group of measures are feeling thermometers, which have previously been used as a proxy for measures of identity strength (Winter & Berinsky 1999). This approach is in line with other scholars (Wilcox, Sigelman & Cook 1989, Mason & Davis 2015), who argue that feelings towards groups represent the best proxy for social identity that we have in historical data. The validity of these thermometers varies, for example, a thermometer differential between one's racial in- and out-group provides a measure of racial identity with reasonably strong face validity, and previous scholars have used this measure (Mason & Davis 2015, 2016). Here, I also use feeling thermometers as measures of class identity (measured as the difference between feelings towards unions and feelings towards big business), sexual identity (measured using feelings towards gays and lesbians), and gender identity (measured using feelings towards feminists). While imperfect measures, they offer the best proxy available in this historical data. The full list of ideological and identity measures included can be found in Table B2 in the appendix.

In order to explore the relationship between affective polarization and sorting, I use the "sorting scores" generated by the ML models as a predictor in a linear regression model of affective polarization. I measure affective polarization as the difference between the feeling thermometer evaluations of a respondent's stated in- and out-parties, in line with previous research (Iyengar et al, 2019), generating a measure running from -100 (indicating total preference for the out-party over the in-party) to 100. Some authors argue affective polarization can be measured as purely increasingly negative feelings towards the out-party (Adams et al. 2020). However, given that social sorting theory tends to focus on the effects of strengthening identity with the in-party (and the necessity to positively distinguish the in- party from the out-party), I use both the difference-based measure and separately model effects on feelings towards in- and out-parties individually. This enables me to explore whether social and ideological sorting increase affective polarization via increasing feelings towards an in-party, or through increasing negative evaluations of the out-party, or both.

Results: Sorting over time

Given the preponderance of evidence that exists that ideological sorting has increased over time, both at the elite and mass level, this analysis serves primarily as a validation of the approach of measuring sorting with this method. An increase in predictive accuracy would indicate that ideological beliefs are becoming more predictive of party identification, i.e. that sorting is occurring at the aggregate level. This would **not** indicate that respondents only held liberal or conservative beliefs, or that certain social identities were only associated with one party or the other. Rather it indicates that certain combinations of beliefs or combinations of social identities which the machine learning model is able to define, are incredibly rarely

Figure 2.1: Over time change in the AUC produced by the sorting machine learning models



(a) Ideological Sorting

(b) Social Sorting

found in one party or the other.

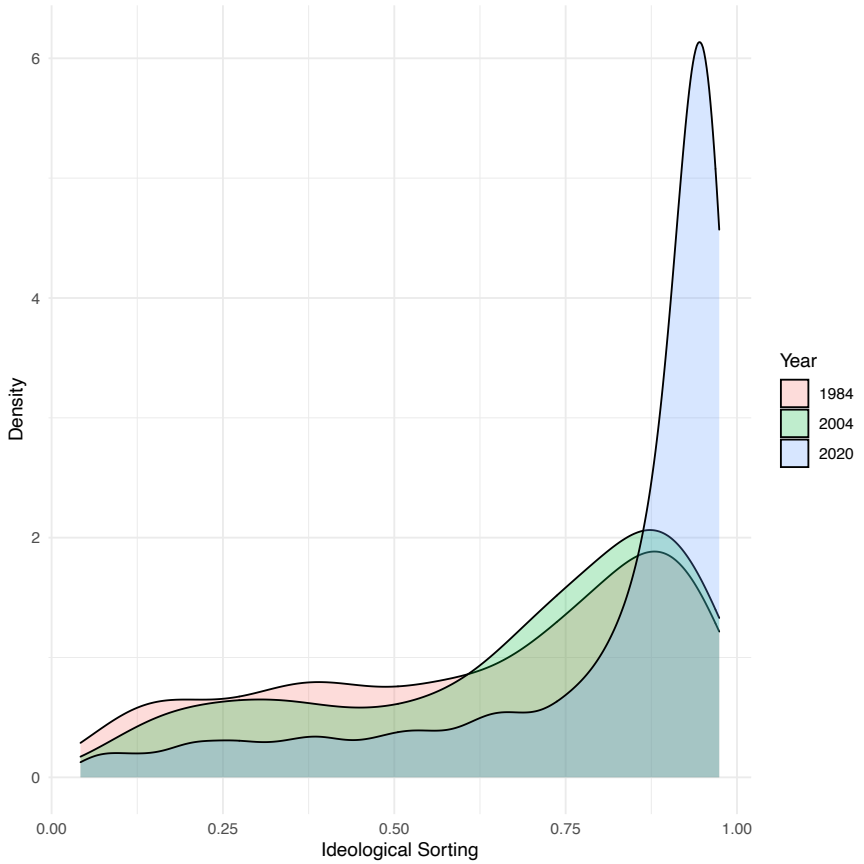
Figure 2.1a demonstrates that over time, ideological sorting has increased by roughly 19% from 1984 to 2020. Each point represents my measure of aggregate sorting: the AUC in the testing set during a given election year. The shaded region represents the 95% confidence interval around the estimate[3]. While the rate-of-change is not consistent overtime, we still see that in general, the machine learning algorithm is better able to predict the party identification of respondents in the latter half of the dataset than the earlier half, indicating that over time, the liberal or conservative beliefs of the individuals became more predictive of their party identification. Thus, consistent with previous research, the aggregate data indicates that ideological sorting within the electorate has increased over time, providing strong support for ideological hypothesis one, and most importantly, validation of this approach to measuring sorting.

In 2.1b, we see that rates of social sorting are comparable to those of ideological sorting. Between 1984 and 2020, the overall level of social sorting within the electorate has increased, by roughly 13%. Again as we might expect, the rate-of-change is not consistent overtime, but the aggregate picture is fairly clear: as time progresses, social identities become more predictive of party identification, with social sorting reaching its highest level in the dataset 2020. These results provide strong support for ideological hypothesis one, and social hypothesis one: both ideological and social sorting appear to have increased over time, to the extent that party identities can now be explained with a high degree of accuracy by the ML model using either ideological beliefs or social identities.

[3]Note that the 95% confidence intervals are calculated using 2000 stratified bootstrap replicates. The size of the CI is somewhat dependent on the number of observations in a survey year with some years having fewer, particularly 2000 and 2004, resulting in fewer observations for this year in the testing set, and larger confidence intervals.

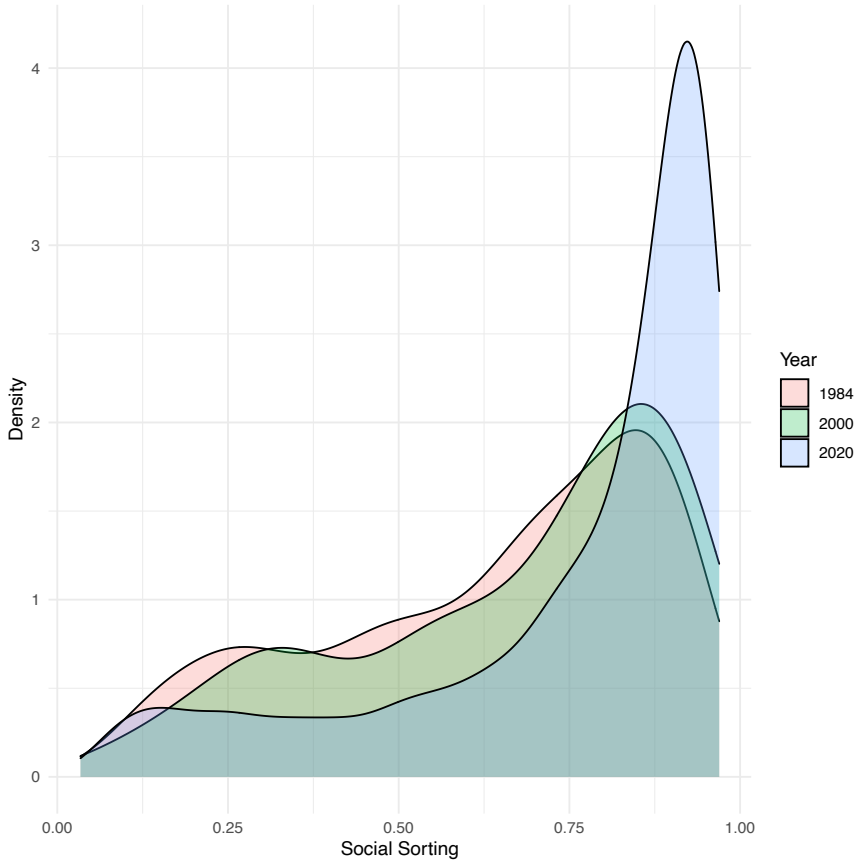
Another way to explore levels of sorting at the aggregate level is to examine the density of the individual predictions generated by the machine learning ensemble algorithm during different election cycles. If social identities or ideological beliefs become more predictive of party identification over time, we would expect to see greater density of higher sorting scores (the predicted probability an individual is a member of their in-party) in later election years. If the electorate is not well sorted, we would expect a roughly normal distribution of sorting scores, with a peak around 0.5. If the electorate is well sorted, we should expect a high density of scores above 0.5, with a taller, narrower peak indicating a very highly sorted electorate.

Figure 2.2: Over time change in the area under the ROC curve produced by the Ideological sorting machine learning model



Figures 2.2 and 2.3 display exactly that. When comparing the 1984 election survey to the 2020 election survey, we find that the sorting scores generated by the ML model are more strongly clustered towards the higher end in the later survey years than the earlier years. This can also be seen by the increase in the median sorting scores over time. In 1984, the median sorting score generated by the model is 0.72 for ideological sorting and 0.69 for social sorting. By 2020, these median scores have increased to 0.91 and 0.86 respectively. The present American electorate is one in which very few individuals appear to be poorly sorted within their

Figure 2.3: Over time change in the area under the ROC curve produced by the Social sorting machine learning model



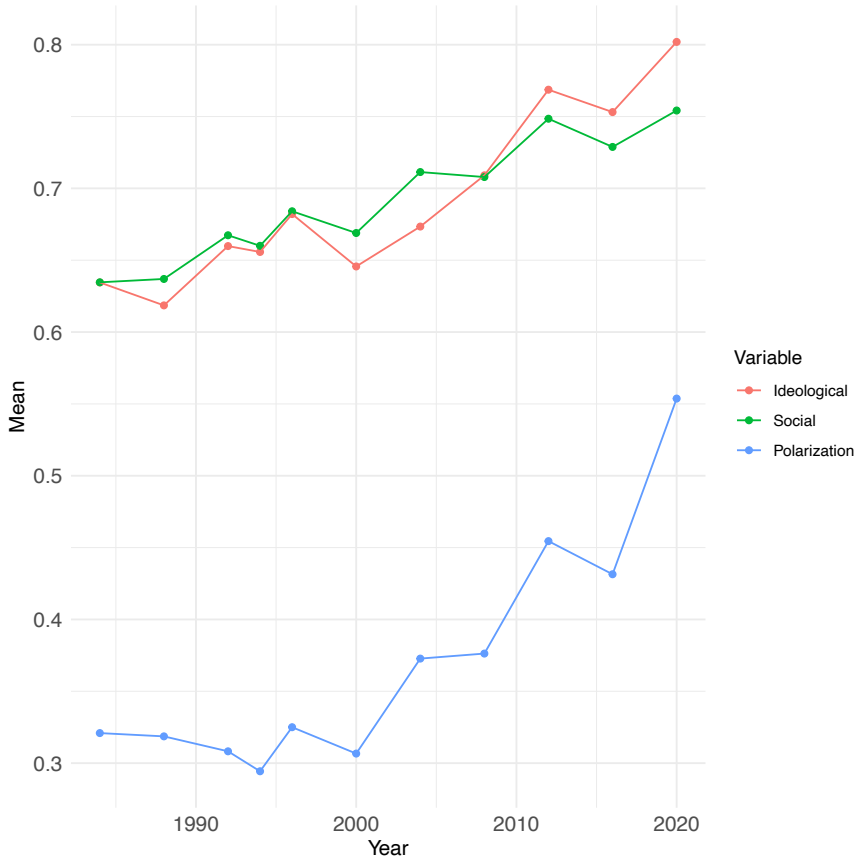
parties either in terms of ideological beliefs or social identities — it has become an electorate which is clearly divided on partisan lines, in terms of both beliefs and identity.

Taken together these results provide strong evidence in favor of HI1, and HS1. Both ideological and social sorting have increased from the period 1984-2020. A final note on these results regards the final predictive capability of the model — by the 2020 election, the machine learning algorithm was almost able to categorize the survey respondents as either Democratic or Republican with a high degree of accuracy, indicating extremely high levels of sorting along both ideological and social lines. It indicates that as time has progressed, the parties have become more homogeneous in terms of the clusters of social identities or ideological beliefs possessed by their supporters, such that while individual identities or beliefs may exist across the partisan divide, the ways in which these beliefs or identities interact with other beliefs or identities is now almost entirely different between the major parties in the US.

Results: Affective polarization

I now turn to an exploration of the effects of sorting on affective polarization. I first explore the bivariate relationships between the different kinds of sorting and levels of affective polarization at and aggregate level. Figure 2.4 shows the mean individual sorting score for both ideological and social sorting, and the mean individual level of affective polarization, for each survey year over the full time period. If a relationship exists, we would expect to see these levels of sorting and polarization display at least some correlation over time.

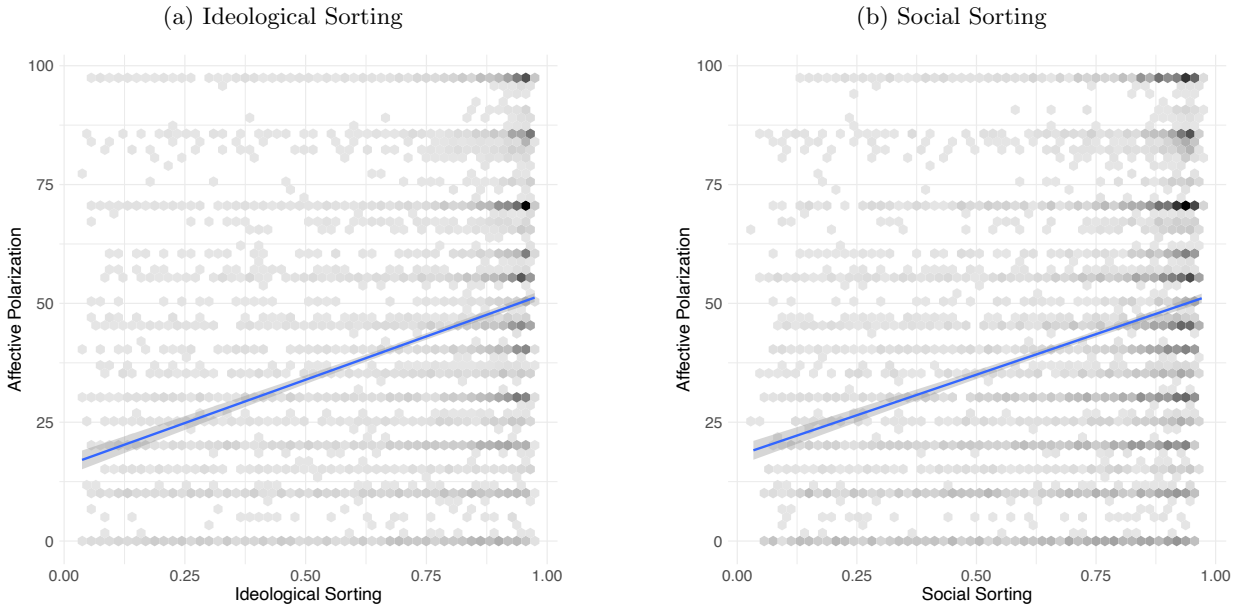
Figure 2.4: Over time change in the aggregate level of sorting and mean level of individual affective polarization



Each point represents the mean for the given variable (individual ideological sorting score, social sorting score, difference in thermometer ratings between the parties) for a given survey year

Looking at Figure 2.4, we see a strong correlation between social sorting (r = 0.89), ideological sorting r = 0.94), and affective polarization. This is consistent with the expectation that sorting is positively associated with affective polarization — the phenomena clearly increase and decrease in tandem over the period studied.

Figure 2.5: Evidence that ideological and social sorting positively predicts affective polarization.



Each hexagon represents a group of observations in an area of the plot, with darker shaded hexagons indicating a larger number of observations in that area.

At the individual level, I evaluate the basic bivariate relationship between individual sorting scores and individual thermometer ratings of the parties. Again, I would expect that those with higher levels of sorting (both ideological and social) should also demonstrate increased levels of affective polarization. Figure 2.5 displays these bivariate relationships, also seen in columns 1 and 2 of Table 2.1. Again we see evidence that a relationship exists between levels of sorting and affective polarization, as predicted by the affective polarization hypotheses. The correlation at the individual level between the social sorting score (r = 0.30) and ideological sorting score(r = 0.32) with individual levels of affective polarization remains positive and statistically significant, as we would expect.

