Measuring Social Sorting and Its Association With Affective Polarization

Olivier Bergeron-Boutin

Department of Political Science

McGill University, Montreal

15 April 2024

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Arts

Contents

Abstract

In the United States, increasing affective polarization – the tendency of supporters of opposing parties to dislike one another – has worried political scientists and attracted much media attention. Among the potential causes of this rise in affective polarization is the phenomenon of social sorting, whereby the social identities of partisan groups become neatly aligned with partisanship. In this thesis, I reassess the evidence regarding the link between social sorting and affective polarization by leveraging historical data from the American National Election Studies. I report two main findings. First, the over-time increase in social sorting is rather modest, unless self-reported ideology is included as a social identity – which, I argue, is theoretically dubious. Second, I find some suggestive evidence that highly sorted individuals tend to be more affectively polarized, which lends some support to the social sorting theory. I conclude by drawing attention to the undertheorized distinction between actual sorting and perceived sorting.

Résumé

Aux États-Unis, la polarisation affective croissante – la tendance des partisans de partis opposés à se détester les uns les autres – inquiète les politologues et attire beaucoup d'attention médiatique. Parmi les causes potentielles de cette montée de la polarisation affective figure le phénomène de tri social, par lequel les identités sociales des groupes partisans s'alignent parfaitement avec l'identification partisane. Dans ce mémoire, je réévalue les preuves concernant le lien entre le tri social et la polarisation affective en exploitant les données historiques du American National Election Studies. Je rapporte deux conclusions principales. Premièrement, l'augmentation du tri social au fil du temps est plutôt modeste, à moins que l'idéologie autodéclarée ne soit incluse comme identité sociale – ce qui, à mon avis, est théoriquement douteux. Deuxièmement, je trouve des preuves suggérant que les individus hautement triés ont tendance à être plus polarisés sur le plan affectif. Je conclus en attirant l'attention sur la distinction sous-théorisée entre le tri social réel et le tri social perçu.

Acknowledgements

My journey to the submission of this thesis has been (too) long and tortuous, but exceptionally rewarding. First and foremost, I want to thank Dietlind Stolle, my supervisor at McGill and the person most responsible for introducing me to the world of academic research. Her support through my many detours and delays has been invaluable. I also wish to thank professors at McGill who were instrumental to my thinking: Aaron Erlich, whose expertise in methods and steadfast adherence to open science influenced me greatly; Krzysztof Pelc, thanks to whom I first explored the world of academic research; Elisabeth Gidengil, who trusted a fresh-out-of-undergraduate Olivier to help on major research projects; Elissa Berwick, whose graduate methods courses were a great opportunity for me to develop my teaching skills; and Leonardo Baccini, whose encouragement during my undergraduate degree at McGill was inspiring. My fellow students at McGill were also an important part of my experience. Sincere thanks to Philippe Chassé, Mathieu Lavigne, Tim Roy, Aengus Bridgman, and Colin Scott. Finally, having spent the last sixteen months at Dartmouth College, I wish to thank the enormously supportive community that has accepted me there – most of all, Brendan Nyhan and John Carey, whose support means a lot to me.

Contribution of author

The analysis and writing of the entire thesis was performed solely by the author.

List of Figures

List of Tables

1 Introduction

In the United States, a lively scholarly debate on the nature of mass polarization continues. After much ink had been spilt on the issue of ideological polarization, scholars of American politics have recently turned their attention to worrying trends in *affective* polarization – the tendency of Democrats and Republicans in the mass public to "dislike, even loathe, their opponents" (Iyengar, Sood, and Lelkes 2012, 405). The potential causes of this growing partisan animosity are many and still hotly debated. The growth of highly partisan media outlets (Lelkes, Sood, and Iyengar 2017), the influence of polarized political elites (Banda and Cluverius 2018), and mass ideological polarization itself (Rogowski and Sutherland 2016) have all received much attention as possible drivers of affective polarization. In addition, one interesting suspect borrowed from social psychology has emerged: the phenomenon of "social sorting." Over the last several decades, the two major parties in the United States have become increasingly homogenous in terms of their social composition (Mason 2018a). Ethnic minorities, residents of urban areas, and non-religious individuals have sided with the Democrats; white, religious men living in rural areas have associated with the Republicans. These are, of course, crude generalizations – but precisely the type of crude generalizations that social identity theory predicts people will make when forming mental images of in-groups and out-groups (see Ahler and Sood 2018). This process of social sorting "has made it much easier for partisans to make generalized inferences about the opposing side" (Iyengar et al. 2019, 134). From the perspective of a socially sorted partisan, the opposing side is emphatically not amorphous: it is composed of precisely the sort of people with whom one does not associate. The result has been the formation of a partisan "mega-identity", "with each party representing not only policy positions but also an increasing list of other social cleavages" (Mason 2018a, 20). Cross-cutting identities, long thought to be a stabilizing force for democratic regimes, consequently wither away as a social chasm between Republicans and Democrats opens up.

Is social sorting truly the driver of affective polarization? The theory of social sorting, primarily developed by Mason (2018a), is elegant and well-grounded in decades of social psychology research. However, I argue that the extant tests of the social sorting hypothesis have not been empirically convincing. On this point, I make two broad claims. First, there is a mismatch between the broad narrative that motivates this area of research (the *growing* social chasm between Democrats and Republicans) and the evidence brought to bear (primarily *cross-sectional* surveys of Americans in the 2010s). Second, the existing individual-level measures of social sorting rely on rather crude coding schemes that lack flexibility: they do not allow the relationship between social identities and party identification to wax and wane through time. As a result, the validity of the measure is compromised in contexts of changing socio-political cleavages.

The aim of this thesis is to provide more solid evidence on the evolution of socio-political cleavages in American politics and its relationship with growing affective polarization. As a remedy, I propose to apply a measure of individual-level cross-pressures emanating from social identities developed by Brader, Tucker, and Therriault (2014) to historical survey data. The intuition behind this measurement scheme is quite straightforward: a voter faces cross-pressures if, in trying to predict their party identification using their social identities, a statistical model returns an ambiguous answer. By contrast, if a statistical model can predict a voter's party identification with high certainty using just their social identities, this voter can be considered socially sorted. I implement this measurement scheme to historical data from the American National Election Study (ANES) in order to provide descriptive evidence on the evolution of social sorting in the United States. If American voters have socially sorted, it should be increasingly easy to predict their party identification on the basis of their social identities. The second step in the empirical analysis will be to examine the relationship between this individual-level measure of social sorting and affective polarization. While I cannot claim to be able to isolate the relevant causal effect, it is worth examining the naturally occurring relationship so as to think clearly about confounding and identify what *would have to be true* in order for the association to be causal.

2 Affective polarization: the extant literature

2.1 A brief history of partisanship in the United States

Before describing trends in affective polarization, I begin by providing some context on the partisan environment in the United States in the decades leading up to the fraught political moment we live in. Barely 25 years ago, the idea of an overpowering partisan identity would have seemed unrealistic and out of step with available data. Indeed, the configuration of the American party system had for long not been amenable to the formation of strong and stable identities. Before the 1960s, the Democratic Party and the Republican Party were "ideologically heterogeneous, with sizable contingents of both liberals and conservatives at the mass and elite levels" (Levendusky 2009, 1). As the civil rights movement caused drastic realignments in the partisan structure of American politics, voters "sorted" and each party became more coherent ideologically (Noel 2012, 159). Nonetheless, even as this process was taking place, the influence of parties on the political behavior of the masses declined. Up to the early 2000s, the conventional wisdom among scholars of American politics held that parties had been in decline for decades (Bartels 2000; Norpoth and Rusk 1982). Data from multiple sources appeared to validate the thesis of parties in decline – the number of leaning Independents had doubled between 1960 and 1980; the proportion of "strong partisans" had declined since the 1950s; and split-ticket voting had reached an all-time high in the 1992 election cycle, which witnessed the most successful third-party candidacy in 80 years (Hetherington 2001, 619–20).

How, then, could affective polarization have arisen from these circumstances? A first piece of the puzzle is the disconnect between elite behavior and mass behavior during much of the 1980s and 1990s. Precisely as political scientists proclaimed that parties were in decline, party elites were polarizing. Measures of ideology based on roll-call votes in Congress revealed a growing chasm between Republicans and Democrats (Schaffner 2011). The ideological overlap between Democratic and Republican members of Congress disappeared: the most conservative Democrat was now less conservative than the most liberal Republican. As a result, the party "brands" became clearer than ever.[1] Insofar as the behavior of political elites is an important cue that influences the behavior of the public (Lenz 2012), the increasing elite polarization that began in the 1970s may help explain subsequent downstream polarization.

Whether or not the public polarized in ideological terms over the same time span is less clear and the subject of intense debate. On the one hand, proponents of the "minimalist" thesis held that despite elite-level polarization, most Americans remained ideologically moderate (Fiorina and Abrams 2008; Fiorina, Abrams, and Pope 2008). On the other hand, proponents of the "maximalist" view argued that the proportion of Americans identifying as ideological moderates decreased rapidly over the second half of the 20th century (Abramowitz and Saunders 2008). Insofar as citizens' affect toward political opponents is driven by (perceived) differences in ideology, mass ideological polarization may help explain the rising political acrimony that I will describe next.

2.2 The rise of affective polarization

While political scientists examined elite ideological polarization and argued about mass ideological polarization, a vast change in Americans' political attitudes went largely unnoticed. In a 2012 article, Iyengar et al. (2012, 406) pointed out that "policy-based division is but one way of defining partisan polarization." While an emerging gap in the ideological content of political parties was of obvious interest, Iyengar et al. argued that

[1] For reference, Figure A1 in the Appendix shows distributions of DW-NOMINATE scores (1st dimension) in the House of Representatives between the 81st Congress and the 116th Congress.

the *affective* distance between parties was also relevant. In other words, even if one were to believe the "minimalist" thesis emphasizing centrist voters that I outlined above, it could still be the case that the American public was polarizing if partisans' views of their opponents were growing more negative. This is precisely what Iyengar et al. (2012, 407) found: historical survey data going back decades clearly showed that "Democrats and Republicans not only increasingly dislike the opposing party, but also impute negative traits to the rank-and-file of the out-party."

Figure 1 presents the evolution of American citizens' feelings toward political parties in a manner similar to Iyengar et al. (2012) and Iyengar et al. (2019). For each edition of the ANES going back to 1980, respondents were asked to rate their feelings toward various groups, including the Democratic Party and the Republican Party, on a scale from 0 to 100 – the so-called "feeling thermometer". Panel A of the figure displays the mean scores attributed to the in-party (the party that the respondent identifies with) and the out-party (the major party that the respondent does not identify with), as well as the difference between the two. Panel B shows the difference in thermometer scores by stated interest in the presidential election for each year. For this analysis, only respondents who identified as Democrats or Republicans are included.[2]

Over the 40 years that this data covers, in-party affect has not changed in a substantial fashion: unsurprisingly, for each election, American partisans feel mostly positively about their own party, with mean ratings hovering beween the high 60s and the mid 70s. By contrast, feelings toward the out-party have grown much more negative, a trend that some scholars have referred to as "negative partisanship" (see Abramowitz and Webster 2016; Abramowitz and Webster 2018). In 1980, respondents' mean ratings of their opponents was just barely under water (46.6), which could be interpreted as apathy more than antipathy. It is only in 2004 that ratings of the out-party fell below 40. Subsequent elec-

[2]Note that prior to 1980, the ANES asked respondents about their feelings toward "Democrats" and "Republicans", rather than the more amorphous parties. Since we know that the specific question wording matters (see e.g. James N. Druckman and Levendusky 2019), the data presented here only extends back to 1980.

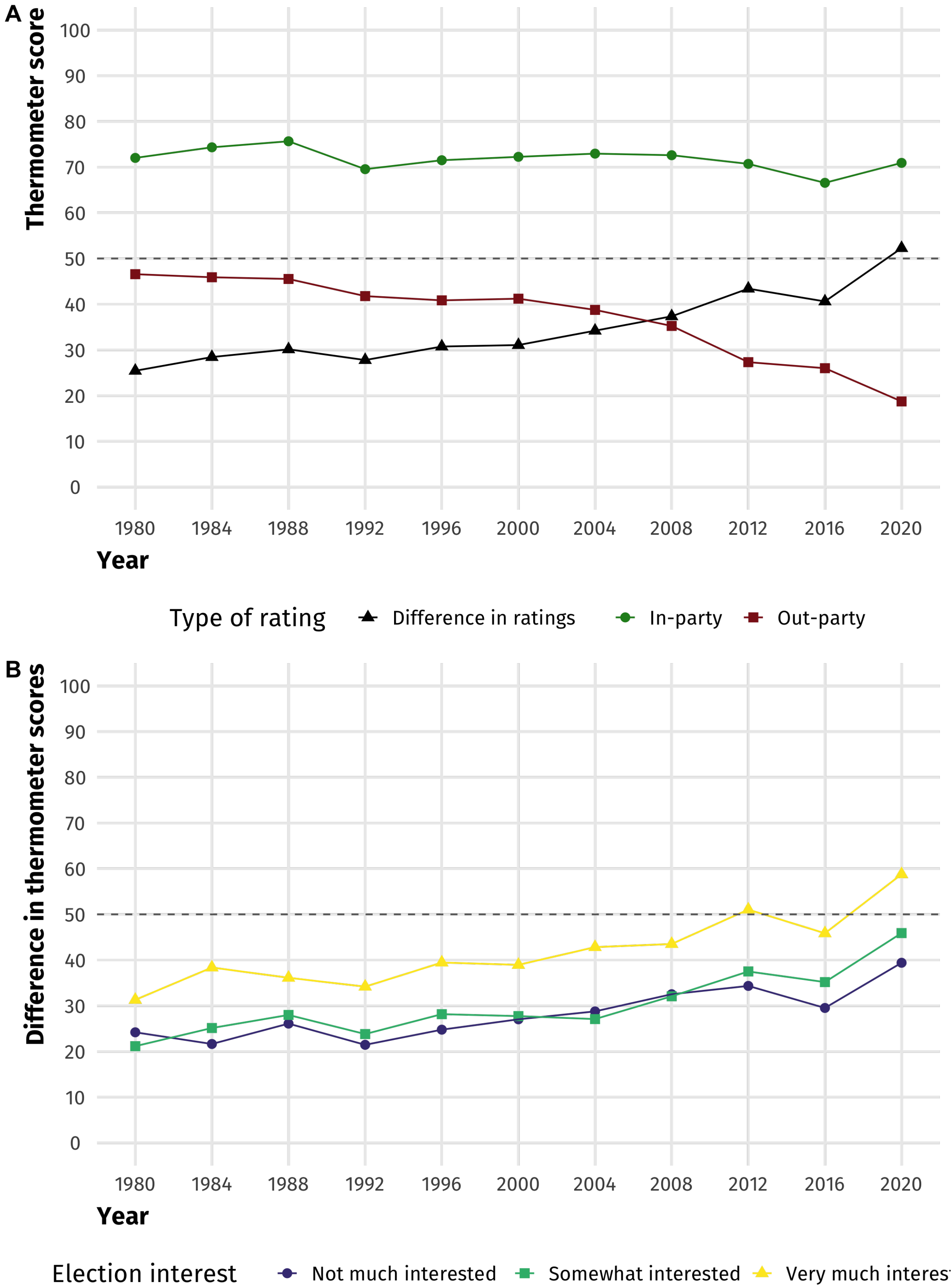

Figure 1: Evolution of thermometer score ratings of political parties, 1980-2020

tion cycles frequently witnessed unprecedented drops in out-party affect. For instance, from 2008 to 2012, the mean rating of the out-party decreased from 35.2 to 27.3, while the period from 2016 to 2020 saw a drop from 26.0 to 18.7.

