

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Sound and Fury, Signifying Something: Polarization, Identity, and Dissatisfaction

Permalink

<https://escholarship.org/uc/item/5xd136sc>

Author

Lauterwasser, Steven William

Publication Date

2022

Peer reviewed|Thesis/dissertation

Sound and Fury, Signifying Something:
Polarization, Identity, and Dissatisfaction

By

Steven William Lauterwasser

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Sociology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kim Voss, Chair

Professor Ann Swidler

Professor Paul Pierson

Summer 2022

Copyright © 2022 by Steven William Lauterwasser
All Rights Reserved

Abstract

Sound and Fury, Signifying Something: Polarization, Identity, and Dissatisfaction

by

Steven William Lauterwasser

Doctor of Philosophy in Sociology

University of California, Berkeley

Professor Kim Voss, Chair

This dissertation consists of three essays on political polarization in the United States which, together, justify and build a new *worldview model* of polarization. I develop this theory to better address the issue of whether polarization is asymmetric at the mass level, a question about which quantitative work has thus far had relatively little to say, despite clear evidence that Republican elites are far more polarized than their Democratic counterparts.

The first essay takes the open ground of asymmetry as a starting point for a theory contest: some work on affective polarization defines polarization solely in terms of out-party animus, while other work sees both in- and out-party affect as constitutive. Given asymmetries at the elite and organizational levels, we expect Republicans to be more polarized than Democrats, but these two theories of affective polarization have clearly different expectations for what the asymmetry looks like. However, this apparently neat theory contest fails. Using data from the American National Election Studies, I look at how affective polarization developed from 1978-2016 and I find that *Democrats* are more polarized, but that that difference is driven by in-party affect (while out-party animus isn't asymmetric). These findings are anomalous for both theories, and I argue that they should disrupt the basic assumption of a straightforward link between affect, identity, and polarization.

The second essay brings together sociological work on cultural schemas, recent advances in the study of attitudinal polarization, and core insights on the role of affective polarization from political science, to propose a new theory which overcomes this disruption: the worldview model of polarization. By defining polarization in terms of persistent patterned relationships between partisan affect, perceptions of threat, and feelings of fear, I am able to approach asymmetry as a question not only of degree, but of *kind*. Using data from the 1978-2016 ANES, I not only clearly demonstrate that Republicans are indeed significantly more polarized at the mass level, but also suggest a mechanism for this asymmetry, namely, that Republicans are caught in a radicalizing feedback loop of fear of the opposition and dissatisfaction with their own party's response to that perceived threat.

The final essay elaborates on this model by turning to the public opinion literature on belief systems. Developing the insight that polarization consists in the persistent patterned

relationships between emotions and judgments, not either alone, it uses belief network analysis to compare the structure of these relationships over time. Comparing 1992 and 2016, it clearly shows the strengthening integration of affective and attitudinal survey items and overall stronger centralization of the networks, indicating both that political belief systems were becoming more emotionally grounded and that grounding was consistently oriented against a small number of issues (i.e. partisanship and ideology). This not only provides another strain of support for the basic insights of the worldview model, the inclusion of affective items in a quantitative assessment of belief systems is a major contribution to the study of public opinion more generally.

For my father, Bruce Lauterwasser. He did not get to see it, but without him I never would have finished. I am happy to have lived up to his faith in me.

Contents

Acknowledgments	iii
1 Introduction	1
2 Negative Partisanship is Not Enough: Affective Polarization, Asymmetry, and a Failed Theory Contest	6
3 Strangers in their Own Party: Dissatisfied Partisans and the Worldview Model of Polarization	28
4 Beyond the Structure of Attitudes: Belief Systems, Affect, and Polarization	51
5 Conclusion	82
References	86
A Face Validity Checks for Chapter 2	95
B Putting Trends in Context	97
C Model Building for Chapter 2	101
D Variable Descriptives for Chapter 3	105
E Model Building for Chapter 3	109

Acknowledgments

I could not have completed this dissertation without the intellectual, moral, practical, and emotional support of a host of people. Each of them share a measure of the credit for the work presented here (insofar as it is successful; errors are, of course, my own) and all of them have my utmost gratitude. To name them all reliably is beyond my abilities, but I attempt here to recognize those whose contributions were most meaningful.

First, this project has been indelibly shaped by every member of my committee. They all pushed me to find the best version of my work and, with their insight and honest engagement, helped me to achieve it. I especially want to thank Ann Swidler for her effervescent confidence in me and also for pointing me towards polarization as a subject in the first place, giving me a spark of inspiration in the wake of the 2016 election. Paul Pierson provided crucial perspective, both intellectual and strategic, from political science, along with a deep knowledge of, and theoretical engagement with, American politics. Kim Voss, my chair, has been incredibly generous with both her time and attention. She continues to be a source of wisdom and guidance and it is really in dialog with her that this project took shape. Working with her has been a joy.

After my committee, the people who have probably seen the most drafts of this work are Adam Storer and Kristin George, the other members of Kim's dissertation dinner group. Their feedback on some of the earliest versions of what would become this dissertation was invaluable, as was their comraderie and intellectual community. The people who heard the most about this work, on the other hand, are almost certainly Katherine Hood, Jeff Weng, and Mary Shi, my longtime officemates in SSB 364. I am grateful for not only their commiseration and encouragement, but also their advice and insight. Most of all, however, I am grateful for their steadfast friendship. The office we shared was the environment which made so much of this work possible.

Neil Fligstein, Heather Haveman, and the entire Culture, Organizations, and Politics Workshop helped me to find the core of every argument and the most effective form for every paper. Cihan Tuğal, Marion Fourcade, Irene Bloemraad, Dave Harding, Joe LaBriola, Andrew Jaeger, Matthew Stimpson, Santiago Molina, Robert Pickett, Emily Ruppel, Matty Lichtenstein, and Carmen Brick all offered advice and support at key moments.

I thank Catherine Norton, Anne Meyers, Laura Enriquez, Martin Eiermann, Tyler Leeds, and Garrett Strain for their assistance in a time of crisis. Without them, this dissertation would not be finished now.

Both George Emblom and the St. Mark's Choir, as well as Perfect Fifth (especially Sung Hyup Lee, Rachel Durney, Madeline Wolf, and Mark Sumner), gave me both community and commitment outside of this work. Singing with them kept me sane.

My family, Karen, Clara, Greg, Hannah, and Jill Lauterwasser provided a haven on the East Coast, from which I always returned refreshed.

Ariana Nedelman, Helen Hailes, and Libby Bova have been my friends for more than a decade. They are people to whom I can always turn. Their presence in my life is fundamental.

Fritjof Capra and Elizabeth Hawk opened their home to me during the pandemic and provided me with a refuge for the final two years of this process. They were a family when I could not be with my own. Their generosity made this possible.

Finally, I owe perhaps the greatest debt to Juliette Capra, my then partner, now friend. Throughout the life of this dissertation, her love and support were a fundamental constant upon which I could rely. I can still name now specific moments, from even the earliest days of the dissertation, when her kindness, encouragement, and guidance were decisive. There is no thanks sufficient, but I offer all I can.

Chapter 1

Introduction

This dissertation presents a theorization and investigation of polarization in the United States from 1978 to 2016. The research was begun long before January 6th, 2020 (Tan, Shin, and Rindler 2021); before the targeted harassment of election officials (So 2021; Wines 2020), educators (Carr 2022; Heinrichs 2022; Lawton 2022; Salgado 2022), and healthcare workers (Jurecic 2022); before the concerted book banning campaigns (Alter 2022; Schwartz 2022); before the violent threats and confrontational, disruptive, protest of children’s LGBTQ Pride events (Crimaldi and Hilliard 2022; Helmore 2022; Ring 2022; Tensley 2022). As such, this dissertation cannot address them directly. However, it is nevertheless an attempt to understand them, or at least the forces behind them. These assaults on democratic institutions and minority acceptance can all be understood as symptoms of polarization. They make plain just how deep that polarization runs, but they also make plain much of what we do not know about it. If these are the symptoms, then we can say two things about the disease. First, it seems to have dramatically different effects on Democrats and Republicans. And second, those effects are partly interpretive in nature, binding partisan identity together with ways of seeing the world, linking together propositional attitudes, evaluative judgments, and emotional responses. But, despite powerful conceptual and qualitative work, neither of these fundamental characteristics of polarization have been well addressed in the existing quantitative literature. This dissertation is an effort to remedy that fact by demonstrating the partisan differences in polarization and by developing a new theory which not only captures those differences, but also helps make sense of them.

Conceptually, polarization is perhaps best understood as the division of society into an “us” and a “them” between whom compromise is less and less possible and conflict is more and more fierce, increasingly understood through Manichean narratives of zero-sum, existential conflict (DellaPosta 2020; Finkel et al. 2020; Lelkes 2016; McCoy, Rahman, and Somer 2018). Despite this, quantitative work on polarization offers not a single strong understanding of polarization, but many. Work across various fields and sub-fields relies upon many different definitions, definitions that are not always immediately reconcilable. These definitions vary in their level of analysis (polarization as distributional vs. group vs. individual property), their techniques of measurement (looking for extremity within vs. correlation across opinion items), their populations of interest (all citizens vs. voters vs. partisans), and their defining outcomes (political attitudes vs. non-political attitudes vs. partisan affect vs. candidate assessments) (DellaPosta 2020; DiMaggio, Evans, and Bryson 1996 for emblematic examples; see, e.g.,

Mason 2015).

This diversity can, however, be divided into two broad categories: *attitudinal* approaches on the one hand and *affective* approaches on the other. Attitudinal polarization understands polarization fundamentally as a matter of disagreement on issues and so looks for it in the characteristics of political attitude distributions (e.g. bimodality, extremity, constraint, etc.) (see, e.g., Baldassarri and Gelman 2008; DiMaggio et al. 1996; Fischer and Hout 2006). Affective polarization, by contrast, sees polarization as basically social and emotional, defining it in terms of affective in/out-group bias: the actual disagreement on issues matters little if two groups hate each other (Finkel et al. 2020; Iyengar et al. 2019; Iyengar, Sood, and Lelkes 2012; Mason 2015). As I will go on to show, although both sets of approaches have accomplished much and have much to recommend them, they ultimately have difficulties substantiating the link between their specific definition and the basic conceptual understanding of polarization as a zero-sum, us versus them, conflict. This difficulty manifests perhaps most clearly in their inability to systematically capture partisan differences in polarization.

The evidence for such differences is extensive and only becomes more substantial as polarization increases and its consequences become more and more obvious. Polarization has not developed symmetrically, impacting Democrats and Republicans equally. Rather, it seems to be fundamentally asymmetric in nature, with Republicans polarizing more than Democrats. Indeed, such an asymmetry is widely regarded as constitutive of polarization at the elite level, where the Republican Party is both measurably more extreme than the Democratic Party and also clearly responsible for polarization in a way the Democrats are not (Bermeo 2019; Grossmann and Hopkins 2015; Hacker and Pierson 2005; Mann and Ornstein 2016; McCoy et al. 2018; Pierson and Schickler 2020; Sinclair 2006). But the asymmetry is also evident at the institutional and popular levels, as the litany of offenses above should begin to make clear. But even aside from these recent patterns, the asymmetry is still evident: Republican popular organizations have a unique role in pushing polarizing narratives; the party itself has seen repeated turnover in leadership due to popular insurgency (see, e.g., Eric Cantor, John Boehner, and Paul Ryan); and continued increases in right-wing extremism coupled with ethno-nationalist resentment, both of which are fundamentally popular in nature (Hacker and Pierson 2020; Skocpol 2020). Yet, despite this, neither of major quantitative approaches have been able to say much about this evident asymmetry.

As I will argue, this is in part because these differences are neither purely attitudinal nor purely affective, but rather interpretive. That is, what matters are not attitudes or affects themselves, but rather what they *mean* to people. But that meaning cannot be found in any single opinion or emotion in isolation. Instead, it is a result of the constellation of relationships among opinions and emotions. It means something very different to dislike President Biden, even to dislike him a great deal, and to dislike him while also believing he stole the election and fearing that his administration may signal the end of the republic. Alternatively, disliking Trump as an odious outlier in modern politics is very different from disliking him as a representative of the current Republican party and its agenda. All these meanings have significant consequences for those holding them, orienting them towards politics in distinct ways, telling them different things about the political world and their place in it. Understanding polarization, then, requires attending to this interpretive power.

In summary, my contention is that the quantitative study of polarization has thus far done too little to recognize either the empirical fact of asymmetry or interpretive power

of polarization. These problems are two sides of the same coin: attitudinal and affective approaches have been unable to capture asymmetry because they miss the importance of the link between attitude and affect and therefore polarization's interpretive nature, but they also miss this interpretive core of polarization because they have not yet attended to asymmetry in a way which would make it obvious. The three papers collected in this dissertation elaborate, demonstrate, and (attempt to) solve these problems.

The first substantive chapter, "Negative Partisanship is Not Enough: Affective Polarization, Asymmetry, and a Failed Theory Contest," demonstrates the existence of an unexpected asymmetry in polarization and shows how its existence fundamentally unsettles current affective approaches. Given the lack of previous quantitative work on, but the clear empirical expectations about, asymmetric polarization, it uses this question as an opportunity for a theory contest between two competing theories of affective polarization. The first posits that polarization is defined purely by out-party animus, while the second sees both in-group favor and out-group dislike as constitutive. Using the American National Election Studies, I track the development of polarization from 1978 to 2016 using both definitions, but I find no asymmetry on out-party animus and an asymmetry towards *Democrats* on in-party affect. These results are fundamentally anomalous: they are either proof of a dramatic mismatch between mass affective polarization and every other piece of evidence, which show an asymmetry towards Republicans, or they indicate that there is something basically wrong with these definitions and demonstrate the need for a new approach.

The second chapter, "Strangers in their Own Party: Dissatisfied Partisans and the Worldview Model of Polarization," explains where this problem comes from, then proposes and validates just such a new approach: the *worldview model* of polarization. Having shown in the first chapter the ways in which affective approaches to polarization fail in grasping the evident asymmetries in polarization, the second chapter elaborates the argument previewed above that the problem is really a theoretical one which lays at the heart of both affective and attitudinal approaches. In essence, the problem is a lack of attention to the interpretive aspects of polarization. I show how this omission undermines the claimed connection between attitudinal or affective measures and the concept of polarization, arguing that the differences which make polarization asymmetric are not differences of *degree*, but differences of *kind*. Rather than extremity of opinion or simple out-party animus, polarization is a coupling of emotions and judgments which makes political division meaningful in a particular way. Specifically, it defines polarization in terms of persistent patterned relationships between partisan affect, assessments of threat, and feelings of fear (of the other party). Again using the American National Election Studies, this new worldview model of polarization reveals that Republican partisans polarized more, faster, from 1978 to 2016 than Democrats, matching Republican elites. Recognizing this coupling of emotions and judgments reveals a radicalizing feedback loop at the heart of Republican polarization, in which fear of the opposition and dissatisfaction with their own party's response to that perceived threat feed off each other.

The worldview model defines polarization in terms of the coupling of affects, attitudes, and judgments, and it focuses specifically on partisan bias, threat, and fear because those have been shown to be important to polarization in previous work. However, such a focused approach is just one option and the final substantive chapter of the dissertation, "Beyond the Structure of Attitudes: Belief Systems, Affect, and Polarization", reinforces and expands on the worldview model by looking at all possible affects and judgements together. In doing so, it focuses

especially on the question of the connection between attitude and affect, bringing emotion into the recent proliferation of work on the structure of attitudes. Despite the recognition of the importance of affect elsewhere (including in much work on polarization), work on public opinion and belief systems has so far looked only at propositional attitudes. But, as seen in the worldview model of polarization and argued throughout this introduction, the connections between propositional attitudes and affective judgments are of central importance to the actual function of any given belief system (or worldview). This chapter uses Boutyline and Vaisey's (2017) belief network analysis to examine the American National Election Studies, comparing 1992 and 2016. But beyond examining just a network of propositional beliefs, I also look at the network of affective items, as well as the combined network of both attitudes and affects. In this way I show, first, that the different kinds of networks are similar enough to be comparable, but different enough to provide analytical leverage; and second, that the integration of affects and beliefs has changed over time, in line with polarization. This increasing integration of emotion and attitude demonstrates a fundamental piece of the worldview model on a larger scale and it opens the way for further research to elaborate on these methods.

Together, these chapters make significant empirical and theoretical contributions, both to the study of polarization specifically and beyond. Empirically, they demonstrate quantitatively the asymmetric nature of mass polarization in the US, helping bring our quantitative understanding of polarization into line with the clear qualitative asymmetry we can observe. In doing this, my dissertation expands our knowledge of the ways in which polarization fundamentally and differentially drives partisans' understandings of the political landscape. My findings help to elaborate the links between party as organization, partisan identity, and ongoing political change, highlighting both the structural role of intra-party conflict and the individual role of emotion. The worldview model itself contributes to ongoing definitional debates in the study of polarization, showing how the particular content of political identities matters and reorienting quantitative definitions away from standard affective measures towards the specifically interpretive functions of identity and emotion.

These findings have impacts beyond the study of polarization alone. The combined investigation of affect and attitudes, especially the empirical and theoretical focus on the ways in which they are linked together, significantly extends previous work on belief systems, showing that affect must be considered as a central part of such systems. This, then, has important implications for the study of pluralism and political conflict. By highlighting the importance of affect, my work gives us a new way to think about how opinion cleavages become politically relevant and opens a new area for empirical investigation using the tools I develop in the dissertation.

Perhaps the most important thing this dissertation does, however, is to substantively improve our understanding of the current fraught political moment. The work presented here has clear explanatory power; although it cannot say exactly why the January 6th insurrection happened, why election officials are being harassed, or why we are seeing militant disruptions of LGTBQ events for children, it can tell us a lot of about why these *kinds* of things are happening and why they mostly involve Republicans. Worldview polarization helps us understand not only that Republican partisans are more polarized, but what that means and, in so doing, helps make clear the connection between polarization and these extreme political actions. The definition of polarization in terms of persistent, patterned,

relationships between partisan affect, threat, and fear allows us to see how Republican identity is increasingly tied up with a worldview which understands political contestation through a lens of threat and fear. It is this shifting orientation towards politics which can help make sense of our ever more precarious political reality.

Chapter 2

Negative Partisanship is Not Enough: Affective Polarization, Asymmetry, and a Failed Theory Contest

2.1 Introduction

Polarization has been a subject of popular and scholarly concern for nearly three decades, driven by fears of democratic dysfunction and societal fragmentation. But progress has been uneven: while the reality of elite polarization became clear quickly, for years there was no consensus on whether the mass public was polarized *at all* Fiorina, Abrams, and Pope (2011). The introduction of affective polarization, defined by how partisans feel about one another (rather than in terms of political attitudes), was a major breakthrough. It provided clear evidence of mass polarization and its consequences, helping bring together our understandings of the elite and mass levels (Iyengar et al. 2019, 2012; Mason 2015). Major mismatches still remain, however; in particular, we know that at the elite level Republicans are substantially more polarized than Democrats (Jordan, Webb, and Wood 2014; McCarty 2019). Indeed, this asymmetry is arguably *constitutive* of polarization in the United States (Bermeo 2019; Hacker and Pierson 2005, 2020; McCoy et al. 2018; Pierson and Schickler 2020). But whether this constitutive asymmetry has a corresponding expression at the mass level remains an open question. And this uncertainty is exacerbated by a major conceptual issue in the study of affective polarization: if polarization is about partisans' feelings, which feelings matter? Specifically, there is ongoing disagreement about whether dislike of the other party alone is enough, or whether positive feelings towards one's own party are a play a constitutive role as well. These two theories, which I'll refer to as "negative partisanship" and "net affect", have thus far enjoyed an uneasy coexistence, but the choice between them has far reaching consequences for how we understand polarization in general and asymmetry in particular.

In principle, it should actually be relatively straightforward to address both these questions at once by using the empirical context of asymmetry as the grounds for a theory contest: given existing evidence of asymmetry towards Republicans, net affect and negative partisanship theories produce clearly divergent expectations about what that asymmetry should look like. However, in carrying out this contest, I find something strikingly unexpected. Drawing on

data from the American National Election Studies (ANES), I examine differences in affective polarization between the parties over time, net of demographic controls, and find asymmetric polarization in favor of *Democrats*. Further surprising both theories, this asymmetry is driven almost entirely by in-party affect; that is, Democrats like their own party more than Republicans like theirs. What's more, I also find an asymmetry in the *relationship* between in- and out-party affect: for Democrats, liking their own party more is associated with liking the Republican Party less, but there is no such relationship *at all* for Republicans. These results substantially challenge both dominant theories of affective polarization: they raise questions about the link between our measures and our concepts and complicate the fundamental claim of a simple and straightforward relationship between affect and polarization.

The argument proceeds as follows: I begin by introducing the net affect and negative partisanship approaches to affective polarization in more depth, detailing the (not always obvious) points of contention and demonstrating the real conflict between them. Then I elaborate on the above discussion of asymmetry, showing how the issue of mass asymmetry provides the perfect grounds to test net affect and negative partisanship against one another and setting the stakes of the contest. Building on this, I then discuss my data and analytical strategy in more detail before presenting my findings. Finally, I lay out the consequences of this unexpected asymmetry in affect towards Democrats. Although this in no way should be taken to undermine the central role affect plays in the study of polarization, nor the major contributions by work based on net affect and negative partisanship approaches, the findings presented here necessarily change how we approach affective polarization. In particular, they show us that we must begin to think more seriously about historical and institutional context, and how this context may play a role in *defining* what exactly we should call polarization. I close by offering one preliminary example of what such contextually oriented thinking might look like.

2.2 Literature Review

2.2.1 Polarization, Affect, and Affective Polarization

After over a decade and a half of work on polarization which focused almost entirely on political attitudes, the introduction of affective polarization by Iyengar, Sood, and Lelkes was a major breakthrough (2012). Attitudinal theories understand polarization as a kind of disagreement, defining it in terms of various distributional properties of political attitudes, e.g. bimodality, extremity, etc. In comparison, affective polarization's innovation is twofold: first, and most obvious, it discounts attitudes in favor of looking at how partisans feel each about other. But this is grounded in a more fundamental difference: seeing partisanship primarily as a *social* identity rather than as an adjunct to individual's political attitudes. As a result, attitudinal distance is no longer the salient issue, *social distance*, the in-group/out-group bias between partisans, is (Iyengar et al. 2019, 2012). This new definition of polarization had an obvious degree of face validity, i.e. the degree of nominal agreement on issues makes little difference if two groups hate each other, as well as strong empirical backing. While attitudinal approaches had produced mixed results and ongoing debate Fiorina et al. (2011), work on affective polarization showed (and shows) clear evidence of increasing polarization

and its consequences (Druckman et al. 2021; Iyengar et al. 2012; Iyengar and Westwood 2015; Mason 2015).