In order to better isolate the relationship between affective polarization and sorting, I model individual levels of affective polarization as a function of both social and ideological sorting, and the individual's self-reported strength of partisanship (measured using the traditional 7-point party ID scale, folded in half), strength of their ideological identity (measured using the traditional 7-point Liberal-Conservative scale, again folded in half), and their self-reported levels of political interest[4]. This allows for easy comparison between the effects of operational ideological sorting (as measured by policy positions) and ideological social identity (as measured by self-reporting their ideological identity on the traditional 7-point scale, see Conover & Feldman 1981, Devine 2015, Mason 2018). A null finding on the effects of ideological *sorting* would imply that affective polarization is driven much more by identity-based issues than policy-based disagreement. It

[4]For precise question wordings, see supplemental materials

Table 2.1: Evidence that ideological and social sorting increase affective polarization at the individual level. N = 6,213

	1	2	3	4	5
Ideological Sorting	29.19**		17.19**		13.00**
	(2.95)		(2.11)		(1.79)
Social Sorting		29.01**		15.65**	10.60**
		(2.67)		(1.88)	(1.46)
Strength of PID			25.04**	24.45**	24.27**
			(1.39)	(1.37)	(1.35)
Strength if Ideological ID			11.73**	12.77**	11.20**
			(1.13)	(1.26)	(1.09)
Political Interest			8.25**	8.60**	7.87**
			(1.06)	(1.01)	(0.96)
Adjusted R^2.	0.169	0.168	0.348	0.345	0.354

* p < 0.05, ** p < 0.01
Note: All models include survey-year fixed effects, with standard errors clustered

also allows us to disentangle the potential effects of strength of partisan identity and the effects of social sorting. Though strength of party identity is potentially informed by social sorting, there is only a modest correlation in this dataset (r = 0.21). A positive coefficient on the social sorting score would indicate that social sorting increases affective polarization above what may be expected given an individual's professed strength of party identity.

Table 2.1 displays the results of five OLS regression models of affective polarization. This analysis is performed on individuals only within the testing set, to ensure the sorting scores are not biased by being part of the training set. Columns 1 and 2 display the bivariate relationships between sorting and affective polarization, and demonstrate that not only are both significant predictors of increased affective polarization, but that the explanatory power of these phenomena alone can account for a substantial amount of the variation in affective polarization, as demonstrated by the adjusted R^2. Column 3 displays a model of ideological sorting and affective polarization, in which we can see that ideological sorting is positively associated with affective polarization even after controlling for ideological *identity*. In other words, holding policy beliefs that are congruent with your party identity increases polarization in affective evaluations, even when holding the expressed ideological identity constant. That is, the fact that Democrats espouse liberal positions and Republicans espouse conservative ones may contribute more to the levels of animosity between the parties than the ideological labels attached to the parties and their supporters. Dislike driven by ideology is not merely an issue of identity, but of policy differences as well. This also accords with previous research

(Devine 2014) and comparative research on the subject of affective polarization, which consistently shows that operational ideological differences between the parties is a key driver in affective polarization (Adams et al. 2020).

Column 4 tells a similar story regarding social sorting. Here, we see that individuals whose identities better align with their party are more polarized in their affective evaluations, even when controlling for their self-reported strength of their party ID. In other words, identification with social groups associated with a political party increases warmth towards that party and/or dislike of the opposing party, no matter how strongly an individual identifies with the party itself.

Column 5 displays a model which includes both ideological and social sorting, allowing us to compare the effects of both kinds of sorting with other important drivers of polarization. We see that the ideological sorting appears to have a slightly larger effect on affective polarization than social sorting, though this difference is not statistically significant. The model also shows that sorting has a similar level of association with affective polarization as both strength of ideological social identity and political interest, two factors that have been shown previously to be associated with affective polarization.

To summarize, these results show strong support for HI2 and HS2. The new measure proposed and validated in this paper enables me to show that increased sorting, both ideological and social, is consistently associated with increasingly polarized party evaluations at both the individual and aggregate level, even when controlling for related constructs, and these relationship are of a sufficient magnitude to be worth further study. That ideological and social sorting have increased in tandem and hold relatively similar weight in relation to affective polarization, mean that any attempts to reduce the risks posed by affective polarization must tackle both the increasing ideological and identity-based divides in the US.

Extensions, Robustness Checks and future research

Given the lack of social identity measures available over time, this work has had to rely on feeling thermometers as measures of identity, an unideal situation. To ensure that the results presented here are not driven solely by the inclusion of the feeling thermometers (i.e. that the feeling thermometers represent more than identity and are thus biasing the results), I re-ran the analyses using two further models, the results of which are presented in Table B3 and Table B7 in Appendix **??**. The first, the "limited" model, excludes feeling thermometers related to individual groups (feminists and LGBT individuals), as these represent the most challenging measures. We see that their exclusion, while changing the absolute level of sorting depicted by the model, does not alter the pattern of increasing sorting over time, or the effect of sorting on affective polarization. The second model, excluding all feeling thermometers, does result in a slightly reduced increase

in sorting over time (roughly 10% rather than 13%), and a reduced size of the effect of sorting on affective polarization. However, both the increase in sorting and the effect on polarization remain substantial and significant, indicating that even when these measures of social identity are excluded, we can detect an increase in sorting, resulting in an increase in affective polarization.

Previous work on affective polarization, particularly by Adams et al. (2020) and others, has tended to focus on only levels of out-party dislike, rather than the overall difference between the feelings towards one's in-party compared to an out-party. While this paper uses primarily the difference-based measure of polarization for a number of reasons, it is also instructive to explore whether sorting affects feelings towards the in-party and out-party equally. Table B5 in Appendix **??** presents the results of models of feelings towards a respondent's in-party and out-party separately. I find that while ideological and social sorting are associated with warmer feelings towards a respondents' in-party, they have a larger association with negative feelings towards the respondents' out-party. These results are consistent with the theory that increasing affective polarization is primarily driven by out-party dislike and that sorting is a key driver of this increase.

I also explore whether these results differ by party. Table B6 displays models of in- and out-party evaluations among Democrats and Republicans respectively. I find that social and ideological sorting increase both in-party favoritism and out-party dislike among Democrats, but that increased sorting of both kinds are associated with only an increase in out-party animosity among Republicans. Further study is required to dissect why the effects of sorting differ between the parties, whether this is due to asymmetric polarization (Leonard et al. 2021), or the relationship between racial evaluations and party evaluations, particularly evaluations of the Democratic party (Hutchings & Valentino 2004), or some other cause.

This paper presents two important avenues for potential future research. The first is replication of this method, both within the US on other samples to validate the findings here, and beyond the US, to explore the extent to which sorting is occurring in other countries, and whether ideological and social sorting are driving changes in affective polarization in other polities. (Harteveld 2021) has already shown some promising evidence that social sorting has occurred in other contexts, and that this form of sorting is associated with changes in party evaluations. Applying this machine learning approach would enable future scholars to expand on these findings by increasing the number of social identities under study, and to compare ideological and social sorting across different contexts.

A second potential direction for future research is to use this approach to more closely examine which kinds of identities or beliefs are the most central to increasing sorting. Various tools exist to determine the most important features to any ML algorithm, and interpreting the importance of different identities and beliefs, or combinations of identities or beliefs, could shed valuable light on the forces that are driving sorting

within the US. For example, this kind of research could explore the relative impact of gender compared to race in how people develop and change their party identifications over time, and the relative weight each consideration has in driving affective polarization.

Conclusion

This paper presents evidence that over the period 1984-2020, both ideological and social sorting have increased within the American electorate. Further, this increase in sorting is associated with an increase in affective polarization between the two parties. Put simply, as individuals become better aligned with their party in terms of their political beliefs and their social identities, they become more likely to view their party favorably, and to view the opposing party in a more negative light. These findings have several important implications for American politics.

Firstly, the findings paint a stark picture of the two parties within the American system. The high degree of accuracy of these machine learning models in later survey years emphasizes the deep ideological and identity-based cleavages between the parties that now exist, such that clusters of each that appear in both parties are harder and harder to find. These findings emphasize the importance of interactions between different identities (or different beliefs), such that, while looking at aggregate sorting of social groups paints a polarizing picture, when taking into account interactions between the different social identities, we find an even more sorted electorate who see fewer and fewer people who are truly "like them" on the other side of the aisle. That is, while it is absolutely still possible to find black Republicans or white Democrats, this task becomes much harder when we move to the sub-group level, such that a younger, black, non-religious, feminist woman, may be far less likely to find similar individuals within both parties. This increasing sorting is likely part of a feedback loop; an increase in sorting further clarifies partisan stereotypes, party cues then become clearer, and sorting increases further.

Secondly, these findings present a fairly bleak prospect for the health of the American polity. While ideological polarization has long been touted as a major issue in the health of the nation's democracy, affective polarization may in fact be a graver threat. Were affective polarization based solely on policy differences, it could perhaps be somewhat easily mitigated through changes in elite behavior. However, the evidence that differences in social identity are equally as impactful in increasing affective polarization poses a much more challenging issue to solve. As the two parties have become more sorted in terms of identity over time, we may now be trapped in a problematic cycle, in which individuals view the two parties as representing two distinct camps of combinations of identities, choose to identify with the party that represents their camp, and thus continue to increase the extent to which the parties are distinct from each

other, sharing fewer and fewer similar groups of identities, resulting in a problematic feedback loop. This could in essence lead to a "balkanization" of the American political system, as the parties drift further from each other in terms of *whom* they represent, regardless of the policy views espoused, and this increased drift causes the two camps of partisans to view each other in increasingly negative light. Reducing ideological polarization requires changes in *what* individuals espouse in terms of policy, but reducing social polarization may require a further more challenging change: changes to *whom* the parties represent.

This paper also offers what I hope will be a useful application of machine learning methods to political science. While the tools of machine learning have increased in number and complexity over recent years, still political science has often found it challenging to put these tools to use in analyzing substantive questions. I hope the method suggested here, using the predictive tools of machine learning to evaluate change over time, and to determine the likelihood of individuals being classified according to certain characteristics, may serve as a useful method for other topics of interests, such as modelling the numbers of independents vs. partisans, swing voters, social ideological identification, and others.

However, the drawbacks of these machine learning methods also provide some useful starting points for future research on both sorting and affective polarization. While the work here establishes a strong association at both the aggregate and individual level between sorting and polarization, more work is required to demonstrate causality. Further, generating both aggregate and individual sorting scores is a useful tool, but does not deliver the kinds of detailed assessment of which kinds of combinations of identities are most well sorted within the parties. More work is also needed to determine how this sorting is perceived by the electorate — while substantial research exists on the ideological perceptions of the parties, still more is needed on the perceptions of group and social identities associated with each party, and the extent to which these perceptions shape the levels of identifications of individuals.

Chapter 3

The Affective Effects of the Gender Makeup of Party Support

Abstract

How does the gender composition of a party's supporter base impact the affective evaluations of the party? Findings in previous studies offer competing hypotheses which have not yet been addressed. This paper finds a larger gender gap between parties increases levels of affective polarization, and that while women feel more warmly towards parties that have a larger number of women supporters, men are relatively unmoved. These findings have significant implications for the causes of and potential solutions to affective polarization.

Introduction

One commonly accepted, though understudied, phenomenon among western political parties is the variation in their ability to attract support from different gender groups. Rarely do we see multiple parties within a system with a perfectly equal number of male and female supporters, or even a gender composition that perfectly reflects that of the electorate. One particularly under studied aspect of this variation, is the effect it has upon the perceptions of citizens. Given that the so-called gender gap in voting is often well established within a party system (Barnes et al. 2021, Abendschn & Steinmetz 2014), alters the way parties are presented in the media (Greene & Lhiste 2018) and is frequently discussed around elections (Gothreau 2021), we may expect that the gender composition of a particular party's supporter coalition will affect how others view the party. While other work (Adams et al 2022) has begun to tackle this question with regards

to the ideological perceptions of parties, there are many reasons that effects may exist beyond the effect on left-right placement of the party. Particularly attention is perhaps needed to the relationship between party support composition and a phenomenon about which there has been increasing concern in Western polities - affective polarization. The consequences of increased distrust, dislike, and animosity between parties are numerous, including economic discrimination (McConnell et al. 2018), interpersonal animosity (Iyengar et al. 2019), and even increasing dehumanization (Matherus et al. 2021) and a greater willingness to tolerate violence for political ends (Kalmoe & Mason 2018, 2022). Given the recent violent events such as the January 6 insurrection in the USA and the protests against the legitimate election of Lula in Brazil, further study of the causes and consequences of inter-party dislike is needed, yet relatively little exists regarding how the composition of party support impacts affective polarization. Here I explore whether the gender composition of party support affects the affective evaluations of parties - put simply, I ask whether people like (or dislike) parties more because of the relative gender balance among their supporters.

To explore this, I draw on the group theory of party identification, social sorting theory, and findings from previous studies on women's representation, to offer competing propositions for how the gender balance gap may impact affective evaluations. I ask: do individuals prefer parties whose supporter composition is similar to their own party in terms of gender? Or do they prefer parties with larger numbers of supporters that match their own gender? Alternatively, do they simply prefer parties that are able to attract a larger number of female supporters?

I find that individuals appear to feel more warmly towards parties that have similar gender makeup as their own party's supporters — that is, they prefer parties with a smaller gender gap to their own party — but also that women may prefer parties with larger proportions of women supporters, at least until the party reaches gender-parity. I show that these results are robust to a number of specifications, though the relationship between women's representation and affective evaluations only holds consistently among women. I also demonstrate that these effects are not a result of using the gender composition of the parties to make inferences about the ideological proximity of the party to either the respondent or the respondent's in-party, and instead must be driven by non-ideological factors.

In summary, I demonstrate that the gender composition of party support does have an impact on the evaluation of those parties, and that this area of research represents a first step in better understanding how the representation of different identity groups within a supporter base affects perceptions of a party. The magnitude of the effects found here is small, but given this comprises only one identity among many which may be expected to have similar effects, this marks an initial exploration of the relationship between the differences in party composition and affective evaluations. Opportunities for future research are discussed, as well as the implications for those concerned with affective polarization. These findings have potentially

conflicting normative implications for those concerned with affective polarization, but point to the notion that in order to improve their image, parties should aim to find a gender balance that is closest to those of other parties within their system and close to gender-parity, rather than necessarily increasing representation of one gender group over another.

Theorizing the Effects of Party Gender Composition

While there is extensive previous literature on the relationship between gender and political evaluations, the vast majority has focused on the relationship at the elite level. Research has shown that women are perceived as being less extreme (O'Brien 2015, 2019), more consensual (Bauer et al. 2017, Cassese and Holman 2017), and caring and compassionate (Bauer 2019). Media coverage of female candidates disproportionately associates them with "feminized" traits and issues (Kittilson & Fridkin 2008), while research has also shown that voters associate candidates of different genders with different issue areas and competencies (Huddy & Terkildsen 1997, Dolan 2010). Most pertinently, research has consistently shown that voters feel more favorably towards institutions and organizations when women are represented (Clayton, O'Brien, & Piscopo 2019, Badas & Stauffer 2022; Ben-Shitrit et al. 2021; Verge et al. 2020), perceive parties lead by women as being less extreme (O'Brien 2015, 2019), and, as Adams et al (2022) demonstrate, voters feel more positively towards political parties that have larger proportions of women MPs. These changes in perceptions may be driven by real differences between the political styles of men and women in elite politics, as women have been shown to be less adversarial (Childs 2004, Sones, Moran, & Lovenduski 2005, Grey 2002, Hargrave & Langengen, 2020), and more collaborative in parliament (Holman & Mahoney 2018). However, these studies have all focused on the presence and perception of women at the elite level. Little research has been undertaken into the role of gender among party supporters. In order to develop theoretical expectations about the role of gender among party supporters on party evaluations, I draw on group theories of party identification, and particularly, on social sorting theory.

The group theory of party identification treats a political party like any other kind of social identity (Tajfel & Turner 1979, Tajfel 1981), and argues that individuals identify with a party by comparing themselves to the 'party stereotype' that exists in their mind before determining if they 'fit' within that particular party (Green, Palmquist & Schickler 2002, Lupu 2014). Changes in the stereotype thus influence changes in party identification over time. Social sorting theory builds on this, and argues that individuals who more closely match the party stereotype - those whose identities have become 'sorted' within one particular party - form stronger attachments to the party, resulting more in polarized affective evaluations of out-parties, and stronger responses to threats to the party (Mason 2016, 2018, Harteveld 2021). This concept that

individuals have some latent measure of 'alignment' between their own identities and those associated with the political parties forms the basis of the hypotheses proposed here. I focus on only one identity, gender, but these theories could be extended and applied to numerous other identities that represent cleavages within party systems, such as race, ethnicity, religion, age, or education. While research on the effects of social sorting on affective evaluations is ample within the United States (Mason 2016, 2018 for example), it is limited in a comparative context. Harteveld (2021) has shown a promising approach to studying the relationship between sorting and affective polarization comparatively, but his focus on multiple identities and the interactions between them limits our ability to determine the effects of each identity individually, particularly when the effect of the sorting of a particular identity may not be simple, based on previous theory and research, such is the case with gender.

Gender is one of many social identities held by supporters that may impact the evaluations that individuals make of out-parties. Gender is the focus of this paper, not only because of the extensive literature on the effects of gender within politics, but also because it represents the most obvious and immediate test of the effects of social sorting on affective polarization, for several reasons.

Firstly, there is extensive evidence that gender represents a political cleavage in many polities around the world, with particular evidence that women tend to support more left-leaning parties (Barnes & Cassese 2017, Barnes et al. 2021, Abendschn & Steinmetz 2014). The existence of this cleavage allows us to develop specific expectations, based on social sorting theory, about how individuals may respond to different parties with different gender balances within their supporter base. Secondly, is among the most visible signifiers of the identity composition of a party supporter group (unlike for example, religious or regional identity), meaning that citizens can procure information about the gender composition of party supporters easily, whether by seeing images of party supporters in the media or by discussing politics in their day to day life. Indeed, research exists that shows that individuals are very able to accurately identify which kinds of social groups are most associated with different party supporters (Busby et al 2021), though they may be inaccurate in their assessment of the levels of representation of particular identities (Ahler & Sood 2018). In many countries, the gender gap in voting is frequently discussed around elections (Gothreau 2021) with pundits and the media referring to gender groups as voting blocs for one party or the other, making it a potentially salient cleavage that should be expected to impact affective evaluations, and one about which individuals may be expected to have more exact information.