The rise of affective polarization appears to be a broad-based phenomenon: while levels of affective polarization vary across groups, the historical trajectory is uniformly in the same direction. For instance, panel B shows that affective polarization among ANES partisans in 2020 who stated they were "not much interested" in the presidential election was higher than affective polarization among ANES partisans in 1980 who stated they were "very much interested" in the election. In 2020, among respondents who stated that they were "very much interested" in the election – 52.2% of the sample and presumably the most politically influential group in the electorate – feeling toward the out-party was a frigid 14.9. To contextualize just how low this figure is, consider that Republicans' mean affect toward illegal immigrants in 2020 stood at a comparatively lukewarm 32.9. After years of elite-level agitation regarding the threats posed by immigration and the presidency of a man whose central policy proposal involved the construction of a border wall to stem the flow of migrants, the Republican rank-and-file *still felt more negatively toward the Democratic Party than illegal immigrants*.

This worrying trend was easily observable from widely available data, but it was not subject to much scrutiny from political scientists until Iyengar et al. (2012) published their influential article. In the 10 years that have since elapsed, affective polarization has emerged as a major research topic for scholars of American politics – indeed, it is now "one of the most influential literatures in contemporary social science" (Broockman, Kalla, and Westwood 2023, 808). The study of affective polarization has also been extended to other contexts, with particular attention focused on European democracies. This comparative evidence shows that the over-time increase in affective polarization in the United States is not an isolated phenomenon, though its magnitude is noteworthy (Boxell, Gentzkow, and Shapiro 2022; Reiljan 2020; Garzia, Ferreira da Silva, and Maye

2023; Wagner 2021).

2.3 Shortcomings to keep in mind

Before going any further, it is worth reviewing two crucial questions that the extant literature on affective polarization has only partly addressed: the meaningfulness of survey answers in measuring affective polarization and the likely effects of affective polarization on politics. This project is not designed to answer these questions – indeed, if survey answers are not meaningful and/or affective polarization has no downstream consequences, one may not care much about the sources of affective polarization. Nonetheless, clearly delineating the relevance of extant empirical evidence can help contextualize this thesis.

2.3.1 Measuring affective polarization: are survey answers meaningful?

The vast majority of research examining affective polarization relies on various survey items to measure the key concept of interest. Typically, respondents are asked to rate how they feel about members of the in-party and members of the out-party, with the gap between the two ratings acting as the measure of affective polarization.[3]

These survey measures provide important evidence, but one must proceed carefully in interpreting self-reported attitudes in the context of a survey. There are two main concerns we must keep in mind.

First, massive changes in the polling industry in recent years may well cast doubt on the validity of over-time comparisons in public opinion. For decades following the emergence of high-quality public opinion surveys, pollsters using random-digit dialing

[3]A related issue is what survey respondents have in mind when they are asked to summarize their feelings toward the opposing party. American citizens tend to dislike parties and party leaders more than rank-and-file partisans. Therefore, asking about the latter produces lower estimates of affective polarization (James N. Druckman and Levendusky 2019). In addition, even when researchers make clear that they are asking about rank-and-file partisans, survey respondents tend to imagine a stereotypical supporter of the out-party. As Druckman et al. (2022, 1106) show, "many individuals express indifference, rather than hostility, once they are asked to evaluate the typical member of the other party."

to contact survey respondents enjoyed high response rates that minimized sampling bias (Hillygus 2011). In recent years, however, response rates have plummeted: the Pew Research Center (2019) reports that a typical telephone survey in 1997 would successfully reach 36% of individuals it tried to contact, compared to just 6% in 2019 (see also Leeper (2019)). Instead of the gold standard of random-digit dialing surveys, firms and academics have increasingly relied on opt-in surveys conducted on the Internet. The potential drawbacks of these non-probability samples have been hotly debated. Clearly, opt-in samples on the Internet tend to skew toward the highly educated and Democrats (in the United States). Nonetheless, it appears that the drawbacks on non-representativity are fairly limited when it comes to *experimental* studies: absent massive heterogeneity in treatment effects across subgroups that are severely over- or under-represented in online samples, estimates obtained from Amazon MechnicalTurk and similar platforms are usually comparable to estimates obtained from more conventional samples (Berinsky, Huber, and Lenz 2012; Coppock 2019; Coppock and McClellan 2019). However, the same is not necessarily true when one attempts to merely *describe* a population-level quantity. This issue has been prominently displayed in recent failures to predict election outcomes such as the Brexit vote in the United Kingdom and Donald J. Trump's victory in the U.S. 2016 presidential election (see, e.g., Kennedy et al. (2018)). The higher the correlation between selection into the survey and whichever population-level estimand we wish to estimate, the higher the risk that a survey estimate will be severely biased. In the case of affective polarization, there are good reasons to be concerned: contemporary surveys often over-represent precisely the sort of individuals who tend to report highly polarized affect: those who are highly politically interested. Indeed, there is evidence that the observed increase in ideological polarization in the United States is partly due to declining response rates over time (Cavari and Freedman 2018, 2023). Of course, this is not to say definitively that there has not been an over-time increase in affective polarization. However, the magnitude of such increase may be overstated. Readers should be keep this caveat in mind when interpreting the results presented below.

The second point I wish to make is that the very relevance of survey answers deserves scrutiny. For survey respondents, answering questions in a way that aligns with their partisanship may represent a relatively costless way of posturing and deriving some emotional benefit. Scholars have called this tendency "partisan cheerleading" or "expressive responding" (Bullock et al. 2015; Prior, Sood, and Khanna 2015; Bullock and Lenz 2019). Furthermore, expressing outrage about political opponents – or affection for political allies – may be seen as socially desirable, in which case surveys would tend to overestimate the prevalence of affective polarization (Connors 2023). Given this, should we be concerned about the substantive importance of survey responses? The evidence on this point is mixed. Survey respondents sometimes report beliefs that are less politically congenial when given monetary incentives to provide correct answers. However, other studies report minimal levels of expressive responding (see, e.g., Berinsky 2018; Graham and Yair 2023). Furthermore, real-world, high-stakes behavior appears to be influenced by partisanship in a way that is consistent with survey results, including romantic dating (Huber and Malhotra 2017). At the risk of oversimplifying a vast literature, we can summarize the measurement of attitudes as: there is more likely a there there, but it may be overestimated and can easily be misinterpreted.

2.3.2 *The consequences of affective polarization: why should we care?*

For all the attention that affective polarization has drawn, what do we actually know about its likely effects on the political landscape? The vast research effort poured into investigating the sources of affective polarization proceeds from a sometimes explicit, sometimes implicit assumption that affective polarization has downstream consequences. In other words, we care about affective polarization because it likely affects political outcomes that we care about.

Recent experimental evidence suggests that many of the assumed consequences of affective polarization may not manifest. Broockman, Kalla, and Westwood (2023) catalogue examples from over "a dozen studies that express concern about such potential political consequences of affective polarization." Political outcomes that are thought to be influenced by affective polarization range from electoral accountability to legislative bipartisanship and respect for democratic norms (Kingzette et al. 2021). There are, however, reasons to be skeptical that affective polarization matters in many cases. In perhaps the most convincing empirical investigation of the causal effects of affective polarization to date, Broockman, Kalla, and Westwood (2023) report that the effects of reducing affective polarization through experimental manipulation "are consistently null across outcomes and approaches for manipulating affective polarization." Using a similar experimental approach, Voelkel et al. (2023) report that depolarization interventions do not clearly reduce anti-democratic attitudes. In a study that encompasses eight European countries and the United States, Harteveld et al. (2022) show that reducing affective polarization – again in an experimental setting – has relatively limited immediate effects on a wide range of relevant outcomes. Of course, while experiments allow for clean causal identification, whether or not their reduction of affective polarization in an artificial setting is a good approximation of real-world processes is unclear. Furthermore, one may well consider affective distance between partisan groups to be inherently undesirable, regardless of its consequences; but it is important to keep in mind that the evidence on this front is quite thin.

3 The social sorting hypothesis

In an effort to better understand the rise of affective polarization and forestall further polarization, political scientists have sought to study its causes. The potential explanations are numerous, but this thesis will focus on the hypothesis centered around social sorting – the increasing socio-demographic homogeneity of political parties.

Partisanship has increasingly been theorized as a *social* identity (Green, Palmquist, and Schickler 2002). Based on foundational work in social psychology by Tajfel and Turner (1979), social identity theory posits that individuals self-categorize as a member of a particular social group and consequently "attempt to maximize differences between in-group and out-group" out of a "need for some positive distinctiveness" (Greene 1999, 394). In a way, affective polarization is an inevitable downstream consequence of humans' propensity to categorize: "the mere act of identifying with a particular group in competitive environments – no matter how trivial the basis for group assignment – is often sufficient to trigger negative evaluations of outgroups" (Iyengar and Westwood 2015, 691). Indeed, in researching the sources of human conflict, the presence of deep-rooted, intractable differences between groups did not appear to be necessary: the "minimal group" paradigm held that even arbitrary distinctions created from whole cloth by researchers could create an in-versus-them mentality and powerfully shape behavior (see e.g. Tajfel 1970).

Citizens' attachment to political parties can be viewed as one such social identity. Political parties play a central role in the organizing of mass democratic politics: as aptly put by Schattschneider (1942), "modern democracy is unthinkable save in terms of the parties." Yet partisanship is frequently blamed for some of most vexing pathologies of democracy. The diagnostic that political scientists have reached could be described in a very rough manner as follows: parties are a net positive and, in any case, inevitable in modern democracies.[4] In the words of Stokes (1999), many "seem to regard political parties as an unpleasant reality, a hardy weed that sprouts up in what would otherwise be the well-tended garden of democratic institutions." However, the structure of the party system must not be left to chance. On the one hand, the parties should maintain a robust national infrastructure and avoid the disorganization that follows from too much decentralization. In 1950, the American Political Science Association published a landmark report making these recommendations. On the other hand, and key to the subject of this

[4] See, e.g., Aldrich (1995) for a rational choice approach.

thesis, mass partisanship should be rooted in cross-cutting cleavages. Cross-pressures, as they are also called, refer to a situation whereby a voter faces conflicting influences as a result of politically relevant identities that draw them in different directions. A staple of early research on voting behavior (Lazarsfeld, Berelson, and Gaudet 1948; Campbell et al. 1960), the study of cross-pressures had "largely disappeared due to an accumulation of negative evidence" by the late 1970s (Mutz 2002, 839).[5]

In recent years, however, interest in cross-pressures has been reinvigorated by the worrying trends in affective polarization I outlined above. The guiding question is "what happens when a partisan social identity is joined by a host of other social identities that are strongly affiliated with the party" (Mason 2016, 353). The answer is a decrease in cross-cutting ties, which "tend to allow partisans to engage socially with their fellow citizens and partisan opponents" (Mason 2018a, 8). The idea that cross-cutting ties allow for more cordial interactions between groups in society is not new. When social groups neatly sort into different partisan groups – when religious individuals overwhelmingly become Republicans and urban dwellers overwhelmingly become Democrats – cross-cutting identities become rarer. As a result, "the various identities work together to drive an emotional type of polarization that cannot be explained by parties or issues alone" (Mason 2016, 353). If cross-cutting identities can act as an "emotional brake" that tames partisan emotions (Mason 2016, 353), the *absence* of cross-cutting identities suggests unbriddled partisanship.

4 Shortcomings of the extant evidence

In this section, I argue that the extant evidence on the role of social sorting in mass affective polarization suffers from several shortcomings. Some of my reservations have to do with the basic premise underlying the social sorting hypothesis, while others are related

[5]Note that early theoretical work on cross-pressures stressed its purported effect in depressing political participation.

to specific details of the most relevant empirical tests. The latter are, in my view, much easier to address and this is what I aim to achieve in this thesis. The former, by contrast, cast more fundamental doubt on the explanatory power of the social sorting hypothesis.

To be clear, social sorting may well be a contributor to affective polarization. However, I argue that the jury is still out. More work is needed before a definitive conclusion can be reached.

4.1 Theoretical shortcomings

Social sorting can only explain an over-time increase in affective polarization if it has itself been on an upward trajectory. The idea that Americans have sorted into immutable tribes has certainly captured the public imagination.[6] But for all of the catchy headlines, a basic, but incredibly important fact has been overlooked: the political behavior of Americans remains difficult to predict on the basis of socio-demographic characteristics alone. In an important recent article – and the closest comparative to this thesis –, Kim and Zilinsky (2022) analyze historic survey data and conclude that "the informativeness of demographic labels [is] surprisingly muted" and that "their predictive power has not increased over time." Their general approach is to fit logistic regression models and tree-based machine learning models on ANES data (1952-2020) in order to predict vote choice and party identification using a series of socio-demographic characteristics: race, education, income, age, and gender. Their models can, at best, predict 64% of two-party vote choices and party IDs in out-of-sample tests. This performance is an improvement over naive guesses based on two-party vote shares, but falls well short of what the neat picture of American political division seems to imply. To be sure, these sorts of analyses, including those presented in this thesis, are inherently limited by historical survey data: questions that have not been asked – in this case, individual-level socio-demographic

[6]For instance, at the time of writing, the Wall Street Journal had recently published a piece titled "Why Tribalism Took Over Our Politics" that featured Lilliana Mason's research.

characteristics that were not measured – can create a mismatch between ideal empirical specifications and what is feasible. It is possible that Americans have polarized along characteristic absent from Kim and Zilinsky (2022)'s analysis. Furthermore, it is plausible to think that Democrats and Republicans are polarized in terms of groups memberships that are more difficult to measure – such as preferences for certain music genres or sports – and that these group memberships deepen the ingroup-outgroup dynamic. However, the most detailed evidence to date shows that these "lifestyle" differences are fairly muted (Praet et al. 2022) and, to explain *rising* affective polarization, they would have to be rising themselves. Absent clear evidence that socio-demographic characteristics are much more predictive of partisanship than in the past or that gaps in lifestyle preferences have widened, the conclusion that hews closest to the underlying data is that Americans have socially sorted to a relatively weak extent.

How can American political behavior remain so difficult to predict? Recent years have, without a doubt, witnessed the emergence of new cleavages – or the intensification of long-standing ones – in American politics. For instance, as I will discuss later, it is only in the 1990s that religiosity became associated with the Republican Party (and a lack of religiosity or atheism became associated with the Democratic Party). Americans have most definitely "sorted" on a variety of dimensions. However, the appearance of these new division lines should not obscure the fact that other cleavages that were once prominent have waned and, in some cases, disappeared altogether. As an example, before the appearance of religiosity as a core dividing line in American politics, voters were instead "divided along mainly denominational lines. . . most Catholics were Democrats and most mainline Protestants were Republicans" (Margolis 2018, 13). To paint an accurate picture of social sorting, one must keep in mind that as cleavages emerge, others tend to weaken or disappear. On balance, the net change on the social roots of partisan identity may not be as strong as assumed. Indeed, it is likely that emerging divisions may come to mind more easily than faltering ones. This is why tracking the bases of party support across time is valuable; it is easy to point to an increasingly important cleavage, but more difficult to show that, on net, it contributes to a more divided political configuration.