Measuring affective polarization was also straightforward and easy to do in much existing quantitative data. Most commonly used are feeling thermometers, which ask respondents to rate their feeling towards a named group (i.e. the Democratic and Republican Parties) on a scale from cold to warm. Also used are trait batteries, where respondents rate how well various positive (honest, intelligent) and negative (hypocritical, selfish) traits fit the parties. Finally, some questions attempt to directly measure social distance by asking respondents about the comfort level being friends or neighbors, or having their children marry, someone from the other party. As Druckman and Levendusky (2019) have shown, these methods, though related, get at slightly different things. Thus, since I am here interested in the general cognitive construct of affective polarization, rather than particular behavioral outcomes, I will focus on feeling thermometer measures.

However, if polarization is about partisans’ feelings, there has been some uncertainty from the start about *which* feelings mattered. Specifically, there are two implicitly (and sometimes explicitly) conflicting views of in-party affect. On the one hand, there is the tendency to see polarization foremost as a function of affect towards the out-party, i.e. how much one side’s partisans dislike the other, seen most explicitly in recent work on “negative partisanship” (Abramowitz and Webster 2016; Finkel et al. 2020; Webster and Abramowitz 2017). On the other hand, there is work which understands polarization as a matter of *both* out-party affect and in-party affect, what I will call a “net affect”¹ approach (Mason 2015). Here, out-party affect is still important, but how much partisans like their own party also plays a constitutive role. This leaves us with a foundational theoretical question which hasn’t fully been dealt with. Indeed, even though this issue is often left implicit, it is absolutely crucial: on it hinges the basic issue of what polarization *is*. This quickly becomes apparent if we consider where these two tendencies come from.

Affective polarization was originally grounded in social identity theory (and has in fact been called “social identity polarization”), and this is where the focus on net affect originates (Iyengar et al. 2019, 2012). Social identities are fundamental to being human, they guide us through the world and play an essential role in our sense of self, but identifying with a group necessarily creates a privileged in-group of which we are a part and a less privileged out-group (Brewer 1999; Tajfel et al. 1971; Tajfel and Turner 1979). This basic “us” and “them” division is at the heart of social identity theory and of affective polarization, with an extensive literature showing that even the most meaningless group identification creates a degree of in-group favoritism, a bias tied up in nothing more than group membership (see, e.g., Billig and Tajfel 1973; Tajfel et al. 1971). And positive feelings towards the in-group are the *sine qua non* of this bias: bias is obviously a matter of comparative evaluation, but previous work has shown that the minimum for in-group favoritism is driven by evaluating the in-group more positively rather than the out-group more negatively, and it is also clear in real world situations that one can have “in-group love” without “out-group hate” (Brewer 1999; Gaertner et al. 1993). However, situations of political conflict and perceived threat link the

¹Because, from a feeling thermometer perspective, combining in- and out-party affect is most straightforwardly done by subtracting the latter from the former, resulting in a measure of “net (positive) affect” towards the in-party.

two together (Brewer 1999), which fits precisely with a net affect understanding of affective polarization (Iyengar et al. 2012; Lu and Lee 2019; Mason 2015). Furthermore, both are linked to the salience or strength of an identity as well, with increasing commitment driving both in and out-group bias (Huddy 2001). This is the basis on which net affect approaches understand in-party affect as playing a constitutive role in affective polarization. And this understanding is not some ignored implication of the underlying theory; rather, it has been routinely recognized and built into specific definitions of affective polarization (see, e.g., Lu and Lee 2019; Mason 2015)

At the same time, however, there has also been longstanding *empirical* uncertainty about the importance of in-group affect. From the very earliest work on affective polarization there has been a recognition that empirically speaking, in the case of US partisan polarization, overall changes in net affect have been driven almost entirely by worsening feelings towards the out-party, while in-party affect has hardly changed at all (Iyengar et al. 2012). This is shown quite clearly in Figure 2.1, which plots mean in-party affect, out-party affect, and net affect (in-party affect minus out-party affect) over time, based on data from the American National Election Studies.² Though mean in-party affect has fluctuated somewhat, there is nothing like the precipitous and consistent decline seen with out-party affect. Indeed, it would be fair to say that this substantial out-party hostility is the thing which is truly *new* to modern politics; if we wanted to point to a single statistic as evidence of increasing polarization, this is would be it. As a result, even much work explicitly grounded in social identity theory focuses solely on negative affect. Often this is a purely empirical matter, but sometimes it is a matter of definition as well. However, this has remained largely unremarked from a theoretical standpoint. That is, much work draws on social identity theory and excludes in-party affect from investigation, but doesn't offer any particular theoretical justification or draw any theoretical conclusions from this decision. To be clear, it is certainly possible, in a social identity theory based understanding, that in- and out-party affect might matter differently for polarization. However, to say that in-group affect doesn't matter at all, that it isn't constitutive of polarization, is simply not compatible with the underlying theory of social identity.

And so recently some work has begun shifting away from using social identity theory as its primary theoretical justification and instead explicitly taken out-group hostility as fully definitive. Abramowitz and Webster (2016) introduce the idea of “negative partisanship”³ as the primary concept of interest, highlighting the fact that partisans increasingly dislike the other party more than they like their own. Though work on negative partisanship often also discusses affective polarization as a separate concept, the actual relationship between them is never really specified and so it frequently functions as a replacement for affective polarization, even if that is not explicitly argued or intended (Abramowitz and McCoy 2019; Abramowitz and Webster 2016; Webster and Abramowitz 2017). Beyond negative partisanship, Finkel et al. have attempted to shift away from the terminology of polarization entirely with the concept of “political sectarianism”, the “moralized identification with one political group and

²I will discuss the ANES in more detail in the methods section below, so for the moment it will suffice to say that it is nationally representative survey whose use is incredibly common in the study of polarization.

³Note that, in the rest of the paper, I am using “negative partisanship” in a slightly more general way to include all those approaches to affective polarization which don't see in-party affect as constitutive, not only referring to Abramowitz and Webster's concept specifically, as here.

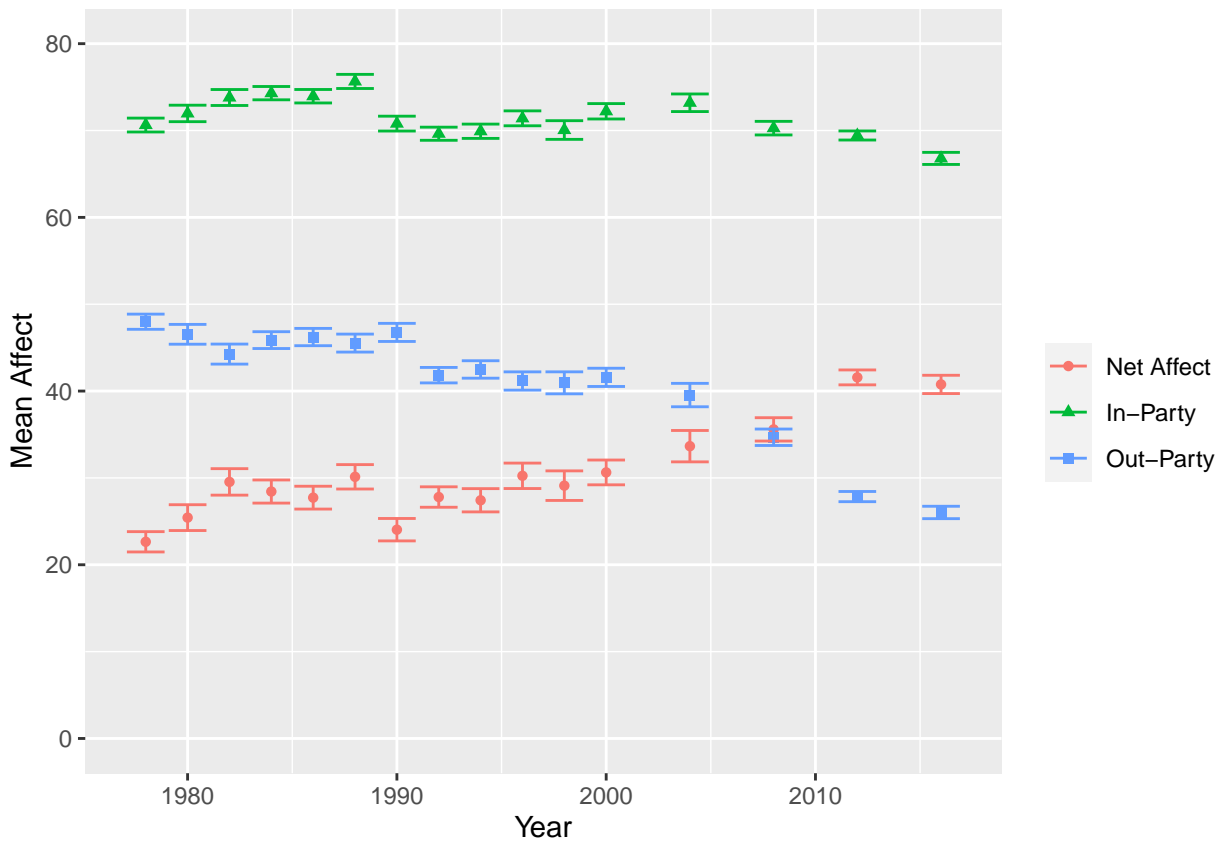


Figure 2.1: Mean Partisan Affect Over time, 1978-2016

against another” (2020:533). Both negative partisanship and political sectarianism maintain the innovations which made affective polarization so useful in the first place (e.g. the focus on social relations, affect, and identity rather than attitudes), but they are nonetheless conceptually and practically distinct.

Of course, we may be tempted to think that these distinctions don’t really matter, that they are largely semantic. Or we might imagine that in and out-party affect simply matter differently, or are connected to different aspects of polarization (a famously multidimensional concept). Indeed, the current state of the literature might seem to indicate this. On the one hand, scholars not only continue to use both negative partisanship and net affect based approaches, but also a variety of approaches in between: from work which looks solely at net affect, to work which looks at net affect but also disaggregates it into in- and out-party affect, to work which empirically focuses on out-party affect alone but still grounds itself theoretically in social identity theory, and so on. On the other hand, we have some empirical evidence that in- and out-party affect do indeed matter differently in terms of both how much and how they matter to different political consequences (see, e.g., Banda and Cluverius 2018; Iyengar and Krupenkin 2018; Iyengar and Westwood 2015; Jardina 2021; Yarchi, Baden, and Kligler-Vilenchik 2020). However, this is in fact further evidence of the problem: finding that in- and out-party affect matter differently is fairly clear evidence that these are *not* irrelevant word game distinctions. And we can’t simply throw up our hands and say that they matter differently. Our question isn’t whether in- and out-party affect both matter in some general sense; rather, our question is a definitional one. The distinction we’re interested in is between *doesn’t matter at all* and *plays a constitutive role*. In other words, when we say that affective polarization is increasing, what does that actually mean?

And the answer to this question matters, because, when we consider their theoretical implications more carefully, we can see that net affect and negative partisanship approaches are not actually compatible. They start from similar places: for both theories, polarization is about partisan identities and how they structure social and political relations. Further, drawing originally on social identity theory, both theories define this division in affective terms. However, whether in-party affect plays a constitutive role or not implies radically different theories of how identity and division actually work. By incorporating in-party affect as constitutive, net affect approaches make two key theoretical claims: first, that in-group affect is linked to the strength or meaning of an identity. And second, that in-group and out-group affect may *interact* in the production of polarization. On the other hand, negative partisanship approaches necessarily imply the opposite: either in-group affect doesn’t have any real bearing on the strength of in-group identity, or the nature of that identity isn’t actually relevant to polarization. These may seem like straightforward differences, but they are straightforward because they are basic, and their theoretical, methodological, and analytical consequences are wide-ranging. There are obvious methodological differences in terms of what to measure, but beyond that these differences change how we should interpret their results, how we can incorporate them into a wider understanding of polarization beyond affect, what kinds of methodological innovations might allow us to make further progress, and much more. Put another way, the issue at stake here is how we substantiate the link between our survey based, quantitative, definitions and the empirical and conceptual phenomenon of polarization that we’re really interested in. This is, in essence, an issue of conceptual validity lurking in the foundations of one of our major understandings of polarization and it is damaging our

ability to make progress.

These problems are becoming more and more visible in the literature. As mentioned above, we are finding more and more evidence that in and out-party affect don't simply differ in how much they matter, but differ in what they matter to. Thus, while we know out-party affect frequently has stronger links to various political and non-political consequences, to other aspects of polarization, etc. (see, e.g., Banda and Cluverius 2018; Iyengar and Westwood 2015), we also have evidence that in-party affect may play a unique role in affective polarization, with performances of positive in-group sentiment being used as a way to mark underlying conflict between groups (Yarchi et al. 2020:17). Similarly, although Lu and Lee (2019) focus analytically on negative affect, they theorize a number of ways that negative emotions towards the out-group may result in more positive feelings towards the in-group, exactly the kind of interactions accommodated by a net affect approach but largely ruled out by a focus on negative partisanship. And Druckman and Levendusky (2019), in their investigation of what it actually means to measure affective polarization, find that neither negative partisanship nor net affect feeling thermometer measures have a clear advantage in terms of how well they correlate with other metrics (e.g. trait batteries, social distance measures). Given all this, even though in-party affect shows very little macro variation and out-party affect is driving changes in overall net affect, we still cannot simply disregard in-party affect on empirical grounds. We must confront the theoretical issues head on.

But despite their centrality (or arguably *because of it*), this is difficult to do. To begin with, we know that we lack the kind of clear aggregate variation on in-party affect that would provide us with an obvious answer; indeed, this lack is precisely one of the motivations for negative partisanship approaches. However, we also have evidence that in-party affect *does matter*, sometimes in ways indicative of unique importance. Thus, there's a significant difficulty in disentangling the clear, evident, well known fact that out-party affect has been the driver of changes in net bias from the question of whether in-party affect is still meaningfully *constitutive* of affective polarization. In other words, we can already see that the variation isn't there to answer the question in the obvious way we might like. That said, what we can do is try and draw inferences about what the two approaches should look like on a smaller scale. For instance, if we were to find low in-group affect in a case we *know* to be polarized (in the general conceptual sense of "division of society into meaningful sides"), then we could fairly say that in-group affect must not be constitutive. Likewise, finding that the more polarized of two groups consistently has stronger positive feelings about the in-group would lend some credence to the idea that in-party affect plays a definitional role. What we need, then, is a situation which would allow us to enact this kind of theory contest. The question of whether mass polarization is asymmetric gives us exactly this.

2.2.2 Asymmetry as a Theory Contest

Although the issue of mass asymmetry remains open, we do have clear evidence of asymmetry at other levels. At the elite level, both congressional voting patterns and party platforms show Republicans stampeding to the right while Democrats saunter leftwards (Jordan et al. 2014; Lewis et al. 2020; Pierson and Schickler 2020). There is also circumstantial evidence of asymmetry at the mass level. This is perhaps most obvious in the January 6th insurrection and the harassment of election officials which both preceded and followed it (So 2021; Tan et

al. 2021; Wines 2020), but it’s also visible in the way that popular insurgency within the party has been able to drive repeated turnover in party leadership (see, e.g., Eric Cantor, John Boehner, and Paul Ryan) and in continued increases in right-wing extremism (Hacker and Pierson 2020; Skocpol 2020).

Indeed, there’s a strong case to be made that this asymmetry is *constitutive* of polarization in the United States. Although assigning responsibility for increasing polarization to any single cause or actor would be farcical, numerous scholars have identified the Republican Party as the primary “polarizing actor” whose actions are crucial in explaining both how we got here and where we’re going (Bermeo 2019; Grossmann and Hopkins 2015; Hacker and Pierson 2005; Mann and Ornstein 2016; McCoy et al. 2018; Pierson and Schickler 2020; Sinclair 2006). This determination is grounded partly in the way the party has pursued (at times) strategies of deliberate intransigence (Mann and Ornstein 2016) and, increasingly, driven the use of “constitutional hardball” (the strategic choice to violate established norms to gain a partisan advantage) (Hacker and Pierson 2014; Tushnet 2004). But this goes beyond the party itself, extending to the unique organizational ecosystem which surrounds it, where organizations act as “surrogates” for the party and push polarizing narratives (Hacker and Pierson 2020; Hochschild 2016; Polletta and Callahan 2017). And this all links back to foundational differences between the parties: the Republican Party is ideologically based in a way the Democratic Party simply is not, an ideological basis which has become more and more extreme (Grossmann and Hopkins 2015; Mann and Ornstein 2016; Skocpol 2020). However, despite all of this, the question of whether mass polarization is asymmetric remains quantitatively understudied.

As such, mass asymmetry is the perfect place to address the role of in-party affect in affective polarization. In short, it is an open empirical question worth answering, about which we have extremely strong expectations, which thus allows us to enact precisely the sort of theory contest introduced above. Specifically, while we know that everyone has grown more polarized since the late 1970s (Iyengar et al. 2019), based on the evidence presented above we clearly expect Republicans to have polarized more than Democrats, and as such we can make clear hypotheses about what we should see in the development of affective polarization over time. Now, to be clear, we have no particular expectations about mass asymmetry based on the role in-party affect does or does not play. However, net affect and negative partisanship approaches do provide clearly different expectations for *how* any asymmetry should manifest.

Work which understands affective polarization as a matter of *net affective bias* obviously expects to see asymmetry on a *net* measure of affect. But more than that, it also expects that asymmetry to persist when the net measure is disaggregated; that is, for there to be asymmetry on *both* in- and out-party affect individually. As discussed above, affective polarization is one case where social identity theory sees in and out-group affect as fundamentally linked: the stronger an identity, the stronger the affective bias against the out-group *and* the affective bias towards the in-group. As such, an asymmetry on one should mean an asymmetry on both. This doesn’t require an *equal* asymmetry on both in- and out-party affect, but it must be present: it could certainly be the case that in-party affect changes less overall, but social identity theory expects that it will change. We can actually take this a bit further and say instead that we would expect a strengthening inverse *relationship* between in-party and out-party affect over time, one which gets stronger faster for Republicans than Democrats.

On the other hand, a negative partisanship approach expects to see asymmetry *only* on

out-party affect. At first, this might seem like a strong claim and that a focus on negative partisanship is really merely *ambivalent* to in-party affect. However, this would ignore the basis for focusing on negative partisanship in the first place. This basis is twofold: first, an essentially *complete* lack of interesting variation on positive in-party affect over time (in the aggregate), and second, an argument that there is a more direct link between negative out-party affect and the substantive, real-world, consequences of polarization (e.g. contempt for members of the out-party). Thus, the negative partisanship perspective expects asymmetry only on negative out-party affect, as asymmetry on positive in-party affect is exactly the kind of meaningful variation which has been implicitly (and sometimes explicitly) ruled out. Indeed, any sort of consistent development in in-party affect over time would be somewhat anomalous from this perspective. Not only should there be no asymmetry on in-party affect, but there should also be no real relationship between in- and out-party affect developing over time.

We can now see how framing our investigation around mass asymmetry solves some of the difficulties inherent trying to address the issue of in-party affect. The main problem is the apparent contradiction between a lack of clear aggregate variation on in-party affect and various evidence that in-group affect plays a (potentially unique) role in affective polarization. Looking at the question of asymmetry over time helps us in a few ways. Most simply, it means comparing subgroups (i.e. Republicans and Democrats) and therefore opens up the possibility of variation at the subgroup level which isn't visible in aggregate. Indeed, our interest isn't in overall change per se, but rather in *relative* change between groups. Furthermore, looking at the development of asymmetry over time means that we're looking at deviations from the overall trend. The question is whether the change over time for Democrats is the same as it is for Republicans, and it is those changes which should correspond to the underlying expectations of a net affect of negative partisanship understanding of polarization. And it's not just that we're comparing groups in the context of an overall trend, it's that it's *these* two groups. The asymmetry we're expecting to see is, based on all evidence, foundational to polarization in America. Because of this, we might expect to see differences here which might not show up along other axes. In other words, if in-party affect plays a constitutive role in affective polarization, then we should see it here, in this comparison. If we don't see it here, then it likely doesn't matter.⁴

2.2.3 But What Does It Mean?

To sum up then, here's what we expect to see: as a baseline, affective polarization should increase for all partisans. However, based on a range of evidence, we expect it to increase asymmetrically (that is, to increase more, faster) for Republicans than Democrats. If net affect approaches are right and how much you like your own party is an essential part of affective polarization, then we should see at a least a small increase in how much Republicans like their own party compared to how much Democrats like theirs, along with a strengthening association between in-party warmth and out-party animus for Republicans. On the other

⁴To be clear, I don't mean at all to imply that disaggregating partisans to compare Democrats and Republicans is some kind of innovation in itself. That would be self-aggrandizing in the extreme. Rather, the point is simply that doing so in this way, over this span, to look at this question, is, if not new, then at least useful.

hand, if negative partisanship approaches are correct and in-party affect is basically irrelevant to the issue of polarization, then we should see no real changes on in-party affect or its relation to out-party animus.

Having such a clear theory contest has both upsides and downsides. With strong expectations built on the two major affective theories of polarization along with extensive empirical evidence of asymmetry, findings which don't fit raise fundamental questions for our current theories of polarization. On the one hand, finding that mass polarization is in fact symmetrical, occurring evenly and equally for Democrats and Republicans, wouldn't be decisive for the issue at hand, but would imply a shocking disconnect between qualitative observations and our quantitative measures of polarization. And that would, arguably, be a best case scenario: finding asymmetry in favor of *Democrats* or asymmetry on in-party affect would wildly contradict any and all of our expectations. It would imply either an even stranger disconnect between empirical observation and quantitative measure or a foundational issue with our understanding of polarization. It would fundamentally undermine the ways in which we have, up to now, been understanding the links between measures of affect and the concept of polarization as such. But this is in fact what we find: asymmetry towards Democrats and asymmetry on in-party affect.