Third, gender makes an effective test as it provides easy cross-national comparison. Unlike racial or ethnic groups who have experienced very different political and social conditions depending on the country, women, on the whole, have been traditionally marginalized in politics across the Western world, until recent decades. This means the effect of gender, unlike race or ethnicity, likely requires less concern about different

national interpretations of the identity, and allows for more effective cross-country comparisons. Finally, the previously discussed literature surrounding gender and politics at the elite level provides interesting competing expectations regarding the effect of the gender composition of party support on affective evaluations, making it a good candidate to test multiple competing theories.

Hypotheses

In applying social sorting theory to the role of gender among partisan supporters, there are two competing possibilities. Firstly, it may be that individuals' preferences regarding the gender composition of out-party supporters is based on the gender makeup of their in-party's supporters. Given that individuals identify with a party based on a satisfactory level of identity alignment, we may expect that they want to see similar ratios of identities in their own parties be present within other parties. That is, the gender composition of a party could act as a signal that the party is similar in terms of its identities to their own party, and so they may prefer the smallest possible 'gender gap' between the parties. That is, the gender composition of a party represents a kind of "identity marker" — a symbol of the groups that the party represents, and given that social sorting theory predicts that individuals feel more warmly towards parties that represent the same groups as their own party, they may feel more positively about parties that have similar gender compositions, and more negatively about those who stray too far from the gender composition of their own party. Alternatively, individuals may feel that the gender composition of one's in-party represents an "acceptable" level of gender balance for the individual, and that they should find other parties that have a similar gender makeup to be equally acceptable — that is, individuals believe that this kind of gender balance represents adequate gender representation, and so prefer parties that can match this, again preferring a smaller gender gap between the parties. This leads to my first hypothesis:

> ***Gender gap hypothesis***: All else being equal, partisans display warmer affect toward out-parties with a similar gender balance to their own party

A second alternative hypothesis which remains consistent with social sorting theory, is that individuals may feel more warmly towards parties with larger proportions of supporters who share the individual's gender. Rather than basing their evaluations on a comparison with their in-party, they may instead simply evaluate each identity individually, and feel more warmly towards parties that better reflect each identity. In this case, women would be expected to feel more warmly towards out-parties with a higher proportion of women, and men would feel more warmly towards a larger proportion of men. Given that these are mutually

exclusive (A larger proportion of men necessarily dictates a lower proportion of women), the reverse would also be expected to be true - individuals may feel less warmly towards those parties that have larger proportions of supporters of the opposite gender to the individual, an equivalent finding though framed differently. This provides hypothesis two:

> ***Gender identity hypothesis:***: All else being equal, partisans display warmer affect toward out- parties with a higher proportion of supporters who share their own gender

Finally, an alternative hypothesis for the way in which the gender composition of party supporters impacts affective evaluations is based on the findings among the representation of women at the elite level. Adams et al (2022) find that individuals feel more warmly towards out-parties who have higher proportions of women MPs. They present several arguments why this may be — the perception of women as more consensual or more moderate, the finding that individuals feel greater levels of trust and satisfaction with institutions that contain more women, and the evidence that women employ more consensual style of leadership. It is possible that these theories also hold at the level of party support, that individuals simply prefer parties with larger proportions of women supporters, either due to their consensual nature, perceptions of women as being more moderate, or some other predisposition towards the representation of women in politics. This leads to my third hypothesis:

> ***Women's bonus hypothesis:***: All else being equal, partisans display warmer affect toward out- parties with a higher proportion of women supporters

These hypotheses present three competing explanations for the effects of gender composition on party evaluations. Each is theoretically motivated by social sorting theory or previous literature on women's representation, but point in different directions regarding the expected findings. One important way in which the gender gap and the gender identity or women's affective bonus hypotheses differ, is that under the gender gap hypothesis, we would expect any kind of difference —even one in which the out-party represents a greater proportion of the individual's own gender – would be looked on unfavorably, as it marks a departure from the level of gender balance that the individual has decided is acceptable (the level they find within their own party). Under the other two hypotheses, the in-party's proportion of women is irrelevant, instead we would expect individuals simply to base their evaluations on the out-party's proportion of women supporters, regardless of the proportion within their own party, but these two differ in regards to their expectation of

evaluations among men. Under the gender identity hypothesis, we would expect increased representation of women among out-party supporters (and thus, decreased representation of men) to lead to more *negative* evaluations of the party, whereas under the affective bonus hypothesis, we would expect both women *and men* to feel more warmly towards out-parties with more women supporters. This work attempts to resolve these potentially conflicting expectations.

Which of these hypotheses do we expect to find the most consistent support? While each has strong theoretical motivations, the gender gap hypothesis is perhaps the most compelling. Given the extensive evidence of the effects of social sorting in previous work, and the salience of gender as an identity within politics, it makes logical sense that we would see the same kinds of effects seen elsewhere in this case. The gender identity hypothesis also relies on social sorting theory, but is perhaps hampered by the imprecision with which we can estimate how much representation of a particular identity an individual may want to see. That is, while individuals may want to see their own gender represented to a greater extent among out-party supporters, does that mean we would expect individuals to prefer parties composed of *only* supporters of their own gender? Would we truly expect to see men favor parties of 100% men, and women favor parties comprised of 100% women? Likely not. Rather, we may expect them to want *at least some base level* of representation of their own gender, but only up to a certain point, after which the party may seem wildly unrepresentative of the electorate, and so may be perceived as failing some diversity goal that the individual may have. Therefore, among the two social sorting hypotheses, we may expect the greater support for the gender gap hypothesis, which provides for representation of both genders in out-party support, but with individuals having some preference that (we assume) is satisfied by their own party, therefore may favor out-parties who also satisfy this gender composition preference.

As for the women's affective bonus hypothesis, there is also potential for caution. This hypothesis, proven already at the elite level by Adams et al (2022), is based heavily on the literature on women in politics at the *elite* level - that individuals favor women's representation because of how women leaders and members of parliament behave. Here, I extend this analysis to the mass level, exploring if individuals favor women's representation at the *supporter* level, but it is not clear if the reasons behind individuals' preference for women members of parliament — namely, that they are perceived as more moderate, more consensual, and more collaborative — apply equally to supporters. That is, the mechanism through which the effect of women's representation at the supporter level may have on affective evaluations, is perhaps less clear here than it is at the elite level. Therefore, the a priori expectations of this paper are that the gender gap hypothesis is the most likely to be consistently supported.

Data and Methods

To test these hypotheses, I use survey data from the Comparative Study of Electoral Systems (CSES) for 25 Western democracies between 1996 and 2020. Tables C1 through C4 list the countries, election years, and the 159 parties studied. I operate at the individual level of observation, using this survey data to compile a dataset in which each observation represents a respondent's evaluation of a single political party in a given survey year, excluding the evaluations of the respondent's own party. I include only those parties for whom full data on all variables (including variables only used in the robustness checks, such as the Comparative Manifesto Project's RILE measure) are available, and exclude parties with fewer than 5 seats at the time of the election survey, to ensure that results are not biased by the inclusion of extremely small parties.[1]. This yields a dataset of 91,166 respondents, who provide 353,101 party evaluations, at an average of 3.87 party evaluations per respondent.

Respondents were classified as belonging to a given party based on their response to the question "Do you usually think of yourself as close to any particular party? If so, which one? Respondents who said no were asked "Do you feel yourself a little closer to one of the political parties than the others?" Individuals are considered as party supporters if they feel both close to one particular party, or a little closer to one party than others. The analysis is limited only to those who identified as partisan in some way, as I have no expectations about how non-partisans would compare the gender of their in-group to the other parties.

The dependent variable for this analysis is the affective evaluation of the out-party, measured as the thermometer score that the individual gave to the out-party in question, on a scale from 0 (indicating intense dislike) to 10 (indicating the maximum level of positive affect) [2]. [3]

The primary independent variables for this analysis are estimates regarding the gender composition of the parties, based on the gender of respondents to the CSES surveys. An estimate was made of the gender composition of each party based on the respondents who identified with that party in that particular wave of the CSES, and weighted using the existing demographic weights[4]. Those individuals who identified as non-binary, other, or refused to respond, were excluded. This value, the percentage of women supporters of a given party, provided the primary independent variable necessary to test the gender identity hypothesis and

[1]To ensure the results do not change substantially based on the parties included, I also conduct analysis on all parties for which data for the primary independent variables are available in Table C5, with the results unchanged aside from slightly larger parameter estimates

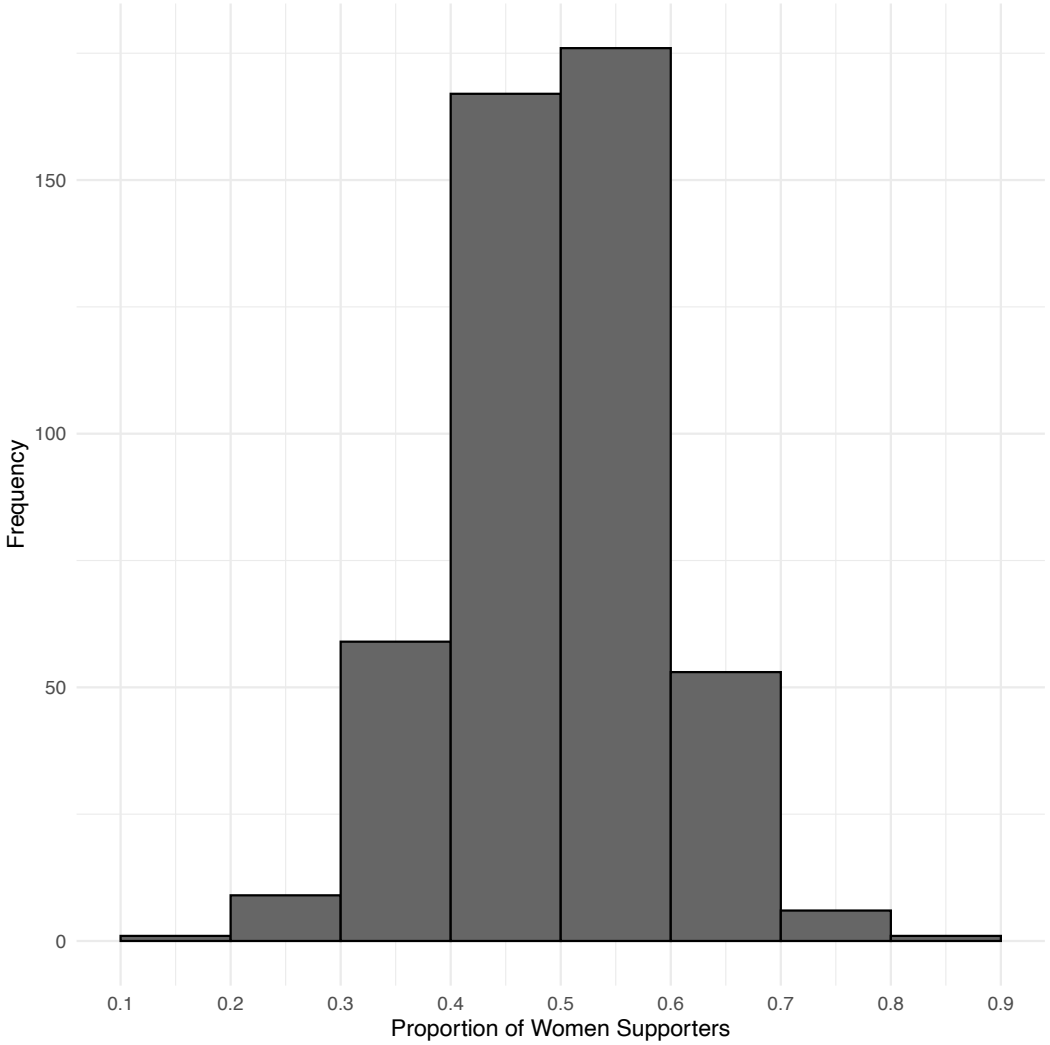
[2]The precise question wording is: "I'd like to know what you think about each of our political parties. After I read the name of a political party, please rate it on a scale from 0 to 10, where 0 means you strongly dislike that party and 10 means that you strongly like that party.

[3]As a robustness check, I also conducted analysis using the thermometer differential used in many other works (e.g. Druckman & Levendusky 2019) between the respondent's in-party and the out-party in question, to control for any baseline differences in thermometer ratings. This analysis is found in the appendix in Table C13, and the use of the differential measure does not alter the substantive findings.

[4]As a robustness check, the analysis was also conducted with the unweighted data, with no significant change to the results

the women's affective bonus hypothesis. The distribution of women supporters can be seen in Figure 3.1, where we see that the majority of the data is clustered around 50% as we would expect, but with more than 25% of the observations being parties with greater than 60% women supporters or less than 40% women supporters, thus providing sufficient variation to determine the effect of out-party women supporters on affective evaluations.

Figure 3.1: The proportion of women supporters in each party constituency in the dataset

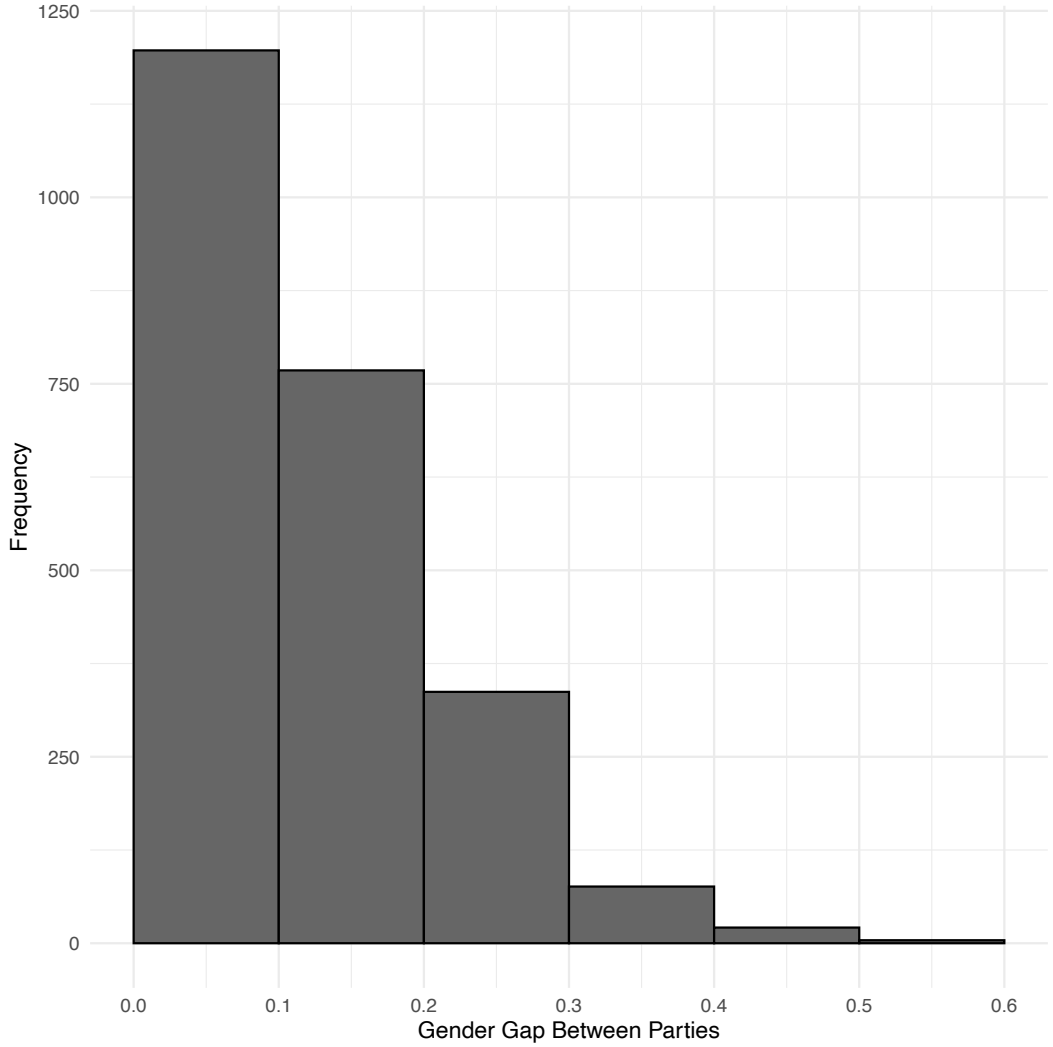


To test the gender gap hypothesis, I computed the absolute difference between the percentage of women supporters for each pair of parties for whom at least one respondent of the in-party evaluated the out-party [5]. Pairs of parties with larger values represent parties with substantially different proportions of men and women supporters. The absolute value was used, as the gender gap hypothesis is not directional — I do not

[5]As a further robustness check, I also conduct the analysis on a dataset which excludes parties with 50 or fewer supporters in the data, in order to determine whether the results are driven by smaller parties having substantially skewed estimates of gender balance, due to lack of data. The results are reported Table C9, and do not differ from those reported in the main text

expect that individuals will feel warmer towards parties that have smaller or larger proportions of a certain gender *than their own party*, but rather, as discussed previously, they will favor parties that match whatever the gender makeup is of their own party. Figure 3.2 presents the distribution of the gender gap between each unique party dyad (a pair of parties in which a member of the in-party is evaluating the out-party) within the dataset. We find that half of the party dyads in the dataset have a greater than 10% gender gap, while roughly 18% have a gender gap greater than 20%, which again should provide sufficient variation to test these hypotheses. It is important to note that these larger differences are not solely clustered among smaller parties, with the mean gender gap among party dyads in which each party has more than 20 seats in the legislature being 9.3%, with a standard deviation of 7.3%.