4.2 Empirical shortcomings

The most comprehensive account of the role played by social sorting in the United States was provided by Liliana Mason in a 2018 book and a series of articles (Mason 2016, 2018a, 2018b; Mason and Wronski 2018). Using a combination of the ANES and original survey data, Mason demonstrates that social sorting is associated with a variety of political attitudes, including affective polarization.[7]

To measure social sorting, Mason focuses on six specific social identities: ideological liberal, ideological conservative, secular, evangelical, black, and Tea Party. For each of these six social identities, Mason asks survey respondents a battery of four items developed by Huddy, Mason, and Aarøe (2015) to measure subjective association with a given identity. The four items are listed in Table 1.

In order to link each of the six identities to a political party, Mason (2018a) uses "connections found in prior research and verified by examining the mean level of each identity for each party separately in the data" (p.66). She concludes that the liberal, secular, and black identities are aligned with the Democratic Party and that the conservative, evangelical, and Tea Party identities are linked with the Republican Party. As a result, a survey respondent who identifies with the Republican party is coded as socially sorted if they indicate strong subjective association with the latter three identities. On the contrary, a non-sorted Republican would indicate strong subjective association with the three identities that are linked to the Democratic Party. A citizen who is somewhere in between would indicate strong subjective association with some of the identities linked to the Re-

[7] To my knowledge, neither the data used in Mason's book and series of articles nor the replication code is publicly available. I exchanged emails with Dr. Mason to have access to the underlying data and code. After replying that she would send me the relevant materials, I never heard back despite a subsequent reminder. As a result, I am not able to conduct an exact replication of her work – or, indeed, to verify my own understanding of her statistical procedures by inspecting the replication code.

Table 1: The four items used by Mason in the social identity scale

Survey item	Answer choices
1. How important is being a [social identity] to you?	Extremely important Very important Not very important Not important at all
2. How well does the term [social identity] describe you?	Extremely well Very well Not very well Not at all
3. When talking about [social identity], how often do you use 'we' instead of 'they'?	All of the time Most of the time Some of the time Rarely Never
4. To what extent do you think of yourself as being a [social identity]?	A great deal Somewhat Very little Not at all

publican Party, but not all. Table 2 below reproduces a stylized example provided by Mason herself in her book's appendix.[8]

Table 2: Reproduction of Table A.3 in Mason (2018)

Individual A: Highest-score Republican	Individual B: Lowest-score Republican
Republican identity = 1 Democratic identity = NA Conservative identity = 1 Liberal identity = NA Evangelical identity = 1	Republican identity = 0 Democratic identity = NA Conservative identity = NA Liberal identity = -1 Evangelical identity = NA
Secular identity = NA Black identity = NA Tea Party identity = 1	Secular identity = -1 Black identity = -1 Tea Party identity = NA
Sorting score = (1+1+1+1)/4 = 1	Sorting score = (0-1-1-1)/4 = -0.75

A major strength of Mason's approach is that she is able to measure the subjective salience of social identities using the four-item battery first developed by Huddy, Mason, and Aarøe (2015). Indeed, as she notes in her book (pp.65-66), "the alignment between identities means nothing if a person does not identify with one or more groups." This

[8]Note that values of "NA" indicate that a respondent does not feel any association with a given identity.

is a point that bears emphasizing. Substantial evidence shows that people make sense of their identities and social affiliations in different ways. Most obviously, not all Latinos consider their Latino identity to be an important component of their politics; not all men consider their sex to be an important component of their politics; and so on. Less obviously, there is evidence that even identities that we may consider unambiguous and identities that are ascribed by the state are endogenous to politics (Egan 2020). Simply put, different people make sense of their group identities in different ways. As such, assigning "objective" affiliations to a given individual risks oversimplifying complex processes of identity formation. The measurement approach that I present in this thesis is, in this sense, fundamentally limited by historical survey data that typically does not ask about subjective identity strength. On this front, Mason's approach remains preferable. Nonetheless, there are other shortcomings to her approach, which I detail below before explaining my revised approach.

Mason uses a rather crude method to establish the identities that are linked to each party. To be sure, the broad characterizations that she makes are directionally sound – it is undoubtedly true that secular, black, and liberal individuals tend to be Democrats, while evangelical, Tea Party supporters, and conservative individuals tend to be Republicans. However, from an empirical point of view, her measurement lacks important nuance.

First, while the individual-level salience of a given identity is allowed to vary, its aggregate-level association with a political party is fixed and assumed to be of equal magnitude to that of other identities. Take, for instance, black identity and born-again Christianity. The latter is, without a doubt, a politically salient identity, but its importance pales in comparison to the enormous significance of black identity (on this point, see Westwood and Peterson 2020). Figure 2 summarizes this point. Between 1980 and 2020, born-again Christianity became an increasingly strong predictor of party identification among White ANES respondents. In the most recent election cycle, 43.1% of White respondents who did not identify as born-again indicated that they identified with the

Democratic Party; this proportion dropped to barely 20.4% among born-again Whites. Among Black respondents, on the other hand, born-again status did not predict partisan affiliation: 92.4% of non-born-again Blacks identified as Democrats, compared to 91.6% of born-again Blacks. Similarly, in the 2016 ANES, among Black respondents who identified as born-again Christians (n = 221), 93.8% identified with the Democratic Party.[9] Assuming a respondent whose answers on the four-item social identity scale reveal great attachment to both their black and evangelical identities, Mason's measurement scheme would categorize them as a weakly sorted voter – a dubious result since an overwhelming majority of such voters in the ANES sample identify with Democrats.[10]

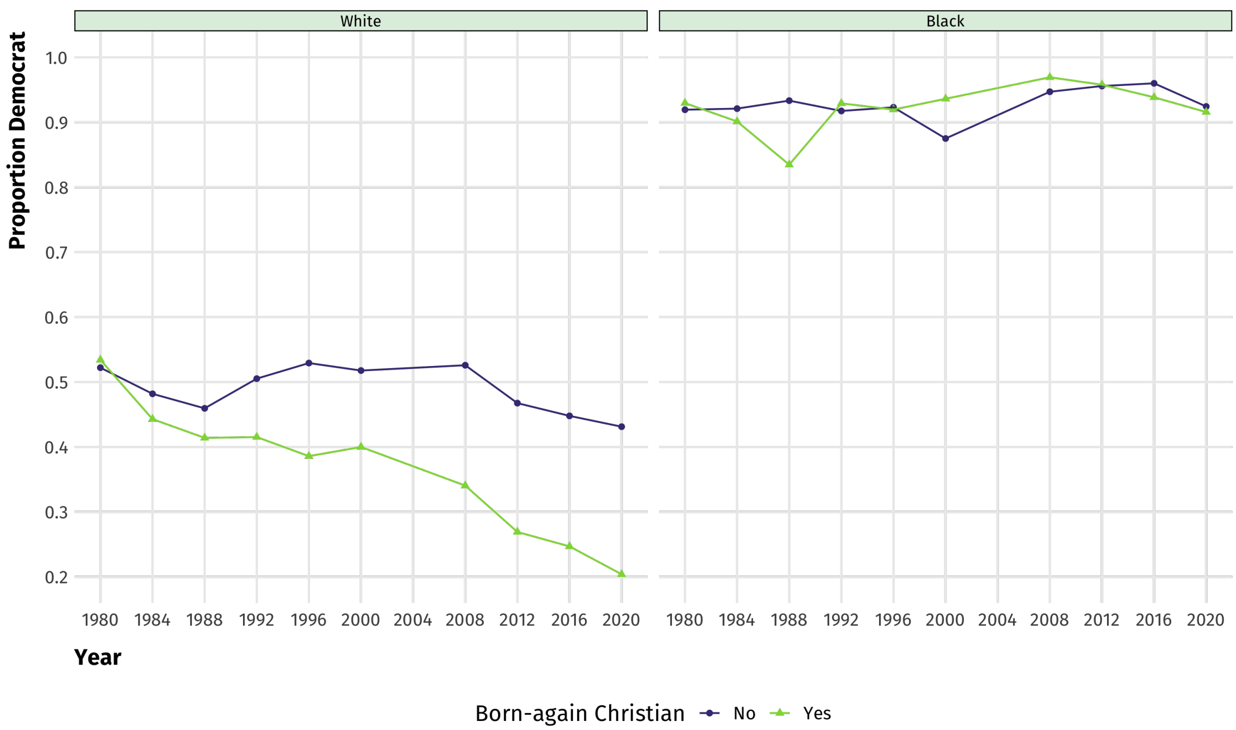

Figure 2: Proportion of Black and White ANES respondents who identify as Democrats, conditional on born-again status (1980-2020)

The conclusion is much the same when it comes to self-reported ideology (Figure 3).[11]

[9] The corresponding figure for White respondents is 24.7%.

[10] Referring to Table 2, a Black evangelical Republican would score 1 on evangelical identity and -1 on black identity, while a Black evangelical Democrat would score -1 on evangelical identity and 1 on black identity.

[11] This figure combines data from the 2012, 2016, and 2020 editions of the ANES in order to capture more

Unsurprisingly, Black and White respondents who indicate that they are "extremely liberal" overwhelmingly identify with the Democratic Party. As we move toward the conservative end of the traditional 7-point scale, White respondents become less and less likely to identify as Democrats; by the time we reach respondents who indicate that they are "conservative" and "extremely conservative" barely any Whites (approximately 3%) identify with the Democratic Party. The corresponding party-ideology gradient among Blacks is notably less steep. Among Blacks who say that they are "conservative" (n = 120), 73% identify with the Democrats. Simply put, *some identities overpower others* (see e.g. White and Laird 2020). Mason's approach cannot account for this simple fact.

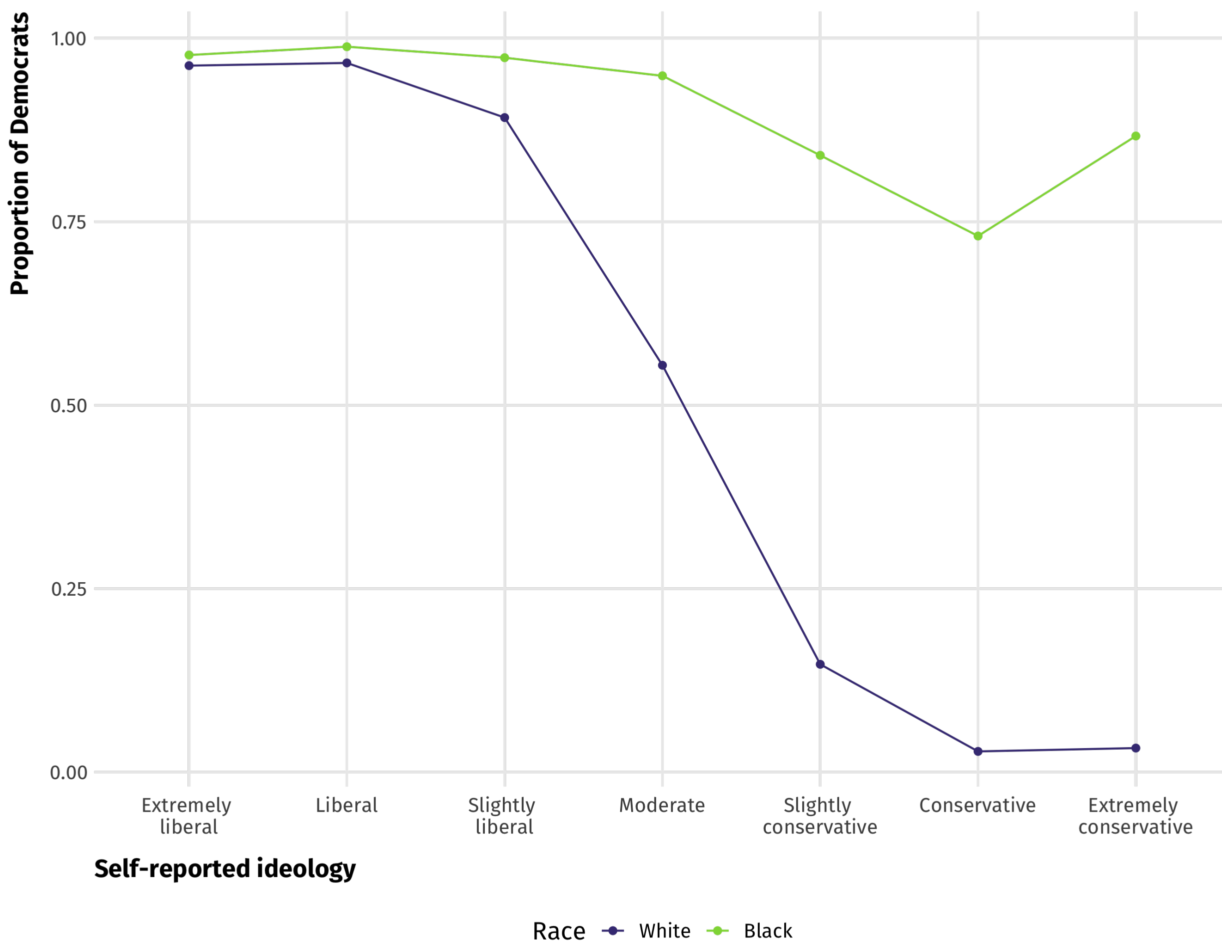

Figure 3: Proportion of Black and White ANES respondents who identify as Democrats, conditional on self-reported ideology (2012-2020)

precise subgroup estimates.

Second, Mason's coding scheme cannot incorporate changing social cleavages over time. The current political fault lines that she correctly identifies have not always been present. For instance, the hardening religious-partisan divide in the United States is a relatively new phenomenon: until the last 30 years or so, there was "virtually no religiosity gap between the parties" (Margolis 2018, 27). Other cleavages have waned over time. If we are to measure the *evolution* of social sorting over time, we must allow the associations between party identification and social identities to wax and wane. Consider the example of a highly religious American voter in the late 1980s. Using Mason's coding scheme, an analyst seeking to measure the extent of social sorting would consider religiosity to be associated with the Republican Party. Consequently, a survey respondent who indicated affinity with the Republican Party and a strong religious identity would be considered highly sorted. This simple categorization, however, obscures the fact that the religiosity gap between Democrats and Republicans – however well established it is now – is a relatively new phenomenon. Until the early 1990s, higher proportions of the Democratic base were composed of Biblical literalists (Margolis 2018, 26). Applying Mason's measurement scheme to historical data would then risk misclassification of respondents due to changing cleavages. In order to trace the evolution of social sorting, one must use a more flexible procedure.

5 Empirical approach

In light of the limitations outlined above, I aim to use a more flexible way of measuring social sorting. To be useful, this new measurement scheme must: a) allow for social cleavages to have varying importance in any given time period; b) allow for social cleavages to change over time.