2.3 Data and Methods

Conceptually, the analytical strategy required here is straightforward: in order to compare the predictions of net affect and negative partisanship approaches about asymmetric polarization, I just need to look at the development of polarization over time, by party, and see how it manifests on net, out-party, and in-party measures of affect. However, things are a little more complex in practice. First, rather than relying simply on graphs of affect over time,⁵ I use interactions and year fixed effects to show how associations between partisan identity affect change over time, net of other factors. Additionally, as discussed above, I'm not interested only in the association between party identification and affect, but how in and out-party affect are related for partisans on each side and whether that changes over time. This introduces additional complexities in modeling and interpretation which I discuss below.

I conduct this analysis using data from the American National Election Studies (henceforth the ANES). The ANES is a nationally representative survey which has run in every presidential election year since 1948 and in midterm years from 1954 to 2002. This extremely long tenure (longer than the General Social Survey by more than twenty years) combined with its focus on elections (and thus politics) make these data extremely well suited to the study of polarization, particularly over time. Of course, the ANES must strike a balance between asking consistent questions over time and changing the survey to match contemporary concerns, preventing me from making use of its full historical scope. Nevertheless, the questions available do allow me to look at the period from 1978 to 2016, which is almost precisely the period during which we know affective polarization developed and thus our period of empirical interest.⁶

⁵Although I do also include straightforward graphs of net, in, and out-party affect over time, by party, in Appendix A.

⁶I provide additional contextual analyses using alternative dependent variables in Appendix B which look at the period from 1954 to 2004. These analyses confirm the affective polarization really only takes off in this

Furthermore, looking at the development of differences over time in this way is important because polarization is a *process*, and this allows me to look for relatively durable differences in trajectories, rather than simply spotting differences in single survey years (DiMaggio et al. 1996).

2.3.1 Variables

To look at asymmetry in mass affective polarization, I first need measures of affect as my dependent variables. The goal here is to measure net affect as well as in- and out-party affect on their own. As mentioned above, *feeling thermometers* for affect towards the Republican and Democratic parties are one of the main ways to do this. Feeling thermometers in the ANES ask respondents to rate how they feel about a group (in this case, each party) on scale from 0 to 100, where 0 is cold and 100 is warm.⁷ I use these to capture in- and out-party affect, based on respondent self-ID, and combine them into a net affect measure by simply subtracting out-party from in-party affect.⁸

For my independent variables, I first use a simple dummy variable for party, 1 if the respondent is a Democrat and 0 if they are a Republican.⁹ Then, to get at development over time, I include the year of the survey, both as an indicator (i.e. as a series of dummies with one for each year) and as an integer. This allows me to look at trends over time *net* of year-to-year fluctuations (Brooks and Manza 2013 use a similar strategy).

I also include a variety of additional controls in the regression models. To focus purely on party identification as such, I control for both the strength of partisan identification (a scale from 1, leans, to 3, strong) and ideological extremity (a scale from 0, moderate, to 3, strong liberal/conservative).¹⁰ These are accompanied by a set of demographic controls: years of education, income, female (dummy), African American (dummy), age, southern residence (dummy), and frequency of church attendance.

2.3.2 Methods

The purpose of the analysis is to demonstrate whether affective polarization has developed asymmetrically for Democratic versus Republican partisans, and further to show the role of in vs. out-party affect in any asymmetry. In order to do this, I present a series of three regressions below, one for each of the feeling thermometer dependent variables: net affect, in-party affect, and out-party affect. In each model, I include the above mentioned set of controls and the dummy for Democratic identification. I allow for the development of partisan

more recent period.

⁷Note, however, that the ANES cumulative data file codes all responses between 97 and 100 as 97 (for convenience of harmonizing across surveys), so the actual data ranges from 0 to 97.

⁸These measures are undefined for true independents, as they have no in or out party. However, they *are* defined for independents who lean towards one of the parties, as I have followed standard practice and counted partisan leaners as identifying with the party they lean towards (Druckman, Peterson, and Slothuus 2013).

⁹As mentioned, I follow previous work and count “leaners” as partisans.

¹⁰Previous work has highlighted the role of partisan sorting, or the degree to which identification and ideology match up (Mason 2015). All results presented here are substantively the same if partisan sorting is used instead of strength of identification and ideology.

asymmetry over time by interacting Democratic identification dummy with year.¹¹ I do the same with the strength of partisan identification and ideological extremity, because previous work leads us to expect these to have stronger effects over time (Mason 2015).

As it stands, these models would allow me to look at the link between party identification and affect over time, but not at the relationship between in- and out-party affect for Democrats and Republicans, which is one of the major ways to not only establish a meaningful asymmetry, but also discover whether in-party affect plays an important role in it. To do this, the in- and out-party affect models each include the other measure of affect (i.e. out- or in-party) as an independent variable in a three-way interaction with both Democratic identification and year. As a result, I see not only any asymmetry on affect, but also whether in- and out-party affect are linked to one another differently for Democrats and Republicans, and whether these links change over time.

However, while two-way interactions are fairly straightforward to interpret, three-way interactions raise additional difficulties for substantive interpretation, even if they are not conceptually difficult. In a standard OLS regression with no interactions, each coefficient represents the change in the dependent variable given a unit change in the predictor. This is the “marginal effect” of that predictor and is a good measure of that predictor’s substantive relationship with the outcome. A two-way interaction means that a predictor’s marginal effect is contingent on the value of another predictor (i.e. the one it is interacting with), and a three-way interaction makes it contingent on the value of yet another predictor.

It is precisely this contingent variation that we are interested in: below, we will focus on two questions. First, how the marginal effect of Democratic identification on net, in, or out-party affect varies over time, and second, how the marginal effect of in/out-party affect on out/in-party affect varies over time for each party. I calculate all of these as a kind of *average marginal effect*, obtained by calculating the marginal effect of the variable of interest at every observed value and taking the average. Specifically, these are *marginal effects at representative values*, in which the above procedure is followed repeatedly with some covariates held at values of interest (i.e. across the range of survey years or Democratic vs. Republican identification). This will result in clearly interpretable values showing the substantive association between our variables of interest, e.g. if Democratic identification had a marginal effect of 1 point on net affect in 1978, but an effect of 7 points in 2016, then we can say that partisan affect has developed *asymmetrically* because the impact of being a Democrat (as opposed to a Republican) has increased.¹²

2.4 Findings and Discussion

Before looking at the marginal effects themselves, Table 2.1 shows the results of the three regressions.¹³ Each column of the table shows the results for the regression on one dependent

¹¹As mentioned above, year is a continuous variable in these interactions, while it is included as a dummy in the main effects model. This can show trends over time net year-to-year fluctuations.

¹²For more on the use of marginal effects and the use of AMEs vs. MEMs vs. MERs, in addition to the already mentioned Hanmer and Ozan Kalkan (2013), Leeper (2018), and Williams (2012), see also Druckman et al. (2021) for an example of a similar empirical application and Long and Mustillo (2018) for a theoretical discussion.

¹³For a comparison of the main effects models with the interaction models presented here, see Appendix C.

variable: first net affect, then in-party affect, then out-party affect. What might first appear most striking is the size and direction of the main effects for Democratic identity, partisan strength, and ideological strength; however, this is simply an artifact of the model specification: the main effects must be large because the interactions are with the year of the survey, whose value ranges from 1978 to 2016. Given that, the interactions become much more interesting, and they are, indeed, where we should focus our attention. Looking at the interaction between year and Democratic identity, we see that, for both the net affect and in-party affect models, the impact of Democratic identity gets stronger over time. On the other hand, there is no significant difference between Republicans and Democrats on out-party affect. At the most basic level, then, these regressions indicate that Republicans and Democrats both dislike the other party increasingly over time to similar degrees, but that the development of in-party affect has been asymmetric in favor of Democrats.

Table 2.1: Effect of Democratic ID on Net Affect, In-Party Affect, and Out-Party Affect

	<i>Dependent variable:</i>		
	Net Affect (1)	In-Party Affect (2)	Out-Party Affect (3)
Democrat	-389.147*** (53.378)	-445.394*** (65.856)	-41.695 (158.644)
Dem x Year	0.197*** (0.027)	0.225*** (0.033)	0.022 (0.079)
Out Therm		1.487 (1.241)	
Dem x Out		4.854** (1.517)	
Out x Year		-0.001 (0.001)	
Dem x Out x Year		-0.002** (0.001)	
In Therm			5.205** (1.715)
Dem x In			1.640 (2.205)
In x Year			-0.003** (0.001)
Dem x In x Year			-0.001 (0.001)
Partisan Strength	-90.066** (33.385)	-42.827* (21.124)	-31.185 (28.160)
Partisan Strength x Year	0.052** (0.017)	0.025* (0.011)	0.013 (0.014)
Ideological Strength	-183.822*** (28.247)	-14.257 (17.973)	148.677*** (21.970)
Ideological Strength x Year	0.095*** (0.014)	0.008 (0.009)	-0.077*** (0.011)
Black	5.816*** (0.526)	4.127*** (0.327)	-1.444*** (0.412)
Income	-0.487** (0.156)	-0.630*** (0.097)	-0.156 (0.121)
Age	0.046*** (0.010)	0.048*** (0.006)	0.002 (0.007)
Gender	1.163*** (0.323)	1.361*** (0.200)	0.195 (0.251)
Education	0.047 (0.199)	-1.264*** (0.124)	-1.308*** (0.155)
South	-0.376 (0.351)	-1.741*** (0.218)	-1.384*** (0.273)
Church Attendance	0.560*** (0.105)	-0.317*** (0.065)	-0.858*** (0.081)
Observations	26,738	26,738	26,738
R ²	0.233	0.204	0.201

Note:

*p<0.05; **p<0.01; ***p<0.001
Year fixed effects not displayed.

2.4.1 How Big an Asymmetry?

In order to see just how asymmetric, we turn now to look at how the marginal effects of Democratic identification vary by year. Figure 2.2 shows the average marginal effect of Democratic identification on net affect in each year. In other words, Figure 2.2 shows us the difference in net affect associated with being a Democrat, rather than a Republican, over time. And it is a substantial difference: while in 1978 there was no significant difference at all, in 2016 the difference had grown to around 7.5 points. Now, this might not initially seem that consequential on a scale with a maximum value of 100; however, the total change in average net affect between 1978 and 2016 was just under 18 points, making this asymmetry larger than a *third* of the total change.

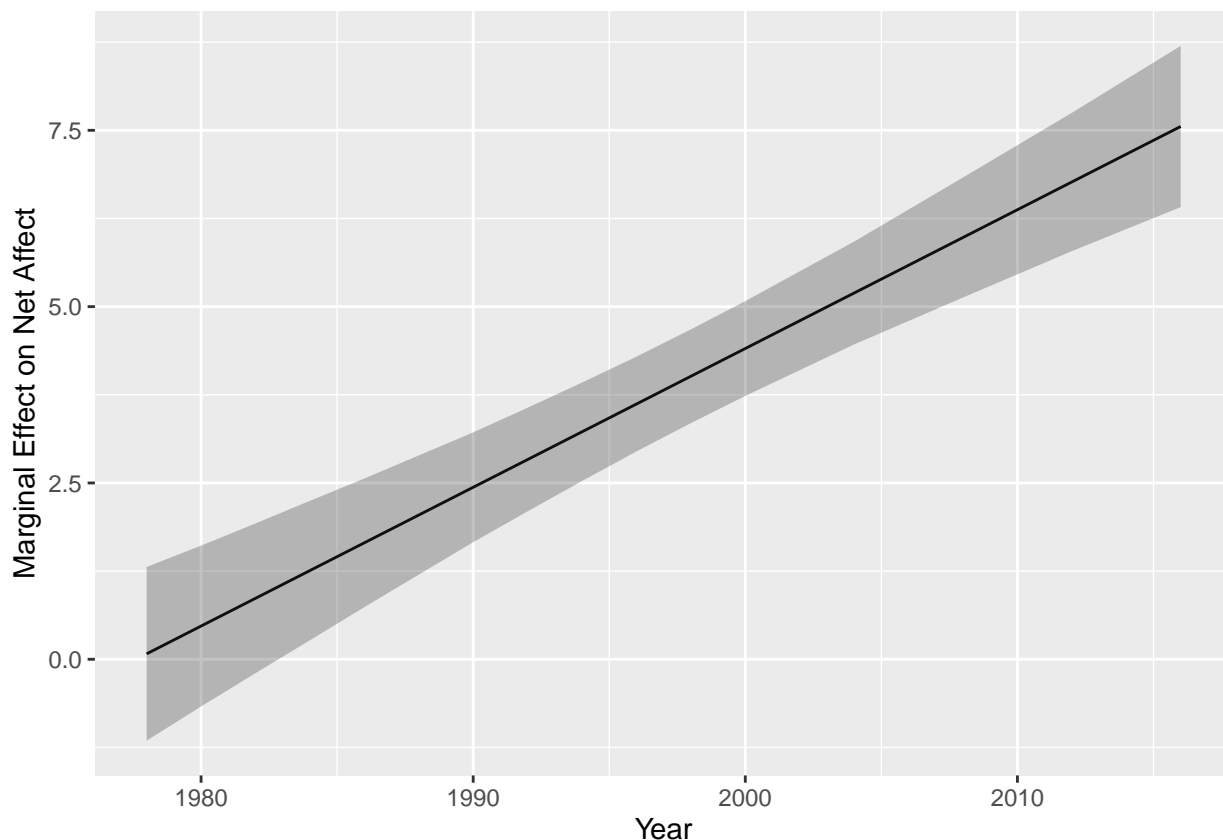


Figure 2.2: Marginal Effect of Democratic ID on Net Affect

That we see an asymmetry towards Democrats, much less such a substantial one, is surprising on its own, but looking at the marginal effects of Democratic identification on the separate in- and out-party affect models in Figure 2.3 really makes clear that this asymmetry is driven nearly entirely by *in-party* affect. While Democratic identification does, by 2016, have a small marginal effect on out-party affect, it's substantively very weak, and is statistically indistinguishable from zero in some alternative specifications of the model.¹⁴ On the other

¹⁴Note that this asymmetry on out-party affect is still an asymmetry *towards* Democrats: increased out-party affect would be increasing warmth towards the out-party, so *decreasing* out-party affect is linked to polarization.

hand, Democratic identification was associated with a 5.5 point increase on in-party affect in 2016, compared with no statistically significant effect in 1978. Once again, while this might at first seem relatively small, it is approaching a third of the total increase in *net affect* over the period.

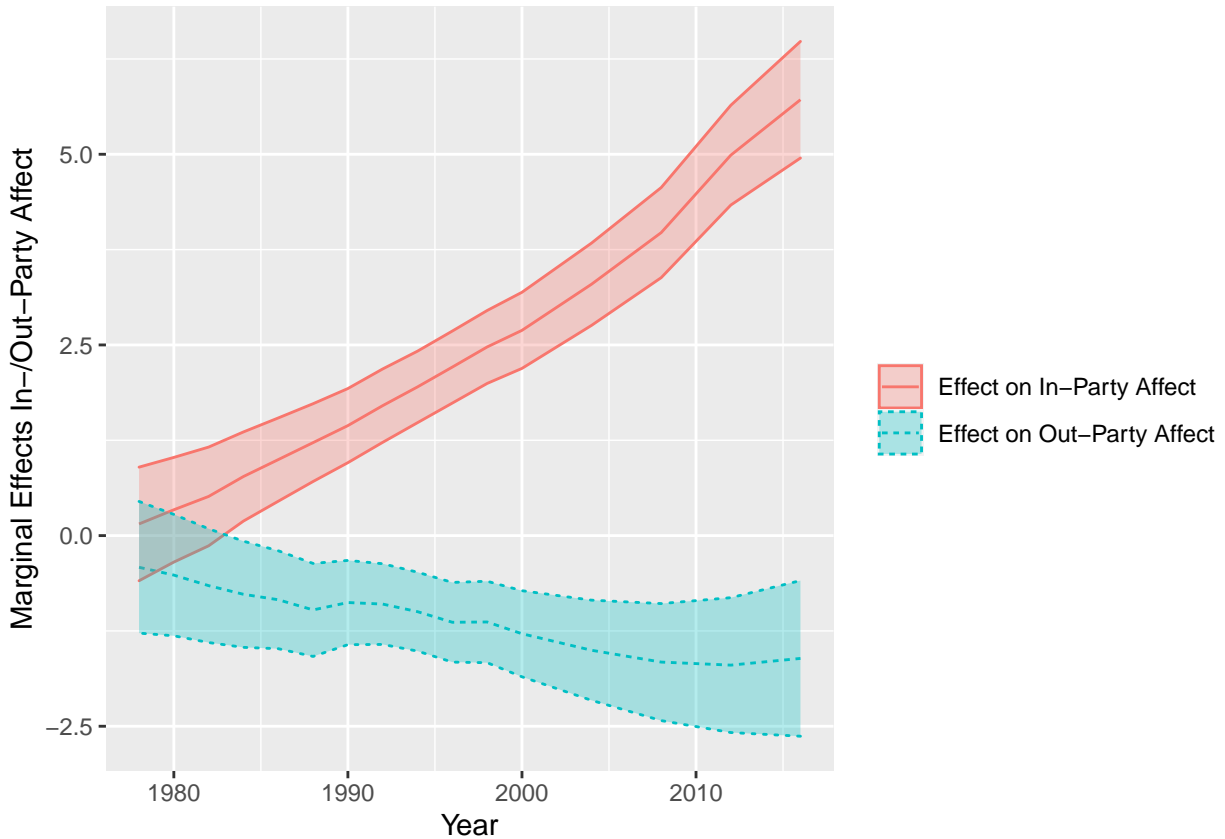


Figure 2.3: Marginal Effects of Democratic ID on In- and Out-Party Affect

The size of these effects is apparent even when compared to the impacts of partisan strength and ideological extremity on net affect, shown in Figure 2.4. Now, to recognize the obvious, both partisan and ideological extremity have larger effects on net affect than Democratic identification, particularly when we consider that marginal effects are not standardized: partisan strength is a three point scale and ideological extremity is a four point scale while Democratic identification is a dummy variable. However, the literature is pretty much unanimous on the point that partisan and ideological strength are key to explaining affective polarization, so it would be truly strange to find them not doing the lion's share of the explanatory work here. Thus, the fact that the effects of Democratic identification are even on the same scale speaks to their importance. This comparison also serves to remind us that the effects of Democratic identification are *net* of other covariates, including partisan strength or ideological extremity.

Based on this, we can clearly say that we do find a substantial asymmetry in affective polarization, but towards Democrats rather than Republicans, driven by in-party affect. This completely reverses our empirical expectations: we had expected an asymmetry towards Republicans, matching the clear asymmetry seen in elite polarization and elsewhere, and for

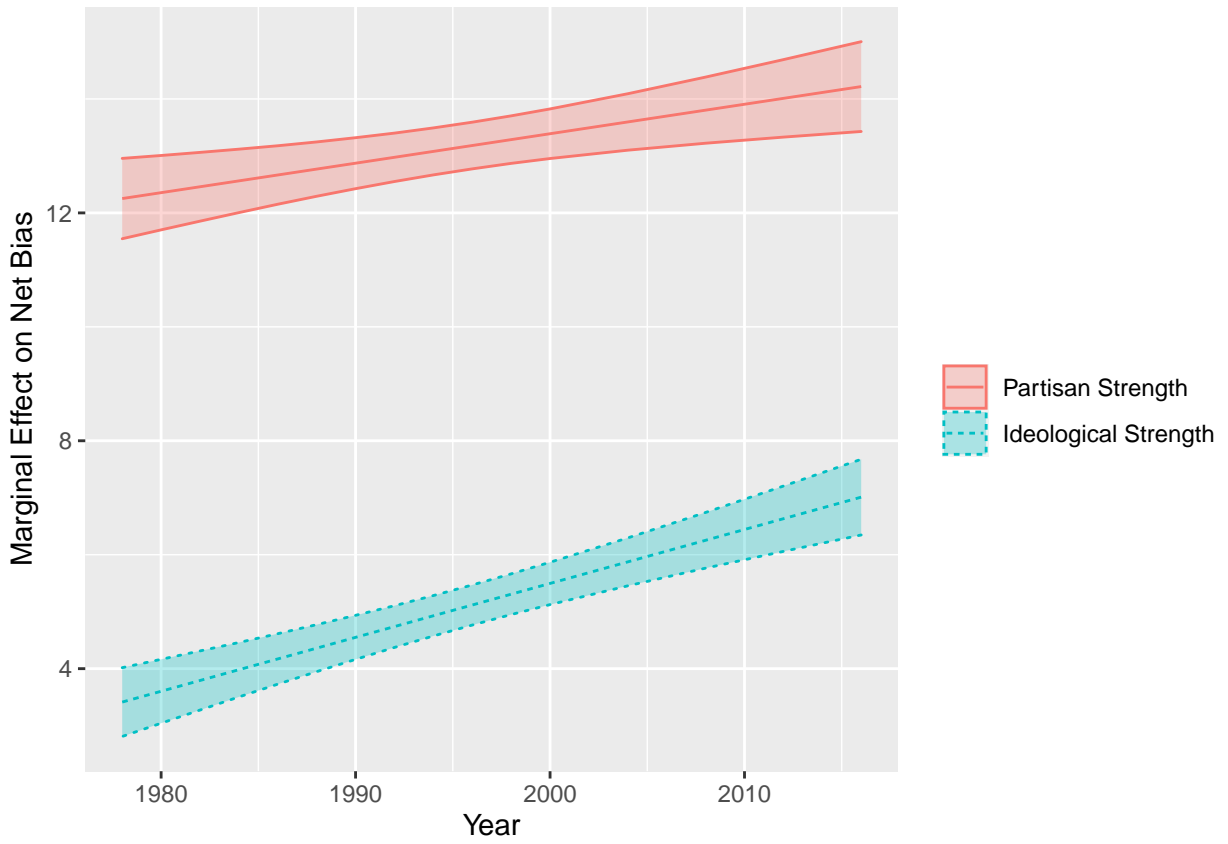


Figure 2.4: Marginal Effects of Partisan and Ideological Strength on Net Affect

that asymmetry to manifest either on both in- and out-party affect or on out-party affect alone. But, while this is a compelling finding already, examining the the relationship between in- and out-party affect strengthens it further.