Figure 3.2: The gender gap between each party constituency in the dataset



I control for other factors that have consistently been shown to impact the affective evaluations of

individuals. Previous work has shown that ideological proximity to a party increases affective evaluations. Respondents in the CSES were asked to place the parties being evaluated on a left-right scale, from 0 (left) to 10 (right). I use this data to control for perceived left-right distance between the out-party and the individual respondent[6]. *Perceived left-right distance* is computed as the absolute difference between the respondent's placements of themselves on the left-right scale and the out-party being evaluated. Given previous research, we would expect increased perceived left-right distance to result in decreased out-party evaluations.

I also control for whether the respondent's own party was in coalition or opposition with the out-party being evaluated. Previous scholarship (Horne et al. 2021) has shown that individuals feel more positively towards those parties with whom their own party has shared power or participated together in opposition.

Since error terms may correlate within elections, I use ordinary least squares regression (OLS) with robust standard errors clustered at the election level. The model also includes country-year fixed effects to capture any factors associated with specific countries and time periods such as economic conditions, or electoral laws, which may be unaccounted for in the model and may otherwise bias the results. Therefore, the parameter estimates from the model capture the within-country and within-election differences in individuals' ratings of the evaluated out-parties. As a further robustness check, I also model the data using individual-level fixed effects and clustering the standard errors at the individual level for each respondent in the dataset, shown in Table C6, and I also include a model using country-level fixed effects to capture the variation with country across different election years, with results shown in Table C14. Neither of these alternative specifications alter the substantive results presented here.

Results

Table 3.1 presents initial tests of the gender gap hypothesis and the women's affective bonus hypothesis. If individuals do indeed feel warmer towards parties that have a similar gender balance to their own, we would expect a negative coefficient on the "gender gap" variable — indicating that as the difference between the gender makeup of the in-party and out-party increases, their feelings towards the out-party become colder. This is supported by columns 1,2 and 3 in Table 3.1. We see a significant negative effect of the gender gap on out-party evaluations, demonstrating that respondents feel more negatively towards those parties whose gender makeup is substantially different to their own party. The magnitude of the effect from the full model in column 5 represents a 1.5 unit decrease in the 10-point thermometer rating of an out-party when moving from two parties that share the same gender makeup, to two parties in which each party represents

[6]As a robustness check, I run model specifications which include the perceived left-right distance between the two parties, rather than the distance between the respondent and party, and the RILE distance between the parties. These results are discussed in the robustness checks section of the paper.

Table 3.1: Effects of Supporter Gender Composition on Party Feeling Thermometer

	1	2	3	4	5
Gender Gap	−2.32**	−1.62**			−1.51**
	(0.82)	(0.39)			(0.43)
Out-Party % Women Supporters			1.30**	1.39**	1.30**
			(0.44)	(0.34)	(0.36)
Perceived right-left distance		−0.47**		−0.47**	−0.46**
		(0.02)		(0.02)	(0.02)
Parties were in coalition		0.55**		0.60**	0.57**
		(0.13)		(0.12)	(0.13)
Parties were in opposition		0.36**		0.36**	0.38**
		(0.08)		(0.08)	(0.08)
Observations	353.101	353.101	353.101	353.101	353.101
Adjusted R^2	0.104	0.303	0.102	0.303	0.305

* p < 0.05, ** p < 0.01

a different gender group. This case is of course unlikely to occur, but a gender gap between parties of two standard deviations represents a 0.24 unit decrease in thermometer scores - about half the size of the effect of two parties being in coalition together, a small, but consistent and non-negligible effect.

However, this effect also appears to exist alongside the effect of increased affective evaluations of parties with more female supporters. Columns 3, 4, and 5 provide a test of the women's affective bonus hypothesis, and show that partisans do evaluate out-parties who have more women supporters more positively, all else equal. The effect size here indicates the change in thermometer scores of an out-party when moving from 0% women supporters to 100% women supporters. To give a more common example, a two standard deviation increase in out-party women supporters corresponds to an increase in thermometer ratings of 0.25 units, almost identical to that of the gender gap effect. This raises several further questions. Firstly, whether this effect is consistent among men and women — thereby providing full support for the women's bonus hypothesis and invalidating the gender identity hypothesis — and secondly, how do these two effects function together in affecting thermometer scores. Table 3.2 provides a test of the first question, while Figure 3.3 explores the second.

If the gender identity hypothesis is true, we would expect to see a strong positive effect of out-party women supporter percentage among women respondents, and a strong negative effect among male supporters. However, if the women's affective bonus hypothesis is correct, we would expect a positive effect among both men and women. Table 3.2 provides our answer - among men, there is a small but positive effect of increasing numbers of women supporters on out-party evaluations. While the effect is substantially smaller than that among women, it is still positive and significant, providing no support for the gender identity hypothesis, and stronger support for the women's affective bonus. Among women, an increase in out-party women supporters

Table 3.2: Effects of Supporter Gender Composition on Party Feeling Thermometer by Respondent Gender

	Women	Men
Gender Gap	−1.45**	−1.47**
	(0.44)	(0.42)
Out-Party % Women Supporters	1.92**	0.82*
	(0.39)	(0.34)
Perceived right-left distance	−0.46**	−0.47**
	(0.02)	(0.02)
Parties were in coalition	0.54**	0.59**
	(0.14)	(0.12)
Parties were in opposition	0.41**	0.36**
	(0.09)	(0.09)
Observations	165.840	187.261
Adjusted R^2	0.311	0.302

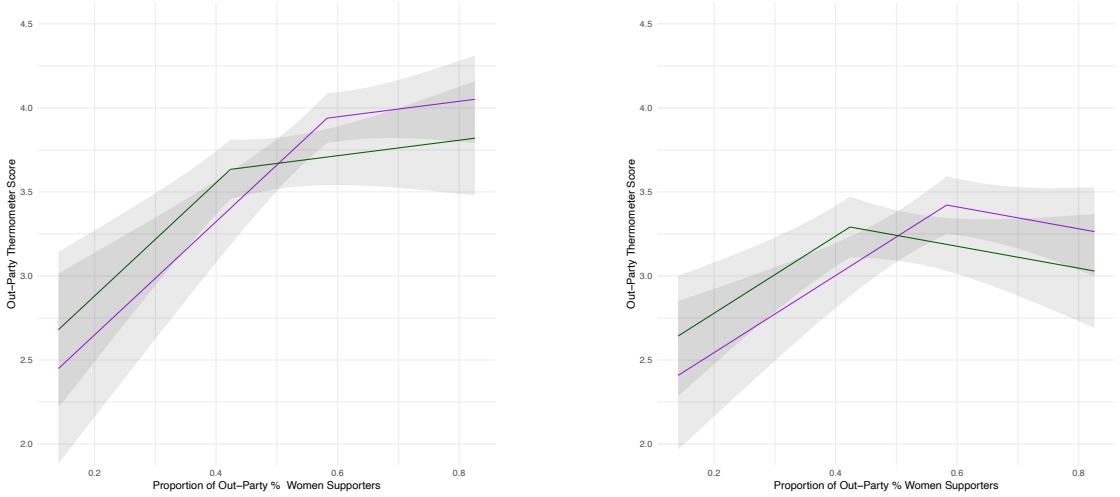
* p < 0.05, ** p < 0.01

of two standard deviations represents an increase of 0.37 units in thermometer scores, whereas among men, this increases by only 0.16 units. Put simply, women are more strongly affected by increases in women's representation than men, but both appear to feel more positive about parties with larger shares of women supporters. This is perhaps an unsurprising finding, research has already shown differences between men and women, in terms of both their absolute levels of affective polarization and what causes it, with women being more impacted by differences in political identity (Ondercin & Lizotte 2020). However, this finding extends this research to show that gender differences also exist in the effects that gender has on affective polarization.

There is, of course, an inherent contradiction between these findings. In many cases, an out-party increasing its share of supporters who are women is also likely to increase the difference in the gender balance of its supporters with other parties — i.e. while increasing the number of women supporters may have a positive effect, it brings with it the negative effect of the gender gap. Not only this, but Table 3.2 shows that while men and women are roughly equally affected by the gender gap, the effect of increasing the percentage of out-party women supporters among women is roughly double that among men.

Figure 3.3 graphically displays the relationship between respondent gender, the number of out-party women supporters, and the gender balance gap between in- and out-parties. Here we see how respondent gender and the gender gap condition the effect of increasing numbers of women out-party supporters. Among both men and women, as the number of out-party women supporters increases towards parity with the respondent's in-party, voters feel more positively towards the out-party. However, once the gender balance between the two parties reaches parity, the similarity ends. Among women, respondents feel more positively towards out-parties with larger numbers of women supporters even beyond parity, but the increase in affective

Figure 3.3: The effect of out-party support gender composition on feeling thermometer ratings



(a) Female respondents

(b) Male respondents

Each line represents the predicted out-party thermometer ratings, based on that party's proportion of women supporters. The green line indicates the predicted evaluations of respondents who belong to a party with a proportion of women supporters that is one standard deviation below the mean (Green), while the purple line shows the predicted evaluations of respondents who belong to a party with a proportion of women supporters that is one standard deviation above the mean, thus providing different levels of the gender gap

evaluations becomes much smaller once that out-party surpasses the percentage of women supporters in the respondent's own party. Women do appear to want more women supporters in out-parties, but are mostly satisfied once the out-party matches the gender balance of their own party. Among men however, once an out-party has the same gender makeup as the male respondent's in-party, any increase in women out-party supporters results in an insignificant change in affective evaluations at best, and *decreasing* affective evaluations at worst.

These findings do not provide full throated support for either the gender identity or the affective bonus hypothesis. It appears that women are indeed attracted to parties with higher levels of women supporters, even above that present within their in-party, but among men, the effect of increasing the proportion of women supporters is highly contingent on the level of support within their in-party. Yet, it is also not clear that men feel more *negatively* towards more women-dominated parties (or indeed, feel more warmly towards more male dominated ones), and thus provides no real support for the gender identity hypothesis, and certainly none for the affective bonus hypothesis. Instead, what we find is strong support for the gender gap hypothesis - both men and women feel more warmly towards parties that match the gender representation of their own party. While women may feel more warmly towards parties that exceed the level of women's representation of their in-party, men are focused only on matching the representation of their chosen parties.

Robustness Checks

To ensure the findings did not represent an anomaly driven by the modelling approach used here, several robustness checks were conducted, the full results of which can be found in the Appendix. Table C9 limits the analysis to only those parties for whom 50 or more respondents were included in the dataset, in order to limit the possibility that the results are driven primarily by extreme cases of proportions of women supporters that may be present in smaller parties. The results are unchanged. Table C6 presents the results when using individual fixed effects, rather than survey-year effects, and models in which the standard errors are clustered at the individual level. Again, the primary results remain unchanged.

Given previous results that have shown that individuals have stronger feelings towards radical right wing parties (Fuller at al. 2022, Helbling & Junkunz 2020; Reiljan & Ryan 2021, Weeks et al. 2021), I ran further analysis to explore whether the effects here were driven primarily by feelings towards these parties. Table C7 limits the analysis by excluding parties who had a RILE score of more than two standard deviations to the right of the mean party in the dataset, with no changes to the results. The results presented in Table C8 exclude all parties considered "Nationalist" by the Comparative Manifesto Project according to their party family classification. Here, we see that while the effects of the gender gap remain consistent, the effects of increasing women's representation among out-party supporters loses its significance, and in fact among men, the direction of the effect is reversed. This poses the question as to whether it is only nationalist parties (as classified by the CMP) who could gain an affective bonus by attracting more women supporters. This deserves further study, though it should be noted that some parties classed as nationalist would not be classed as radical right wing under most other criteria (for example, the Scottish Nationalist Party or Plaid Cymru in the United Kingdom), making it more challenging to determine exactly what characteristic of these parties may result in the loss of the effect of women's representation among the remaining parties. The effects of the gender gap remain significant and directionally consistent.

One robustness check which yielded somewhat mixed results is presented in Table C10. This analysis uses the Comparative Manifesto Profject's measure of the left-right placement of party manifestos, to measure the distance between the respondents' in- and out-parties. Here, the effect of the gender gap becomes significant only at the level of $p < 0.1$, while the effect of increases in out-party women supporters is no longer significant among men, though the effect among women remains as expected. This somewhat mirrors the effects of excluding radical right or nationalist parties — in both cases, the effect of increasing the proportion of women supporters above that of the respondent's in-party no longer has an effect on the evaluations of male voters.

One potential explanation for these results is that individuals use the cues provided by the gender composition of a party's supporters to make inferences about their ideology. Given that prior research and

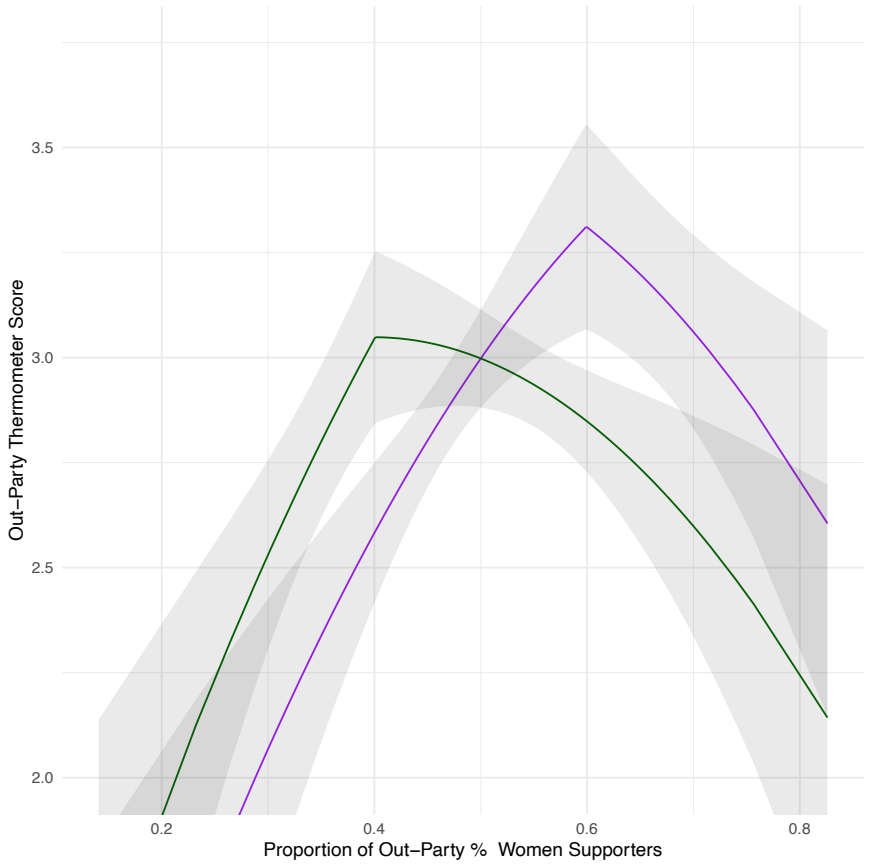
the previous results which demonstrate that ideological distance between the individual and the out-party results in increasingly negative evaluations of the out-party, any effects that *reduce* affective polarization may act by reducing ideological distance. In Table **??** I explore whether the proportion of out-party women supporters, or the gender gap, alter the perceived distance between the respondent and the out-party, or between the respondent's own party and the out-party. In both cases, the answer is negative - while gender composition of party support does influence affective evaluations, it does not do so through the mechanism of ideological proximity.

One alternative explanation for the effects of the gender gap may be that respondents do not necessarily favor parties that match the gender makeup of their own party, but instead favor parties that have a gender balance that approaches parity — an equal proportion of men and women. Verge et al. (2020) show that citizens in Spain and Portugal report higher levels of stratification with institutions that maintain some level of gender balance, and it is plausible this could also extend to political parties. Given that most larger parties have a gender composition that approaches some kind of gender parity, it is easily possible to conflate the effects of an increased gender gap with the effect of an out-party moving away from the desired 50-50 men-women split. In effect, in most cases where the gender gap is low (particularly among larger parties that comprise the majority of respondents in the dataset), the gender balance in each party is generally also closer to 50-50. Essentially the challenge here is to understand the inflection point - the point at which an increase in out-party women supporters no longer increases affective evaluations. Is it when the out-party approaches a 50-50 balance between men and women, or is it when the out-party approaches the same gender composition as the in-party?

In order to explore this possibility, I conducted analysis on a set of cases in which the gender gap between the two parties was larger than 10%. This provides a set of cases for party dyads in which an out-party consisting of 50% women represents a substantial difference from the in-party, while meeting the needs of an individual who favors gender balance. In essence, subsetting the data in this way limits us to the cases in which there is substantial variation between the gender gap and gender-parity. This is an inelegant solution to this challenge, but preferable given the data available. I include a new variable in the analysis *the proportion of out-party women squared*, to explore if the effect of increasing the proportion of out-party women significantly diminishes once the party reaches a 50-50 parity. If this alternative gender parity hypothesis is correct, I would expect to see the effect of the gender gap reduce or disappear as individuals only look for an out-party to achieve gender parity, not to match their in-party, and thus the inflection point should be located at 50% women for all parties, with a large and significantly positive coefficient on the squared term. However, if the gender gap hypothesis is true, we should see the inflection point located at the percentage of women supporters in the in-party, and that any increase above this rate, even if it means

the out-party better approaches 50%, should decrease rather than increase out-party affective evaluations. Table C15 in the appendix, and Figure 3.4 presented here display the results.