I propose to use historical data from the American National Election Study (ANES) in order to build individual-level estimates of cross-pressures that emanate from different social identities, based on work by Brader, Tucker, and Therriault (2014).[12] The procedure is as follows. The first step is to estimate regression models that relate socio-demographic characteristics – language, region, social class, ethnicity... – to party identification.[13] Given that I will be working with historical data, these models will be fully interacted with a *year* variable in order to allow the estimated coefficients to vary by year – in other words, in order to allow the effect of social identities on party identification to wax and wane. The *year* variable is specified as a factor variable, such that marginal effects unique to each year is estimated and there is no assumption regarding the functional form of the over time differences in marginal effects. With this specification, we would expect to see, e.g., the marginal effect of religiosity to be quite weak, or even close to 0, in the 1970s and 1980s, before rising in the 1990s and beyond. The second step is to produce predicted probabilities of supporting different parties for each survey respondent. Finally, I will use the variance in predicted probabilities for a given respondent in order to estimate the intensity of cross-pressures that emanate from the social group memberships that were used as input to the regression models.[14]

The intuition behind this model is quite straightforward. If a respondent is highly socially sorted, it should be quite easy to predict with reasonable certainty with which political party they identify on the basis of the social groups they belong to. By contrast, if a respondent is not socially sorted, the model should produce ambiguous results: we will be unsure, on the basis of the respondent's social identities, which party they identify with. As a simple example, imagine a respondent who is assigned, based on the statistical model outlined above, a 0.55 probability of identifying as a Democrat and a 0.45 proba-

[12]Note that I consider cross-pressures to be inversely related to social sorting. If voters are socially sorted, their social identities should produce few cross-pressures; if voters are not socially sorted, their social identities should produce more cross-pressures.

[13]I use socio-demographic characteristics because these are widely available in historical survey data and come reasonably close to the theoretical construct of social identity. Of course, some salient social identities are not rooted in socio-demographic characteristics. Nonetheless, many relevant identities can be measured using widely-used survey items. Additionally, note that the ANES data does not allow me to measure the subjective salience of social group memberships. This is a limitation of using historical data that must be kept in mind.

[14]In a two-party context, taking the simple difference in probabilities is sufficient.

bility of identifying as a Republican. This respondent faces greater conflicting pressures on the basis of his or her social identities than another respondent who is assigned a 0.90 probability of identifying as a Democrat and a 0.10 probability of identifying as a Republican (see Table 3 below for a stylized example). Accordingly, we will consider the latter a socially sorted individual and the former a non-sorted individual.

Table 3: Stylized example showing one highly sorted voter and one cross-pressured voter

Race	**Gender**	**Religiosity**	**Education**	**Pr(Democrat)**	**Pr(Republican)**	**Δ probability**
White	Male	Very religious	High school	0.10	0.90	0.8
Black	Female	Fairly religious	High school	0.55	0.45	0.1

By leveraging ANES data going back decades, I am able to trace the history of cross-pressures emanating from social group memberships. I use the following measures from the cumulative ANES data file:

1. Income, measured in five discrete percentile buckets
2. Educational attainment
3. Race
4. Religious attendance
5. Union membership

In some models, I also consider self-reported ideology measured on the classic 7-point scale. The final step will be to directly examine the key relationship that makes social sorting a relevant object of study: the idea that highly sorted partisans are less likely to be able to engage socially with their political opponents and therefore more likely to be affectively polarized. I will retain the cross-pressure scores introduced above as a key variable and introduce it on the right-hand side of a regression model that seeks to predict affective polarization, measured as the difference in feeling thermometer scores between a respondent's in-party and out-parties. I will run a series of cross-sectional linear regression models in order to establish whether the relationship between social sorting and affective polarization: a) is substantively and statistically significant; b) fluctuates through time.

6 Results

6.1 Descriptive evidence

In order to get a sense of "what goes with what", I begin by displaying the average marginal effects (AMEs) of the various social characteristics utilized on the right-hand side of the logistic regression model. Since the regression model interacts each characteristic with a year indicator, the quantities depicted in Figure 4 are specific to each year. The darker points represent AMEs from earlier years and the lighter points represent AMEs for more recent editions of the ANES. The AMEs are computed using "counterfactual" versions of the underlying survey data: for each observation $i \in [1, 2, 3, ..., n]$, I use the regression model to estimate the change in probability of Democratic identification when moving from the reference category of a given covariate to some other value. The quantities depicted in Figure 4 are the mean estimated counterfactual changes across all observations.[15]

Overall, the results presented in Figure 4 show a party system whose structure changed substantially over the years. While contemporary political cleavages may seem like unmovable features of a calcified political landscape, it is evident that the social bases of party support are in fact quite fluid. The first noteworthy takeaway is that race and attendance to religious services have become vastly more predictive of party identification over time. In 1972, the first election year for which we have data on religiosity, moving from the reference category of "never" attending religious services to attending religious services "every week" is estimated to decrease the probability of Democratic identification by 0.03, a substantively small quantity. The association between religiosity and party identification remained weak throughout most of the 1970s and 1980s, but by 1996, the estimated effect of moving from lowest to highest religiosity had ballooned to -0.20. In the

[15] See Figure 8 in the appendix for an alternative presentation of the same results.

2020 election, the effect is estimated to be a 0.30 decrease in the probability of Democratic identification – a substantively enormous effect. As stated previously, a major strength of the modeling approach employed in this thesis is its flexible nature: contra Mason (2018a), I do not have to arbitrarily decide which social attributes go with which party. The estimated effects quoted above demonstrate the advantages of this flexibility. Relatedly, *overspecification* is much less of a concern with this approach. If a particular social characteristic is not a strong predictor of party identification, including it in the regression models should not skew our results.

The story is similar, though less drastic, when it comes to race. Black and Hispanic identification is increasingly predictive of Democratic partisanship, which implies that Caucasian identification is increasingly predictive of Republican partisanship (White and Laird 2020). Note, however, that the strength of the Black constituency within the Democratic Party was already well established in the 1970s. The patterns observed here have more to do with reinforcement of an existing cleavage rather the creation of a new one, as observed with religiosity.

When it comes to gender, early elections in the dataset show little explanatory power of gender, while more recent electoral cycles show that women have become associated with the Democratic Party. This pattern tracks with comparative evidence of a changing ideological gender gap across Western democracies, with women growing relatively more liberal over time (Dassonneville 2021).

Have Americans' social identities truly sorted with their partisan identities? For each ANES respondent, I computed the mean difference in the estimated probability of identifying with the Democratic Party and the estimated probability of identifying with the Republican Party. This quantity can be expressed as follows:

$$E[Pr(PID_i = Democrat|X_i)] - E[Pr(PID_i = Republican|X_i)],$$

where X_i is a vector of socio-demographic characteristics included in the logistic re-

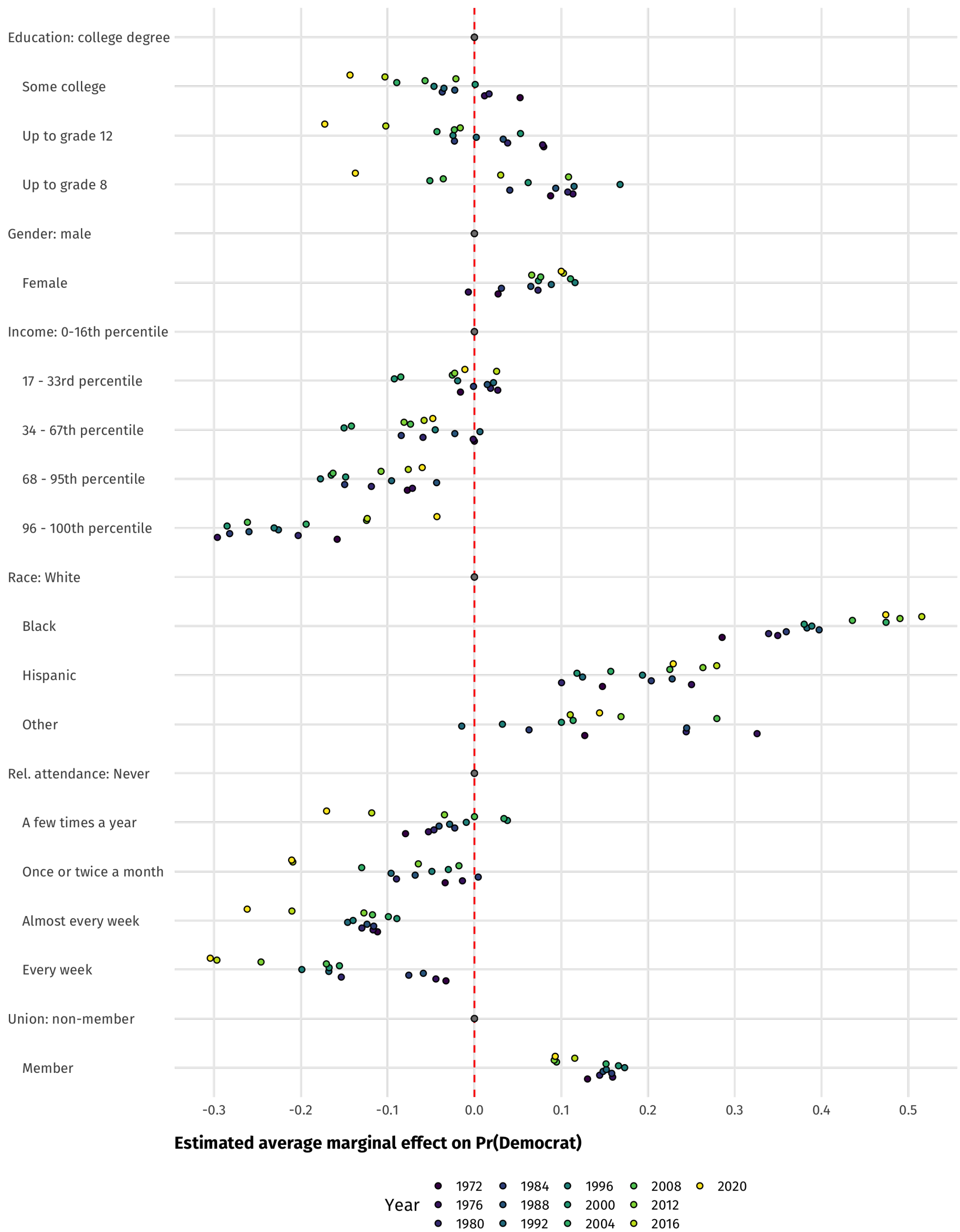

Figure 4: Estimated average marginal effects of social attributes on the probability of Democratic identification, 1972-2020

gression model described above. The range of this quantity is bounded by 0, in which case the regression model estimates an equal probability that some respondent i identifies with the two parties, to 1, in which case the regression model estimates with certainty that some respondent i identifies with some party.

In order to demonstrate what the results look like and examine face validity, I divided respondents into four quartiles of social sorting (with the quartiles computed separately for each edition of the ANES). Then, I drew one random respondent from each quartile for each survey year. Table 4 shows the respondents that were randomly drawn. As an example, the first row in the table shows a random respondent in the 1st quartile of sorting in 1972: a white male respondent with some college education, average income, no union affiliation, and infrequent attendance to religious services. The regression model estimates with 0.51 probability that this respondent will identify with the Democratic party, which mechanically implies a 0.49 probability of identifying with the Republican party. Taking the difference, between these two probabilities, the model estimates a "sorting score" of 0.01 for this particular respondent[16], meaning that it is quite difficult to predict their partisan affiliation. In the terminology of social sorting theory, the socio-demographic characteristics of this ANES respondent do not pull them in any obvious political direction. By contrast, the fourth row of Table 4 shows a respondent in the highest quartile of social sorting in the 1972 ANES. This respondent is a white woman with a high school education, average income, a union affiliation, and frequent religious attendance. The regression model estimates with 0.74 probability that this respondent will identify with the Democratic party, and therefore a 0.26 probability of identifying with the Republican party. Again, taking the difference, we arrive at a social sorting score of approximately 0.49. Are these predictions "correct"? Of course, it is impossible to say: we only observe each respondent once and the model may be misspecified in myriad ways. Nonetheless, the results demonstrate high face validity: given background knowledge of American

[16]Note that due to rounding, the difference in probability quoted is not exact.

political parties and their bases of support, most the examples presented below make intuitive sense.

Table 4: Example respondents from the ANES

Education	Gender	Income	Race	Rel. attendance	Union	Pr(D)	Pr(R)	Sorting
1972								
Some college	Female	68-95th p.	White	Every week	0	0.50	0.50	0.01
Up to grade 12	Female	0-16th p.	White	A few/year	0	0.57	0.43	0.14
Up to grade 12	Male	68-95th p.	White	A few/year	1	0.60	0.40	0.19
Up to grade 12	Female	68-95th p.	White	Every week	1	0.67	0.33	0.35
1976								
Up to grade 12	Male	34-67th p.	White	Every week	0	0.58	0.42	0.16
Up to grade 8	Male	0-16th p.	White	Never	0	0.67	0.33	0.35
Some college	Female	96-100th p.	White	Never	0	0.23	0.77	0.53
Up to grade 12	Female	34-67th p.	White	1-2/month	1	0.77	0.23	0.54
1980								
Some college	Female	68-95th p.	White	A few/year	0	0.51	0.49	0.03
Up to grade 12	Female	34-67th p.	White	Never	0	0.66	0.34	0.32
College degree	Male	96-100th p.	White	Almost every week	0	0.24	0.76	0.51
Up to grade 12	Male	0-16th p.	Black	A few/year	0	0.92	0.08	0.84
1984								
College degree	Male	68-95th p.	White	Never	0	0.43	0.57	0.14
Up to grade 12	Male	34-67th p.	White	A few/year	1	0.63	0.37	0.26
Up to grade 12	Male	0-16th p.	Other	Never	0	0.64	0.36	0.27
Some college	Male	34-67th p.	White	Almost every week	0	0.34	0.66	0.33
1988								
College degree	Female	68-95th p.	White	Every week	1	0.57	0.43	0.14
College degree	Female	34-67th p.	White	1-2/month	0	0.41	0.59	0.17
College degree	Male	68-95th p.	White	Every week	0	0.33	0.67	0.34
Up to grade 12	Female	17-33rd p.	Black	Every week	0	0.88	0.12	0.76
1992								
Some college	Male	34-67th p.	White	Never	0	0.53	0.47	0.07
College degree	Female	34-67th p.	White	Never	0	0.67	0.33	0.34
Some college	Male	96-100th p.	Other	Almost every week	1	0.27	0.73	0.46
Some college	Male	68-95th p.	Black	Never	1	0.93	0.07	0.87
1996								
Up to grade 12	Female	17-33rd p.	White	Every week	0	0.47	0.53	0.07
Some college	Female	34-67th p.	White	1-2/month	0	0.59	0.41	0.17
College degree	Male	0-16th p.	White	Every week	0	0.38	0.62	0.23
Up to grade 12	Female	68-95th p.	White	Never	1	0.72	0.28	0.43
2000								
Some college	Male	17-33rd p.	White	Never	0	0.52	0.48	0.04
College degree	Male	34-67th p.	White	Never	0	0.45	0.55	0.10
Up to grade 12	Male	34-67th p.	White	A few/year	0	0.56	0.44	0.12
Up to grade 12	Female	0-16th p.	White	A few/year	0	0.79	0.21	0.57
2004								
Up to grade 12	Male	34-67th p.	White	Never	0	0.41	0.59	0.17
College degree	Female	34-67th p.	White	A few/year	0	0.64	0.36	0.28
Some college	Male	17-33rd p.	White	Every week	0	0.26	0.74	0.48
Up to grade 12	Female	0-16th p.	Black	Every week	0	0.96	0.04	0.92
2008								
Some college	Female	68-95th p.	White	Never	0	0.46	0.54	0.08
Some college	Female	34-67th p.	White	Never	0	0.58	0.42	0.16
Up to grade 12	Female	34-67th p.	White	1-2/month	1	0.72	0.28	0.45
Up to grade 12	Female	0-16th p.	Black	A few/year	0	0.98	0.02	0.96
2012								
Some college	Female	68-95th p.	White	Never	0	0.52	0.48	0.04

Table 4: Example respondents from the ANES *(continued)*

Education	Gender	Income	Race	Rel. attendance	Union	Pr(D)	Pr(R)	Sorting
Up to grade 12	Male	68-95th p.	Other	Every week	1	0.43	0.57	0.15
College degree	Female	34-67th p.	Hispanic	Almost every week	1	0.80	0.20	0.60
College degree	Male	34-67th p.	Black	Never	1	0.98	0.02	0.96
2016								
Up to grade 12	Female	34-67th p.	White	A few/year	1	0.56	0.44	0.11
Some college	Male	34-67th p.	White	Never	0	0.43	0.57	0.14
Up to grade 12	Female	34-67th p.	White	Every week	0	0.23	0.77	0.55
College degree	Male	68-95th p.	Hispanic	Never	0	0.81	0.19	0.62
2020								
Up to grade 12	Female	68-95th p.	White	Never	0	0.48	0.52	0.03
College degree	Male	68-95th p.	White	Never	0	0.57	0.43	0.14
College degree	Male	34-67th p.	White	Almost every week	0	0.29	0.71	0.42
College degree	Female	17-33rd p.	White	Never	0	0.74	0.26	0.47

Note:
For each year, a random respondent was drawn from each quartile of the sorting score.