2.4.2 In-Group Love and Out-Group Hate

Where the marginal effect of Democratic identification on affect showed us a basic asymmetry, looking at how the marginal effect of out-party affect on in-party affect varies over time and by party provides more foundational evidence of the role of in-party affect.¹⁵ On the one hand, a negative partisanship understanding of affective polarization implies that in-party affect and out-party affect shouldn't have any particular relationship at all, while a social identity theory based net affect understanding would expect them to develop a reciprocal relationship as polarization grows. Looking at variations in the marginal effect of out-party affect on in-party affect lets us look at this directly, which is what Figure 2.5 does.



Figure 2.5: Marginal Effect of Out-Party Affect on In-Party Affect, by Party

Based on Model 1 in Table 2.1, Figure 2.5 shows the marginal effect of out-party affect on in-party affect in each year, for each party. Thus, we can see that in 1978 a one point decrease in out-party affect was associated with a (statistically insignificant) .012 point decrease in

¹⁵I'm presenting only the out-party on in-party marginal effect here for the sake of simplicity. The results of the reverse (the in-party on out-party marginal effect) are substantively the same and result in essentially identical interpretation with no analytic gain.

in-party affect for Republicans and a (significant) .037 point decrease for Democrats. And we can see how these associations have developed over time, such that, in 2016, out-party affect had *no* statistically significant relationship with in-party affect for Republicans, but it did for Democrats, with each point of change in out-party affect being associated with a .084 change in-party affect. In short, for a Democrat in 2016 in-party and out-party affect were reciprocally related in a way they simply were not for Republicans.

This may seem like a minor point, given the size of the effects, but these effects are larger than they look. In terms of a substantive comparison, if we imagine a Democrat and a Republican in 1978 and in 2016, and then look at how big of a difference going from fully loving the other party to fully hating the other party would make on in-party affect, we see the following: in 1978, a Democrat who went from loving the other party to hating it would *like their own party less* by about 3.7 points. On the other hand, the same change in 2016 would result in an over 8 point *increase* in in-party affect. In other words, over the time period considered, the potential impact of out-party affect changed for Democrats by about 12 points, or more than a tenth of the scale. For Republicans, on the other hand, there was never a statistically significant relationship between out-party affect and in-party affect. Indeed, even assuming the point estimates are accurate, we see impacts of less than two points, even across the whole scale of variation.

That said, the point is not that these are huge effects; indeed, making an argument about magnitude based on the maximum possible effect size would be questionable. Instead, the point is rather that these effects are present and meaningful in magnitude, and that they cannot and should not be dismissed as too small to matter. Fundamentally, the simple fact that we see *any* meaningful relationship between in-party love and out-party hate, that it develops over time, and that we see it for only Democrats, is the finding.

2.5 Conclusion

The results presented here strongly indicate asymmetric polarization in favor of *Democrats*, based almost entirely in the asymmetric development of in-party affect. Furthermore, they also show the development of a reciprocal relationship between in- and out-party affect for Democrats and not for Republicans. This is both empirically and theoretically surprising: empirically, this was an attempt to confirm that Republicans are more polarized than Democrats at the mass level, in line not only with what we know about elite polarization, but also with a variety of circumstantial and qualitative evidence. And theoretically, this investigation was motivated by the ongoing debate about the role of in-party affect in affective polarization: on the one hand, net affect approaches see in-party affect as playing a constitutive role, while negative partisanship discounts in-party affect and sees out-party animus as definitive. In the context of expected asymmetric polarization towards Republicans, these two theories make clear predictions about what that asymmetry should look like and what role in-party affect should play in it. Thus, looking at mass asymmetry was intended as a kind of theory contest between net affect and negative partisanship understandings of affective polarization based on the strength and clarity of our empirical expectations.

Obviously, these findings rather put paid to that neat idea. Not only do they contradict our clear empirical expectations, but on a theoretical level, they are akin to a bull in a china

shop: unexpected and destructive. At a basic level, finding this kind of meaningful variation on in-party, but not out-party, affect contradicts one of the basic ideas behind negative partisanship theories. But net affect approaches also lack a convincing explanation. It is true that social identity theory, on which net affect approaches are based, generally gives theoretical primacy to in-group affect and doesn't necessarily find this kind of asymmetry problematic as a rule (see, e.g., Brewer 1999); however, political polarization is one of the cases where we would expect a reciprocal relationship between in-party and out-party affect, based on both social identity theory itself and previous work on net affect approaches (Brewer 1999; Mason 2015). So this is a case where, theoretically, we precisely would not expect an unmooring of in- and out-party affect, but this is exactly what we see for Republicans.

Now, we do see a reciprocal relationship between in- and out-party affect develop among Democrats, but this actually raises more questions than it answers. To begin with, this link develops without any asymmetry on out-party affect. So, while Republicans and Democrats both like each other less and less, to roughly equivalent degrees, it is only for Democrats that this dislike is linked with how they feel about their own party. Even given that this were theoretically sensible, it is still wildly out of line with our empirical expectations. And this issue actually gets worse: if net affect approaches are right, then this kind of link between positive and negative affect should be present for *everyone who's polarized*. So we could conclude from this that only Democrats are polarized, and whatever is happening with Republicans isn't really polarization. Indeed, there are arguments from social identity theory that the kind of purely negative affect driven change seen among Republicans is not, properly, a case of group-based behavior at all (Brewer 1999). However, while this would perhaps be a theoretically tenable position from a social identity theory perspective, it isn't really plausible by the lights of any previous work on affective polarization. All that said, the reciprocal relationship for Democrats does support the net affect idea that in-party affect plays a constitutive role in polarization. It even undermines one of the basic motivations for negative partisanship approaches in the first place, namely that negative affect is more substantively connected to polarization. This may still be true, but we can also say that, at least for Democrats, any connection out-party affect has to polarization, in-party affect *also* has. It may be weaker, but it is *there*.

The impact of these findings is exacerbated by the role played by partisan identification. Both negative partisanship and net affect approaches point to partisan identification as *the* fundamental object of concern, and so the fact that this unexpected asymmetry is linked directly to partisan identification as such, net of other factors, is doubly troubling.

These anomalous findings are, I think, tremendously consequential, especially for net affect and negative partisanship approaches. But before elaborating those consequences, I want to clarify that the point here is not to suggest we throw out net affect and negative partisanship and their contributions, far from it. While, as I have and will argue, the findings above should prompt theoretical innovation, they do not erase the basic insights won by previous work on affective polarization. Indeed, this paper is admittedly limited in important ways. Most obviously, the methods here are incapable of causal identification, and so it's possible that the partisan asymmetry found here is grounded in some other factor. Beyond that, I have not looked directly at any "polarization related" outcomes, e.g. willingness to discriminate against the out-group, assessments of governmental legitimacy, etc. This is one of the areas where negative partisanship approaches have made some of their strongest claims,

and it is an area I cannot address directly. Finally, while the feeling thermometer measures of affect that I use are standard in the literature, and have been found to generally correspond well to other ways of measuring affective polarization, they do have limitations (Druckman and Levendusky 2019; Mason 2018).

But these findings are important despite their limitations, and we should not give in to an easy temptation to minimize them. In particular, it may seem that the magnitude of the asymmetry doesn't justify such strong conclusions: surely a 5.5 point asymmetry on in-party affect isn't really meaningful in comparison to, for instance, the 22 point drop in average out-party affect from 1978 to 2016. And, as said, I certainly am not arguing that the changes in out-party affect aren't important, or even central. However, that is also not the point: given that we were expecting either no asymmetry on in-party affect or no asymmetry without a matching asymmetry on out-party affect, the question isn't "which of these is larger", but rather "is this asymmetry substantial". In other words, the proper point of comparison isn't 22, it's 0. More basically, one could argue that neither this asymmetry nor the reciprocal relationship between in- and out-party affect for Democrats are actually related to polarization, that in-party affect is (following negative partisanship) irrelevant, and that whatever is going on here is something else entirely. But that wouldn't change how empirically unexpected these findings were in the first place, nor would it be terribly convincing in the face of the argument above. These findings demand explanation; they require us to somehow reevaluate our understanding of affective polarization.

To do this, we must start by bringing institutional and historical context into the core of our theories: the issues raised by my findings make apparent the basically "institutionally 'thin'" nature of our quantitative theories of affective polarization (Pierson and Schickler 2020:52). Pierson and Schickler argue that we need a "developmental analysis" of polarization, which considers polarization as a historical process enmeshed in and responding to various institutional configurations (2020:38). This is something which net affect and negative partisanship approaches, in their basic forms, have trouble doing, as they explicitly ignore contextual factors like the content of different identities, their social positions, etc. in order to focus solely on the fact of identification itself (for an elaboration of this critique of social identity theory specifically, see Huddy 2001; for examples of this issue in the literature, see, e.g., Mason 2015). But, as our discussion of elite and organizational asymmetry above makes clear, the content of identities matters, their unique institutional ecosystems matter. And it is not simply that attention to context should inform our expectations, but rather that context should play a role in defining what we call polarization in the first place. This is clear when we consider, for a moment, the use of feeling thermometers: the in- and out-group affect they measure is understood to be definitive of polarization when looking at Democrats and Republicans, but not with any other identity pair. Up to now, we've been silently relying upon assumptions about the current context to justify this (along with, obviously, some unique facts about partisan identities). The unexpected results here don't directly "prove" that we cannot use feeling thermometers in this way, but they should show us that we ought to consider whether or previous assumptions are reliable. Put another way, one of the reasons these findings are difficult for previous theories to assimilate is because we lack a rubric for establishing what they actually mean.

In fact, the issue is deeper than that of historical or institutional context: that the asymmetry found here not only doesn't match our expectations, but almost perfectly contradicts

them in a way which cannot really be fit into either of the dominant theories, demonstrates that we need to reconsider what exactly affect actually means. Generally speaking, affect has been taken to be simply representative of bias, either in-group favoritism or out-group prejudice, and therefore of (affective) polarization. This is not an unreasonable assumption, grounded as it is in social identity theory (see, e.g., Brewer 1999; Tajfel and Turner 1979), and it has a degree of support in previous work looking at the impact of affect on biased outcomes (see, e.g., Iyengar and Westwood 2015). Indeed, I suspect that in many cases such an assumption may well be justified. However, in neither social identity theory, nor the various theorizations of negative partisanship, is affect per se taken to be the fundamental issue. Rather, the fundamental issue in both cases is *identity*, with affect merely being a way to conceptualize and measure the degree of bias associated with that identity. Indeed, the operative question for social identity theory is really about how group members are able to establish their identity's positive distinctiveness vis-a-vis relevant out-groups (Brewer 1999; Huddy 2001; Tajfel and Turner 1979). But, as just discussed, the specifics of different identities matter, and so we need to consider the possibility that affect (either positive or negative) may function differently for the different partisan identities or in different situations. For instance, the fact that in- and out-party affect are connected for Democrats but not for Republicans could be because only Democrats are really polarized. But it could also be because in-party affect means something different for Democrats than it does for Republicans. In terms of social identity theory, perhaps the generalized sense of “warm” affect measured by a feeling thermometer captures a key part of the “positive distinctiveness” of Democratic identity, but has no such meaning for Republicans.

If affect doesn't have a perfectly fixed meaning, this then serves to make the specifics of institutional and historical context all the more important. A developmental analysis of polarization is necessary not only because we need to better understand the assumptions which undergird our measures, as I suggest above, or to understand the self-regulating versus self-reinforcing developmental contexts, as Pierson and Schickler (2020) originally argue, but because our measures may not have a fixed meaning at all. It is precisely through careful attendance to context, to the particular meaning and content of different identities, their unique institutional environments, and their specific histories, that we can make empirically and theoretically well grounded arguments about how to measure polarization. The point is not that we should throw out previous work, but rather that the surprising findings presented above complicate the fundamental claim of a simple and straightforward relationship between affect and polarization.

To try and make concrete just how and why this matters, I want to close by suggesting a possible explanation for the anomalous findings presented here, one grounded in the theoretical arguments above. Taking as our starting point the basic idea that we can no longer rely upon a direct correspondence between affect and polarization, we can explore a previously understudied theoretical prediction from within the literature on polarization, namely that polarization should lead to increasing *intra*-party, rather than only inter-party, conflict (McCoy and Somer 2019:246–47). This conflict should arise partly for strategic reasons, but also from a growing commitment to group uniformity in the context of a political struggle which feels increasingly existential (Somer and McCoy 2019:14). Now, because the meaning of affect is not fixed, neither over time nor by party, we can hypothesize that we should see this intra-party conflict manifested on in-party affect. Thus, rather than in-party

affect acting as a measure of positive in-group image, I want to suggest that it may act partly as a metric of *dissatisfaction*. That is, under conditions of increasing polarization, in-group affect may convey less about the comparative evaluation of the in-group versus the out-group than it does a disappointment with the actual existing in-group (as distinct from the prototypical ideal of the in-group, for more on this see Huddy 2001).

In theory, the basic story would look something like this: as we've discussed, Republican elites are more polarized than Democrats and are indeed acting as a polarizing actor; in doing this, they have managed to convince their base of the seriousness of the the political struggle, that Democrats are a threat, potentially an existential one. Of course, Republican elites have not managed to defeat that threat, however, and as a result their base is *dissatisfied* with what the party is doing to address it, a dissatisfaction which manifests in the asymmetry on in-party affect I have shown above. Indeed, this would also explain the lack of any link between in-party love and out-party hate for Republicans: if this dissatisfaction hypothesis is correct, then we naturally wouldn't expect the partisans who are happiest with their own party to be those who most dislike the Democrats. Now, to be clear, this explanation is purely hypothetical at this point, and it would still leave some questions unanswered, e.g. why don't Democrats suffer from dissatisfaction in the same way. But it would also suggest plausible and testable answers: it may be that the behavior of political elites is critical, or that Democrats are actually *less* polarized. The point here is not to provide a definitive explanation, but rather to show how the theoretical consequences of the findings above can help show us a way forward. Whether dissatisfaction is actually driving things, it is a promising hypothesis with clear implications and shows how rethinking polarization in light of the failed theory contest presented here actually opens up exciting new possibilities.

Chapter 3

Strangers in their Own Party: Dissatisfied Partisans and the Worldview Model of Polarization

3.1 Introduction

On January 6th, 2021, hundreds of Trump supporters stormed the Capitol in an attempt to overturn President Biden’s victory in the November 2020 presidential election (Tan et al. 2021). At the same time, 61 Congressional Republicans attempted to do the same as they objected to the certifying of Arizona’s electoral college results (Tan et al. 2021; Wagner et al. 2021). This kind of assault, both physical and institutional, on the federal electoral process is unprecedented in modern American politics. And yet it was not only true that, as one rioter said, “our president wants us here”, but that much of their party did too (Barry, McIntire, and Rosenberg 2021). Indeed, this was only the natural outgrowth of Republican attacks on the electoral process dating from long before the election itself and continuing to this day (Epstein and Lerer 2021).

This larger assault is, itself, just one symptom of the increasing polarization of American politics: as polarization ramps up, it sharpens political contention and poisons the political process, ultimately threatening the very contingent consent upon which democracy relies (Hacker and Pierson 2005; McCoy et al. 2018). But this danger is fundamentally one-sided: though both parties have been impacted by growing polarization, there is simply no Democratic equivalent to any part of the story of January 6th. The reasons for this are, in some ways, very clear: asymmetry is one of the constitutive facts of polarization at the elite level. Not only is the Republican Party measurably more extreme than the Democratic Party, it is also clearly *responsible* for polarization in a way the Democratic Party largely is not: in the development of polarization in the US, it is the Republican Party that has played the role of the polarizing actor (Bermeo 2019; Grossmann and Hopkins 2015; Hacker and Pierson 2005; Mann and Ornstein 2016; McCoy et al. 2018; Pierson and Schickler 2020; Sinclair 2006). Likewise, the asymmetry is clear qualitatively at a more popular level, in the unique role of Republican popular organizations in pushing polarizing narratives, the repeated turnover in party leadership due to popular insurgency (see, e.g., Eric Cantor, John Boehner, and Paul

Ryan), and the continued increases in Republican right-wing extremism (Hacker and Pierson 2020; Skocpol 2020). This makes it all the stranger, then, that quantitative work on mass polarization is essentially *ambivalent* on the question of asymmetry; indeed, the common approaches neither capture nor explain it.

In this paper, I argue that this ambivalence stems from the fact that this asymmetry is not merely a difference of *degree*, but a difference in *kind*. Previous approaches are powerful but, due to the assumptions on which they rely, they have been limited in their ability to grasp an asymmetry of this kind. In response, I propose a new theory of polarization, one which builds on recent insights from sociological approaches to polarization grounded in the relational nature of meaning (Boutyline and Vaisey 2017; DellaPosta 2020) with the concept of affective polarization from political science (Iyengar et al. 2012; Mason 2015) and defines polarization in terms of the persistent patterned relationships between affect, attitudes, and judgments, bound to particular political identities. In essence, polarization is a population level consequence of shared evaluative schemas (cf. Brooks and Manza 2013). This is the *worldview model* of polarization. If polarization is fundamentally about the increasing division of society into meaningful political “sides” (usually two),¹ an “us” and a “them” between which compromise and understanding is increasingly difficult (DellaPosta 2020; Finkel et al. 2020; Iyengar et al. 2012; Lelkes 2016; McCoy et al. 2018), then the worldview model focuses on assessing how those sides view one another and the political or social struggles they’re involved in. Looking at relationships in this way not only fits well with polarization on a conceptual level, it also provides the analytical leverage necessary to more directly operationalize specific insights from conceptual (or qualitative) definitions of polarization.

I then demonstrate the worldview model by using data from the American National Election Study to track the relationship between partisan identity, affect, threat, and fear and thereby assess the development of mass polarization (and asymmetry) over time. In essence, I look at the degree to which each party’s partisans 1) see the other side as a threat, 2) fear the other side, 3) hate the other side, and whether those things are related. Previous work on polarization has posited these sets of related judgments as *definitional* of a polarized identity. And so it is these relationships that are key: polarization consists not merely in disliking the other side, but in that dislike coming coupled with further evaluations and judgments, in other words, with a worldview. Thus, the more perceptions of threat, fear, and hate/dislike hold together, the more evidence for a polarized worldview in aggregate. Asymmetry, then, will then appear in the degree to which these things are more or less strongly related for Democrats vs. Republicans. In the same way, I also examine how threat and fear relate to how much partisans *like* their own party, an area where predictions from previous work are much less clear. In this way, I both *validate* this new approach against a clear set of expectations, and *demonstrate* the additional analytical leverage it provides by intervening in an ongoing definitional question.

Ultimately, I *do* find the expected asymmetry in polarization, with Republicans developing a far stronger and more consistent connection between feelings of threat, fear, and dislike of the out-party. However, I find nearly the *opposite* with regard to in-party affect, where it is *Democrats* who see these relationships growing stronger, faster. This surprising finding

¹However, see, e.g., Lauka, McCoy, and Firat (2018) or, for a related point, Fischer and Hout (2006) and Fischer and Mattson (2009) on “fragmentation”.

suggests a potentially critical explanation: that Republicans are *dissatisfied* with their party, judging it for not taking the Democratic threat seriously enough. This would explain the overall differences in trends between the parties and help explain *why* there has been no Democratic equivalent to the January 6th insurrection, or the institutional assaults on the federal election process which preceded and followed it: this dissatisfaction may be part of a *feedback loop of radicalization*, in which Republican elites so effectively present the Democrats as a true and dire threat that their base is dissatisfied with the party’s response and demands yet stronger statements, yet stronger actions, statements and actions which then further *demonstrate* the severity of the threat. Though it is outside the scope of the paper to fully validate it, this theory of dissatisfied partisans caught in a feedback loop *does* fit with both specific theoretical predictions about polarization (McCoy and Somer 2019), and the empirically visible pattern of increasing radicalism and popular insurgency which seem to have possessed the Republican Party. Both the worldview model of polarization and the empirical findings presented here, then, are significant contributions to the study of polarization specifically, and the study of popular political identity more generally.

3.2 Context and Theory

3.2.1 What’s Polarization and What’s Wrong With It?

In order to understand why I’m proposing a new theory of polarization, we first need to understand why previous quantitative theories can’t really capture asymmetry. Broadly speaking, previous quantitative approaches fall into one of two camps: *attitudinal polarization* understands polarization as basically connected to political attitudes and thus defines it in terms of characteristics of political attitude distributions (e.g. bimodality, variance, correlations across opinions) to define division (Abramowitz 2010; Baldassarri and Gelman 2008; DiMaggio et al. 1996; Fiorina and Abrams 2009). On the other hand, *affective polarization* see polarization as social and defines it in terms of affective in/out-group bias (i.e. how the sides feel about each other) (Iyengar et al. 2019, 2012; Mason 2015). And each of these approaches has its own, sound, theoretical justifications linking their own phenomenon of interest and particular measures to the broader concept of polarization: attitudinal approaches draw on theories of pluralism and cleavage, while affective approaches build on social identity theory.

These approaches have had significant success, and each has substantial upsides. Early results from attitudinal work were admittedly mixed, leading to a long running debate about the *existence* of polarization at all (see, e.g., Abramowitz and Saunders 2008; Baldassarri and Gelman 2008; DiMaggio et al. 1996; Fiorina et al. 2011; Fischer and Mattson 2009). However, more recent work that focused on *constraint*, or the degree to which different opinion items are correlated with (“constrain”) each other, has found more consistent results (DellaPosta 2020; Kozlowski and Murphy 2021). These constraint-based approaches are also fairly direct ways of looking at the degree of aligning vs. cross-cutting opinion cleavages, a central question in theories of pluralism and political conflict (Baldassarri and Gelman 2008; Blau 1974; Dahl 1961; Lipset and Rokkan 1967). On the other hand, affective approaches have shown clear and consistent results from the beginning, even as debate continued around attitudinal measures (Abramowitz and Webster 2016; Iyengar et al. 2019, 2012; Mason 2015). Furthermore, looking

at division in terms of the affective bias between social identities is intuitively satisfying and valid on its face: attitudinal agreement or disagreement make little difference if two groups hate each other. However, despite these upsides, both theories have had trouble adequately capturing the asymmetry we see.