Figure 3.4: The gender gap between each party constituency in the dataset



Each line represents the predicted out-party thermometer ratings, based on that party's proportion of women supporters. The green line indicates the predicted evaluations of respondents who belong to a party with a proportion of women supporters that is one standard deviation below the mean (Green), while the purple line shows the predicted evaluations of respondents who belong to a party with a proportion of women supporters that is one standard deviation above the mean, thus providing different levels of the gender gap

The results clearly show that the gender gap remains a significant predictor of affective evaluations, even when controlling for the squared effect of out-party women supporters. We find that individuals do indeed have some preference towards a 50-50 balance in gender composition of party support (the effect of out-party proportion of women supporters is positive and significant, while the coefficient on the squared term is negative and significant) but that this is secondary to their preference for a match between the in-party and out-party gender composition. We find that the effect of increasing out-party women supporters drops more rapidly after the out-party has achieved a 50-50 balance, but that this drop *begins* when the out-party has matched the composition of the in-party. That is, under this analysis, we see that differences between men and women shown previously begin to disappear here, and among both men and women, the

effect of increased women's representation becomes negative once the the gender gap between the parties has been eliminated, but only becomes sharply negative once the proportion of out-party women supporters has substantially exceeded gender parity. The effect of increasing women's representation among a party's supporter base on affective evaluations here is contingent not only on the the level of representation among other parties, but also on whether the party has already achieved gender parity.

To be clear, this analysis is conducted under the unideal scenario of focussing on a more extreme subset of cases. The results here are indicative, but by no means conclusive, and so further work — potentially involving experimental analysis to provide a wider range of potential scenarios than is available here – is required. However, from this analysis it appears that the gender gap between the in and out-party is a more important predictor of out-party affective evaluations than the extent to which the out-party reaches gender parity.

Taking these robustness checks together, it is clear that the gender composition of party supporters does play a role in the affective evaluations of other partisans. Increasing the gender gap between two parties is consistently associated with more negative evaluations of an out-party in all specifications of the model. Increasing the proportion of women out-party supporters is consistently associated with more positive evaluations of an out-party among women, with limited support among men, and potentially less support once the out-party achieves gender parity. Given the data available here, it is not possible to conclusively determine whether the effect of the gender gap is driven by the differences between the parties or the desire for out-parties to reach some kind of gender parity, but the results indicate that it is the former.

Conclusion and Future Research

The results here paint a complex picture of the effects of the gender composition of out-party supporters on affective evaluations. We see a consistent effect that individuals feel more positively about out-parties whose supporter base resembles the individual's own party in terms of gender. Yet we also see a positive effect of the proportion of women in an out-party's supporter constituency among women supporters. While these effects are small in magnitude, they point to a greater truth, that the identity makeup of partisan supporter bases does have substantial impacts on the affective evaluations made by citizens. The results here provide support for the predictions driven by social sorting theory, that individuals prefer out-parties that look like their own party. However, the picture is more complex than that. While the gender gap is indeed a polarizing force, the simple increased presence of women in an out-party supporter constituency can have a moderating effect on this, at least among women, and the extent to which a party has achieved gender-parity can also limit these effects. While there is no simple way for parties to seek to appeal to

out-parties by managing the gender composition, it appears that one tactic that may yield the greatest benefit, is to try and maintain a supporter gender balance that hews closely to those of other parties within a system, and does not dramatically differ from a 50-50 gender split, in order to avoid affective punishments for a gender gap or from exceeding gender parity.

This paper has benefited from the relative ease with which gender can be compared cross-nationally and over time in order to make claims that can now be analyzed in more detail on a polity-by-polity basis. Further questions that will need to be answered include how visible is the gender gap, and does the respondent's knowledge of the gap mediate the effect found here? Do voters overestimate this gap, and if so, does correcting them reduce the polarizing effects on evaluations? How is information about this gap conveyed to citizens, and does the manner of conveyance matter?

Given the effects found here of just one identity, one which varies in its political salience between countries and over time, it begs the question of how great the effects of multiple kinds of identities may be on affective evaluations, particularly when considering the effect of interactions between them. Some work (Harteveld 2021) has explored social sorting (along the lines of education, income, religion, and region) in comparative context and found significant results, but we do not yet have good estimates of the impact of each of the various identities at play, or the ways in which different interactions between different identities may impact affective evaluations. If individuals form more negative evaluations of a party simply based on the gender gap, how much more negative do they feel towards a party when this is combined with a racial gap, a religious gap, an age gap, an education gap, an income gap, a regional gap, and a class gap. This work will hopefully spur further study on these kinds of effects.

Furthermore, these findings pose questions as to whether there are other less polarizing effects of representation within party supporter constituencies. Given the evidence presented here that the presence of women in a party's supporter base may have some positive effect on affective evaluations, it begs the question as to whether the presence of other traditionally politically marginalized groups has an equal effect. Do voters favor increased presence of racial, ethnic, religious, or other minorities in out-party supporter bases? It is very likely these kinds of effects will vary by the political context of each identity group within a given polity. Even the results here regarding gender may not generalize to all cases - Klar (2018) finds that emphasizing gender can actually increase partisan hostility between women of different parties if they share a dramatically different conception of the role of women. Clearly, more study is needed to explore exactly what kinds of groups voters want to see in the supporter composition of out-parties.

Affective polarization continues to be a prominent topic of discussion within Western democracies, particularly in the wake of events such as the January 6th insurrection in the United States, and the more recent political protests in Brazil. This paper offers another set of factor that scholars should consider when

studying affective polarization; how the identity groups present within the supporter bases of parties relate to each other, how the differences in supporter composition impacts party evaluations, and what kinds of groups voters feel more or less positively about. The findings here could be interpreted in many ways. One could focus on the polarizing effects of the gender gap and lament the effects that increases in this gap may bring in different contexts. Alternatively, one could look at the slow and steady increase in women's participation within politics across the Western world, and political parties increased incentives to cater to women in order to gain their support, as a potential force to reduce affective polarization. Given that research has already shown that radical right parties are using gender as a tool to shift perceptions of their policy platforms (Ben-Shitrit et al. 2021), the findings here could easily serve parties in trying to alter their images in the eyes of voters. I argue that the most consistent conclusion found here gender gap remains a potent signifier of the identity gaps between parties, and retains the potential to further increase affective polarization among those parties who explicitly cater to one gender group over another, but that given the right incentives, it is possible for parties to wish to seek a gender-balance among their supporters that would not only be more representative of their citizens, but also reduce the differences between parties, and lessen partisan tensions.

Appendices

A Chapter 1 Appendix

Study 1 & 2: Demographic / Identity Measures:

What is your age?

18-30
31-40
41-50
51-60
61-70
71 years or older

Do you identify as...?

Male
Female
Other / Non-Binary
Prefer not to say

Would you describe yourself as religious or non-religious?

Religious
Non-religious
Prefer not to say

While people may have multiple races/ethnicities they identify with, which one of the following do you most identify with?

Black of African-American
Hispanic or Latino
Asian or Asian-American
Native American
Middle Eastern
Mixed Race
Other

Generally speaking, do you consider yourself a Democrat, Republican, or Independent?

Democrat
Republican
Independent

Other/Don't know

Do you think of yourself as closer to the Republican Party or the Democratic Party?
Democratic party
Republican Party
Other / Don't know

Study 1: Party Identity Scale:

Earlier, you said you identified (Party). Thinking about (Party), would you call yourself a strong (Partisan) or a not very strong (Partisan)?

How important is being (Partisan) to you?

How well does the term (Partisan) describe you?

When talking about (Party), how often do you use "we" instead of "they"?

To what extent do you think of yourself as being a (Partisan)?
Thinking about the (Party), how much do you agree with these statements? (7-point scale, Strongly Agree to Strongly Disagree)

I am interested in what other people think about this party.
I have a lot in common with other supporters of the (Party).
When I meet someone who supports this party, I feel connected to this person.
When people praise this party, it makes me feel good.

Study 1: Identity Manipulation Check Scale:

You are now going to be asked a few questions about your views of a political party or group in the U.S. You will be given a party or group, and asked to respond with your perceptions of that party or group.

Thinking about (Partisans), do they tend to be

Black or African American
White
Hispanic or Latino
Asian
Other

Thinking about (Partisans), do they tend to be aged...

18-30
31-40
41-50
51-60
61-70
71 years or older

Again thinking about (Partisans), do they tend to be...

Religious
Non-religious

Finally, thinking about (Partisans) do they tend to be...

Male
Female

Study 1: Recall Question wording:

You will now be asked a few short questions about the article you read, to measure how much you remember about the article, and your reactions to it.

In the article you read, at which university was the study conducted?

UC Berkeley
Stanford University
Harvard University
The University of Michigan
Don't know

In which journal will the study be published?

Journal of Politics
American Political Science Review
American Journal of Political Science
Political Studies Journal
Don't know

Which group is the focus of the study?

The Democratic Party
The Republican Party
The Green Party
Independents
Don't know

Roughly how many individuals were contacted as part of the study?

1,000
5,000
10,000
100,000
Don't know

Study 1: Policy Cue Question wording:

Right-to-work laws are pieces of legislation that guarantee that no employee can be required to join, or not join, a union or be required to pay dues to a labor union as a condition of employment. Do you support or oppose right-to-work laws?

Recently, California passed a ballot measure raising income taxes on Californians who earn more than $250,000 per year for seven years. Specifically, the measure raised income taxes by 1% on income between $250,000 and $300,000, 2% on income between $300,000 and $500,000, and 3% on income above $500,000. Do you support or oppose these changes in income tax policy?

Do you support or oppose increasing the federal minimum wage from $7.25 an hour to over $12 an hour?

Do you support or oppose allowing universities to increase the number of minority students studying at their schools in order to increase diversity, by considering race along with other factors when choosing students?

Do you support or oppose the Affordable Care Act (often called Obamacare) passed in 2010? This law requires all Americans to buy health insurance and requires health insurance companies to accept everyone

Cue condition includes: "The Democratic/Republican Party opposes these policies, while the Republican/Democratic Party supports them."

Study 2: Evaluations Question wording:

To what extent would you support (Insert Organization) opening up a UC Davis chapter in the new future?

If (Insert Organization) were to open up a UC Davis chapter, how likely do you believe you would be to join this organization?

If (Insert Organization) were to open up a UC Davis chapter, how likely do you believe you would be to donate money to this organization?

To what extent do you agree with the following statement: "UC Davis should provide financial and operational support to (Insert Organization) to open up a chapter on the UC Davis campus"

Thinking about (Insert Organization), how much do you agree with these statements?

I am interested in what other students think about this group.
I support the mission of this group
I feel I would have a lot in common with supporters of this group

Study 2: Ideological Placement Question wording:

In politics people sometimes talk of left and right. Where would you place yourself on the following scale? — 1 (Left) - 10 (Right)

Where would you place the groups you read about earlier on the same left-right scale? — 1 (Left) - 10 (Right)

Figure A1: Sample treatment text for Study 1

Stanford | News

Home Find Stories For Journalists Contact

New study sheds light on who supports the Democratic Party

Nationwide study of over 100,000 Democrats may change our understanding of the party's supporters.

By Natalie Broad

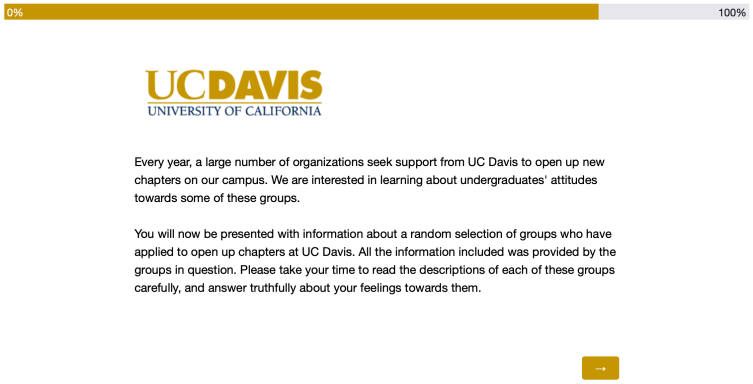
A new study from Stanford University promises to reveal the most in depth look at supporters of the Democratic Party in history. As part of the study, undertaken by Professors Stan Englehoff and Katie Dillman, telephone interviews were conducted with over 100,000 Democratic Party supporters in all fifty states. The study promises to share some intriguing discoveries about the characteristics of supporters of the Democratic Party. For example, men comprise a larger share of the party's supporters than the researchers had anticipated.

The authors believe they now have a more complete understanding of the demographics of Democratic Party support - including the age, race, gender, and religion of the average Democrat - than at any time in history. Some of the top line findings the authors have shared include that Democrats tended to be *more* likely to identify as White than previous studies had suggested, while the average Democratic Party supporter was more likely to fall into the 31-40 age category than had been expected.

Professor Englehoff commented "It's a really interesting dataset, we've never had such a deep well of information to work from before, and we're incredibly excited to share our findings with the world." The study promises to tackle some of the most important issues in American politics, such as the policy views of the party's supporters, their views on the President and the recent elections, and their perspective on the role of religion in American life. The authors have already shared that Democrats showcased lower levels of religiosity than the authors expected, with more respondents identifying as non-religious.

The findings of the study will be published next year in the *American Journal of Political Science*.

Figure A2: Sample treatment text for Study 2



A Voice for All

A Voice for All (AVA) is an international organization which campaigns for greater inclusion in politics. Founded in 2015, the group deals with issues such as voter registration, participation in local, state, national, and international politics, and aims to be a source of unbiased information regarding the political process.

The organization offers financial support for travel to those who attend their yearly conference, as well as providing ongoing assistance to chapters to run their own events.

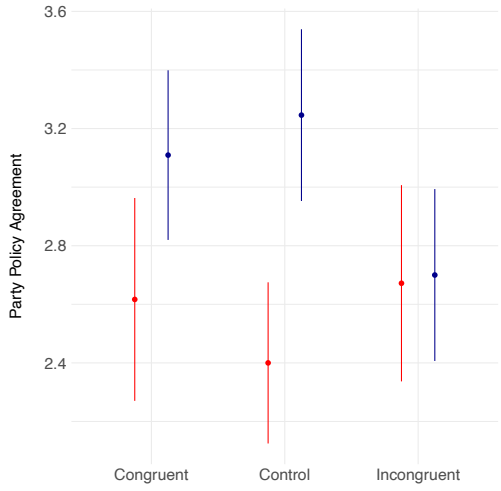
AVA has proven popular with younger men of a non-religious background, but has struggled to attract support from racial minorities.

Movement for Consumer Freedom

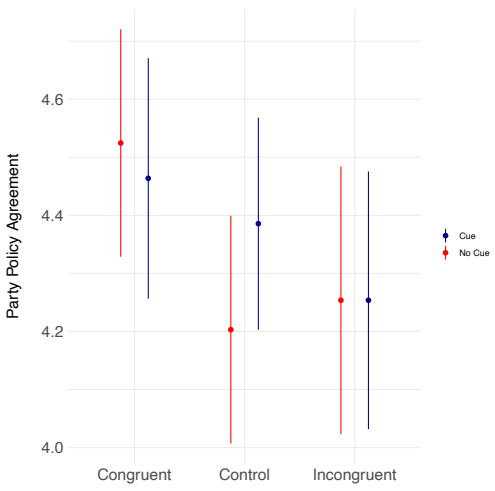
Founded as an international movement for deregulation in 2004, the Movement for Consumer Freedom encourages individuals to campaign for less government intervention in the market in their country. Activities in local chapters include hosting speakers, communicating with local and national governments, and attendance at international conferences and events.

The group hopes to open a new UC Davis chapter to encourage students to get involved in local and national issues regarding consumer choice, including by offering funding and support to interested students to participate in local, national, and international events.