Figures 5 presents, for each presidential election year, the mean social sorting score estimated using the procedure described above. Panel A is based on regression models that do not include self-reported ideology, while Panel B is based on regression models that do include self-reported ideology.

The first point I wish to emphasize is a crucial one: for the typical ANES respondent, across most of the election years analyzed here, it is actually quite difficult to predict party identification on the basis of socio-demographic characteristics alone and without taking self-reported ideology into account. Take the estimates for the most recent presidential election as an example: a mean difference in estimated probabilities of around 0.36 implies that the typical respondent is assigned a 0.68 probability of identifying with one of the two major parties and, on the reverse, a 0.32 probability of identifying with the other major party. Clearly, socio-demographics characteristics are informative, but even in the highly polarized times we live in, party identification is far from a deterministic function of group memberships (or at least the ones measured and considered here). By contrast, if we take self-reported ideology into account, it becomes much easier to sort voters: a mean difference in probabilities of around 0.75 in the 2020 cycle implies the typical respondent is assigned a 0.88 probability of identifying with one of the two major parties and, on the

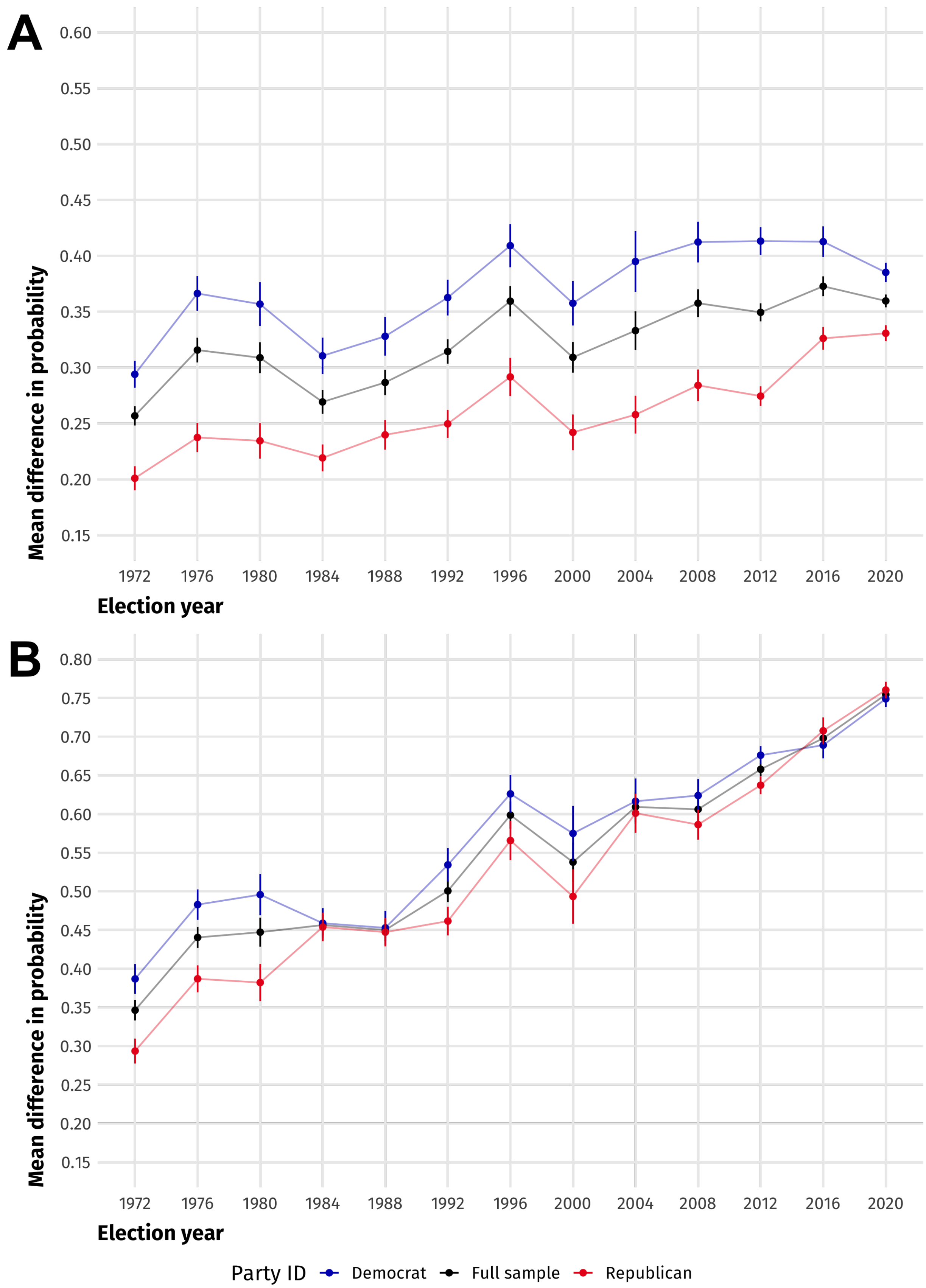

Figure 5: Mean difference in estimated probabilities, by election year and party ID (1972-2020)

reverse, a 0.12 probability of identifying with the other major party.

The second interesting result is the trend over time. The narrative of substantial social sorting in the last few decades is not obviously confirmed. It is true that estimates from recent elections tend to be higher than older estimates. However, all estimates produced using models without self-reported ideology lie between 0.26 (1972) and 0.36 (2020). The former estimate represents a 0.63-to-0.37 split, while the latter estimate represents a 0.68-to-0.32 split. Given modeling uncertainty[17], it is not clear that this difference is meaningful, even though it reaches conventional levels of statistical significance. To be clear, I do not claim that this is definitive evidence against the social sorting hypothesis. However, the results presented here cast doubt on the more extreme version of the descriptive claim that underlies the social sorting theory.

In models that take into account self-reported ideology, the conclusion is notably different. A clear upward trajectory is visible throughout the years. In 1972, the model returns a mean difference in probabilities of 0.35, which is approximately equivalent to a fairly ambivalent 0.62-to-0.38 split. By contrast, in the most recent election cycle, the model returns a mean difference in probabilities of 0.75, which is equivalent to a 0.88-to-0.12 split.

Moving beyond measures of central tendency, Figure 6 shows the boxplots and density plots of the social sorting measure for each year of the ANES survey. We can discern a recurrent pattern: the distribution of social sorting scores is, year after year, right-skewed. Most ANES respondents are located in the lower end of the scale, between 0 and 0.30. A minority of respondents at the upper end of the scale bring the mean in their direction.

Who are these respondents populating the upper end of the social sorting measure? For the most part, this upper end is composed of Black Americans. As stated previously, race has long been a central organizing feature of American politics. Figure 7 shows mean social sorting scores by self-reported race, by ANES edition. Republicans and White

[17] For instance, a cleavage that was more relevant in 1972 than in 2020 may have been omitted, which would lead to an underestimation of sorting in 1972, relative to 2020.

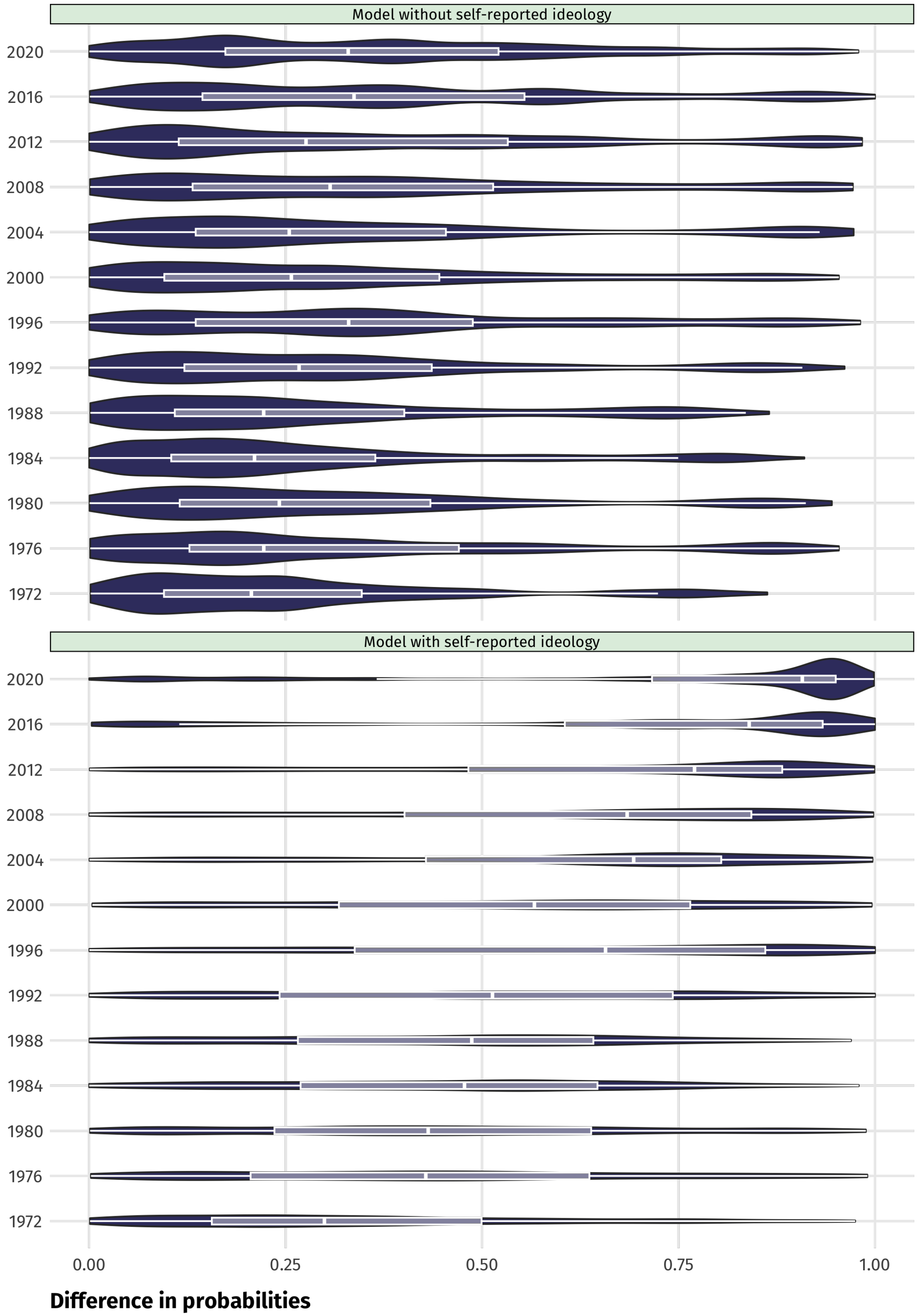

Figure 6: Boxplots and violin plots of the distributions of estimated differences in probabilities, conditional on election year

Democrats are, across all years, quite weakly sorted, with scores ranging from approximately 0.16 to 0.34. A weak upward trend is visible. By contrast, Black Democrats are consistently highly sorted, a consequence of the fact that the overwhelming majority of Black Americans support the Democratic Party. In the case of Black Democrats, there is no obvious increase or decrease through time.

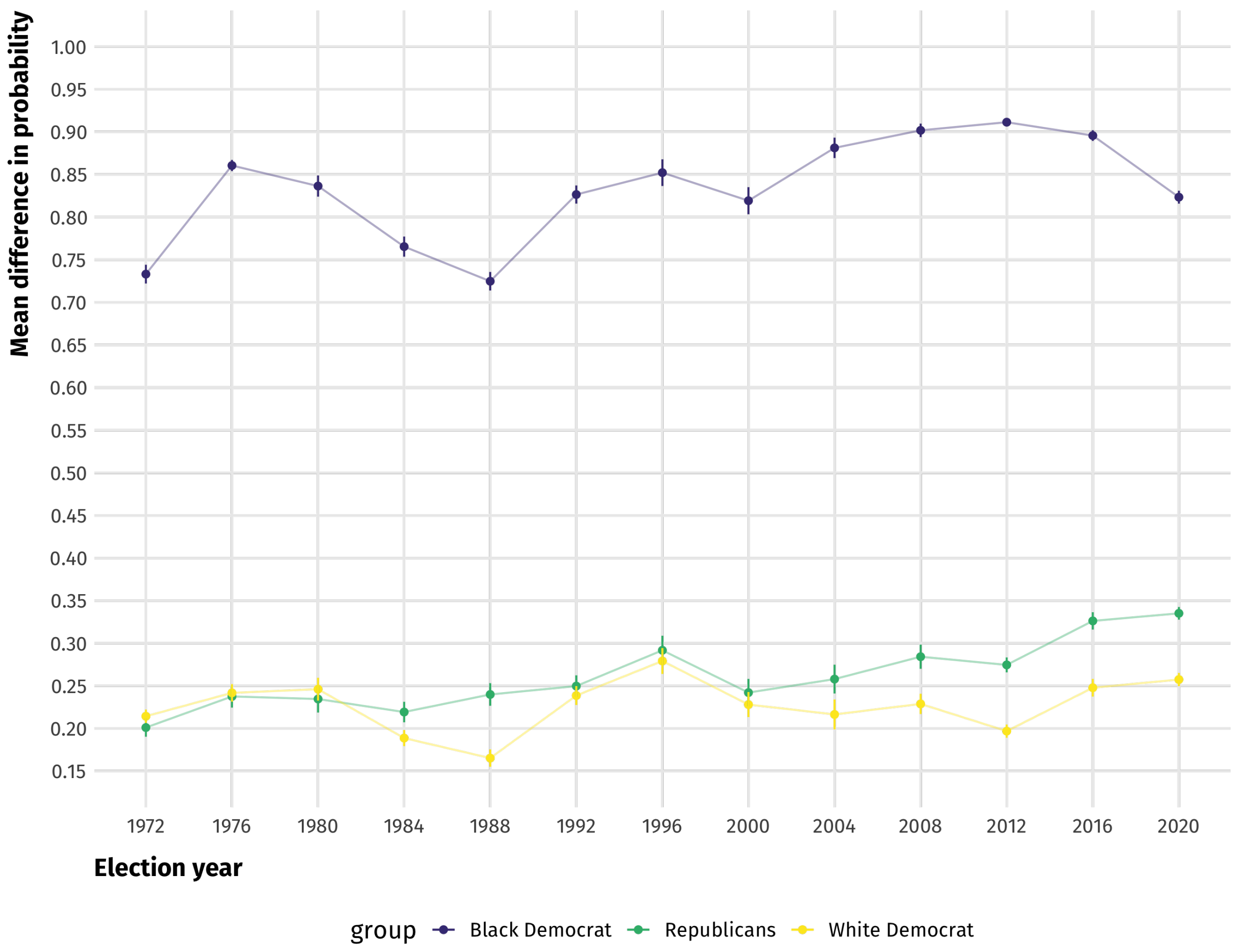

Figure 7: Mean social sorting score by year among different partisan and racial groups

What should we make of the large differences between models that include the measure of self-reported ideology compared to models that exclude it? It is not clear what our substantive conclusion should be. On the one hand, ideology is a salient social identity. As Americans have increasingly sorted into ideologically homogeneous parties, it is likely that people's mental images of the parties' compositions has been affected: a conservative Republican may well dislike Democrats because "they" are all liberals, while a liberal Democrat may well dislike Republicans because "they" are all conservatives. It is plausible that ideological sorting reinforces to perceive the opposing party as the "other."