This is most obvious with attitudinal theories, simply because so much of what is concerning in asymmetric polarization is fundamentally non-attitudinal. Now, to be clear, there are absolutely attitudinal aspects to asymmetry. Indeed, this is one of the more obvious divisions at the elite level, where party platforms and congressional voting records make it clear that Republicans have moved right far faster than Democrats have moved left (Jordan et al. 2014; Lewis et al. 2020; Pierson and Schickler 2020). However, much of what is most worrying is not attitudinal in nature. The argument that the Republican Party should be identified as a polarizing actor, for instance, is not grounded in the extremity of policy positions, but rather in the way the party has pursued (at times) strategies of deliberate intransigence (Mann and Ornstein 2016) and, increasingly, driven the use of “constitutional hardball” (the strategic choice to violate established norms to gain a partisan advantage) (Hacker and Pierson 2014; Tushnet 2004), or in the unique organizational ecosystem which surrounds it, where organizations act as “surrogates” for the party in pushing polarizing narratives (Hacker and Pierson 2020; Hochschild 2016; Polletta and Callahan 2017). Likewise, while there is certainly an attitudinal element in opinions about validity of the 2020 election, what is concerning is less an attitudinal disagreement about electoral policy, and more that Trump’s claims of a stolen election, and the support those claims received from across the party, led to the January 6th insurrection and “sparked a campaign to terrorize election officials nationwide” (So 2021). Election workers at all levels have been intimidated and harassed physically, verbally, via text message, phone call, and email (So 2021; Wines 2020) and a survey in spring 2021 found that “one in three election officials feel unsafe because of their job, and nearly one in five listed threats to their lives as a job-related concern” (Brennan Center for Justice and Bipartisan Policy Center 2021:4). Now, attitudinal theories of polarization would counter that the point isn’t that all consequences of polarization are attitudinal, just that their causes basically are. But this would be more convincing if attitudinal measures of polarization were showing a clear of asymmetry of causes matching these asymmetric consequences. In fact, there’s at least some evidence from the study of social movements that the link between attitudes and action is a weak one at best, and so we precisely shouldn’t expect that kind of connection between attitudes and consequence (Munson 2008). So, while attitudes likely play a role, it seems implausible that they are the whole story. Instead, our focus should perhaps be on behavior, or a broader approach to politics.

At first glance, affective theories appear to do a much better job of this, but they ultimately run into similar problems. Shifting to understanding polarization as a matter of social identities, rather than ideological positions, is almost explicitly aimed at solving the basic issue with attitudinal approaches discussed above. To put it somewhat idiomatically, attitudinal agreement doesn’t matter if people hate each other’s guts. And, indeed, this basic idea has largely been borne out by the empirical results: defining polarization in terms in and out-group bias has not only found clear evidence of increasing polarization over time, but strong links between affective polarization and the kinds of behaviors polarization tells us to worry about (Iyengar et al. 2019). So-called “social distance” measures of affective polarization push this even further, by providing quasi-behavioral measures (e.g. how willing

would you be to marry someone from the opposing group) and further strengthening the already powerful face-validity affective polarization (deservedly) enjoys (Mason 2018).

However, while this is in some ways better, it still doesn't solve the problem. Though we have stronger face validity (and, indeed, attested, measured validity) for the connection between affective polarization and biased behavior, the measures we use depend on some deep assumptions which go substantially beyond their explicit theoretical claims. Indeed, the basic idea that in and out-group bias per se are definitive of polarization is problematic when placed in comparative context: in/out-group bias is not sufficient for polarization with any other identity pair. Only in the particular context of partisan identities are we calling them measures of polarization, and it is only in this particular context that they seem valid as such. For instance, if we were to take feeling thermometer bias as a definitive measure of polarization, then the US was "polarized" around gay/straight conflict in the 80s and 90s, when feeling thermometer scores towards gays were around the level out-party thermometer scores are now (Hetherington 2009:435–36). There is perhaps an argument to be made that this was the case, but most definitions of polarization clearly exclude this possibility. In some cases this is based on *a priori* assumptions that polarization is necessarily about *political* identities,² but in others it is for more coherent theoretical or empirical reasons, e.g. the totalizing nature of polarized partisan conflict or the connection between polarization and the construction of truth, etc. (Finkel et al. 2020; McCoy and Somer 2019; Somer and McCoy 2019). Ultimately, the point is that this comparison makes clear that a simple measure of in/out-group bias cannot *alone* define polarization. As with attitudinal measures, while there is certainly a connection between affective measures and the manifestations of polarization we care about, it is limited and incomplete.

But really what this is all showing us, in different ways, is that affective and attitudinal theories share the same problem: their conceptions of polarization are well specified, but limited, and as a result they can really only conceive of asymmetries of *degree* and can't deal with asymmetries in *kind*. This is a problem because, as should likely be apparent based on the examples discussed so far, we know that asymmetries in kind are foundational to polarization in the US. That the Republican Party can be fairly identified as a polarizing actor, surrounded by a unique organizational ecosystem, and driven partly by continued increases in right-wing extremism and popular insurgencies within the party, while the Democratic Party cannot, is not a difference of degree along a scale but a difference in kind. And going deeper, we know that the parties are different in nature, with Grossmann and Hopkins (2015) arguing that, while the Democratic Party is best understood as a pragmatic coalition of social groups, the Republican Party is fundamentally ideologically based. And this ideologically based party has become the party of radical right-wing extremism, including both ethno-nationalist resentment and extreme anti-government sentiment (Mann and Ornstein 2016; Skocpol 2020). This ethno-nationalist resentment, especially, has been fundamentally *popular* in nature, and was the source of the Tea Party: a major popular insurgency in the party which overturned leadership both locally and nationally (Skocpol 2020:18). Now, to be clear, it's not that a theory of polarization needs to capture every individual aspect of this wide array of differences

²Of course, the question of what exactly makes an identity "political" is very much in play; however, as alluded to above, these definitions almost universally mean *partisan* identities (Finkel et al. 2020; Hetherington 2009; Iyengar et al. 2012; Mason 2015; McCoy and Somer 2019).

directly, especially as some of them are only contingently related to what we would normally recognize as polarization. Rather, a theory of polarization must be able to capture the kinds of impacts these differences might have. Especially given the constitutive role of this asymmetry in American politics, it is difficult to imagine that it wouldn't have impacts. But it is certainly possible that those impacts might not show up in ways that attitudinal or affective theories can grasp.

Indeed, attitudinal and affective theories can generally only capture the impact of these differences insofar as they affect a relatively limited selection of attitudinal and affective survey questions. And we can easily imagine ways that, e.g. January 6th, might be caused by asymmetric polarization even if we see no asymmetries on feeling thermometers or political attitudes. Put another way, I don't think that we can assume that the harassment of election officials would necessarily show up as lower out-party affect, as measured by survey questions. Part of the issue here is a basic problem underlying survey questions generally: there are limits to the degree to which we can take survey responses as meaningfully true and accurate representations of respondents' actual ways of being in the world (Abbott 1988; de Leon, Desai, and Tuğal 2015; Vaisey 2009). The same attitudinal question may be answered in the same way by two respondents despite vast differences in how much the two are concerned with or care about the issue in general or their position specifically, and a respondent might have rated the Democratic Party a zero on a 100 point feeling thermometer for decades before they disliked them enough to attend a "Stop the Steal" rally. The point here isn't that this makes survey data unusable (though some would argue that point, see, e.g., de Leon et al. 2015,), but that these are uncertainties which contribute to the difficulty attitudinal and affective theories have in getting at asymmetries in kind.

This issue of asymmetry in kind rather than degree is, however, only a symptom of a more fundamental problem, which is at the heart of all of the specific issues I've discussed so far: disagreement is not inherently meaningful; it must be made so. Part of the fundamental difficulty of survey questions discussed above is that meaning is necessarily *relational*. So, for instance, a respondent's opinion about government spending is meaningful, both with regard to the literal meaning and in terms of what we ought to make of it, only by way of its relation to other beliefs. Indeed, this basic fact is at the heart of constraint based definitions of attitudinal polarization, which take those relations as the object of interest (Boutyline and Vaisey 2017; DellaPosta 2020). But it is not only true of semantic meaning, but also more broadly of affects, judgments, etc. (Hennion 2007; Wimmer 2013). And it is perhaps especially true for the way all of these things relate to politics. This is what it means to say that disagreement is not inherently meaningful, but rather is made so. This is the essential point raised by work on political articulation: no division has political meaning by nature or necessity (de Leon et al. 2015). Rather, it is precisely a political project to take some line of division, distinction, or disagreement, and not only make it politically salient, but make it politically salient in particular ways which require particular responses.

We can see this quite clearly in practice if we consider the development of politics and public opinion around abortion since *Roe v. Wade*. Disagreements over abortion mean something very different now than they did 50 years ago, and have meant different things over that span. Focusing on the pro-life side of things, when *Roe v. Wade* was decided, being anti-abortion wasn't a partisan position, and the clearest religious valence it had was *Catholic*. But by 1980, the growing Christian (Evangelical) Right was increasingly committed, and

the Republican Party included a pro-life position in their national party platform (Munson 2008:85–87). As time passed, early activism aimed at an expected quick reversal of *Roe* became more radical, as participants adopted more militant tactics (street protests, clinic blockades) from the mid 1980s, and engaged in violent, terrorist attacks through the 1990s. That level of violence has largely disappeared from the current abortion debate, but other forms of disruption (i.e. picketing clinics) have continued and increased (Munson 2008:88–90). This has all taken place along with one of the few examples of true opinion polarization visible in the 90s and early 2000s (even as partisan affective polarization was far lower than it is now) (DiMaggio et al. 1996; Evans 2003). And yet, overall public opinion on abortion has actually remained remarkably *stable* for most of the period (Munson 2008:90). So over the past half century, being anti-abortion would put one in a very different relation to party politics, religious commitments, tactical questions, and even political violence, while the literal “political attitude” as measured by a survey question remained the same. While being pro-life in the 90s didn’t entail approving of terrorism, it did require thinking about one’s relationship to that violence in a way which an *identical* position today does not, because of the ways in which the sides of this conflict have evolved, changed, and developed to be meaningful or not at different ways in different times (see Abbott 1988 for a general formulation of a similar point). These changes were part of an explicit political project by which abortion was made a partisan issue and affectively charged in new ways

3.2.2 The Worldview Model

My response to these challenges is a new theory of polarization which brings together the strengths of attitudinal and affective theories with insights from the study of culture and a grounding in the basic conceptual definition of polarization. As said above, polarization fundamentally refers to the increasing division of society into *meaningful political sides* between which political conflict becomes increasingly intense, dangerous, and unavoidable. There are two parts to this basic conceptualization: on the one hand we have the idea of division, and on the other we have the question of what that division *means*. Put another way, the cognitive schema which makes the categorical division into “us” and “them” is logically and practically separate from the set of schemas which make up what it means to be one of “us” and determine how “us” and “them” are supposed to relate (i.e., the schemas that make the distinction *matter*) (Wimmer 2013:9).³ So in the case of the US, the first question is about the magnitude of division between Republicans and Democrats:⁴ the strength of boundary markers, the clarity of the categories, etc. The question of “meaning” is about whether the partisan divide is in fact linked to other schemas. The partisan identities are the sides, but it is these other schemas which may make the sides *meaningful*. Insofar as polarization matters, to distrust of elections, to radicalization, to fear of the other side, to January 6th, it is because of these schemas. Without them, what appears to be polarization is really just a distinction without a difference.

³To put this in Bourdieusian terms: we can easily imagine that two groups with a similar class position might be carefully delineated, and yet be relatively similar in *habitus* (Bourdieu 1984).

⁴Polarization does not *need* to be defined in terms of the parties (see, e.g., DiMaggio et al. 1996; Fischer and Hout 2006; Stavrakakis 2018), but as a practical and empirical matter, the vast majority of work on (and results about) polarization in the US takes political parties as the division of interest.

Therefore, I argue that polarization is best understood as a set of evaluative schemas, linked with a political identity. This is the *worldview model* of polarization: a polarized society is one in which people have to see politics and each other in a certain way (namely a particularly antagonistic way). Seen in this way, assessing polarization means asking whether there are distinct sides, how people feel about the other side (and their own), and what people think that feeling *means*. Affective and attitudinal theories of polarization get at this in important ways, but only partially (as seen above). The worldview model follows affective approaches by making polarization a matter of identity and how identity shapes political conflict. But I go beyond them by building on insights from recent constraint based attitudinal theories (Boutyline and Vaisey 2017; DellaPosta 2020) and taking the relationships between survey items as the key question of interest, recognizing the problems of meaning discussed above. Thus, in this view, polarization requires not only out-party dislike (for instance), but for that dislike to hold together in a persistent, patterned, way with other judgments about the other (and/or one’s own) party.

Put another way, polarization is a particular “orientation towards politics” and it is worth laying out how this distinguishes the argument I am making here from previous work on asymmetric partisan thinking. An extensive literature in social psychology has looked at whether there are differences in how liberals and conservatives (and more recently, Democrats and Republicans) think. Although there is ongoing debate on some specifics, including the degree to which conservatives and liberals may differ in their vulnerability to motivated cognition Garrett and Bond (2021), there is nonetheless substantial evidence of significant “elective affinities” between psychological predispositions and needs (e.g. cognitive rigidity, the need for closure or the management of threat, etc.) and conservative ideology (Hibbing, Smith, and Alford 2014; Jost et al. 2007; Jost 2017).⁵ The discussion of partisan differences and asymmetries of kind, above, owes much to this literature, as does my basic contention that polarization is a particular way of thinking about politics; however, my theory of polarization ultimately runs orthogonal to it. Despite the importance of partisan differences in helping to establish our expectations of asymmetry and the idea of asymmetry in kind, the mere existence of divergent ways of thinking among partisans is not, itself, sufficient for (or necessarily particularly relevant to) polarization. Polarization is a specific concept, which I am defining in terms of a particular worldview, namely one which entails a particular antagonistic orientation towards politics (cf., e.g., Stavrakakis 2018). And, crucially, there is no *necessary* relationship between the kind of asymmetries previously investigated and such an antagonistic orientation. For instance, the fact that liberals may exhibit a weaker negativity bias and conservatives may be more sensitive to perceptions of threat (Hibbing et al. 2014) cannot alone tell us anything about who is more polarized, because the way in which such general differences manifest is highly contingent. This, then, clarifies my focus on relationships: a propensity to feel threatened only matters in conjunction with a set of schemas which determine how that threat is responded to.

In addition to the argument above, hopefully the face validity case for this is fairly clear: returning to the abortion example above, there is an intuitive difference between believing

⁵Though there is certainly complexity here as well, with cognitive rigidity being associated with right-wing ideologies specifically, but also with ideological extremity in general, regardless of the side (Zmigrod 2020; Zmigrod, Rentfrow, and Robbins 2020).

abortion is wrong, believing abortion is murder, believing abortion is murder which pro-choice Democrats are functionally aiding and abetting, and believing that abortion is murder which Democrats are aiding and abetting and that if Democrats win the White House again they will usher in the death of the republic. The basic attitude held is the same in each example, but the *meaning* of the attitude (or affect) varies wildly different based on the context of the *worldview* in which it is held.

Attempting to get at this in a comprehensive way would be extraordinarily difficult, and also potentially cause problems for the kind of over-time analysis of asymmetry which is my empirical goal here. Thus, what I will do here is lay out a minimal set of affects and judgments which make up a polarized worldview. Fortunately, there are three things in particular which both previous work on asymmetry and conceptual work on polarization highlight as being core to a polarized worldview: affect, threat, and fear. While affective approaches to polarization have their issues (discussed in detail above), the basic insight that affective bias is key to polarization still holds (and is supported by other approaches to polarization, see, e.g., Finkel et al. 2020; McCoy et al. 2018; Somer and McCoy 2019). Indeed, affective polarization, as previously defined, is an absolutely a necessary element of a polarized worldview; it's just not sufficient on its own. Rather, we must consider whether it meaningfully holds together with other political judgments.

Seeing the other party as a threat, on the one hand, and being afraid of it, on the other, are the judgments I focus on here.⁶ Both threat and fear of the “them” (in the polarizing “us” vs. “them”) are integral parts of a polarized worldview (Iyengar and Westwood 2015; Lu and Lee 2019; McCoy et al. 2018). They are crucial to the increasingly moralized and vicious nature of political conflict in polarized conditions. As the other side is seen as more threatening, and fear of the consequences of their potential victory increases, individual moments of conflict become linked in a larger and increasingly Manichean narrative of zero-sum, life and death struggle (Finkel et al. 2020; McCoy et al. 2018; Ward and Tavits 2019). Threat and fear also play a central role in social identity theory, which understands threats to the in-group as particularly emotive and motivating (Brewer 1999; Mackie, Devos, and Smith 2000; Mason 2015). And because commitment to an identity changes how one makes evaluations, as an identity grows stronger, threats become easier and easier to perceive, especially from the out-group (Banda and Cluverius 2018; Huddy 2001; McCoy and Somer 2019). Threat can also itself motivate identification (Gorman and Seguin 2018; Tajfel and Turner 1979). Finally, threat can provoke reactions of fear, which can then motivate (and sometimes suppress) political activity (Almeida 2018; Azab and Santoro 2017; Jasper 2011; Lu and Lee 2019; Mason 2015).

Threat and fear are also key to what we know about the asymmetric polarization between Democratic and Republican elites, and the visibly asymmetric actions taken by the Republican Party. To begin with, they are clearly visible in the unique organizational ecosystem highlighted by Hacker and Pierson which operates to stoke outrage and, in the case of the Christian Right, feelings of existential threat (2020). This outrage has built into a “deep story” of “us vs. them” in which conservatives are positioned as *victims* of liberal policy priorities

⁶Threat and fear can be difficult to differentiate, but here I follow the literature on emotion in social movements and understand the difference in this way: threat is an evaluation of some external situation (Goldstone and Tilly 2001; Goodwin 2011), while fear is an internal affective response (Jasper 2011). Fear may be felt in response to threat, but other emotions may be as well (Jasper 2011; Mora et al. 2018).

(Hochschild 2016; Polletta and Callahan 2017). And this story is ultimately linked to status threat, i.e. a dominant group’s fear of the loss of that dominant status and attendant privileges to an undeserving and dangerous other, which again matches predictions from theories of polarization (McCoy and Somer 2019:241). Status threat plays a clear role in Republican politics generally, but was especially visible (and effective) in first the success of the Tea Party, then Trump’s mobilization of xenophobia and racial resentment in the 2016 election and thereafter (Abramowitz and McCoy 2019; Bermeo 2019; Mason, Wronski, and Kane 2021; Mutz 2018; Willer, Feinberg, and Wetts 2016). Finally, we can turn again to the harassment of election officials, the events of January 6th, and broader Republican contestation around the election, which can clearly be framed as a polarized response to perceptions of threat and fear of consequences, in line with McCoy and Somer’s theory of “pernicious polarization” (2019:258).

3.2.3 Empirical Expectations

Given all this, our basic expectations are clear: based on the theoretical relationship between partisan affect, threat, and fear, the question is whether and to what degree fear and threat are linked with affect empirically. The stronger that link, the more evidence of a meaningful way of seeing the world (a way which sees one’s own side as threatened by an evil opponent), and thus the more polarized that side is. Put another way, we are looking for increasing constraint between identities and a set of judgments in a way which deepens conflict between those identities. Then, given both the asymmetry in elite polarization among Republicans in general, and the specific ways that asymmetry has related to affect, threat, and fear in particular, we clearly expect that asymmetry to replicate at the mass level. That is, while the links between threat, fear, and affect should strengthen for partisans on both sides, they should get stronger, faster for Republicans.

However, there is one area where our expectations are not so clear. While I have thus far not differentiated between in and out-party affect in discussing affective polarization, there is in fact disagreement over whether positive in-party affect plays a role in affective polarization, or whether affective polarization might be better understood as *negative partisanship* (Abramowitz and Webster 2016; Finkel et al. 2020; Iyengar and Krupenkin 2018; Mason 2015; Ward and Tavits 2019). This is another place where a new approach can be useful: up to now the literature’s focus on negative out-party affect has primarily reflected a comparative lack of variation in in-party affect (i.e. negative partisanship is responsible for the lion’s share of the overall change in affective bias). But there are still good theoretical reasons to expect in-party affect to play a substantive role in polarization (Mason 2015; Ward and Tavits 2019). Indeed, there are really three theoretically justified possibilities which we can help adjudicate among. The first two are straightforward: either in-party affect *is* a key part of polarization, and so it proceeds asymmetrically in just the same way as out-party affect described above or in-party affect *isn’t* a key part of polarization, and so we find some combination of no relationship between threat, fear, and (in-party) affect, flat relationships which don’t change meaningfully over time, and a lack of asymmetry.

The third, and most interesting, possibility goes beyond the affective polarization/negative partisanship debate by drawing on specific qualitative empirical and theoretical findings on polarization, namely, that polarization should lead to *intra*-party, as well as inter-party,

division (McCoy and Somer 2019:246–47). This division can be partly strategic, but it also flows from a commitment to group uniformity in the context of a political struggle which feels increasingly existential (Somer and McCoy 2019:14). Building on this idea, I propose a new role for in-party affect, namely, as a marker of *dissatisfaction*. For partisans who develop a polarized worldview, driven by growing feelings of fear and threat from the out-group, the question of what their own party is doing to respond to that threat is *critical*. As polarization increases, partisans will demand more extreme responses, and may be frustrated or disappointed at what they see as insufficient action on the part of party elites. This would result in a *weakening* relationship between threat and fear with in-party affect for Republicans, and thus an asymmetry in the *opposite* direction (i.e. with threat and fear more strongly related to in-party affect for *Democrats*).

3.3 Data and Methods

In order to measure the relationships between partisan affect, threat, and fear over time (and thereby assess polarization and asymmetry), I use data from the American National Election Study in a series of regressions, covering the period from 1978-2016. In each regression, I take affect (toward either the in-party or the out-party) as the outcome, and focus on either fear or threat as the independent variable. I then look at how the impact of each independent variable varies over time, by party. Before turning to specific details and justifications of the regression strategy used here, I must first discuss the data and variables upon which the regressions are built.