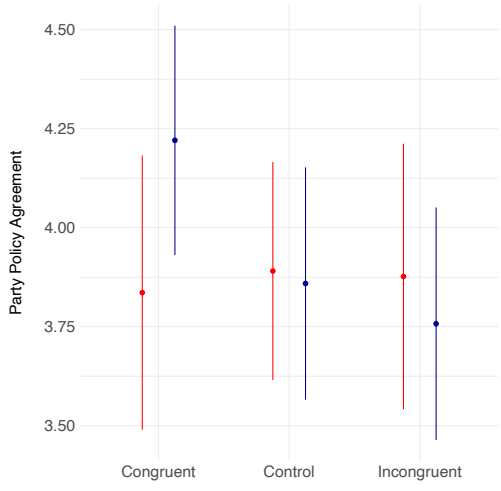
Figure A3: Cue and Treatment effects by policy area



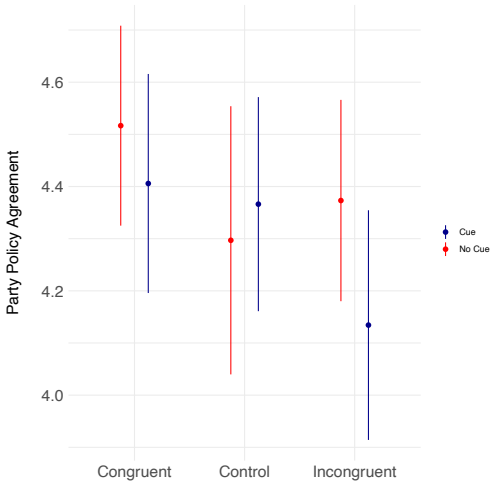
(a) Right to Work



(b) "Obamacare"



(c) Affirmative Action



(d) Increasing the Minimum Wage

B Chapter 2 Appendix

Table B1: Independent variables for ideological sorting measure

Variable	ANES Cummulative File Measure
Guaranteed Jobs	VCF0809 - Guaranteed Jobs and Income Scale
Isolationism	VCF0823 - Better off if US unconcerned with the rest of the world
Aid to Blacks	VCF0830 - Aid to Blacks Scale
Abortion	VCF0838 - By law, when should abortion be allowed
School Spending	VCF0890 - Federal Spending - Public Schools
Government Services Spending	VCF0839 - Government Services - Spending scale
Social Security	VCF9049 - Federal spending - social security
Healthcare	VCF0806 - Government health insurance scale
Equal opportunity	VCF9013 - Society ensure equal opportunity to succeed
Egalitarianism	VCF9016 - Not big problem if some have more chance in life than others
Equality worry	VCF9017 - Should worry less about how equal people are
Treated equal	VCF9018 - US would have fewer problems if everyone treated equally

Table B2: Independent variables for social sorting measure

Variable	Measurement
Race	1= White, 2= Black, 3 = Hispanic. Others excluded due to lack of feeling thermometer measures
Strength of Racial Identity	Feeling thermometer towards Racial in-group minus out-group
Age	Feelings towards Feminists
Gender	Binary, Male/Female, others excluded.
Class	Feeling thermometer towards Unions minus Big Business
Feminist Identity	Feeling thermometer score towards feminists
Sexual Identity	Feeling thermometer score towards Gays and Lesbians
Education	7-category variable, 'grade school or less' to 'advanced degrees incl. LLB'
Religious Denomination	4 categories: Protestant, Catholic, Jewish, Other and none
Marital status	Married, Never Married, Divorced, Separated, Widowed, Partners
Union Household membership	Binary, Yes / No
Employment status	Employed, Not Employed, Retired, Homemaker, Student
Native born status	Were both parents were born in this country? Binary, yes / no
Region	Census region: North East, North Central, South, West

Table B3: Machine Learning Model specifications

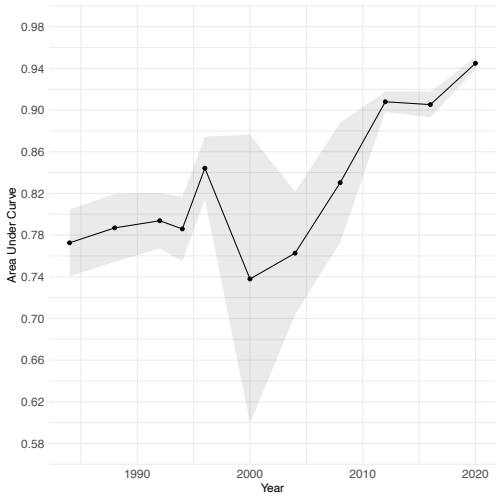
	ROC	Weight	ROC	Weight
Generalized Linear Model	0.857	3.91	0.878	3.82
Random Forest	8.68	0.09	0.883	0.73
Selection Vector Machine	0.867	-7.67	0.880	-5.91
CART	0.77	-1.57	0.776	0.612
Neural Network	0.842	-0.02	0.882	-0.18
K-Nearest Neighbors	0.768	0.11	0.846	-1.92
Gradient Boosting machine	0.870	-2.00	0.883	-1.48
Parallel Random Forest	0.864	3.19	0.880	0.43

Table B4: Effects of sorting on feeling thermometer scores
with equal sample sizes (N=15,236)

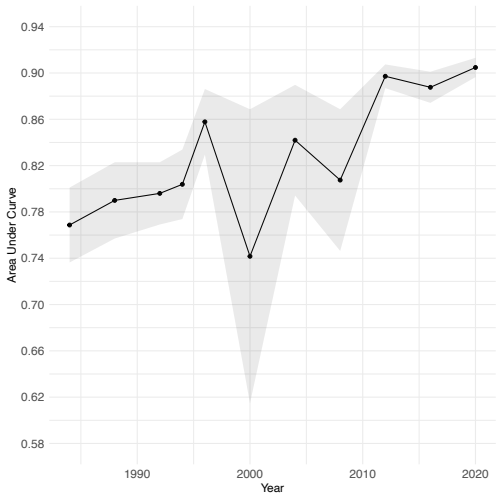
	1	2	3	4	5
Ideological Sorting	31.96**		19.02**		14.53**
	(2.61)		(2.36)		(1.84)
Social Sorting		29.48**		16.25**	10.44**
		(2.64)		(2.50)	(1.94)
Strength of PID			24.50**	24.11**	23.85**
			(0.99)	(0.97)	(0.91)
Strength if Ideological ID			11.58**	12.79**	11.06**
			(1.29)	(1.54)	(1.35)
Political Interest			9.33**	9.72**	9.02**
			(0.62)	(0.75)	(0.71)
Adjusted R^2	0.172	0.165	0.346	0.342	0.352

* p < 0.05, ** p < 0.01

Figure B1: Over time change in the AUC produced by the machine learning models
with equal sample size

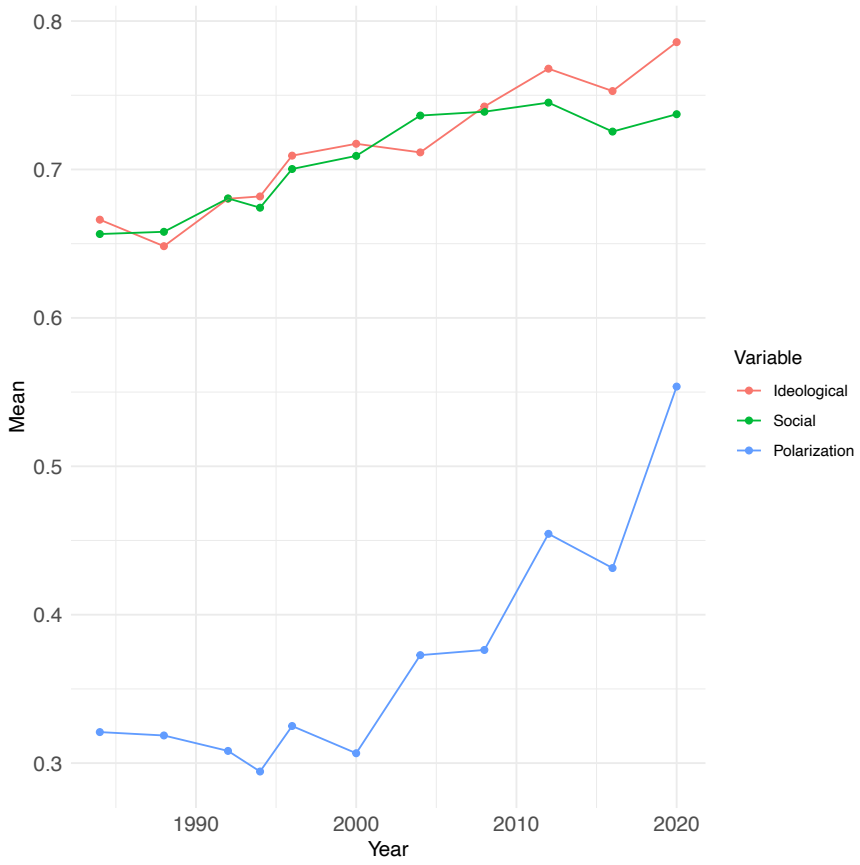


(a) Ideological Sorting



(b) Social Sorting

Figure B2: Over time change in the aggregate level of sorting
and mean level of individual affective polarization
with equal sample sizes



Each point represents the mean for the given variable (individual ideological sorting score, social sorting score, difference in thermometer ratings between the parties) for a given survey year

Table B5: Effects of sorting on in- and out- party feeling thermometer scores (N=6,213)

	In-Party			Out-Party		
	1	2	3	4	5	6
Ideological Sorting	12.44**		4.99**	−27.35**		−16.74**
	(1.61)		(1.42)	(2.62)		(2.18)
Social Sorting		13.77**	4.93**		−23.46**	−10.10**
		(1.47)	(1.18)		(1.72)	(1.39)
Strength of PID			17.17**			−7.64**
			(0.90)			(1.11)
Strength if Ideological ID			0.93			−10.33**
			(0.92)			(0.84)
Political Interest			2.63**			−4.24**
			(0.40)			(1.11)
Adjusted R^2	0.039	0.045	0.197	0.270	0.249	0.338

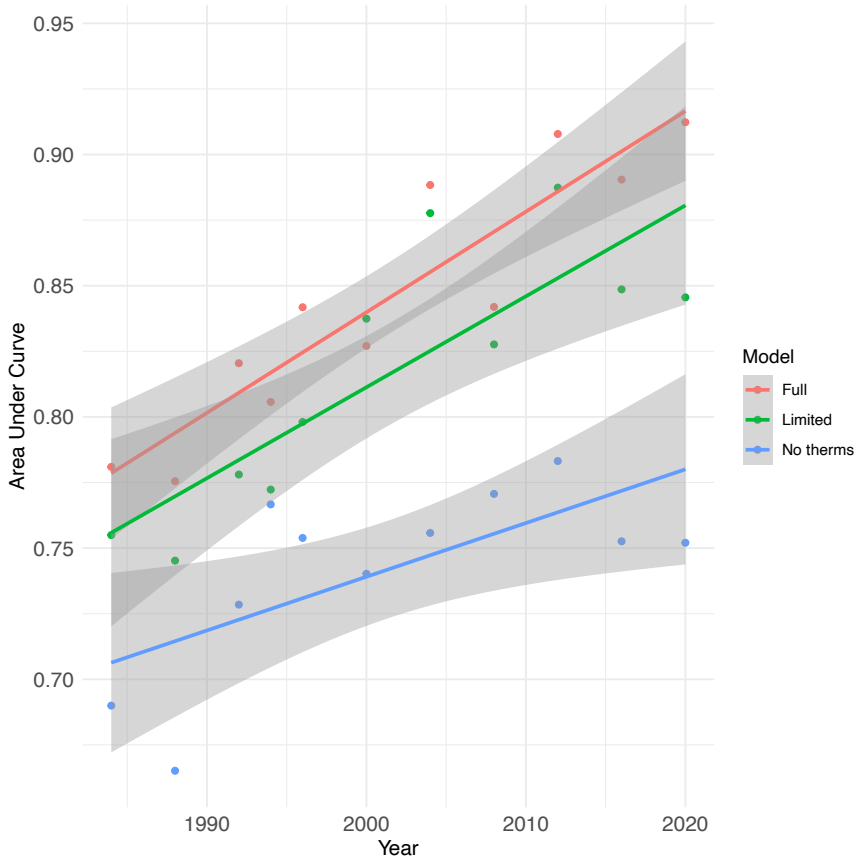
* p < 0.05, ** p < 0.01

Table B6: Effects of sorting on feeling thermometer scores by party

	Democrats			Republicans		
	In-Party	Out-Party	Differential	In-Party	Out-Party	Differential
Ideological Sorting	5.53**	−17.05**	15.85**	2.45	−16.05**	8.01*
	(1.17)	(2.56)	(2.26)	(2.05)	(2.80)	(2.91)
Social Sorting	8.98**	−8.45**	13.30**	−0.65	−13.09**	7.70**
	(1.65)	(1.95)	(1.65)	(1.39)	(1.65)	(1.99)
Strength of PID	16.60**	−7.31**	23.16**	16.92**	−7.87**	24.69**
	(0.90)	(1.72)	(1.37)	(0.95)	(0.79)	(1.47)
Strength if Ideological ID	−0.79	−8.91**	8.53**	6.09**	−11.43**	16.56**
	(1.20)	(1.19)	(1.59)	(1.85)	(0.91)	(2.31)
Political Interest	2.81**	−5.21**	8.46**	2.83**	−3.05*	7.78**
	(0.68)	(1.54)	(1.34)	(0.53)	(1.08)	(1.32)
Observations.	3430	3430	3430	2783	2783	2783
Adjusted R^2	0.209	0.298	0.327	0.202	0.395	0.402

* p < 0.05, ** p < 0.01

Figure B3: Over time change in the AUC produced by the different sorting machine learning models



Each point represents the area under the curve in a given survey year for each model. The lines display the basic regression line for each model over time, with a 95% confidence interval

Table B7: Comparing effects on thermomter differential of full model, limited feeling thermomter, and no feeling thermometer models (N=6,213)

	1	2	3	5	4	6
Ideological Sorting				13.00**	13.83**	15.53**
				(1.79)	(1.97)	(2.24)
Social Sorting	15.65**			10.60**		
	(1.88)			(1.46)		
Limited Feeling therms		14.61**			10.16**	
		(1.97)			(1.62)	
No feeling therms			12.22**			7.80*
			(2.64)			(2.48)
Strength of PID	24.45**	24.62**	25.09**	24.27**	24.32**	24.61**
	(1.37)	(1.27)	(1.23)	(1.35)	(1.27)	(1.23)
Strength if Ideological ID	12.77**	13.29**	14.47**	11.20**	11.39**	11.94**
	(1.26)	(1.27)	(1.36)	(1.09)	(1.05)	(1.03)
Political Interest	8.60**	8.93**	9.52**	7.87**	8.01**	8.29**
	(1.01)	(1.09)	(1.18)	(0.96)	(1.00)	(1.03)
Adjusted R^2	0.345	0.343	0.336	0.354	0.354	0.350

* p < 0.05, ** p < 0.01

C Chapter 3 Appendix

Table C1: Countries and Parties analyzed by year

Country & Years & Parties	Country & Years & Parties
Australia 1996, 2004, 2007, 2013	**Denmark 1998, 2001, 2007, 2019**
Australian Labor Party	CD Centre Democrats
Australian Greens	KF Conservatives People's Party
Liberal Party of Australia	SD Social Democratic Party
National Party of Australia	SF Socialist People's Party
Australian Democrats	V Liberal Party
Palmer Party	EL Red-Green Unity List
Austria 2008, 2013, 2017	RV Danish People's Party*
GA Green Alternative	KrF Christian People's Party
SPO Austrian Social Democratic Party	**Finland 2003, 2007, 2011, 2019**
OVP Austrian People's Party	KD Christian Democratic Party
KPO Communist Party of Austria	KESK Centre Party
VdU League of Independents	KOK National Coalition Party
FPO Austrian Freedom Party*	RKP/SFP Swedish People's Party
BZO Alliance for the Future of Austria*	SSDP Social Democratic Party
NEOS New Austria and Liberal Forum	VAS Left Alliance
TS Team Stronach for Austria	VIHR Green League
JETZT Pilz List	PS True Finns*
Belgium 2019	**France 2002, 2007, 2012, 2017**
N-VA New Flemish Alliance	EELV Green Party
VB Flemish Importance	UDF Union for French Democracy
CD&V Christian Democratic & Flemish	PS Socialist Party
OPEN-VLD Open Flemish Liberals and Democrats	FN National Front*
sp.a Socialist Party Differently	RPR Rally for the Republic
Green	MoDem Movement for Democracy
PVDA Workers Party of Belgium	UMP Union for a Popular Movement
Canada 1997,2004, 2008, 2011, 2015	PG Left Party
BQ Bloc Quebecois	FDP Liberal Democrats
CP Conservative Party	GP Green Party
LP Liberal Party	SP Social Democrats
PC Progressive Conservatives	
ND New Democratic Party	
GP Green Party	

Table C2: Countries and Parties analyzed by year

Country & Years & Parties	Country & Years & Parties
Germany 1998, 2002, 2005, 2009, 2013, 2017	**Iceland 1999, 2003, 2007, 2009, 2013**
CDU Christian Democrats	VGB Left Green Movement
FDP Free Democratic Party	FF Liberal party
GRUNEN Green Party	Sj Independence Party
PDS/LINKE Party of Dem Socialism	F Progressive Party
SPD Social Democratic Party	**Ireland 2002, 2011, 2016**
AfD Alternative for Germany*	SF Sinn Fein
Pirates	
Great Britain 1997, 2001, 2005, 2015, 2017	FG Fine Gael
LAB Labour Party	GP Green Party
LibDem Liberal Democrats	LP Labour Party
CON Conservative Party	FF Fianna Fail
PC Plaid Cymru	SD Social Democrats
SNP Scottish National Party	**Israel 1996, 2003, 2006, 2013, 2020**
UKIP United Kingdom Independence Party*	HaAvoda Labour Party
GP Green Party	MERETZ Mapam-Ratz
Greece 2009, 2012, 2015a 2015b	There is a Future
KKE Communist Party of Greece	Shinui Change
SYRIZA Coalition of the Radical Left	MAFDAL National Religious Party
PASOK Panhellenic Socialist Movement	SHAS Sephardi Torah Guardians
ND New Democracy	Likud Union
ANEL Independent Greeks*	National Union
LS-XA Golden Dawn*	The Jewish Home*
DIMAR Democratic Left	Movement for Civil Rights and Peace
KINAL The River	
P Pirate Party	
So United Socialist Party	
Citizens' Movement	

Table C3: Countries and Parties analyzed by year

Country & Years & Parties	Country & Years & Parties
Spain 1996, 2000, 2004, 2008	**Turkey 2018**
PP People's Party	AKp Justice and Development Party
IU United Left	CHP Republican People's Party
PSOE Socialist Workers' Party	HDP Peoples' Democratic Party
CiU Convergence and Union	MHP Nationalist Action Party
PNV/EAJ Basque Nationalist Party	IYI Good Party
ERC Republican Left of Catalonia	SP Felicity Party
EA Basque Solidarity	**USA 1996, 2004, 2008, 2012, 2016, 2020**
CDS Centre Democrats	Democratic Party
CC Canarian Coalition	Republican Party
BNG Galician Nationalist Bloc	
Sweden 1998, 2002, 2006, 2010, 2014, 2018	
SAP Social Democrats	
FP People's Party	
MP Green Party	
M Moderate Party	
SD Sweden Democrats*	
KD Christian Democarts	
C Centre Party	
V Left Party	
Switzerland 1999, 2003, 2007, 2011	
CVP Christian Democrats	
FDP Liberal Democrats	
GP Green Party	
SP Social Democrats	
SVP Swiss People's Party*	
EVP Evangelical People's Party	
GLP Green Liberal Party	
LT Ticino League	