On the other hand, I would argue that ideological sorting carries fundamentally different implications compared to more strictly *social* sorting. Some amount of ideological sorting, most political scientists would agree, is desirable: having political parties with a well-defined ideological "brand" can help voters make better electoral choices. A party system without any clear *ideological* cleavages is, in a sense, unstable at its core – falling apart at the seams due to competing internal factions. A well-functioning party system presupposes at least some degree of ideological segregation. By contrast, strictly *social* sorting is neither inevitable nor necessary.

In sum, the evidence so far suggests that social sorting may have increased by some moderate amount over time, but our ultimate conclusion depends on whether we consider ideological sorting to be akin to social sorting. An over-time increase in social sorting, however, is just the first step in the larger argument that links social identities to partisan animosity. In the next section, I examine the association between individual-level social sorting and affective polarization.

6.2 Associational evidence

In the preceding section, we established that the process of partisan-ideological sorting whereby members of the American public come to identify with the party that is closest to their ideological leanings is mostly responsible for any "social sorting" that has occurred, if any. Is this sorting associated with affective polarization, as the social sorting hypothesis posits?

I measure affective polarization using the standard thermometer ratings from the ANES. A respondent who identifies with the Democratic Party is affectively polarization insofar as their 0-100 rating of the Democratic Party exceeds their rating of the Republican

Party (and vice-versa).

Figure 8 shows the relationship between the two measures using a scatterplot, with the social sorting scores on the x axis and affective polarization on the y axis. As a reminder, social sorting ranges from 0 to 1 while affective polarization ranges from 0 to 100. Note that for this analysis, only respondents who identify as Democrats or Republicans are included, since the affective polarization measure can only be computed for those who identify with an in-party. Naturally, Independents are also politically consequential and we must be careful not to draw conclusions that may be biased due to compositional effects. Table 7 in the appendix shows the percentage of ANES respondents who identify as Democrats, Independents, and Republicans.

In Figure 8, the raw data at the respondent-level is shown in small gray points. Given the density of the data – thousands of respondents for each ANES edition, with some substantial overlap on both the x and y axes – I also plotted binned conditional means (a so-called "binscatter") using the `binsreg` function in R (orange points). This procedure is quite simple: it divides the data into $J < n$ bins using the empirical distribution of x, then computes the mean outcome within each bin. The number of bins is determined algorithmically in order to optimize integrated mean square error, hence why the number of bins varies by ANES edition (for more details on this procedure, see Cattaneo et al. 2023). The figure also shows predictions from a linear regression model fit to the full data with a second-order polynomial (blue line). Social sorting theory usually does not predict a specific functional form for the relationship between social sorting and affective polarization. For the most part, it appears reasonable to assume a roughly linear association. However, I nonetheless introduce a squared term since this is a first look at the data and I want to be able to detect any non-linearity.

The regression take the following form:

$$Polarization_i = \beta_0 + \beta_1 Sorting_i + \beta_2 Sorting_i^2 + \epsilon_i$$

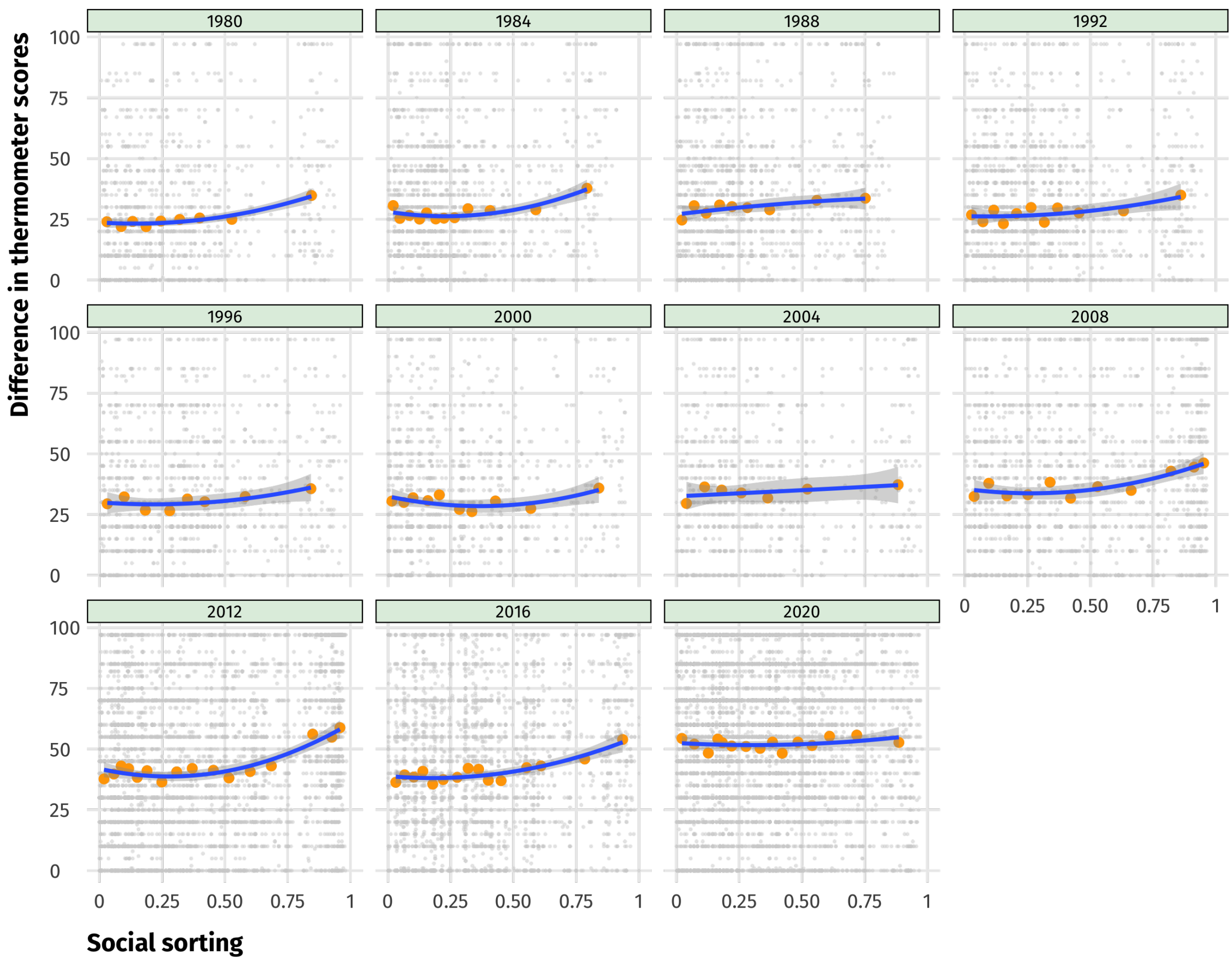

Figure 8: Scatterplot showing the relationship between social sorting and affective polarization, 1980-2020

Overall, the relationship between my measure of social sorting and affective polarization appears fairly weak: most years show a positive and, importantly, a possibly non-linear association. Moving from the most cross-pressured respondents (the least socially sorted) to those who are moderately cross-pressured yields relatively little movement on the affective polarization measure. As an example, let's examine the data from 1980 more closely. The mean level of affective polarization in the first bin of social sorting (which covers social sorting scores falling between 0 and ~0.061) is approximately 23.98. Between the second and eighth bins – which cover social sorting scores that range from ~0.061 to ~0.695 – the mean level of affective polarization vacillates very little, with a minimum of 21.99 (2nd bin) and a maximum of 25.53 (7th bin). The 9th and final bin, however, stands out, with a mean affective polarization of 34.77 – nearly 10 points higher than any other bin. The same pattern can be found in other years. In 1984, for instance, the affective polarization of the 1st through 11th bins varies from 25.06 to 30.63, while the 12th and final bin registers 37.76. The more recent election cycles of 2012 and 2016 show a similar pattern. In other election cycles, the relationship is very weak. For instance, in 2020, the line of best fit from a regression with a second-order polynomial is essentially flat.

To sum up, putting aside concerns about causality, it appears that insofar as there is an effect of social sorting on affective polarization, it manifests mostly at the upper end of the social sorting scale. In other words, the moderately sorted individuals do not differ appreciably in their attitudes than the weakly sorted individuals. However, the highly sorted individuals – those for whom predicting party identification on the basis of socio-demographic characteristics is quite easy – tend to be more affectively polarized.

The results are presented in tabular form in Table 5. Model 1 shows a linear regression model pooling respondents from all ANES editions with year fixed effects. Model 2 shows a similar model, but this time including a battery of covariates: education, race, income, religiosity, party identification, age, and interest in the election.[18] Model 3 adds

[18]Note that I avoid presenting the coefficients attached to the various covariates because of the so-called "Table 2 fallacy" – social scientists' propensity to interpret the coefficients attached to potential confounders

Table 5: Linear regression models relating social sorting to affective polarization

	(1)	(2)	(3)
(Intercept)	21.920 [20.113, 23.727]	13.203 [10.626, 15.780]	7.619 [−45.321, 60.559]
Sorting	10.519 [9.129, 11.909]	5.011 [2.859, 7.163]	3.325 [0.716, 5.934]
Num.Obs.	26 764	25 542	14 831
Year FE	✓	✓	✓
RMSE	30.45	29.54	26.94
Standard controls		✓	✓
PID strength control			✓

* $p < 0.05$, ** $p < 0.01$

strength of partisan identity, which is undefined for Independents and the reason why the sample size for this model decrease drastically.[19] Across the three models, social sorting is positively and statistically significantly associated with affective polarization. However, the magnitude of the coefficient[20] is much smaller in models 2 and 3. Model 1 estimates that a one-unit increase in social sorting – that is, a movement from minimum to maximum – is expected to increase affective polarization by approximately 10.5 points – quite a substantively large change given that mean affective polarization ranges from 25.4 in 1980 to 52.3 in 2020. However, once we add covariates in models 2 and 3, the coefficient on social sorting decreases to 5.0 and 3.3, respectively, though it retains statistical significance at conventional levels ($p < 0.001$ and $p < 0.01$, respectively).

Figure 9 and Table 6 repeat the same analysis, but this time using social sorting scores derived from models that include self-reported ideology. The results are notably different: the association between my measure of social sorting and affective polarization is much stronger and can be detected across *all* ANES editions. The non-linearity of the association is somewhat less obvious: in can be seen in certain years (e.g. 1992, 2016, and 2020) but not in others (1984, 2000).

As Table 6 shows, a one-unit increase in social sorting – that is, a movement from

as, themselves, causal quantities (see e.g. Westreich and Greenland 2013).

[19]Given that this model only examines partisans, its interpretation should be adjusted accordingly.

[20]Note that I avoid using the word "effect" given its causal connotation.

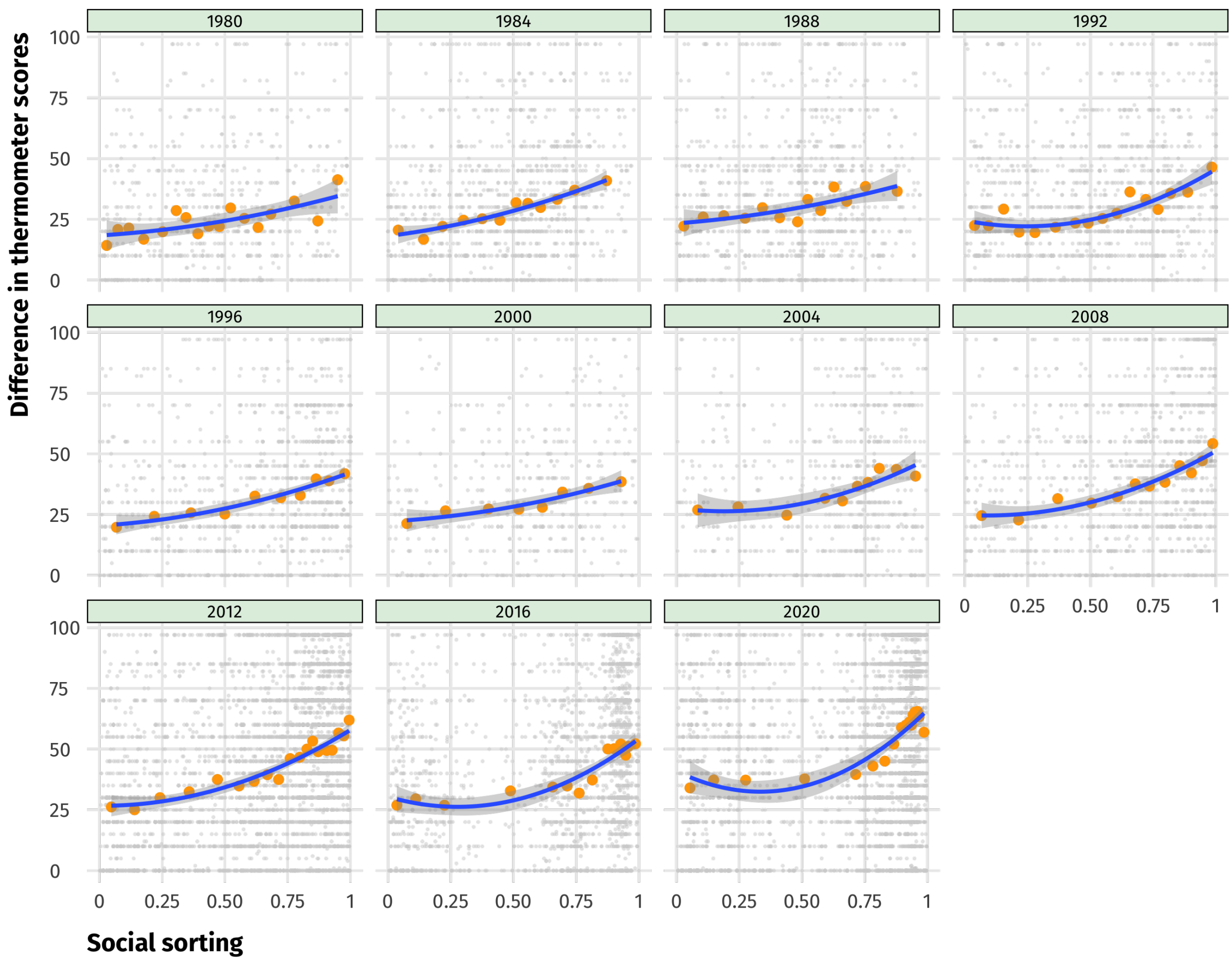

Figure 9: Scatterplot showing the relationship between social sorting and affective polarization, 1980-2020

Table 6: Linear regression models relating social sorting to affective polarization

	(1)	(2)	(3)
(Intercept)	7.619 [−45.321, 60.559]	11.438 [9.310, 13.565]	1.448 [−1.456, 4.353]
Sorting		28.572 [27.203, 29.941]	25.963 [24.524, 27.402]
Num.Obs.	14 831	21 276	20 381
Year FE	✓	✓	✓
RMSE	26.94	29.03	28.26
Standard controls		✓	✓
PID strength control			✓

* $p < 0.05$, ** $p < 0.01$

minimum to maximum – is expected to increase affective polarization by approximately 28.6 points in a model with no covariates and by 17.5 points in a model with a battery of covariates (listed above). This is, without a doubt, a substantively strong association that appears to lend support to the social sorting hypothesis. However, as explored earlier, whether ideological sorting should be thought of in the same vein as more strictly *social* sorting is unclear.