I use the American National Election Study (hereafter the ANES) in order to cover the scale and the span required to address mass polarization. Use of the ANES is extremely common in the study of polarization, for obvious reasons: given its focus on politics, it provides a variety of data which can be used to measure polarization and, perhaps more importantly, it provides relatively comparable data over time. Comparability over time is obviously critical for establishing trends and setting baselines, and the ANES dates back to 1948. The ANES has run surveys in every presidential election year since then, as well as in midterm years from 1954 to 2002. This makes the ANES one of the longest consistently running nationally representative surveys, outdoing the GSS (for instance) by over twenty years. The ANES has to split its purpose between asking consistent questions over time and asking questions of the moment, however, and so the number of questions asked consistently across the period is not nearly as high as we might like. Even so, we can still track polarization from 1978 to 2016. While earlier data would naturally be welcome, starting in 1978 does match up fairly well with previous work on when elite polarization first becomes visible, meaning we can have some confidence of covering the whole period of interest (Jordan et al. 2014).⁷

⁷This is not to say that the earlier period could not be interesting (indeed, comparisons to earlier periods of polarization have the potential to show us what is unique about the current period, see, e.g., Pierson and Schickler 2020); rather, I simply mean that 1978-2016 should cover the whole period of this *modern* incarnation of polarization.

3.3.1 Variables

To examine polarization in this period, I take in/out-party affect as my dependent variables and fear or threat as independent variables in the regressions which follow. Following previous work, I measure in and out-party affect with feeling thermometers, which ask respondents to rate how they feel towards given named group on a scale from 0 to 100 (Abramowitz and Webster 2016; Banda and Cluverius 2018; Iyengar et al. 2019; Mason 2015).⁸ With one thermometer for the Republican and one for the Democratic Party, I then straightforwardly create measures of in and out-party affect based on respondent party identity.⁹

To get at fear of the out-party, I use one of the ANES’s set of yes/no candidate affect questions: “has [presidential candidate] ever made you feel afraid?”,¹⁰ once again following previous work (Mason 2015).¹¹ Fear of the out-party candidate also has an upside that is worth mentioning: because the candidate is representing the opposing party in *a particular election*, there should be a more substantial connection between the response and a sense of the *stakes* of any particular electoral contest. This has already been touched on above, but a sense of the increasing stakes of political conflict is a consistent and major aspect of polarization (Finkel et al. 2020; McCoy et al. 2018). Thus, having a measure of fear which is linked more directly to *specific* stakes allows me to see how specific situations are or are not linked to partisan affect more generally.

The increasing stakes of the conflict are also closely linked with seeing the out-party as a threat. Based on this connection, I use the respondent’s perceptions of out-party ideological extremity as a proxy for perceptions of threat. The specific question asks respondents to place the parties on a 7-point scale, ranging from 1 (Extremely Liberal) to 7 (Extremely Conservative). I then center the scale on zero and take the absolute value, resulting in a value between 0 (moderate) and 3 (extreme). Using this question as a direct proxy for threat is new, but the basic connection is well evidenced. Simply put, the sense of increased stakes and increased threat are co-constituting: the opposing party is threatening *because* the stakes are high, and the stakes are high *because* the other party is perceived as a threat (Finkel et al. 2020; McCoy et al. 2018; Ward and Tavits 2019), and perceptions of out-party *extremity* have been linked to both (Lelkes 2016; Rogowski and Sutherland 2016; Webster and Abramowitz 2017). Furthermore, the nature of this connection is such that it holds regardless of the accuracy of the perception: if we imagine an “accurate” assessment of extremity, then these are connected to real policy differences, which are known to raise the stakes of conflict (Rogowski and Sutherland 2016), while we can also imagine perceptions of extremity as the

⁸It should be mentioned, however, that the ANES cumulative data file codes all responses between 97 and 100 as 97, so in practice the actual range is 0 to 97.

⁹These measures are undefined for true independents, as they have no in or out-party. They *are* defined for independents who lean towards one of the parties, as I have followed standard practice and counted partisan leaners as identifying with the party they lean towards (Druckman et al. 2013). This is useful for a number of reasons, among them that, when including independent “leaners”, there has been no real change in the proportion of the population which identifies with one of the parties.

¹⁰In 2016, the ANES changed this (along with the other candidate affect questions) to ask “how often has...” with responses on a five point scale from “never” to “always”. Following the cumulative data file, I include the 2016 data by collapsing the five categories to match the original yes/no question.

¹¹Note that, for obvious reasons, this question could only be asked in presidential election years, so models using this independent variable begin in 1980 rather than 1978 and only use data from presidential election year studies.

outcome of an increased sense of stakes (Ward and Tavits 2019).¹²

In order to assess these relationships by party (and thus examine asymmetry), I include a simple dummy variable for party, 1 if the respondent is a Democrat and 0 if a Republican.¹³ I also include the year as both a fixed effect (i.e. a series of dummies) and as a numerical variable. This allows me to look for overall trends over time, net of unrelated year-to-year variation (see Brooks and Manza 2013 for a similar strategy).

Likewise, in order to look at polarization net of differences in party composition, I control for both the strength of partisan identification (from 1, leans, to 3, strong) and respondent ideological extremity (using the same 0 to 3 scale as perceived party extremity). Finally, each model also includes a set of demographic controls: years of education, income, female (dummy), African American (dummy), age, southern residence (dummy), and frequency of church attendance.¹⁴

3.3.2 Analytical Strategy

The actual analysis, then, consists of four weighted OLS regressions, one looking at each combination of dependent variable (affect towards either the in- or out-party) and threat (via perceived out-party ideological extremity) or fear (via fear of the out-party presidential candidate). In each of them, I interact threat or fear with *both* year and party identification. In this way, I can assess not only how in- and out-party affect are related to threat or fear of the out-party, but also whether those relationships develop over time, and whether those relationships develop differently (and potentially asymmetrically) by party.

However, while interaction effects are conceptually straightforward, when more than two terms are involved they can become somewhat difficult to substantively interpret. In this case, I am interested in the marginal effects of the variables of interest, that is, the change in the outcome given a unit change in the variable of interest (Brambor, Clark, and Golder 2006). But because of the interactions, these predictors don't have consistent marginal effects. For instance, the marginal effect of fear of the out-party candidate changes with both time and partisan identity.

We could get a good overall sense of the impact of fear by looking at its *average marginal effect* (AME), obtained by calculating the marginal effect at every observed value and taking the average (Leeper 2018).¹⁵ However, this would not tell us anything about the

¹²This conceptualization is agnostic on the question of accuracy in general, and I don't believe it matters for the argument I'm making: the question at issue here is whether *feelings* of threat are connected to in-/out-party affect. That said, this is not to disregard the question of whether *actually being meaningfully threatened* vs. merely feeling threatened may matter more generally; it is absolutely a question worthy of further investigation.

¹³As mentioned above, I follow standard practice and include "leaners" (that is, those who say only that they "lean towards" one party or the other) as partisans, excluding only so-called "true independents".

¹⁴Results are robust to a variety of alternative model specifications, including several different time period models (instead of full year fixed effects) and the use of partisan sorting instead of partisan identification and ideological extremity (see Mason 2015 for more on partisan sorting).

¹⁵This is different from obtaining the *Marginal Effect at Means*, where the marginal effect of a variable is calculated once with covariates held at their means. The AME is a better summary statistic because it incorporates variability across covariates into its estimate, rather than somewhat arbitrarily assuming that the marginal effect at means is necessarily representative (see Hanmer and Ozan Kalkan 2013 for more on

aforementioned variation with time and party identification. Therefore, I instead look at the *marginal effects at representative values* (MER): these are calculated in the same way as average marginal effects, except that one or more covariates are held at values of theoretical interest.¹⁶ This makes MERs analytically useful in a way other kinds of marginal effects cannot be, as it allows us to compare the impact of a variable of interest across different subgroups (Williams 2012:326).¹⁷

For the analysis that follows, I calculate the marginal effect of the variable of interest (either fear or threat) for every combination of party ID and year. In other words, I calculate the AME of fear on out-party affect with year held at 1980 and party ID held first as Democrat, then as Republican. I then repeat the process for each year, resulting in a set of values which describe the effect of fear on out-party affect in each year, for each party. *These* are the real results of interest, because they are a powerful way to assess both the statistical and *substantial* significance of the relationships of interest. AMEs/MERs are clearly interpretable and comparable, both within a variable and across variables: if the AME of fear on out-party affect in 1980 were -3 for Democrats and -5 for Republicans, that would mean that the relationship between fear of the out-party candidate and out-party affect was 2 points stronger for Republicans than for Democrats. Likewise, if the AME of living in the south were -1, then we could meaningfully say that living in the south has a *weaker* relationship with out-party affect than either.

3.4 Findings and Discussion

These relationships between fear or threat and in/out-party affect are the core object of interest. On their own, each of these variables may provide some evidence of polarization, but none of them show any real evidence of an asymmetry which would match the real differences between the parties.¹⁸ But looking at the relationships between them allows us to go further: if the marginal effects of threat and fear on partisan affect strengthen over time, then we are seeing persistent, increasingly patterned, relationships between affect and judgment. As such, we could say that we really are seeing “meaningful sides” holding together more and more as the population-level manifestations of their shared evaluative schemas become more visible. To do this, I use four regression models, presented in Tables 3.1 and 3.2. Table 3.1 presents the models showing the impact of fear (as fear of the out-party presidential candidate), interacted with year and party identification, on out- and in-party affect, while Table 3.2 does the same

this point).

¹⁶Put another way, an MER is simply the AME *at* a given set of values and thus they can be referred to as such, i.e. the average marginal effect of fear for Democrats in 1980.

¹⁷For more on average marginal effects and marginal effects at representative values and their use analytically (including a comparison with marginal effects at means), see Hanmer and Ozan Kalkan (2013), Leeper (2018), Long and Mustillo (2018), and Williams (2012). See also Druckman et al. (2021) for an example of a similar empirical application.

¹⁸An overview of the descriptive trends for in-/out-party affect, fear of the out-party candidate, and perceived out-party extremity is provided in appendix D and a comparison of the full models versus main effects models is provided in appendix E.

for threat (measured as perceived out-party extremity).¹⁹

Table 3.1: Effect of Fear of the Out-Party Candidate on Out and In-Party Affect

	<i>Dependent variable:</i>	
	Out-Party Affect	In-Party Affect
	(1)	(2)
Fear	337.16*** (76.10)	-124.13* (62.68)
Democrat	393.22*** (72.94)	-354.95*** (60.08)
Dem x Year	-0.20*** (0.04)	0.18*** (0.03)
Fear x Dem	-399.75*** (98.86)	-163.96* (81.42)
Fear x Year	-0.17*** (0.04)	0.06* (0.03)
Fear x Dem x Year	0.20*** (0.05)	0.08* (0.04)
Black	-1.62*** (0.49)	4.06*** (0.40)
Income	-0.11 (0.14)	-0.67*** (0.12)
Age	0.00 (0.01)	0.05*** (0.01)
Gender	0.75* (0.30)	1.22*** (0.24)
Education	-0.91*** (0.18)	-1.53*** (0.15)
South Church	-0.80* (0.32)	-1.90*** (0.26)
Attendance	-0.80*** (0.10)	-0.25** (0.08)
Partisan Strength	-4.95*** (0.19)	7.90*** (0.16)
Ideological Strength	-3.81*** (0.16)	0.82*** (0.13)
Observations	18,592	18,592
Adjusted R ²	0.25	0.22

Note: *p<0.05; **p<0.01; ***p<0.001
Year fixed effects not displayed.

Looking at just the tables shows promising results, insofar as we see generally strongly significant relationships. However, as said, that is necessary but not sufficient; it doesn't really tell us anything on its own. Instead, the real results of interest are the average marginal effects over time, by party. Specifically, we have two different sets of expectations for the two different dependent variables: for out-party affect, we expect generally strengthening (inverse) relationships with the relationships getting stronger, faster, for Republicans than Democrats.²⁰

¹⁹Note that all models include year fixed effects, but the coefficients have been omitted for space and readability.

²⁰Note that, because higher values of out-party affect mean liking the opposition more and lower values mean liking them less (i.e. *disliking* them more), we're expecting to see *negative* relationships which get *more negative* over time between threat/fear and out-party affect. Thus *decreasing* marginal effects of threat or fear on out-party affect are what we expect to see. In the discussion which follows, I attempt to focus on the *magnitude* of effects or changes so as to avoid the confusion of referring to an effect which decreases but thereby gets *stronger*.

Table 3.2: Effect of Perceived Out-Party Extremity on Out and In-Party Affect

	<i>Dependent variable:</i>	
	Out-Party Affect	In-Party Affect
	(1)	(2)
Perceived Extremity	207.38*** (34.85)	24.29 (28.74)
Democrat	457.72*** (94.79)	-240.53** (78.15)
Dem x Year	-0.23*** (0.05)	0.12** (0.04)
Perc Extr x Dem	-159.86*** (45.60)	-80.61* (37.59)
Perc Extr x Year	-0.11*** (0.02)	-0.01 (0.01)
Perc Extr x Dem x Year	0.08*** (0.02)	0.04* (0.02)
Black	-0.90* (0.43)	4.19*** (0.35)
Income	-0.05 (0.12)	-0.76*** (0.10)
Age	0.02* (0.01)	0.04*** (0.01)
Gender	0.43 (0.25)	1.43*** (0.21)
Education	-0.99*** (0.16)	-1.44*** (0.13)
South	-1.43*** (0.28)	-1.89*** (0.23)
Church		
Attendance	-0.82*** (0.08)	-0.31*** (0.07)
Partisan		
Strength	-4.63*** (0.16)	7.66*** (0.14)
Ideological		
Strength	-3.47*** (0.14)	0.93*** (0.12)
Observations	23,448	23,448
Adjusted R ²	0.26	0.21

Note: *p<0.05; **p<0.01; ***p<0.001
Year fixed effects not displayed.

This would be the asymmetry in increasing polarization which we have every reason to suspect, but which previous approaches have been unable to tell us much about. For in-party affect, on the other hand, there is some uncertainty, with three plausible outcomes, each of which would have different theoretical implications for our understanding of polarization: 1) results the same as out-party affect (but positive); 2) nothing of interest (e.g., no relationships at all, flat relationships over time, etc.); 3) the *opposite* of the results for out-party affect.

I now turn to the results themselves, beginning with out-party affect.

3.4.1 Out-Party Affect: Validating the Approach

The results for out-party affect are presented in Figure 3.1 which presents two graphs, one for fear and one for perceived extremity. The graphs plot the average marginal effect of either fear or perceived extremity on out-party affect over time, by party, along with the 95% confidence intervals for the estimate. In other words, they show how many points difference in affect towards the out-party fear or perceived extremity made for a Democratic versus a Republican respondent in each year. So, we can see that in 1980/1978²¹ the impact of both fear and threat were nearly indistinguishable for Democrats versus Republicans, though in both cases they were significant and negative (i.e. fear or threat were associated with *lower* out-party affect, *disliking* the out-party more). However, that changed quickly and substantially: by 2016 a Republican respondent who was afraid of the out-party candidate liked the Democrats 13 points less than a Republican respondent who wasn't, a drop of over 6 points since 1980. On the other hand, the equivalent value for Democrats in 2016 was 7.5 points, which is not only smaller, but is *statistically indistinguishable* from what it was in 1980 (and if there is a trend, it's *positive*, i.e. fear doing *less* to drive negative out-party affect).

The results for perceived extremity, our proxy for threat, are smaller but still significant: in 1978 a one point change in perceived out-party extremity reduced out-party affect by around 4.5 points for either party. By 2016, that value had increased to 5.73 for Democrats, but for Republicans it *nearly doubled* to 8.27 points. These results are not only indicative of asymmetric polarization towards Republicans, they may even call into question just how much Democrats are polarizing at all.

Crucially, these results are not only substantial in comparison to each other, but remain significant (and arguably become *more* impressive) when placed in context. The most obvious place to start is with the overall trend in out-party affect, which has decreased *dramatically* across the period, falling from a mean of about 48 points in 1978 to about 26 points (and thus been labeled variously as affective polarization or negative partisanship). Even in comparison to this precipitous decline, the findings above are notable. If we look at fear, its effect on Republicans in 2016 was about 5.5 points stronger than on Democrats, a net 7.2-point swing compared to 1980, and either figure is more than a *quarter* of the total drop in out-party affect over the period. Likewise, a one-point increase in perceived extremity reduced out-party affect by 2.5 points more for a Republican in 2016 than for a Democrat, a net swing of about 3 points. This seems like a smaller difference, but given that perceived extremity is a four point scale,²² this ultimately results in a 7.5 point gap between a Republican and a Democrat

²¹Recall that “has the out-party candidate every made you feel afraid?” was of course only asked in presidential election years, and so the first year for those models is 1980 rather than 1978.

²²AMEs, for clarity, are not standardized. In other words, the marginal effect of a variable is about the

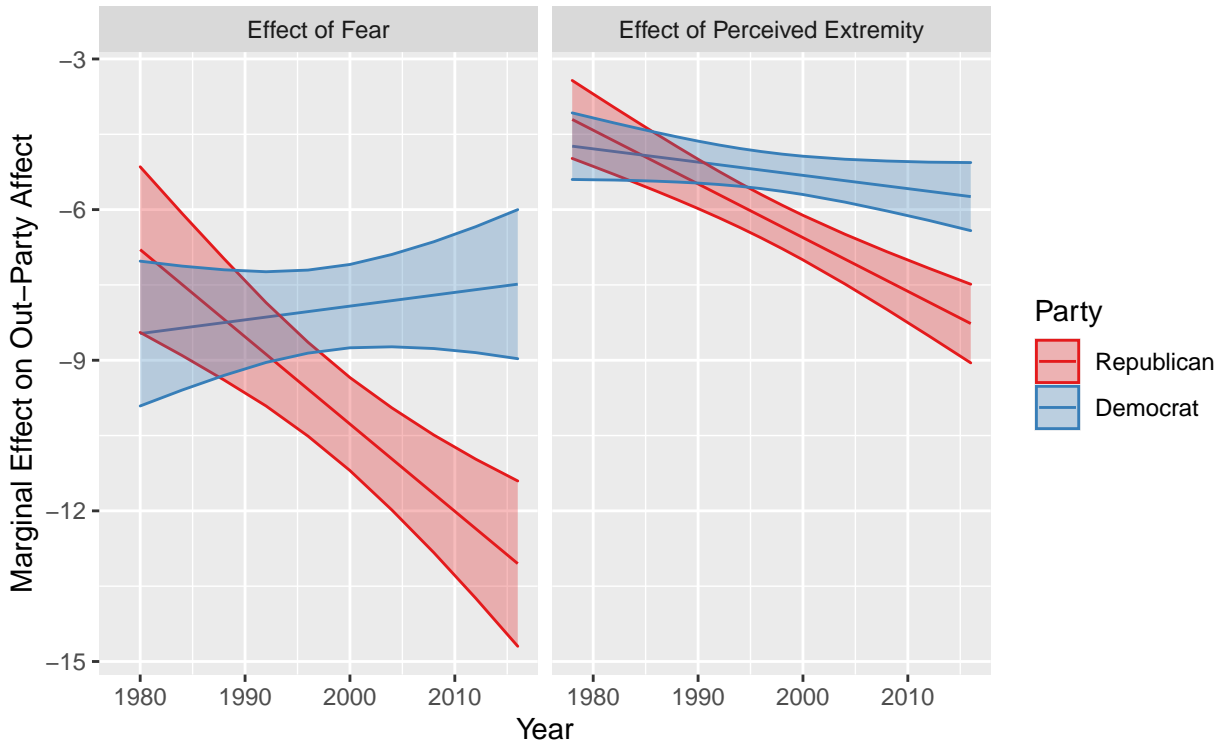


Figure 3.1: Marginal Effects of Fear and Threat on Out-Party Affect

who perceived the opposing party as ideologically extreme in 2016, or a *9 point* net gain for Republicans compared to 1978 (when the effect was slightly stronger for Democrats). In either case, the difference is *more than a third* the total change in average out-party affect since 1978.

This is *almost* exactly what we expected to find. Overall, this provides very strong evidence of connections between out-party affect, threat, and fear for Republicans, connections which have strengthened dramatically over the period such that they are now significantly stronger than the same connections for Democrats. We have, then, evidence that Republican identity holds together now meaningfully in a way it did not at the start of the period, evidence for a more and more coherent meaning to the partisan identity which is demonstrated by a concordance between identity and judgment. What is somewhat unexpected are the results for Democrats: the relationship between threat (perceived extremity) and out-party affect looks exactly as predicted: we see evidence of polarization in a strengthening relationship, just weaker polarization than seen for Republicans. However, the relationship between fear and out-party affect is a different story: we see *no* strengthening relationship, in fact, the trend (though not statistically significant) points towards a *weakening* relationship, if anything. This does not, I believe, undermine the basic takeaway here, which is a finding of asymmetry and a basic validation of this worldview theory of polarization, but it does deepen some of the theoretical questions raised above about whether polarization might function in fundamentally different ways for different identities, a point I turn to empirically now, and will return to theoretically in the conclusion.

impact of a *unit increase* of that variable.

3.4.2 In-Party Affect: Polarization and Dissatisfaction

Turning to the link between in-party affect and fear or threat brings us to the question of the role of in-party affect in polarization: whether in-party affect matches out-party affect, bears no relation at all, or in fact acts in *reverse*. As we can see in Figure 3.2, the third possibility in fact seems to be the case. As with out-party affect, the results are presented in two graphs, one for fear and one for perceived extremity, plotting the AME of each on in-party affect over time, by party. For both fear and perceived extremity, the average marginal effects begin significantly stronger *for Republicans* and wind up statistically indistinguishable. What's more, while for fear this is the result of a differential rate of increase (that is, the AME increased more for Democrats than Republicans), for perceived extremity it results from both a strengthening effect for Democrats and a *weakening* (although not statistically significantly so) effect for Republicans.