Table C4: Countries and Parties analyzed by year

Country & Years & Parties	Country & Years & Parties
Italy 2018	**Norway 1997, 2001, 2005, 2009, 2013, 2017**
M5S Movimento 5 Stelle	SV Left Socialists
PD Partito Democratico	DNA Labour Party
LN Lega	V Liberal Party
GG Forza Italia	KrF Christian People's Party
FdI Fratelli d'Italia	H Conservative Party
LeU Liberi e Ugual	SP Centre Party
Netherlands 998, 2002, 2006, 2010, 2017	Red Electoral Alliance
CDA Christian Democratic Appeal	FrP Progress Party*
SGP Political Reformed Party	GP Green Party
D66 Democrats 66	**Portual 2002, 2005, 2009, 2015, 2019**
GL Green Left	CDS-PP Dem. & Soc Centre+People's Party
PvdA Labour Party	PSP Socialist Party
SP Socialist Party	PSD Social Democratic Party
VVD People's Party for Freedom & Dem	BE Left Bloc
CU Christian Union	**Slovakia 2016**
LPF List Pim Fortuyn*	Direction - Social Democracy (Smer)
PVV Party of Freedom*	Freedom and Solidarity (SaS)
New Zealand 1996, 2002, 2008, 2011, 2014	Christian Democratic Movement (KDH)
ACT New Zealand	Slovak National Party (SNS)
GP Green Party	Bridge (MH)
LP Labour Party	Ordinary people (OLaNO)
MP Maori Party	

Table C5: Results Including All Parties With Data Available

	Full	Women	Men
Gender Gap	−1.88**	−1.80**	−1.82**
	(0.42)	(0.42)	(0.42)
Out-Party % Women Supporters	1.72**	2.30**	1.26**
	(0.33)	(0.34)	(0.32)
Perceived right-left distance (self)	−0.48**	−0.48**	−0.49**
	(0.01)	(0.01)	(0.01)
Parties were in coalition	0.55**	0.53**	0.57**
	(0.12)	(0.12)	(0.12)
Parties were in opposition	0.28**	0.30**	0.25**
	(0.07)	(0.07)	(0.07)
Observations	408.559	192.753	215.806
Adjusted R^2	0.296	0.304	0.291

* p < 0.05, ** p < 0.01

Table C6: Results using Individual Fixed Effects

	1	2	3	4	5	6
Gender Gap	−1.68**	−1.58**	−1.43**	−1.68**	−1.58**	−1.43**
	(0.41)	(0.47)	(0.44)	(0.07)	(0.10)	(0.09)
Out-Party % Women Supporters		1.90**	0.91*		1.90**	0.91**
		(0.39)	(0.35)		(0.08)	(0.07)
Perceived right-left distance (self)	−0.58**	−0.58**	−0.58**	−0.58**	−0.58**	−0.58**
	(0.03)	(0.03)	(0.03)	(0.00)	(0.01)	(0.01)
Parties were in coalition	0.45**	0.45**	0.49**	0.45**	0.45**	0.49**
	(0.13)	(0.13)	(0.12)	(0.02)	(0.03)	(0.02)
Parties were in opposition	0.29**	0.35**	0.32**	0.29**	0.35**	0.32**
	(0.10)	(0.11)	(0.11)	(0.01)	(0.02)	(0.02)
Observations	353.101	165.840	187.261	353 101	165 840	187 261
Adjusted R^2	0.417	0.416	0.423	0.417	0.416	0.423

* p < 0.05, ** p < 0.01

Table C7: Excluding Radical Right Parties

	Full	Women	Men
Gender Gap	−1.57**	−1.48**	−1.51**
	(0.42)	(0.48)	(0.45)
Out-Party % Women Supporters		2.02**	0.92**
		(0.37)	(0.34)
Perceived right-left distance (self)	−0.47**	−0.46**	−0.47**
	(0.02)	(0.02)	(0.02)
Parties were in coalition	0.53**	0.51**	0.57**
	(0.13)	(0.14)	(0.12)
Parties were in opposition	0.36**	0.40**	0.36**
	(0.09)	(0.09)	(0.09)
Observations	341.830	160.609	181.221
R2	0.302	0.310	0.302
Adjusted R^2	0.302	0.309	0.301

* p < 0.05, ** p < 0.01

Table C8: Excluding Nationalist Parties

	Full	Women	Men
Gender Gap	−1.18**	−0.98*	−1.21**
	(0.44)	(0.45)	(0.41)
Out-Party % Women Supporters		0.28	−0.55
		(0.41)	(0.36)
Perceived right-left distance (self)	−0.46**	−0.45**	−0.46**
	(0.02)	(0.02)	(0.02)
Parties were in coalition	0.48**	0.46**	0.50**
	(0.13)	(0.14)	(0.12)
Parties were in opposition	0.47**	0.50**	0.43**
	(0.09)	(0.09)	(0.10)
Observations	309.958	145.748	164.210
Adjusted R^2	0.311	0.316	0.309

* p < 0.05, ** p < 0.01

Table C9: Excluding Parties with Fewer than 50 Supporters from Analysis

	1	2	3
Gender Gap	−1.67**	−1.47**	−1.51**
	(0.39)	(0.44)	(0.42)
Out-Party % Women Supporters		1.96**	0.83*
		(0.40)	(0.34)
Perceived right-left distance (self)	−0.47**	−0.46**	−0.47**
	(0.02)	(0.02)	(0.02)
Parties were in coalition	0.54**	0.53**	0.58**
	(0.13)	(0.14)	(0.13)
Parties were in opposition	0.37**	0.42**	0.37**
	(0.08)	(0.09)	(0.09)
Observations	348.819	163.830	184.989
Adjusted R^2	0.305	0.312	0.304

* p < 0.05, ** p < 0.01

Table C10: Results using Alternate RILE measure of party distance

	Full 1	Full 2	Women	Men
Gender Gap	−1.54*	−1.44	−1.32	−1.42
	(0.69)	(0.78)	(0.79)	(0.75)
Out-Party % Women Supporters		1.22**	2.03**	0.56
		(0.43)	(0.48)	(0.40)
Parties were in coalition	1.01**	1.02**	1.00**	1.05**
	(0.17)	(0.16)	(0.17)	(0.16)
Parties were in opposition	0.57**	0.59**	0.65**	0.54**
	(0.11)	(0.11)	(0.11)	(0.11)
Observations	353.101	353.101	165.840	187.261
Adjusted R^2	0.154	0.155	0.163	0.151

* p < 0.05, ** p < 0.01

Table C11: Results using Alternate Perceived Party ideological Distance measure

	Full	Women	Men
Gender Gap	−1.46**	−1.42**	−1.39**
	(0.46)	(0.47)	(0.43)
Out-Party % Women Supporters	1.48**	2.17**	0.93**
	(0.38)	(0.43)	(0.34)
Parties were in coalition	0.57**	0.54**	0.60**
	(0.12)	(0.13)	(0.12)
Parties were in opposition	0.37**	0.40**	0.35**
	(0.08)	(0.08)	(0.08)
Observations	353 101	165 840	187 261
Adjusted R^2	0.277	0.287	0.272

* p < 0.05, ** p < 0.01

Table C12: Ideological Distance not Impacted by Out Party % Women

	Women Self	Men Self	Women Party	Men Party
Out-Party % Women Supporters	−0.76*	0.29	−0.20	0.62
	(0.37)	(0.42)	(0.39)	(0.42)
Parties were in coalition	−1.00**	−0.98**	−1.12**	−1.10**
	(0.15)	(0.16)	(0.17)	(0.17)
Parties were in opposition	−0.50**	−0.38**	−0.59**	−0.44**
	(0.11)	(0.10)	(0.12)	(0.12)
Observations	160.609	181.221	160.609	181.221
Adjusted R^2	0.152	0.151	0.172	0.173

* p < 0.05, ** p < 0.01

Table C13: Results using Alternate Thermometer Differential measure of Out-Party Evaluations

	Full	Women	Men
Gender Gap	1.45**	1.28**	1.51**
	(0.43)	(0.45)	(0.42)
Out-Party % Women Supporters	−1.21**	−1.82**	−0.70
	(0.40)	(0.40)	(0.40)
Perceived right-left distance (self)	0.57**	0.57**	0.57**
	(0.02)	(0.02)	(0.02)
Parties were in coalition	−0.51**	−0.50**	−0.53**
	(0.11)	(0.12)	(0.11)
Parties were in opposition	−0.20*	−0.21*	−0.19
	(0.09)	(0.08)	(0.11)
Observations	353.101	165.840	187.261
Adjusted R^2	0.274	0.280	0.270

* p < 0.05, ** p < 0.01

Table C14: Results including Country-level fixed effects and clustered standard errors

	Full	Women	Men
Gender Gap	−1.60**	−1.57**	−1.52**
	(0.36)	(0.34)	(0.36)
Out-Party % Women Supporters	1.24**	1.87**	0.74
	(0.41)	(0.41)	(0.42)
Perceived right-left distance (self)	−0.47**	−0.46**	−0.47**
	(0.03)	(0.03)	(0.03)
Parties were in coalition	0.56*	0.54*	0.59**
	(0.21)	(0.23)	(0.19)
Parties were in opposition	0.38**	0.41**	0.36**
	(0.09)	(0.10)	(0.09)
Observations	353.101	165.840	187.261
R2	0.298	0.303	0.296
Adjusted R^2	0.298	0.303	0.295

* p < 0.05, ** p < 0.01

Table C15: Effects of Supporter Gender Composition on Party Feeling Thermometer by Respondent Gender

	Women	Men	Full	Women 2	Men 2
Gender Gap			−2.33**	−2.41**	−2.12**
			(0.66)	(0.80)	(0.59)
Out-Party % Women Supporters	12.06**	9.38**	6.29**	7.61**	5.61**
	(3.11)	(2.54)	(1.99)	(2.64)	(1.83)
Out-Party % Women Supporters Squared	−9.98**	−8.74**	−4.97*	−5.59*	−4.80*
	(2.91)	(2.51)	(1.96)	(2.54)	(1.84)
Perceived right-left distance (self)	−0.46**	−0.47**	−0.46**	−0.46**	−0.46**
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
Parties were in coalition	0.58**	0.62**	0.57**	0.55**	0.59**
	(0.19)	(0.17)	(0.18)	(0.19)	(0.17)
Parties were in opposition	0.34**	0.34**	0.34**	0.34**	0.34**
	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)
Observations	70 750	81 693	152 443	70 750	81 693
Adjusted R^2	0.320	0.309	0.314	0.321	0.311

* p < 0.05, ** p < 0.01

References

Abendschn, A., and Steinmetz, S. (2014) The Gender Gap in Voting Revisited: Women's Party Preferences in a European Context, *Social Politics: International Studies in Gender, State & Society*, 21(2), Summer, 315344, https://doi.org/10.1093/sp/jxu009

Abramowitz, A. I., and Webster, S. (2016) The Rise of Negative Partisanship and the Nationalization of U.S. Elections in the 21st Century. *Electoral Studies* 41, 1222.

Achen, C.H. (2002) Parental Socializations and Rational Party Identification. *Political Behavior*, 24, 151-170.

Achen, C.H. and Bartels, L.M. (2016) *Democracy for Realists: Why Elections Do Not Produce Responsive Government*, New Haven: Princeton University Press.

Adams, J., Bracken, D., Gidron, N., Horne, W., OBrien, D. and Senk, K. (2022). Cant We All Just Get Along? How Women MPs Can Ameliorate Affective Polarization in Western Publics. *American Political Science Review*, 1-7

Ahler, D. J. and Sood, G. (2018) The parties in our heads: Misperceptions about party composition and their consequences. *Journal of Politics*, 80(3), 964981. doi: 10.1086/697253.

Ahn, S. Bergan, D. E., Barry, R. and Ulusoy, E. (2021) Out-Party Cues and Factual Beliefs in an Era of Negative Partisanship, *Journal of Political Marketing*, 20(3), 269-288. DOI: 10.1080/15377857.2021.1939570

Badas, A., and Stauffer, K. E. (2022) Descriptive Representation, Judicial Nominations, and Perceptions of Presidential Accomplishment. *Representation* 1(22).

Baldassari, D. and Gelman, A. (2008) Partisans without Constraint: Political Polarization and Trends in American Public Opinion. *American Journal of Sociology* 114(2), 408-446.

Barnes, T. (2016). *Gendering Legislative Behavior*. Cambridge University Press.

Barnes, T. D., Beall, V. D., and Cassese, E. C. (2021). American Party Women Redux: Stability in Partisan Gender Gaps. *PS: Political Science & Politics*, 54(3), 406410. http://doi.org/10.1017/S1049096520002012

Barnes, T. D., and Cassese, E. C. (2017). American Party Women: A Look at the Gender Gap within Parties. *Political Research Quarterly*, 70(1), 127141.

Bartels, L. M. (2002) Beyond the Running Tally: Partisan Bias in Political Perceptions. *Political Behavior* 24 (2), 117-50.

Bauer, N. M. (2019). Gender Stereotyping in Political Decision-Making *Oxford Research Encyclopedia of Politics*

Ben-Shitrit, L., Elad-Strenger, J. and Hirsch-Hoefler, S. (2021). Pinkwashing the Radical-Right: Gender and the Mainstreaming of Radical-Right Policies and Actions. *European Journal of Political Research* 61(1), 86-110.

Boudreau, C., and MacKenzie, S. (2014). Informing the Electorate? How Party Cues and Policy Information Affect Public Opinion about Initiatives. *American Journal of Political Science*, 58(1), 48-62.

Box-Steffensmeier, J., De Boef, S. and Lin, T. (2004). The Dynamics of the Partisan Gender Gap. *The American Political Science Review*. 98(3), 515-528

Boxell, L., Gentzkow, M. and Shapiro, J. (2020) Cross-Country Trends in Affective Polarization. *National Bureau of Economic Research*, (January). doi: 10.3386/w26669.

Boydstun, A. E., A. Ledgerwood, and Sparks, J. (2017). A Negativity Bias in Reframing Shapes Political Preferences Even in Partisan Contexts. *Social Psychological and Personality Science*. 10(1), 53-61 https://doi.org/10.1177/1948550617733520

Busby, E. C., Howat, A. J., Rothschild, J. E., and Shafranek, R. M. (2021). The Partisan Next Door: Stereotypes of Party Supporters and Consequences for Polarization in America. Cambridge: Cambridge University Press. http://doi.org/10.1017/9781009086462

Brady, H.E. and Sniderman, P.M. (1985). Attitude Attribution: A Group Basis for Political Reasoning. *American Political Science Review* 79 (4), 1061-78.

Broniecki, P., Leemann, L., & West, R.L. (2021). Improved Multilevel Regression with Poststratification through Machine Learning (autoMrP). *The Journal of Politics*, 84, 597 - 601.

Campbell, A, Converse, P. E., Miller, W.E. and Stokes, D.E. (1960) *The American Voter* New York: John Wiley & Sons.

Campbell, D. E., Layman, G.C., Green, J.C., and Sumaktoyo, N.G. (2018) Putting Politics First: The Impact of Politics on American Religious and Secular Orientations. *American Journal of Political Science*, 62(3), 551565. doi: 10.1111/ajps.12365.

Carsey, T. M. and Layman, G.C. (2006). Changing Sides or Changing Minds? Party Identification and Policy Preferences in the American Electorate. *American Journal of Political Science*, 50, 464-477.

Childs, S. (2004). A Feminised Style of Politics? Women MPs in the House of Commons. *The British Journal of Politics and International Relations* 6(1), 3-19.

Claassen, R. L., Djupe, P. A., Lewis, A. R., and Neiheisal, J. R. (2019). Which Party Represents My Group? The Group Foundations of Partisan Choice and Polarization. *Political Behavior* 43, 615-636. doi: 10.1007/s11109-019-09565-6.

Claassen, C., Tucker, P. and Smith, S. S. (2015) Ideological Labels in America, *Political Behavior*. 37(2), 253-278

Clayton, A., Piscopo, J.M. and O'Brien, D.Z. (2019). All Male Panels? Representation and Democratic Legitimacy. *American Journal of Political Science* 63(1), 11329.

Caughey, D. and Schickler, E. (2016) Substance and Change in Congressional Ideology: NOMINATE and Its Alternatives, *Studies in American Political Development*. 30 (2), 128-146

Cohen, G (2003) Party over Policy: The Dominating Impact of Group Influence on Political Beliefs. *Journal of Personality and Social Psychology* 85, 80822.

Conover, P.J. 1988. The Role of Social Groups in Political Thinking. *British Journal of Political Science*. 18(1), 51-76

Davis, N. T. (2018) Religion and Partisan-Ideological Sorting, 19842016', *Social Science Quarterly*, 99(4), 14461466. doi: 10.1111/ssqu.12508.

Davis, N. T. and Mason, L. (2016) Sorting and the Split-Ticket: Evidence from Presidential and Subpresidential Elections', *Political Behavior*. Springer US, 38(2), 337354. doi: 10.1007/s11109-015-9315-7.

Davis, N. T., Klar, S. and Weber, C. R. (2019) Affective Consistency and Sorting', *Social Science Quarterly*, 100(6), 24772494. doi: 10.1111/ssqu.12695.

DellaPosta, D. (2020) Pluralistic Collapse: The Oil Spill Model of Mass Opinion Polarization, *American Sociological Review*, 85(3), 507536. doi: 10.1177/0003122420922989

Delli Carpini, M.X. and Keeter, S. (1996). *What Americans Know About Politics and Why It Matters*. New Haven: Yale University Press.