7 Discussion and conclusion

This thesis aimed to measure the history of partisan social sorting in the United States and to assess its relationship with affective polarization. The analyses presented above point to two main findings.

First, despite claims that Americans have sorted into two dissimilar political camps that have little in common, historical data from the ANES shows that it remains quite difficult to predict partisan identity on the basis of socio-demographic characteristics, *unless* we also include self-reported ideology as a predictor. The implications of this finding are unclear, since some degree of partisan-ideological sorting may be desirable.

Second, there is some suggestive evidence that socially sorted individuals tend to be more affectively polarized. The evidence is much stronger when using a measure of social sorting that includes self-reported ideology.

Additionally, there is a crucial point that is worth keeping in mind in order to contextualize these findings and chart a path forward for research on social sorting and affective polarization: whether *objective* social sorting matters at all is very much up for debate. It is entirely possible for individuals to perceive partisan groups as fundamentally different even if cross-cutting cleavages are, in fact, abundant. The psychological mechanisms posited by social sorting theory – "these people are unlike me, and therefore I feel distant from them or outright dislike them" – concern *subjective* cognitive processes that are internal to each individual. In a way, social identity theory – and the closely associated minimal group paradigm – assume that humans will *always* perceive large differences between groups that they identify with and groups that they think of as outsiders. How the objective social composition of political parties relates to these perceptions is unclear and, to my knowledge, the functional form of this relationship is always left unspecified. For illustrative purposes, Figure 10 depicts two different possibilities. The dashed horizontal line shows perfect calibration: for all hypothetical levels of actual social sorting (x axis), people perceive the exact right amount of sorting.[21]

It is under this unrealistic scenario that the empirical analyses presented in this article should work best. However, we may think that the scenario in blue is most likely given social identity theory: perceived social sorting is high regardless of actual social sorting; and as actual social sorting increases, perceived social sorting barely budges. If this is correct, then we would not expect the empirical analyses presented above to yield much of anything. The "parties in our heads," to quote Ahler and Sood (2018), could be completely unmoored from the actual parties. The scenario in green is similar, but with more responsiveness to real-world conditions.

Clearly, there are reasons to believe that people do not have perfectly calibrated perceptions. Nonetheless, the theory of social sorting as developed so far rests just as much

[21] For the sake of simplicity, I do not consider between-cases heterogeneity in calibration.

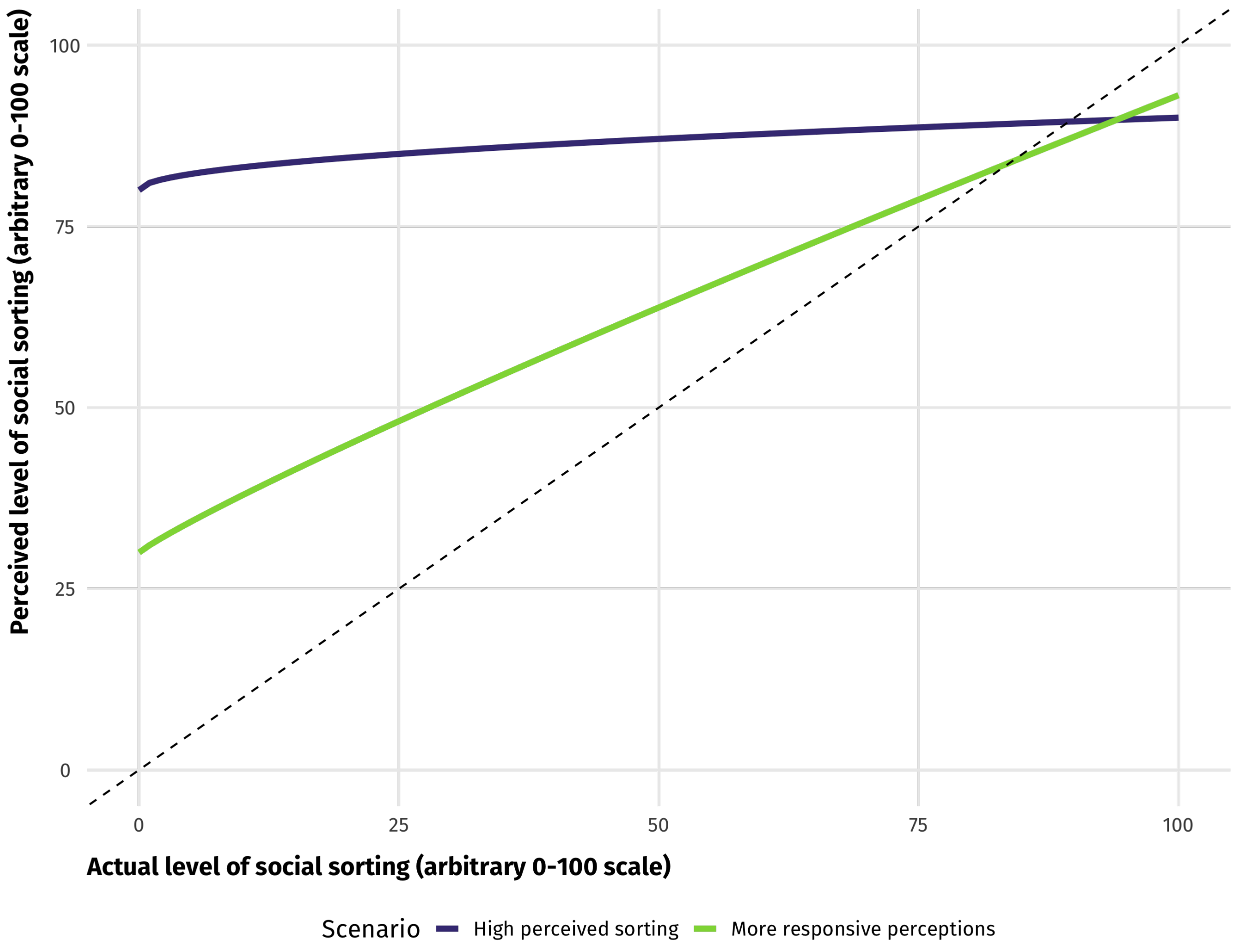

Figure 10: Different hypothetical relationships between the objective level of social sorting and perceived level of social sorting

on perceptions as on empirically observable movements in the structure of the party system. To what extent each is relevant, and how one relates to the other, is an important task for future research. Within the bounds of this thesis, I can only conclude that the social sorting hypothesis is not particularly convincing if perceptions of social sorting are well-calibrated.

Finally, I conclude with some broader thoughts about the contribution of this thesis to the literature on partisan polarization.

First, in order to move forward, the literature should address the concern I raised above: whether subjective or objective partisan sorting matters more and what is the relationship between them. Absent a strong theoretical framework that addresses these questions, the absence of an empirical regularity (e.g. the weak association between objective social sorting and affective polarization) may be mistaken as proof against the social sorting theory. Studies of the type conducted by Ahler and Sood (2018) should be conducted to examine the relationship between objective conditions and subjective perceptions.

Second, the literature on the causes and consequences of polarization would greatly benefit from additional cross-national evidence. This is especially true when it comes to social sorting: while party systems vary wildly across countries, the logic of social identity theory should hold. To my knowledge, there is no cross-national study that systematically examines whether the decline of cross-cutting ties in the United States and its association with affective polarization is reflected in Canada and Western European countries. The empirical approach employed in this thesis can easily be transferred to other contexts.

Third, empirical social scientists should engage with normative questions regarding the desirability of distinctive party compositions. As this thesis documents, rising affective polarization is likely caused by a combination of social identities and genuine ideological disagreement, with the latter possibly a larger contributor.

8 Appendix

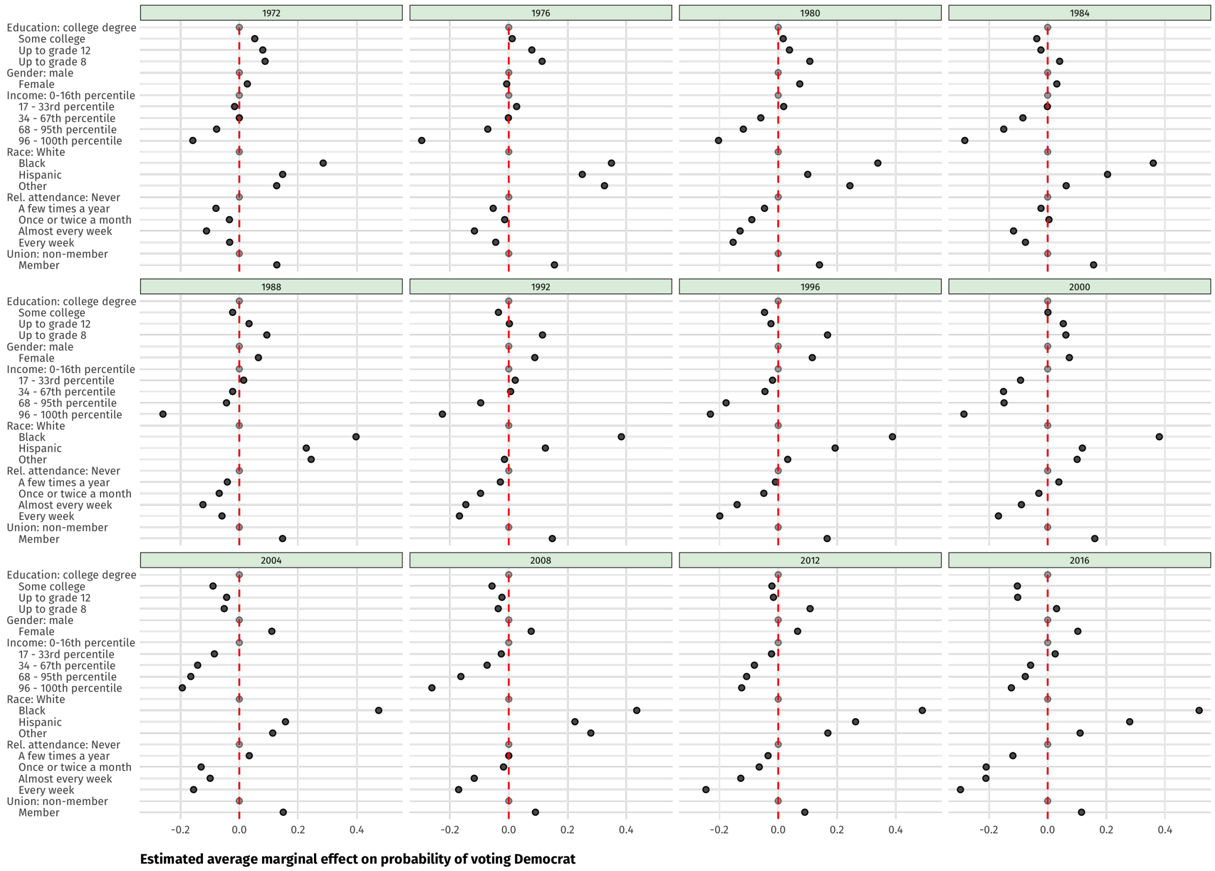

Figure 11: Estimated average marginal effects of social attributes on the probability ofDemocratic identification, 1972-2020 (many facets)

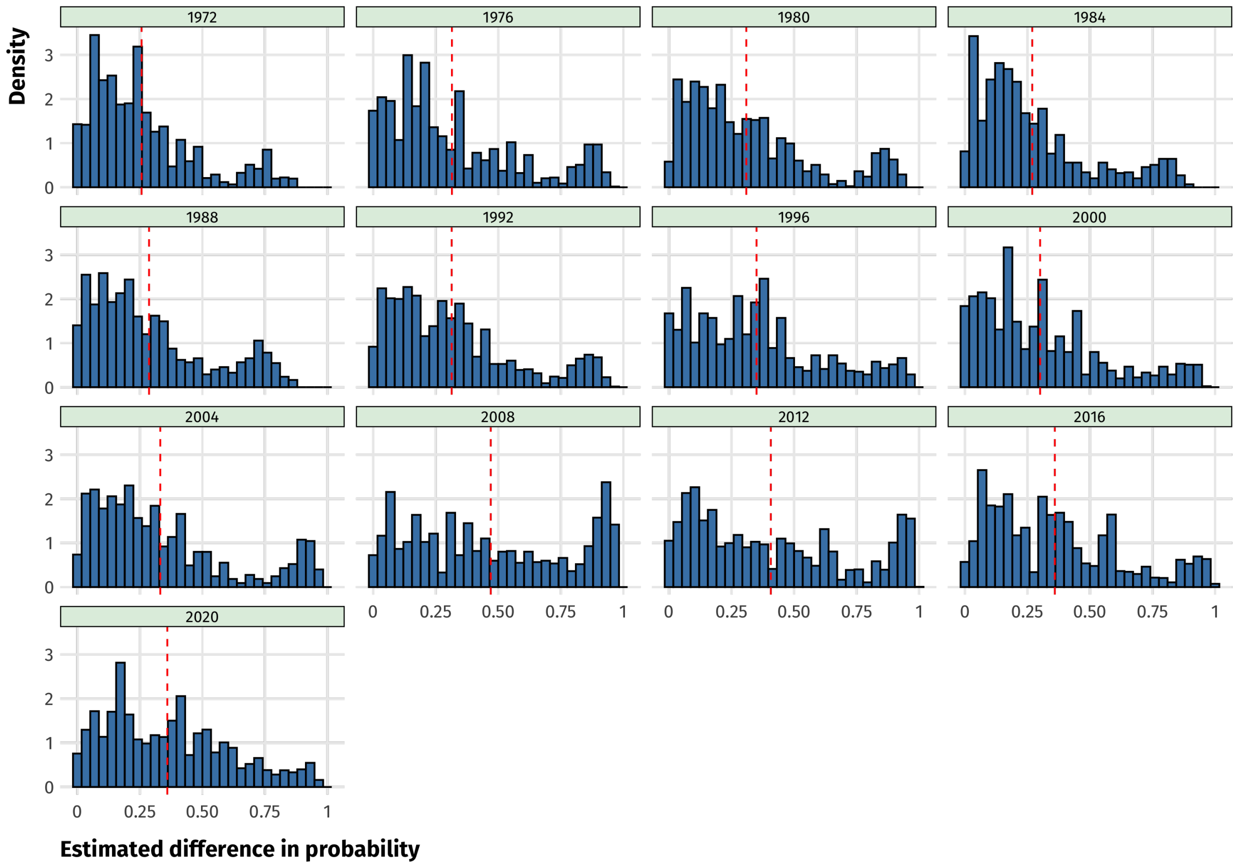

Figure 12: Histograms of estimated differences in probabilities, by year (1972-2020)

Table 7: Distribution of 3-point party identification in ANES surveys (unweighted)

	Democrat	**Independent**	**Republican**
1972	51.2	14.8	33.9
1976	51.1	15.3	33.6
1980	52.2	15.1	32.8
1984	47.8	12.6	39.6
1988	47.1	12.0	40.9
1992	49.7	12.8	37.5
1996	52.5	9.2	38.3
2000	49.6	12.5	37.9
2004	49.5	10.1	40.4
2008	59.8	11.6	28.6
2012	52.7	13.4	33.9
2016	45.7	13.6	40.7
2020	46.5	11.7	41.7

References

10 Abramowitz, Alan I., and Kyle L. Saunders. 2008. "Is Polarization a Myth?" *The Journal of Politics* 70 (2): 542–55. https://doi.org/10.1017/S0022381608080493.