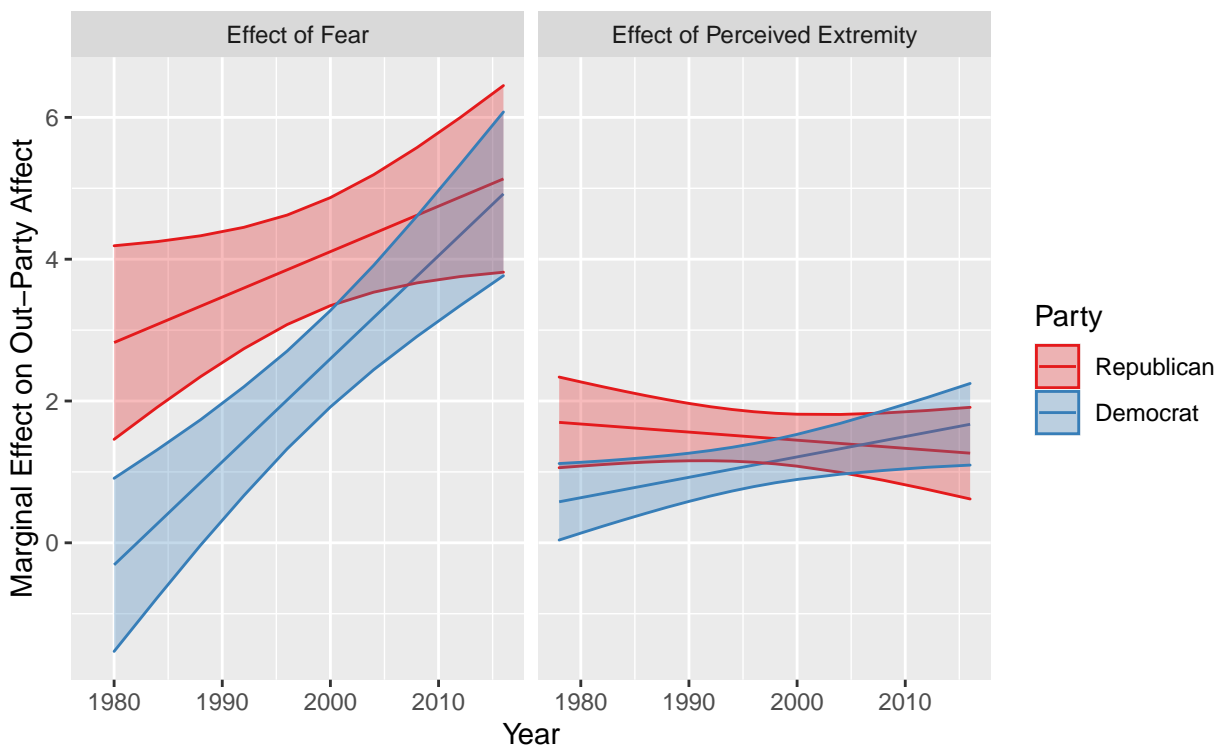


Figure 3.2: Marginal Effects of Fear and Threat on In-Party Affect

These results are of notably smaller magnitude than the results for out-party affect, and may seem like minor changes, potentially minor enough to fit with our second possibility, that of no relationship. But this is simply not the case, for two reasons. First, the (sometimes implicit) argument of work on negative partisanship, from which we have an expectation that in-party affect should be essentially unrelated to polarization, is that there is basically *no* meaningful variation on in-party affect over the period (Abramowitz and Webster 2016; Finkel et al. 2020). But that is not what we see: fear goes from having no statistically significant impact on Democrats' in-party feeling in 1978 to having a 5 point impact in 2016. So the impact for Democrats changes by 5 points, starting from zero, while the impact for Republicans changes by only 2 points. This is absolutely meaningful variation. Second, while

we don't see statistically significant differences between the parties by the end of the period in either case, the trends that got us there are important. In the case of fear, the 5-point increase for Democrats is nearly as big as the 6-point change in the AME on out-party affect for Republicans discussed above. Even more interesting, Republicans see no change at all in the AME of perceived extremity on in-party affect over nearly forty years (or, if we take the point estimates as at least indicative, despite their non-significance, a *decline*), while Democrats *do* see a significant increase.

Though not a perfect fit, these most closely support the third option, namely that we are seeing a *reversed* asymmetry on in-party affect. Against both previous approaches to affective polarization, my relational approach finds meaningful asymmetry on in-party affect *over time*. This matters because it is further evidence that the relationship between partisan affect, threat, and fear, and therefore polarization as such, has developed in fundamentally different ways for the different parties.

3.5 Conclusion

These results strongly suggest an asymmetry in mass polarization towards Republicans. Thus, by redefining polarization as a worldview, shown in the persistent patterned relationships between partisan affect, threat, and fear, I have found evidence of an asymmetry which has thus far been largely invisible to quantitative studies of mass polarization, but which matches the documented asymmetry in elite polarization, the Republican Party's status as a polarizing actor in American politics, and extensive circumstantial evidence of mass asymmetry. With these findings I have validated the worldview model of polarization I developed above. This theory also resulted in more surprising empirical findings, in particular with regard to the role of in-party affect, which can expand our understanding of polarization yet further. And ultimately, both the expected and unexpected findings, along with their theoretical implications, allow us to better understand pressing empirical issues in modern American politics.

That understanding is based, partly, in the new approach presented here. While both attitudinal and affective theories of polarization are powerful, a combination of the scope of their definitions, the assumptions they rely on, and the fundamental challenge of the contingent connection between division and meaningful division, mean that both have been unable to really address the asymmetry found. The problem, in essence, is their difficulty in capturing asymmetry in kind rather than degree. Polarization is a matter of the division of society into meaningful sides, and previous approaches have been too limited in their ability see polarization in the unique characteristics of those sides. I attempt to get around this problem by understanding polarization as a matter of worldview. By defining polarization in terms of persistent patterned relationships between affect, attitudes, and judgments, bound to particular political identities, I am able to incorporate recent insights from sociological approaches to attitudinal polarization on the relational nature of meaning (Boutyline and Vaisey 2017; DellaPosta 2020) in a synthesis with the concept of affective polarization from political science (Iyengar et al. 2012; Mason 2015). This synthesis is further based in a wide range of sociological work on the connection between identity and judgment (Bourdieu 1984; de Leon et al. 2015; Hennion 2007; Wimmer 2013), which links polarization and identity

together in a new way, going beyond the more limited (though still important) attempts of social identity theory or motivated reasoning (Druckman et al. 2021; Druckman et al. 2013; Iyengar et al. 2019; Mason 2015). It also deepens and expands on other work linking differences in public opinion with differences in partisan identity (see, e.g., Brooks and Manza 2013; Morisi, Jost, and Singh 2019), and brings the idea of identity as a heuristic (see, e.g., Boutyline and Vaisey 2017) back to the study of polarization, thereby re-framing the ongoing debate about the primacy of ideology or identity in driving affective polarization (Lelkes 2018; Mason 2018; Rogowski and Sutherland 2016; Webster and Abramowitz 2017). Finally, in making clear the fundamental importance of the specific, and potentially unique, content of political identities, the worldview model provides more analytical leverage by both allowing a stronger coupling between detailed conceptual theories and the empirical investigation of polarization, as well as by analyzing the ways polarization may *work differently* for different identities, beyond the simple matter of magnitude.

This additional analytical leverage was crucial to the question investigated here, and to the strong findings of asymmetry which ultimately validate the approach. The link between partisan affect, threat, and fear is a central part of conceptual work on polarization (Finkel et al. 2020; McCoy and Somer 2019). Furthermore, it is clear that a polarized worldview, one which is linked to the identity of a “meaningful side”, requires a strong link between negative out-party affect and both seeing the other side as threatening, as well as having feelings of fear towards them. And this is precisely what I found. The findings of asymmetry for Republicans on out-party affect are unequivocal, substantial, and meaningful in the context of overall trends in out-party affect in this modern period of polarization. This asymmetry strongly matches expectations from previous work, but has previously been invisible to quantitative work on mass polarization. However, there was one way in which this finding was surprising: the fact that there was no strengthening of the relationship between out-party affect and fear for Democrats. This is not necessarily indicative of a lack of polarization, but rather provides preliminary evidence of at least one way that polarization does seem to work differently for different identities. Put another way, if polarization is a matter of division into meaningful sides, there is no *a priori* reason that the meanings of those sides need be equal and opposite, and this is one way in which they are not.

This idea is further supported by my surprising findings with regard to in-party affect. To recap, the role of in-party affect is an open question even within affective approaches to polarization, with some incorporating it as a part of overall partisan bias (Mason 2015), while others have begun to argue that affective polarization is better understood purely as a matter of *negative partisanship* (Abramowitz and Webster 2016; Finkel et al. 2020; Webster and Abramowitz 2017). However, in place of either replicating findings for out-party affect or finding no relationships at all, I found an *inverse* relationship: the effect of threat and fear on in-party affect increased *for Democrats* while it either increased much more slowly or held steady for Republicans. Now, the results here are admittedly less striking in magnitude than those for out-party affect, but they are still clearly there, and are arguably more striking in their direction. Again, this suggests that something fundamentally different is going for Democrats and Republicans, and provides further evidence that previous affective approaches are insufficient, as it does not fit with either set of predictions.

However, it does fit with one potential explanation rooted in extremely specific, largely ignored, predictions from within the polarization literature. Building on the idea that polar-

ization should lead to increased *intra-*, as well as *inter-*, party conflict, we can understand asymmetry of in-party affect as a marker of *dissatisfaction* (McCoy and Somer 2019:246). If polarization is connected to threat and fear, then, for partisans with a polarized worldview, the dynamics of polarization will necessarily raise the question of whether the party is *doing enough* to respond to the rising threat. Precisely because perceptions of threat and fear are increasing, however, partisans are likely to demand more and more extreme responses and be frustrated or disappointed by the apparently insufficient actions of the party elite.

Now, this theory is obviously preliminary, and is only one potential explanation for the findings here. However, as an explanation for these findings, it would have some major implications. For one, it highlights the difference between affect towards and commitment to a party or identity, two things which are frequently conflated in the literature. In this way it also further clarifies the importance of the *meaning* of different identities: not only that different identities may function in different ways, but that people may stand in relationship to different identities in different ways. For instance, a dissatisfied Republican may be a deeply committed Republican (and might even consider themselves a “real” Republican, in comparison to so-called “Republicans in name only”), while a dissatisfied Democrat might feel *less* committed. Indeed, that same Republican might *give up* their identity as, say, a basketball fan due to dissatisfaction with how the NBA has handled player protests. The same kind of affective understanding can be connected to different identities (or even the *same* identity in different contexts) in radically different ways. Additionally, the link between threat, fear, and dissatisfaction may ultimately build into a *feedback loop*, in which perceptions of threat and feelings of fear make current action seem insufficient, creating demands for more extreme action. This action then proves the validity of those feelings (by recognizing them as actionable) and strengthens them further, ultimately making the new extreme action seem again insufficient. This notion of a feedback loop in particular is potentially relevant for the study of radicalization, both of individuals and, especially, of organizations.

But perhaps most important is that these findings, together with the worldview model as a whole, have clear empirical explanatory power. Put simply, following the logic of the findings and theory through, we can say quite easily why there has been no Democratic equivalent to the storming of the capitol on January 6th, and why it *was* carried out by Republicans. Not only that, we can also contribute to an explanation of the rise of extremism and radicalization of the party and the profound consequences for American politics. The findings above describe a Republican Party that is substantially more polarized than the opposition, but more than that, they describe a Republican Party whose partisans increasingly share a worldview in which their feeling towards that opposition are increasingly driven by perceptions of threat and feelings of fear. Moreover, if we take the strong version of the dissatisfaction hypothesis above, these partisans, though potentially committed to their party, do not trust it to adequately protect them against this increasingly existential threat. January 6th, then, can be understood as what happens when this particular form of polarization and dissatisfaction erupt into action. But we can also frame the historical story of how we got there in these terms as well: in this view, first the Tea Party and then Donald Trump become steps on this treadmill of dissatisfaction, simultaneously the extreme wing of and an insurgency against the Republican Party, simultaneously radicalizing the party and ensuring

that its base will radicalize further.²³ To be clear, I am not claiming that the explanation given here is the only explanation, nor that it has yet been proven. Rather, the point is that the explanation is plausible, powerful, and consistent, that the approach used to arrive at it is valid and analytically useful, and that both might justifiably help set the agenda moving forward.

²³Indeed, we do know that the Tea Party was started by people deeply disillusioned with the Republican Party, and that Donald Trump entered the Republican presidential contest at time when a majority of Republicans had “little faith in party leaders” (Skocpol 2020:18).

Chapter 4

Beyond the Structure of Attitudes: Belief Systems, Affect, and Polarization

4.1 Introduction

The recent focus on the structure, rather than just the content, of beliefs in survey research is an exciting development. Though enabled, in part, by new methodological advances (e.g. belief network analysis and relational class analysis), the consideration of structure in fact aligns with theoretical understandings from much earlier (see, e.g., Converse 1964; DiMaggio 1997; Jost, Federico, and Napier 2009; Martin 2002). Though they used a variety of terms (e.g. belief systems, ideologies, schemas), these theories all understood that an individual belief isn't truly meaningful on its own; rather, it is made meaningful by its relationship to other beliefs. Thus, what really matters is not the distribution of individual opinions in a population, or the correlation of a particular attitude with other factors, but the existence, strength, and structure of belief systems.

This focus is particularly important for the study of political conflict because of its resonance with theories of pluralism, in which stability really is a matter of belief structure. Successful pluralistic societies are not made stable by an absence of disagreement, but rather by the structure of that disagreement. The question is whether issue cleavages align with each, creating unbridgeable fault lines, or crosscut, creating stability by maintaining avenues of agreement and compromise (Blau 1974; Dahl 1961; Durkheim 1984; Lipset and Rokkan 1967). This idea of aligning issue cleavages is really a question of belief structure, and recent work in this vein has offered somewhat sobering assessments of the state of American pluralistic democracy in particular (DellaPosta 2020). Indeed, this is one of the major avenues pursued in the study of polarization, using both structural and other methods, making the examination of political belief structures all the more pressing (Baldassarri and Gelman 2008; Kozlowski and Murphy 2021)

Other work on polarization, however, has focused on the importance of affect and emotions rather than attitudes or beliefs, arguing that polarization (and thus stability) is better understood in terms of social identities (especially partisanship) and how those identities

feel about each other (Iyengar et al. 2019, 2012; Mason 2015; Mason et al. 2021). Likewise, although work on belief structures more generally has also looked almost exclusively at propositional attitudes, work on emotions in variety of sub-fields across psychology and sociology has shown that affects and attitudes can, and often do, influence one another (see, e.g., Clifford 2019; Clore and Schiller 2018; Ladd and Lenz 2008; Mackie et al. 2000; Smith, Seger, and Mackie 2007; Webster and Albertson 2022; Wilson-Mendenhall and Barsalou 2018). But attempts to bring these approaches together in the context of politics have thus far been limited to comparisons or contests, rather than synthesis (see, e.g., Lelkes 2018; Rogowski and Sutherland 2016; Webster and Abramowitz 2017). This paper is the beginning of such a synthesis.

Rather than trying to adjudicate primacy in some way, I simply ask whether affective items can be incorporated into belief structures and investigate the consequences of doing so. I do this by building on Boutyline and Vaisey’s belief network analysis (BNA), constructing not only belief (attitudinal) networks, but affective and combined networks as well (2017). Specifically, I consider attitudinal, affective, and combined networks for matching items in the 1992 and 2016 American National Election Studies, making comparisons both across the different types of network and between the two years of the survey. These comparisons validate the inclusion of affective items by showing, first, that the networks produced by their inclusion are not wildly dissimilar to those produced by attitudinal items alone and, second, that the variation which is there is meaningfully explicable, a source of real analytical leverage. Indeed, the inclusion of affect and the comparisons this enables provides information and, perhaps more importantly, axes of analysis which would not otherwise be available. Thus, beyond simply validating the use of affect, my analysis also helps us to better understand the development of polarization over the period as the increasing enmeshing of attitudinal and affective networks. Ultimately, I argue that we must incorporate affects into our study of belief structures and, furthermore, that the nature of this incorporation is itself a meaningful axis of variation. These findings are not only important methodologically for the study of belief structures, they are substantively consequential for the study of polarization, offering a way towards reconciling affective and attitudinal approaches and pointing the way towards a new definition, as well as for the study of pluralism, presenting a new way to think about which cleavages are important and what exactly cleavage alignment means

4.2 Literature Review

4.2.1 Why a System?

In order to understand why the role of affect in belief systems matters, we must first understand why the perspective of belief systems matters in the first place. We often tend to think that beliefs or ideas are meaningful in themselves, e.g. two people who believe that abortion is wrong necessarily believe substantially the same thing and, furthermore, that this belief should be consistent in its relation to outcomes or causes of interest (Abbott 1988). The fundamental insight of a belief systems perspective is that this is not the case. Instead of standing alone, beliefs depend on one another and are largely meaningful in relation to one another. This core idea is fairly old and has been developed in different forms in psychology, sociology,

political science, and linguistics. For instance, Converse, in his foundational study of American political opinion, defines a belief system by the “constraint or functional interdependence” among its elements (1964:207), while DiMaggio et al. (2018:32) are more explicit about the principle of relationality, according to which “meaning emerges not from single entities but out of relations among them”. Regardless of the precise source, terminology, or justification, all the different versions share this same core idea.¹

This idea has pretty dramatic methodological implications: if beliefs are interdependent, then to study them our analyses must take into account the structure of the overall system, rather than the just the content of individual beliefs. But this is substantially different from the perspective taken by most survey research which tends to focus on individual beliefs, or sets of beliefs, as separate dependent or independent variables (Abbott 1988). Indeed, this has been true even of much work which self-consciously adopts a belief system perspective (Baldassarri and Goldberg 2014). There are, however, a variety of new methods which do allow us to study belief systems and their structure effectively. For instance, while latent class analysis (LCA) identifies groups of respondents with similar patterns of responses, relational class analysis (RCA) groups respondents based on the similar *relationships between* responses (DiMaggio et al. 2018; Goldberg 2011). In other words, rather than finding groups of people who agree with each other, it finds groups of people who organize the space of agreement/disagreement in similar ways. Compared to a standard regression approach, this captures the fact that, though opposing ideologues may disagree completely on the substance of every issue, they understand the field in exactly the same way. And, in sharing this understanding, they are in a sense more meaningfully similar than people without a consistent ideology who nominally agree with them on a few issues. Boutyline and Vaisey’s (2017) belief network analysis (BNA) offers a direct formalization of the common metaphor of belief systems as networks of beliefs, treating survey questions as nodes linked by the strength of the correlations between them. In this way they are able to show the structural importance of conservative/liberal identification to the overall belief system by using network measures of centrality .

4.2.2 Systems of Polarization

As the examples of RCA and BNA hopefully make clear, the perspective of belief systems is especially useful in the study of political conflict, and particularly in work on polarization. Conceptually, polarization is perhaps best defined as the division of politics, and then society, into an “us” and a “them” which are locked in an increasingly existential conflict (DellaPosta 2020; Finkel et al. 2020; Iyengar et al. 2012; Lelkes 2016; McCoy et al. 2018). Such division can have disastrous consequences, poisoning the political process, bringing partisan animosity into everyday life, and ultimately undermining the basic democratic functioning of society (Hacker and Pierson 2005; McCoy et al. 2018). The question, then, is how such division happens and how the “us” and “them” come to be defined.

So called *constraint* based definitions of polarization answer by drawing on studies of pluralism, according to which stable societies are stable not because of a lack of conflict, but

¹Throughout this paper, I will use the term “belief system” for consistency’s sake, but it should be understood here that there is significant overlap with terms like ideology, cognitive/culture schema, etc.

because of the structure of that conflict. Specifically, Pluralistic societies are stable when conflict is expressed along multiple, cross-cutting issue cleavages. So, for example, two people may disagree about the legality of abortion, but agree about funding for schools. Cross-cutting cleavages like this create areas of compromise and limit the depth of conflict (Baldassarri and Gelman 2008; Blau 1974; Dahl 1961; Durkheim 1984; Lipset and Rokkan 1967). Polarization, then, is a matter of aligning issue cleavages which reinforce each other and ultimately lead to all political conflict functioning as a part of the same single overarching conflict (Baldassarri and Gelman 2008; DellaPosta 2020; Kozlowski and Murphy 2021). Fundamentally, then, this is a belief system based definition of polarization; indeed, “constraint” is a structural property of belief systems: it refers to the degree to which different beliefs “constrain” (or correlate with) each other (Converse 1964). High constraint means lots of aligning cleavages, such that, if I know your position on one issue, I likely know your position on others as well. This is exactly the kind of structural property which a belief systems perspective calls attention to, and which RCA and BNA can usefully identify.

However, there are other ways to define the division between “us” and “them” at the heart of polarization: namely, affective (or “social”) polarization. Affective approaches define polarization in terms of partisan animosity or bias, i.e. how partisans feel about each other (Iyengar et al. 2012). Rather than looking for substantive disagreement, affective definitions focus on partisanship as a social identity, drawing on work in social psychology on inter-group conflict to look for the totalizing “us” vs. “them” conflict in moralized animosity between social groups (Finkel et al. 2020; Iyengar et al. 2019, 2012; Mason 2015). These definitions have the advantage of both strong face validity (i.e. it doesn’t matter whether two groups nominally agree on issues if they hate each other) and strong external validity, because we know that inter-party affect has measurable impacts on other indicators of polarized conflict (Iyengar et al. 2019; Iyengar and Westwood 2015).

Partly because of their substantial differences, the relationship between constraint and affect in polarization remains wide open. Work in this area has thus far been basically comparative, attempting to adjudicate or untangle causal primacy (see, e.g., Lelkes 2018; Rogowski and Sutherland 2016; Webster and Abramowitz 2017). And this work has largely not, in fact, focused on constraint, but rather other attitudinal approaches to polarization. Now, this is absolutely valuable work, but it is also limited. It is not as though constraint and affect are by nature separate. They are certainly different, but it remains to be seen whether those differences are not per se incompatible. Perhaps which aligning cleavages are relevant for defining the in/out-group are defined by the affect around them, or alignment occurs *both* through shifting opinions and shifting affect. In other words, rather than arguing primacy, there may be an opportunity for synthesis here.

4.2.3 Belief and Affect

This synthesis, however, would go beyond polarization, because the relationship between constraint and affect in polarization forcefully raises the question of affect in belief systems more broadly. Although affect (and/or emotion)² has certainly been discussed in the wide and

²Throughout this paper, I make no fine-grained distinction between affect and emotion (see, e.g., Wilson-Mendenhall and Barsalou 2018) and understand both to mean, essentially, “embodied feelings in response to

varied literatures on belief systems, broadly construed (see, e.g., Bourdieu 1984; DiMaggio 1997; Hennion 2007; Wimmer 2013), in the context of survey based research, especially on politics, its role remains largely undiscussed. In particular, I am not aware of any work using the more recent, structurally oriented, methods (e.g. LCA or BNA) which engages systematically with affect.³ The combination of this lack of attention and the open question of constraint versus affect in polarization, we have now a very basic question: do affective items belong in our studies of belief systems? Or, to put it a bit more analytically: we need to understand whether affective evaluations similar enough to propositional opinions that they play meaningfully similar roles in belief systems.