Devine, C. J. (2015) Ideological Social Identity: Psychological Attachment to Ideological In-Groups as a Political Phenomenon and a Behavioral Influence, *Political Behavior*, 37(3), 509-535

Devroe, R., Wauters. N. (2018). Political Gender Stereotypes in a List-PR System with a High Share of Women MPs: Competent Men versus Leftist Women?. *Political Research Quarterly* 71(4), 788-800.

Dolan, K. (2010). The Impact of Gender Stereotyped Evaluations on Support for Women Candidates. *Political Behavior* 32(1), 69-88.

Druckman, J., Levendusky. M. (2019). What Do We Measure When We Measure Affective Polarization? Public Opinion Quarterly 83(1), 114122.

Egan, P. J. (2019) Identity as Dependent Variable: How Americans Shift Their Identities to Align with Their Politics', *American Journal of Political Science*, 64, 699-716. doi: 10.1111/ajps.12496.

Feldman, S. and Conover, P. J. (1983). Candidates, Issues and Voters: The Role of Inference in Political Perception. *The Journal of Politics*. 45(4), 810-839

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

Fuller, S., Horne, W., Adams, J. and Gidron, N. (2022). Populism and the Affective Partisan Space in Eight European Publics: Evidence from a Cross-National Survey. *Frontiers in Political Science* 4

Franklin, C.H. and Jackson, J.E. (1983). The Dynamics of Party Identification, *The American Political Science Review* 77(4), 957-973

Garrett, R.K., Long, J.A. and Seon Jeong, M. (2019). From Partisan Media to Misperception: Affective Polarization as Mediator, *Journal of Communication* 69, 490-512

Gerber, A., Huber, G., and Washington, E. (2010) Party Affiliation, Partisanship, and Political Beliefs: A Field Experiment. *The American Political Science Review*, 104(4), 720-744.

Gidron, N., Adams, J. and Horne, W. (2019) Toward a Comparative Research Agenda on Affective Polarization in Mass Publics', *APSA-CP Newsletter*, 29(1), 3036.

Gidron, N., Adams, J. and W. Horne. 2019. 'How Ideology, Economics, and Institutions Shape Affective Polarization in Democratic Polities.' Working Paper.

Gidron, N., Adams, J., & Horne, W. (2022). Who Dislikes Whom? Affective Polarization between Pairs of Parties in Western Democracies. *British Journal of Political Science*, 119.
http://doi.org/10.1017/S0007123422000394

Goren, P. (2005) Party Identification and Core Political Values. *American Journal of Political Science* 49 (4), 881-96.

Gothreau, C. (2021) Everything You Need to Know About the Gender Gap. *Center for American Women and Politics* https://cawp.rutgers.edu/blog/everything-you-need-know-about-gender-gap

Gunderson, J. R. (2022). When Does Income Inequality Cause Polarization? British Journal of Political Science, 52(3), 13151332. http://doi.org/10.1017/S0007123421000053

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

Greene, Z., Lhiste., M. (2018). Symbols of Priority? How the Media Selectively Report on Parties Election Campaigns. *European Journal of Political Research* 57(3), 717-739.

Grey, S. (2002). Does Size Matter? Critical Mass and New Zealand's Women MPs. *Parliamentary Affairs* 55(1), 1929.

Grimmer, J., Messing, S., and Westwood, S. J. (2017). Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods. *Political Analysis*, 25(4), 413434. http://doi.org/10.1017/pan.2017.15

Hare, C., & Kutsuris, M. (2022). Measuring Swing Voters with a Supervised Machine Learning Ensemble. *Political Analysis*, 117. http://doi.org/10.1017/pan.2022.24

Hargrave, L., Langengen, T. (2020). The Gendered Debate: Do Men and Women Communicate Differently in the House of Commons? *Politics & Gender*, 1-27.

Harteveld, E., Mendoza, P., Rooduijn, M. (2021). Affective Polarization and the Populist Radical Right: Creating the Hating?. *Government and Opposition*, 1-25.

Henrich, J., S.J. Heine, and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences* 33, 6183.

Hogg, M. A., Hardie, E.A., and Reynolds, K.J. (1995). Prototypical similarity, self-categorization, and depersonalized attraction: a Perspective on Group cohesiveness. *European Journal of Social Psychology* 25(2), 15977

Hogg, M. A., Abrams, D., Otten, S., and Hinkle, S. (2004). The social identity Perspective: intergroup relations, self-conception, and small Groups. *Small Group Research* 35(3), 24676.

Holman, M.R. and Mahoney, A. (2018). Stop, Collaborate, and Listen: Women's Collaboration in US State Legislatures. *Legislative Studies Quarterly* 43(2), 179-206.

Huber, G.A., and Malhotra, N. (2017) Political Homophily in Social Relationships: Evidence from Online Dating Behavior. *The Journal of Politics* 79(1), 269-283

Huddy, L., and Carey, T. (2009). Group politics redux: Race and gender in the 2008 Democratic primaries. *Politics and Gender*, 5, 116.

Huddy, L., Mason, L., and Aarøe, L. (2015). Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity. *American Political Science Review*, 109(1), 1-17.

Huddy, L., Mason, L., and Horwitz, N. (2016). Political identity convergence: On being Latino, becoming a Democrat, and getting active. RSF: *The Russell Sage Foundation Journal of the Social Sciences*, 2(30), 205228.

Huddy, L., Terkildsen, N. (1993). Gender stereotypes and the perceptions of male and female candidates. *American journal of political science* 37(1), 119-147.

Iyengar, Sood, S.G. and Lelkes, Y. (2012) Affect, Not Ideology: A Social Identity Perspective on Polarization, *Public Opinion Quarterly*, 76(3), 405431

Iyengar, S., and Westwood, S. 2015. Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, 59(3), 690-707

Iyengar, S., Lelkes, S. Levendusky, M., Malhorta, N., and Westwood, S. (2019). The Origins and Consequences of Affective Polarization in the United States.' *Annual Review of Political Science*. 22, 129-146

Jackson, M. S. (2011). Priming the Sleeping Giant: The Dynamics of Latino Political Identity and Vote Choice. *Political Psychology*, 32(4), 691716

Johnston, R. (2006). "Party Identification: Unmoved Mover or Sum of Preferences?" *Annual Review of Political Science*. 9(1), 329351

Kalin, M. and Sambanis, N. (2018) How to Think About Social Identity', *Annual Review of Political Science*, 21(1), 239257. doi: 10.1146/annurev-polisci-042016-024408.

Kalmoe, N., and Mason, L. (2018). Lethal Mass Partisanship: Prevalence, Correlates and Electoral Contingencies. *Paper presented at Midwest Political Science Association*.

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

Kimball, D. C., Anthony, J. and Chance, T. (2016) Political identity and party polarization in the American electorate', *State of the Parties Conference*.

Kittilson, M.C. and Fridkin, K. (2008). Gender, candidate portrayals and election campaigns: A comparative perspective. *Politics & Gender* 4(3): 371-392.

Kuklinski, J., and Hurley, N. (1994). On Hearing and Interpreting Political Messages: A Cautionary Tale of Citizen Cue-Taking. *The Journal of Politics*, 56(3), 729-751

Lau, R.R., and Redlawsk, D.P. (1997). Voting Correctly. *American Political Science Review* 91 (3), 585-98.

Ledgerwood, A. and Boydstun, A. E. (2014). Sticky Prospects: Loss Frames Are Cognitively Stickier Than Gain Frames. *Journal of Experimental Psychology: General*. 143(1), 376-85.

Lelkes, Y. (2018). Affective Polarization and Ideological Sorting: A Reciprocal, Albeit Weak, Relationship. *The Forum* 16(1), 67-79.

Lenz, G. (2009). Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis. *American Journal of Political Science* 53(4), 821-837.

Leonard, N.E., K. Lipsitz, A. Bizyaeva, A. Franci and Lelkes, Y. (2021). The nonlinear feedback dynamics of asymmetric political polarization. *Proceedings of the National Academy of Sciences* 118(50), doi:10.1073/pnas.2102149118,

Lodge, M., and Taber, C.S. (2013) *The Rationalizing Voter*. Cambridge University Press.

Lupia, A. (1994). Shortcuts versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections. *American Political Science Review* 88(1), 63-76.

Lupia, A., and McCubbins, M.D. (1998). *The Democratic Dilemma.* Cambridge: Cambridge University Press.

Lupton, R. N., Myers, W. M., and Thornton, J. R. (2015). Political Sophistication and the Dimensionality of Elite and Mass Attitudes, 19802004. The Journal of Politics, 77(2), 368380. https://doi.org/10.1086/679493

Lupu, N. (2014). Brand Dilution and the Breakdown of Political Parties in Latin America. *World Politics*66(4), 561-602.

Magnum, M. (2013). The Racial Underpinnings of Party Identification and Political Ideology. *Social Science Quarterly*, 94(5), 1222-1244

Mark, N. (2003). Culture and Competition: Homophily and Distancing Explanations for Cultural Niches. *American Sociological Review*, 68, 319-345

Mason, L. (2013) The Rise of Uncivil Agreement: Issue Versus Behavioral Polarization in the American Electorate', *American Behavioral Scientist*, 57(1), 140159. doi: 10.1177/0002764212463363.

Mason, L. (2015) I disrespectfully agree: The differential effects of partisan sorting on social and issue polarization', *American Journal of Political Science*, 59(1), 128145. doi: 10.1111/ajps.12089.

Mason, L. (2016) A cross-cutting calm: How social sorting drives affective polarization', *Public Opinion Quarterly*, 80(S1), 351377. doi: 10.1093/poq/nfw001.

Mason, L. (2018) Ideologues without issues: The polarizing consequences of ideological identities', *Public Opinion Quarterly*, 82(S1), 280301. doi: 10.1093/poq/nfy005.

Mason, L. and Davis, N. T. (2015) A New Identity Politics : How Socio-Partisan Sorting Affects Affective and Ideological Polarization in American Politics, *Presented at the American Political Science Association 2015*.

Mason, L. and Wronski, J. (2018) One Tribe to Bind Them All: How Our Social Group Attachments Strengthen Partisanship', *Political Psychology*, 39, 257277. doi: 10.1111/pops.12485.

Martherus, J. L, Martinez, A.G., Piff, P.K. and Theodoridis, A.G. (2021) Party Animals? Extreme Partisan Polarization and Dehumanization', *Political Behavior*, 43, 517-540 doi: 10.1007/s11109-019-09559-4.

Matland, R. E. (1994). Putting Scandinavian Equality to the Test: An Experimental Evaluation of Gender Stereotyping of Political Candidates in a Sample of Norwegian Voters. *British Journal of Political Science*

24(2), 273-292.

McClain, P. D., Johnson Carew, J.D., Walton Jnr, E. and Watts, C.S. (2009) Group Membership, Group Identity, and Group Consciousness: Measures of Racial Identity in American Politics?', *Annual Review of Political Science*, 12(1), 471485. doi: 10.1146/annurev.polisci.10.072805.102452.

McConnell, C., Margalit, Y., Malhotra, N. and Levendusky, M.S. (2018). The Economic Consequences of Partisanship in a Polarized Era. *American Journal of Political Science* 62(1), 518.

McCoy, J. and Somer, M. (2019). Toward a Theory of Pernicious Polarization and How It Harms Democracies: Comparative Evidence and Possible Remedies. *The Annals of the American Academy of Political and Social Science* 681(1), 23471.

McPherson J.M. (1981). A dynamic model of voluntary affiliation. *Social Forces* 59, 70521

McPherson J.M. (1982). Hypernetwork sampling: duality and differentiation among voluntary associations. *Social Networks* 3, 22549

McPherson, M., Smith-Lovin, L. and Cook, J.. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology.* 27, 415-30

Milem, J. F. (1998). Attitude Change in College Students: Examining the Effect of College Peer Groups and Faculty Normative Groups. The Journal of Higher Education, 69(2), 117140.

Montgomery, J. M. and Olivella, S. (2018). 'Tree-Based Models for Political Science Data'. *American Journal of Political Science.* 62(3), 729-744

Moskowitz, D. J., Rogowski, J. C. and Snyder, Jr, J. M. (2019). 'Parsing Party Polarization in Congress'. *Working paper.*

Mullainathan, S. and Spiess, J. (2017) Machine learning: An applied econometric approach', *Journal of Economic Perspectives*, 31(2), 87106. doi: 10.1257/jep.31.2.87.

Newcomb, T. M. (1943). *Personality and social change: Attitude formation in a student community.* New York: Dryden Press.

Nicholson, S. (2011). Dominating Cues and the Limits of Elite Influence. *The Journal of Politics*, 73(4), 1165-1177.

Newcomb, T. M. (1968). Attitude development as a function of reference groups: The Bennington study. In H. H. Hyman & E. Singer (Eds.), *Readings in reference group theory and research.* New York: The Free Press.

Nicholson, S. P. (2012), Polarizing Cues. *American Journal of Political Science*, 56, 52-66

O'Brien, D. Z. (2015). Rising to the Top: Gender, Political Performance, and Party Leadership in Parliamentary Democracies. *American Journal of Political Science* 59(4), 1022-1039.

O'Brien, D. Z. (2019). Female Leaders and Citizens' Perceptions of Political Parties. *Journal of Elections, Public Opinion and Parties* 29(4), 465-489.

Ondercin, H. L., and Lizotte, M. K. (2021). Youve Lost That Loving Feeling: How Gender Shapes Affective Polarization. American Politics Research, 49(3), 282292. https://doi.org/10.1177/1532673X20972103

Petersen, M., Skov, M., Serritzlew, S., and T. Ramsy. (2013). Motivated Reasoning and Political Parties: Evidence for Increased Processing in the Face of Party Cues. *Political Behavior*, 35(4), 831-854.

Reiljan, A. (2020). Fear and Loathing across Party Lines (also) in Europe: Affective Polarisation in European Party Systems. *European Journal of Political Research* 59(2), 376-396.

Reiljan, A. and Ryan, A. (2021). Ideological Tripolarization, Partisan Tribalism and Institutional Trust: The Foundations of Affective Polarization in the Swedish Multiparty System. *Scandinavian Political Studies*, 44, 195-219

Riggle, E.D., Ottati, V.C., Wyer, R.S, Kuklinski, J. and Schwartz, N. (1992) Bases of political judgments: The role of stereotypic and nonstereotypic information. *Political Behavior* 14(1), 67-87

Robison, J. and Moskowitz, R. L. (2019) The group basis of partisan affective polarization', *Journal of Politics*, 81(3), 10751079. doi: 10.1086/703069.

Rogowski, J. and Sutherland, J.L. (2016). How Ideology Fuels Affective Polarization. *Political Behavior*, 38, 485-508

Rothschild, J.E., Howat, A.J., Shafranek, R.M. and Busby, E.C. (2019). Pigeonholing Partisans: Stereotypes of Party Supporters and Partisan Polarization. *Political Behavior* 41, 423-443

Selway, J.S. (2011). The measurement of cross-cutting cleavages and other multidimensional cleavage structures *Political Analysis* 19(1), 48-65

Shively, W.P. (1979). The Relationship Between Age and Party Identification: A Cohort Analysis. *Political Methodology* 6(4), 437-446

Sones, B., Moran, M. and Lovenduski, J. (2005). *Women in Parliament: The New Suffragettes*. Politicos.

Stauffer, K. E. (2021). Public Perceptions of Womens Inclusion and Feelings of Political Efficacy. *American Political Science Review*, 1-16.

Strickler, R. (2018) Deliberate with the Enemy? Polarization, Social Identity, and Attitudes toward Disagreement. *Political Research Quarterly* 71(1), 3-18

Tajfel, H. (1981). *Human Groups and Social Categories*. Cambridge: Cambridge University Press

Tajfel, H., and Turner, J. C. (1979). An integrative theory of intergroup conflict. In W.G. Austin and S Worchel (Eds.), *The social psychology of intergroup relations* (33-48). Monterey, CA: Brooks/Cole

Turner, J. C. (1981). Towards a cognitive redefinition of the social group. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 1(2), 93-118.

Turner, J. C., Hogg, M. A., Oakes, P., Reicher, S., and Wetherell, M. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. New York, N.Y.: Blackwell.

Turner, J.C. (1999). 'Some current issues in research on social identity and self-categorization theories.' in Ellemers, N., Spears, R., and Doosje, B. (eds), *Social Identity: Context, Commitment, Content*. Malden, Mass.: Blackwell.

Valentino, N.A. and Zhirkov, K. (2018) Blue is Black and Red is White? Affective Polarization and the Racialized Schemas of U.S. Party Coalitions', *Presented at Midwest Political Science Association*. doi:

10.1017/CBO9781107415324.004.

Verge, T., Wiesehomeier, N. and Esprito-Santo, A. (2020). Framing Symbolic Representation: Exploring How Womens Political Presence Shapes Citizens Political Attitudes. *European Journal of Politics and Gender* 3(2), 257-276.

Wagner, Markus. 2021. Affective Polarization in Multiparty Systems. *Electoral Studies* 69.

Weeks, A. C., Meguid, B.M, Kittilson, M.C., and Coff, H. (2021). When Do Männerparteien Elect Women? Radical Right Populist Parties and Strategic Descriptive Representation. https://researchportal.bath.ac.uk/en/publications/when-do-m%C3%A4nnerparteien-elect-women-radical-right-populist-parties

Westwood, S. and Peterson, E. (2019) Compound Political Identity: How Partisan and Racial Identities Overlap and Reinforce', *SSRN Electronic Journal*. doi: 10.2139/ssrn.3417476.

Wilcox, C., Sigelman, L. and Cook, E. (1989) Some Like It Hot: Individual Differences in Responses to Group Feeling Thermometers, *The Public Opinion Quarterly*. 53 (2), 246-257

Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.