Abramowitz, Alan I., and Steven Webster. 2016. "The Rise of Negative Partisanship and the Nationalization of U.S. Elections in the 21st Century." *Electoral Studies* 41 (March): 12–22. https://doi.org/10.1016/j.electstud.2015.11.001.

Abramowitz, Alan I., and Steven W. Webster. 2018. "Negative Partisanship: Why Americans Dislike Parties But Behave Like Rabid Partisans." *Political Psychology* 39 (S1): 119–35. https://doi.org/10.1111/pops.12479.

Ahler, Douglas J., and Gaurav Sood. 2018. "The Parties in Our Heads: Misperceptions about Party Composition and Their Consequences." *The Journal of Politics* 80 (3): 964–81. https://doi.org/10.1086/697253.

Aldrich, John. 1995. *Why Parties? The Origin and Transformation of Political Parties in America*. Chicago: University of Chicago Press.

Banda, Kevin K., and John Cluverius. 2018. "Elite Polarization, Party Extremity, and Affective Polarization." *Electoral Studies* 56 (December): 90–101. https://doi.org/10.1016/j.electstud.2018.09.009.

Bartels, Larry M. 2000. "Partisanship and Voting Behavior, 1952-1996." *American Journal of Political Science* 44 (1): 35. https://doi.org/10.2307/2669291.

Berinsky, Adam J. 2018. "Telling the Truth about Believing the Lies? Evidence for the Limited Prevalence of Expressive Survey Responding." *The Journal of Politics* 80 (1): 211–24. https://doi.org/10.1086/694258.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351–68. https://doi.org/10.1093/pan/mpr057.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2022. "Cross-Country Trends in Affective Polarization." *The Review of Economics and Statistics*, January, 1–60. https://doi.org/10.1162/rest_a_01160.

Brader, Ted, Joshua A. Tucker, and Andrew Therriault. 2014. "Cross Pressure Scores: An Individual-Level Measure of Cumulative Partisan Pressures Arising from Social Group Memberships." *Political Behavior* 36 (1): 23–51. https://doi.org/10.1007/s11109-013-9222-8.

Broockman, David E., Joshua L. Kalla, and Sean J. Westwood. 2023. "Does Affective Polarization Undermine Democratic Norms or Accountability? Maybe Not." *American Journal of Political Science* 67 (3): 808–28. https://doi.org/10.1111/ajps.12719.

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber. 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10 (4): 519–78. https://doi.org/10.1561/100.00014074.

Bullock, John G., and Gabriel Lenz. 2019. "Partisan Bias in Surveys." *Annual Review of Political Science* 22 (1): 325–42. https://doi.org/10.1146/annurev-polisci-051117-050904.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. Chicago: University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/A/bo24047989.html.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng. 2023. "On

Binscatter." arXiv. https://doi.org/10.48550/arXiv.1902.09608.

Cavari, Amnon, and Guy Freedman. 2018. "Polarized Mass or Polarized Few? Assessing the Parallel Rise of Survey Nonresponse and Measures of Polarization." *The Journal of Politics* 80 (2): 719–25. https://doi.org/10.1086/695853.

———. 2023. "Survey Nonresponse and Mass Polarization: The Consequences of Declining Contact and Cooperation Rates." *American Political Science Review* 117 (1): 332–39. https://doi.org/10.1017/S0003055422000399.

Center, Pew Research. 2019. "Response Rates in Telephone Surveys Have Resumed Their Decline." Pew Research Center. https://www.pewresearch.org/short-reads/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/.

Connors, Elizabeth C. 2023. "Social Desirability and Affective Polarization." *Public Opinion Quarterly* 87 (4): 911–34. https://doi.org/10.1093/poq/nfad053.

Coppock, Alexander. 2019. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* 7 (3): 613–28. https://doi.org/10.1017/psrm.2018.10.

Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6 (1): 2053168018822174. https://doi.org/10.1177/2053168018822174.

Dassonneville, Ruth. 2021. "Change and Continuity in the Ideological Gender Gap a Longitudinal Analysis of Left-Right Self-Placement in OECD Countries." *European Journal of Political Research* 60 (1): 225–38. https://doi.org/10.1111/1475-6765.12384.

Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. 2022. "(Mis)estimating Affective Polarization." *The Journal of Politics* 84 (2): 1106–17. https://doi.org/10.1086/715603.

Druckman, James N, and Matthew S Levendusky. 2019. "What Do We Measure When We Measure Affective Polarization?" *Public Opinion Quarterly* 83 (1): 114–22. https://doi.org/10.1093/poq/nfz003.

Egan, Patrick J. 2020. "Identity as Dependent Variable: How Americans Shift Their Identities to Align with Their Politics." *American Journal of Political Science* 64 (3): 699–716. https://doi.org/10.1111/ajps.12496.

Fiorina, Morris P., Samuel A. Abrams, and Jeremy C. Pope. 2008. "Polarization in the American Public: Misconceptions and Misreadings." *The Journal of Politics* 70 (2): 556–60. https://doi.org/10.1017/S002238160808050X.

Fiorina, Morris P., and Samuel J. Abrams. 2008. "Political Polarization in the American Public." *Annual Review of Political Science* 11 (1): 563–88. https://doi.org/10.1146/annurev.polisci.11.053106.153836.

Garzia, Diego, Frederico Ferreira da Silva, and Simon Maye. 2023. "Affective Polarization in Comparative and Longitudinal Perspective." *Public Opinion Quarterly*, February, nfad004. https://doi.org/10.1093/poq/nfad004.

Graham, Matthew H., and Omer Yair. 2023. "Expressive Responding and Belief in 2020 Election Fraud." *Political Behavior*, July. https://doi.org/10.1007/s11109-023-09875-w.

Green, Donald P., Bradley Palmquist, and Eric Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.

Greene, Steven. 1999. "Understanding Party Identification: A Social Identity Approach." *Political Psychology* 20 (2): 393–403. https://doi.org/10.1111/0162-895X.00150.

Harteveld, Eelco, Lars Erik Berntzen, Andrej Kokkonen, Haylee Kelsall, Jonas Linde, and Stefan Dahlberg. 2022. "The (Alleged) Consequences of Affective Polarization: A Survey Experiment in 9 Countries." OSF. https://doi.org/10.31219/osf.io/64uwd.

Hetherington, Marc J. 2001. "Resurgent Mass Partisanship: The Role of Elite Polarization." *American Political Science Review* 95 (3): 619–31. https://doi.org/10.1017/S0003055401003045.

Hillygus, D. Sunshine. 2011. "The Evolution of Election Polling in the United States." *Public Opinion Quarterly* 75 (5): 962–81. https://doi.org/10.1093/poq/nfr054.

Huber, Gregory A., and Neil Malhotra. 2017. "Political Homophily in Social Relationships: Evidence from Online Dating Behavior." *The Journal of Politics* 79 (1): 269–83. https://doi.org/10.1086/687533.

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

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22 (1): 129–46. https://doi.org/10.1146/annurev-polisci-051117-073034.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (3): 405–31. https://doi.org/10.1093/poq/nfs038.

Iyengar, Shanto, and Sean J. Westwood. 2015. "Fear and Loathing Across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3): 690–707. https://doi.org/10.1111/ajps.12152.

Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D. Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, et al. 2018. "An Evaluation of the 2016 Election Polls in the United States." *Public Opinion Quarterly* 82 (1): 1–33. https://doi.org/10.1093/poq/nfx047.

Kim, Seo-young Silvia, and Jan Zilinsky. 2022. "Division Does Not Imply Predictability: Demographics Continue to Reveal Little About Voting and Partisanship." *Political Behavior*, August. https://doi.org/10.1007/s11109-022-09816-z.

Kingzette, Jon, James N Druckman, Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. 2021. "How Affective Polarization Undermines Support for Democratic Norms." *Public Opinion Quarterly* 85 (2): 663–77. https://doi.org/10.1093/poq/nfab029.

Lazarsfeld, Paul D., Bernard Berelson, and Hazel Gaudet. 1948. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York: Columbia University Press.

Leeper, Thomas J. 2019. "Where Have the Respondents Gone? Perhaps We Ate Them All." *Public Opinion Quarterly* 83 (S1): 280–88. https://doi.org/10.1093/poq/nfz010.

Lelkes, Yphtach, Gaurav Sood, and Shanto Iyengar. 2017. "The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect." *American Journal of Political Science* 61 (1): 5–20. https://doi.org/10.1111/ajps.12237.

Lenz, Gabriel S. 2012. *Follow the Leader? How Voters Respond to Politicians' Policies and Performance*. Chicago: Chicago University Press.

Levendusky, Matthew. 2009. *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. Chicago: University of Chicago Press.

Margolis, Michele F. 2018. *From Politics to the Pews: How Partisanship and the Political Environment Shape Religious Identity*. University of Chicago Press. https://doi.org/10.7208/9780226555812.

Mason, Lilliana. 2016. "A Cross-Cutting Calm: How Social Sorting Drives Affective Polarization." *Public Opinion Quarterly* 80 (S1): 351–77. https://doi.org/10.1093/poq/nfw001.

———. 2018a. *Uncivil Agreement: How Politics Became Our Identity*. Chicago: University of Chicago Press.

———. 2018b. "Losing Common Ground: Social Sorting and Polarization." *The Forum* 16 (1): 47–66. https://doi.org/10.1515/for-2018-0004.

Mason, Lilliana, and Julie Wronski. 2018. "One Tribe to Bind Them All: How Our Social Group Attachments Strengthen Partisanship." *Political Psychology* 39 (S1): 257–77. https://doi.org/10.1111/pops.12485.

Mutz, Diana C. 2002. "The Consequences of Cross-Cutting Networks for Political Participation." *American Journal of Political Science* 46 (4): 838. https://doi.org/10.2307/3088437.

Noel, Hans. 2012. "The Coalition Merchants: The Ideological Roots of the Civil Rights Realignment." *The Journal of Politics* 74 (1): 156–73. https://doi.org/10.1017/S0022381611001186.

Norpoth, Helmut, and Jerrold G. Rusk. 1982. "Partisan Dealignment in the American Electorate: Itemizing the Deductions Since 1964." *American Political Science Review* 76 (3): 522–37. https://doi.org/10.2307/1963729.

Praet, Stiene, Andrew M. Guess, Joshua A. Tucker, Richard Bonneau, and Jonathan Nagler. 2022. "What's Not to Like? Facebook Page Likes Reveal Limited Polarization in Lifestyle Preferences." *Political Communication* 39 (3): 311–38. https://doi.org/10.1080/10584609.2021.1994066.

Prior, Markus, Gaurav Sood, and Kabir Khanna. 2015. "You Cannot Be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions." *Quarterly Journal of Political Science* 10 (4): 489–518. https://doi.org/10.1561/100.00014127.

Reiljan, Andres. 2020. "'Fear and Loathing Across Party Lines' (Also) in Europe: Affective Polarisation in European Party Systems." *European Journal of Political Research* 59 (2): 376–96. https://doi.org/10.1111/1475-6765.12351.

Rogowski, Jon C., and Joseph L. Sutherland. 2016. "How Ideology Fuels Affective Polarization." *Political Behavior* 38 (2): 485–508. https://doi.org/10.1007/s11109-015-9323-7.

Schaffner, Brian F. 2011. "Party Polarization." In *The Oxford Handbook of the American Congress*, edited by George C. Edwards, Frances E. Lee, and Eric Schickler, 527–49. New York: Oxford University Press. https://academic.oup.com/edited-volume/34381/chapter/291576814.

Schattschneider, E. E. 1942. *Party Government: American Government in Action*. New York: Routledge. https://doi.org/10.4324/9781315126050.

Stokes, S. C. 1999. "Political Parties and Democracy." *Annual Review of Political Science* 2 (1): 243–67. https://doi.org/10.1146/annurev.polisci.2.1.243.

Tajfel, Henri. 1970. "Experiments in Intergroup Discrimination." *Scientific American* 223 (5): 96–102. https://doi.org/10.1038/scientificamerican1170-96.

Tajfel, Henri, and John Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *Intergroup Relations: Essential Readings*, 94–109. Key Readings in Social Psychology. New York, NY, US: Psychology Press.

Voelkel, Jan G., James Chu, Michael N. Stagnaro, Joseph S. Mernyk, Chrystal Redekopp, Sophia L. Pink, James N. Druckman, David G. Rand, and Robb Willer. 2023. "Interventions Reducing Affective Polarization Do Not Necessarily Improve Anti-Democratic Attitudes." *Nature Human Behaviour* 7 (1): 55–64. https://doi.org/10.1038/s41562-022-01466-9.

Wagner, Markus. 2021. "Affective Polarization in Multiparty Systems." *Electoral Studies* 69 (February): 102199. https://doi.org/10.1016/j.electstud.2020.102199.

Westreich, Daniel, and Sander Greenland. 2013. "The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients." *American Journal of Epidemiology* 177 (4): 292–98. https://doi.org/10.1093/aje/kws412.

Westwood, Sean J., and Erik Peterson. 2020. "The Inseparability of Race and Partisanship in the United States." *Political Behavior*, October. https://doi.org/10.1007/s11109-020-09648-9.

White, Ismail K., and Chryl N. Laird. 2020. *Steadfast Democrats*. Princeton: Princeton University Press. https://press.princeton.edu/books/hardcover/9780691199511/steadfast-democrats.