Now, our immediate answer might well be no. This would be in line with the longstanding division between rational propositional and irrational emotions, a division which is indeed replicated by work on affective polarization and reasoning (see, e.g. Dias and Lelkes 2021). This distinction makes the exclusion of affect seem straightforwardly plausible. For instance, some theories of belief systems see the relationships between beliefs as logical or linguistic; if this is the case, then we might imagine that emotions wouldn't participate in these relations (see, e.g., Converse 1964). Alternatively, perhaps they shouldn't participate in the same way: it may be the case that affect "inflects" or "orients" beliefs, and that these ways of relating are sufficiently different from normal belief system relations that it would be analytically dubious to combine them. It is also possible that affects are connected, meaningfully, in belief systems, but that including them would nonetheless make no contribution because they provide no new information because they are too tightly coupled to the relevant attitudes. For instance, we might imagine that a respondents who are angry about abortion are angry *because* they have strong beliefs, so including an affective item on abortion would only serve to add measurement error. Or, in terms of belief network analysis, affect is too tightly bound to attitudes, such that its addition doesn't add any meaningfully novel connections to the network.

However, it has become clear in more recent research that affect and attitude, emotion and propositional opinion, are far more similar than we tend to think. To begin with, it is at this point obvious that there is no firewall separation between rational propositional thought and irrational emotion. On the contrary, we know now that affect significantly impacts our cognition in a variety of ways across multiple domains (Clore and Schiller 2018; Hibbing et al. 2014; Munro, Weih, and Tsai 2010). At an even more basic level, though, the study of emotion and affect make clear that, while they may function differently from attitudinal reasoning, they are not "irrational" in the sense of being nonsensical. Indeed, affect and emotion bear specific and useful propositional content insofar as they serve as evaluations of a given situation, which help us organize and respond to our sense perceptions (Clore and Schiller 2018). And the similarity goes still further, as it is becoming increasingly clear that emotions are concepts which develop in memory using the same cognitive machinery used by our brains in developing any other concept (Wilson-Mendenhall and Barsalou 2018).

some stimulus". For more, including some potential lines of demarcation, see Clore and Schiller (2018).

³There are two minor exceptions I am aware of: first, in Boutyline and Vaisey's (2017) original application of BNA they do include two affective survey items, but they give no particular discussion to their inclusion. Second, work with LCA, including Goldberg's original statement of the method (2011), has often engaged with questions of taste, which we could plausibly distinguish from propositional attitudes, but these are not affective items, nor are they political.

In keeping with this similarity, we have substantial specific evidence of affect influencing beliefs in a variety of ways. Crucially, this influence is not universal or generic, rather, it is relational and context dependent in much the way propositional beliefs have been shown to depend on one another within a belief system. In fact, much of this evidence comes specifically from the study of affective polarization, where partisan bias at the individual level has been linked to a variety of both specific attitude differences and processes of reasoning and evaluation (Dias and Lelkes 2021; Druckman et al. 2021; Lu and Lee 2019; Webster and Albertson 2022). Importantly, these findings show evidence of specific, relevant, affects and emotions influencing specific relevant beliefs in ways which are predictable and theoretically interpretable. This sort of influence looks exactly like the sorts of heuristic judgments which are important for ideological development (Boutyline and Vaisey 2017).

But there's an even more foundational point here: the division between beliefs and emotions may well be wrongheaded in the first place. Returning to the idea, discussed above, that affect helps us organize our perceptions,, it has been argued that no perception comes to us without some kind of affect attached and that, in this way, affect "provides a common currency for evaluating the variety of otherwise incommensurate objects that people encounter" (Clore and Schiller 2018:533). Given this, we might imagine that likewise every propositional belief will have some affect attached to it, and thus that it is partly by way of these affects that belief systems are held together. Indeed, just as theorists of social constraint have argued that ideological belief systems are built into coherence because of an underlying ideological commitment which enables various shortcuts in the formation of new beliefs (heuristic reasoning, clear rubrics for which sources of information to follow, etc.) (see Boutyline and Vaisey 2017:1377–78 for a brief overview), an overarching affective orientation could easily function in the same way. Even aside from any specific theory of influence, it seems increasingly clear that affect and attitudes are not cleanly separable and that their separation in our understandings of belief systems is perhaps more an artifact of the surveys we use and their limitations.

Ultimately, I am arguing for a more inclusive understanding of belief systems and a more thorough version of a relational theory of meaning. As discussed above, the essential insight of a belief systems perspective is that beliefs do not stand alone, but rather are (at least partly) made meaningful by their relation to one another. However, there are different versions (or perhaps different degrees) of this insight. On the one hand, it may be that beliefs stand in various functional relationships with each other (see, e.g., Converse 1964) or be related by way of elective affinities (see, e.g., Jost et al. 2009). On the other hand, there is the version of the insight grounded in structural linguistics and more common in the sociology of culture which holds that the very literal linguistic meaning of the propositions involved is determined by the larger network (Mohr 1998). That is the perspective I want to argue for here and bring to the analysis of political belief systems: belief systems are not just a matter of functional interdependence of propositional beliefs, but rather they are structures of meaning whose members are only made meaningful (in any real sense) by their relationships with one another, propositional beliefs and affective orientations both. For instance, although we imagine ideologies as basically propositional, an ideological commitment to lower taxes means something substantively different (both to you and to others) if you also feel that taxes are repugnant. Indeed, a belief that less government is better or that government is ineffectual are almost entirely different from those same beliefs *tied up with hatred* towards

the government.

In summary, then, I am arguing that attitude and affect are similar enough in how they function in the context of belief systems that any consideration of belief systems ought to include both. Doing so will not only better fit our wider theoretical understanding of affect and emotion, it will also provide a more complete picture of any given belief system than we would have without including affect. Furthermore, in the particular case of constraint and affect based definitions of polarization, doing this can help us better understand the relationship between these two definitions and even begin to offer a meaningful synthesis. In order to demonstrate this empirically, I apply Boutyline and Vaisey’s method of belief network analysis (mentioned in brief above) in a comparative analysis of attitudinal, affective, and combined “belief” systems

4.2.4 Empirical Strategy

Belief network analysis is perhaps the most flexible of the recent belief system oriented methods, in part because it most directly mobilizes the basic insights of that perspective. Belief systems are often described as “networks” of beliefs connected by ties of mutual influence. Belief network analysis formalizes this observation by constructing literal belief networks out of survey data: each question is a node and each pair of questions is connected by an edge whose weight is equal to the square of their correlation (Boutyline and Vaisey 2017).⁴ Various network analysis methods can then be used to assess the structure of the network. In particular, Boutyline and Vaisey treat the edge weights as inverse distances, such that edges are shorter when the correlation between two questions is stronger. This results in a network in which all nodes (beliefs) are connected to each other, but where the shortest route between any two nodes is unlikely to be their direct connection. They then use this idea of shortest paths to identify the central organizing belief of the network by looking for the belief through which the greatest number of shortest paths passes, i.e. shortest-path betweenness.

As with any method, BNA relies on a number of assumptions. Most substantially, Boutyline and Vaisey ground BNA in a fairly rigid theory of belief systems, in which systems develop outward from a some general organizing belief which then stochastically influences what positions get taken on more and more specific issues (2017). The result is a clear center-periphery structure where more central beliefs play a more central role. It is this assumption which justifies, mathematically, their shortest-path betweenness to identify the central belief. Fortunately, there is good evidence that (at least political) belief systems do work in this way, both from Boutyline and Vaisey’s own study and other work on constraint (e.g., Converse 1964). Additionally, DellaPosta is able to apply BNA profitably under a looser set of assumptions, drawing on work by Goldberg and Stein (2018) which shows that aggregate belief structures can help shape individual beliefs even for individuals who do not clearly share the aggregate structure.

Crucially, based on the preceding theoretical overview, adding affective items to a political belief network should not change the plausibility of even Boutyline and Vaisey’s stricter

⁴The square of the correlation is used in order to limit the influence of statistically insignificant correlations (Boutyline and Vaisey 2017). There is some variation here, however, as DellaPosta (2020) instead uses absolute correlations.

set of assumptions. The basic center-periphery structure they describe is precisely the structure we would expect for a network of political affects, based on the implications of affective polarization and social identity theory (Huddy 2001). Indeed, Boutyline and Vaisey’s assumptions are explicitly grounded in the social constraint theory of ideological development, which we just discussed above as being perfectly consonant with our theoretical understanding of affect.

Furthermore, while these assumptions are not insubstantial, any method requires accepting a set of assumptions and trade-offs and, in this case, the trade-offs are worth it. First, as said above, BNA is perhaps the most encompassing method for looking at belief systems, because it allows you to capture and analyze the whole of a belief system, with the main limitation being the quality of the survey being used. More than that, it possesses strong face validity, with a basically direct correspondence to the underlying theory. As a result, standard networks characteristics and analytic techniques have clear and meaningful implications for the belief system. This is perhaps most obvious with Boutyline and Vaisey’s already discussed use of shortest-path betweenness centrality to identify the organizing, keystone, belief in the system. Measures of centrality like this can then be used assess the overall *centralization* of a network or the degree to which the network is has one well-defined center. In a belief network, high centralization is an indication of the kind of center-periphery structure hypothesized by Boutyline and Vaisey. Likewise, there are a number of ways to detect “communities” of nodes in a network, sets of nodes more connected with each other than the rest of the network, which DellaPosta labels as cohesive “belief modules” and whose nature, number, and significance (compared to the connectedness of the rest of the network) can all tell us about the overall coherence of the system and the strength of larger cleavages within it (DellaPosta 2020:519). Even one of the simplest metrics in a weighted network, density, has useful implications here. Density is the average edge weight over the whole network and increasing density is indicative of the kind of overall increase in constraint standard constraint-based definitions of polarization are concerned with. The higher the density, the stronger the overall system.

I intend to take full advantage of the analytical leverage these techniques provide to examine the similarities and differences between political attitudinal and affective networks, to show that their combination makes sense and provides new information, and furthermore to look at how these networks changed from 1992 to 2016. In this way I want to demonstrate both the analytical usefulness of including affective items in the study of belief systems and to provide further evidence of a combined constraint and affective polarization over time.

4.2.5 Empirical Expectations

Based on the theoretical discussion above, we have some fairly clear empirical expectations for what this should look like. First, our most basic claim is that affective and attitudinal networks should look broadly similar, as should the combined network. In essence, if the networks look incomparably different then there is no way that affect functions similarly enough to attitudes to justify its inclusion in analysis of belief systems, and the central contention fails at the first hurdle. To be clear, this isn’t to say there should be no meaningful differences at all, nor even that the differences shouldn’t be systematic or consistent. Rather, the point is that the differences should be relatively restricted in magnitude. These should be like differences between species in a genus, rather than across kingdoms; the differences

between a wolf and a coyote, not a wolf and mushroom, so to speak. Indeed, there are some differences that we explicitly should expect: given the manner in which affect can serve as bridge between otherwise dissimilar items or situations, the affective network should be overall more densely connected than the attitudinal network. Likewise, we should expect the combined network to sit in the middle, as affective items help bridge between otherwise dissimilar attitudes. For the same reason, the affective network should have a somewhat less differentiated structure: fewer identifiable “communities” of items with a higher percentage of items included in the largest of those groups. On the other hand, the attitudinal networks should have more “communities” with fewer items in the largest ones. What the combined network looks like, however, depends on how exactly the attitudinal and affective networks integrate.

And this brings us to our expectations for differences over time. The differences I have just laid out are fairly foundational, grounded in the basics of what we have established about affects and attitudes, but not all differences between network types will be so consistent. Plenty of differences will instead be basically historical in nature. For instance, we have good evidence that polarization increased from 1992 to 2016 (Iyengar et al. 2019; Lelkes 2016). This increase was larger for affective polarization (Lelkes 2016; Mason 2015), but recent more work has indicated that attitudinal, constraint based, polarization has increased as well (DellaPosta 2020; Kozłowski and Murphy 2021). A number of expectations for differences over time flow directly from this. First, given previous work on pluralism and constraint, as well as on belief network analysis, the most obvious expectation is that we should see an increase in overall density across all networks. And, given Boutyline and Vaisey’s argument in their original application of BNA, we should also expect to see changes in the centralization of the networks and, especially, changes in which nodes are most central, as the belief system becomes more consistently structured around a single polarizing cleavage. However, we may not see increases in centralization across the board, because betweenness centrality functions less well in highly constrained networks (Boutyline and Vaisey 2017). In an attitudinal network, decreasing centralization would imply a kind of decreasing organization, that some consistent ideology or organizing principle has become less important. But in the affective network, given the level of connection already present, it may in fact be evidence of the opposite. As the affective network becomes more strongly connected, the relative importance of the central organizing nodes decreases because *all nodes* are more organized.

Finally, and perhaps most importantly, the degree to which the attitudinal and affective networks are enmeshed (that is, interconnected) should increase over time. The question of integration is an exciting new axis of variation which including affect in belief systems provides. Just because affect should be included doesn’t mean that the relationship between affective items and attitudes will always be the same. Even the argument above, that all attitudes have some affect bound up with them, doesn’t imply that we will see consistent connections in an aggregate belief network. In fact, inconsistent links between beliefs and emotions may well be the norm. As such, we should expect to see variation in the structure of combined affective and attitudinal belief systems: in some, affect and attitude will have little to do with each other, while in others they will be tightly integrated. This degree of integration is an important structural variable, telling us a great deal about the role affect plays in a system. Now, the practical meaning of this variation may not be clear in the general case, but in the case political attitudes and affects, it absolutely is. The polarization

literature consistently understands the increasing role of emotion in politics as a component of polarization (McCoy and Somer 2019; Somer and McCoy 2019). Part of what is concerning in polarization is the way in which partisan animosity touches more and more things, the way in which all of society eventually reorients around a consistent political conflict (DellaPosta 2020; Finkel et al. 2020). And we can also link this to a form of constraint: the more enmeshed affect and attitudes are, the stronger a hold affect has on propositional politics and, ultimately, the more those propositions *matter* in a consistent way (on a related point, see de Leon et al. 2015 on political articulation). As such, along with increasing density of the network and consistent centralization around keystone beliefs or affects, the degree of integration of affective and attitudinal items in the network should be seen as definitive of polarization in the combined network.

4.3 Data and Methods

In order to construct and compare these networks, I draw on the American National Election Studies in 1992 and 2016. Extending the work of Boutyline and Vaisey (2017), I use this data to generate an attitudinal, affective, and combined belief network based on a shared subset of items for each year. The analysis then consists of comparisons both across network types and within each network type over time using a combination of network metrics previously used in belief network analysis (Boutyline and Vaisey 2017; DellaPosta 2020), along with measures made possible only by the addition of affective items. Throughout the analysis, I use a nonparametric bootstrap to establish the statistical robustness of these comparisons.

I use the American National Election Studies (ANES) here simply because few (and possibly no) other survey datasets satisfy the requirements of the analysis. The argument being pursued here relies on having survey data which 1) cover a time span long enough to enable an interesting comparison, 2) include both affective and attitudinal *political* items, and 3) have enough of each type of question which are the same in at least two years (across the aforementioned sufficiently wide time span). The ANES satisfies all three of these requirements. It is one of the longest running surveys in the US, having been conducted every four years from 1948 onward (and every two years between 1954 and 2002). It follows that schedule because it focuses on politics and is structured around the US election cycle and is commonly used in research on polarization for these reasons, as well as for the fact that it includes (in addition to attitudinal items) a fairly wide variety of affective question. The other major contender would be the General Social Survey (GSS), which has been used profitably for belief network analysis (including over time comparison) by DellaPosta (2020). However, while it has some political items, that is not the focus of the survey (which is, in fact, precisely why it was so useful for DellaPosta). Likewise, it largely lacks affective items (especially political affective items) across its span.⁵ The ANES, on the other hand, meets all three criteria perfectly.

That said, though the ANES was first administered in 1948, I am looking here at 1992 and 2016, because this offers the best balance of the first and third requirements. We know

⁵The GSS does include items which are non-attitudinal, e.g., a variety of items related to matters of taste (which Goldberg 2011 used in the original application of latent class analysis). However, these items are neither political nor clearly affective.

that both elite polarization and affective polarization have been increasing from roughly the late 1970s (Lelkes 2016), making a 1978 to 2020 comparison arguably the most ideal. However, such a pairing would provide a total of just 17 items (including both affective and attitudinal items) shared between the two years. Given that the object is to create a set of networks which cover the relevant belief systems as completely as possible, that is simply not enough. By the 1992 to 2016 comparison still very much encompasses a period of significant political change (polarization included), while also providing enough survey items which are present in both years to make for a plausible analysis.⁶

Table 4.1: Attitudinal Variables

Short Name	Description
Partisan Id	PARTISANSHIP: Party Identification of Respondent- 7-point Scale
Fed Few	SYSTEM SUPPORT: Federal Government Run by Few Interests or for the Benefit of All
Waste Tax	SYSTEM SUPPORT: How Much Does the Federal Government Waste Tax Money
Official Care	SYSTEM SUPPORT: Government Officials Care What People Like Respondent Think
No Say	SYSTEM SUPPORT: People like Respondent Have Any Say in What the Government Does
Election Matter	SYSTEM SUPPORT: How Much Elections Make Government Pay Attention to People
Ideo Selfid	IDEOLOGY: Liberal-Conservative Scale
Health Insurance	ISSUES: Government Health Insurance Scale
Jobs Income	ISSUES: Guaranteed Jobs and Income Scale
Aid To Blacks	ISSUES: Aid to Blacks Scale
Abortion	ISSUES: By Law, When Should Abortion Be Allowed
Services Spending	ISSUES: Government Services-Spending Scale
Def Spend	ISSUES: Defense Spending Scale
Stay Home	ISSUES: Better off if U.S. Unconcerned with Rest of World
Gay Job Disc	ISSUES: Strength of Position on Law to Protect Homosexuals Against Discrimination
Gay Adoption	ISSUES: Should Gays/Lesbians Be Able to Adopt Children
Immigrant Numbers	ISSUES: Increase or Decrease Number of Immigrants to U.S.
Spend Schools	ISSUES: Federal Spending- Public Schools
Spend Crime	ISSUES: Federal Spending- Dealing with Crime
Spend Welfare	ISSUES: Federal Spending- Welfare Programs
Spend Childcare	ISSUES: Federal Spending- Child Care
Spend Poor	ISSUES: Federal Spending- Poor/Poor People

⁶Unfortunately, the 2020 survey specifically doesn't work due to some major changes in previously long-running question, changes which hit affective questions in particular.

Table 4.1: Attitudinal Variables (*continued*)

Short Name	Description
Lifestyles	MORAL TRADITIONALISM: Newer Lifestyles Contribute to Society Breakdown
Adjust Morals	MORAL TRADITIONALISM: Should Adjust View of Moral Behavior to Changes
Trad Values	MORAL TRADITIONALISM: Should be More Emphasis on Traditional Values
Tolerance	MORAL TRADITIONALISM: Tolerance of Different Moral Standards
Religion Guidance	RELIGIOSITY: How Much Guidance from Religion
Bible Authority	RELIGIOSITY: Authority of the Bible
Aapref Hiring	ISSUES: Affirmative Action in Hiring/Promotion
Child Cur	Child trait more important: curiosity or good manners
Child Obed	Child trait more important: obedience or self-reliance
Child Consid	Child trait more important: considerate or well-behaved
Child Ind	Child trait more important: independence or respect
Spend Ss	Federal Budget Spending: Social Security
Spend Enviro	Federal Budget Spending: protecting the environment
Egal Equal	Society should make sure everyone has equal opportunity
Egal Worryless	We'd be better off if worried less about equality
Egal Notbigprob	Not a big problem if some have more chance in life
Egal Fewerprobs	If people were treated more fairly would be fewer probs
Penalty Favopp	Favor or oppose death penalty
Resent Workway	Agree/disagree: blacks shd work way up w/o special favors
Resent Slavery	Agree/disagree: past slavery make more diff for blacks
Resent Deserve	Agree/disagree: blacks have gotten less than deserve
Resent Try	Agree/disagree: blacks must try harder to get ahead
Divgov Splitgov	Party Control or split government
Govrole Big	Govt bigger because too involved OR bigger problems
Govrole Market	Need strong govt for complex problems OR free market
Govrole Lessmore	Less govt better OR more that govt should be doing
Stype Hwkwhite	Stereotype: Whites hardworking
Stype Hwkblack	Stereotype: Blacks hardworking
Stype Hwkhispanic	Stereotype: Hispanics hardworking
Stype Hwkasian	Stereotype: Asians hardworking
Stype Violwhite	Stereotype: Whites violent
Stype Violblack	Stereotype: Blacks violent
Stype Violhispanic	Stereotype: Hispanics violent
Stype Violasian	Stereotype: Asians violent

There are a total of 80 items, 56 attitudinal and 24 affective, which are displayed in Tables 4.1 and 4.2. The selection of items was proceeded in several steps. Beginning with all questions appearing in both 1992 and 2016, I followed previous work with BNA (Boutyline and Vaisey 2017; DellaPosta 2020) and considered only what Alwin (2007) would label “non-factual” questions and excluded “factual” questions, the distinction being based on whether a question’s answer could be verified by reference to objective records. Thus I excluded administrative, demographic, and behavioral questions. Then, I excluded any questions with no plausible political relevance. The remaining questions were then categorized as either attitudinal, affective, or neither. This categorization was fairly straightforward, as I was able to follow previous work on, e.g., affective and attitudinal polarization (Druckman and Levendusky 2019; Iyengar et al. 2019; Lelkes 2018; Mason et al. 2021). In short, “feelings” are affect while “propositions” are attitudes (Clore and Schiller 2018; Wilson-Mendenhall and Barsalou 2018). There was a small set of questions excluded at this final stage: what we might label non-affective (or not primarily affective) evaluations. This included things like how the respondent thought the economy was doing, their vote choice, or their ideological placement of the parties. These are questions which are neither directly affective, nor grounded in (relatively) consistent attitudes; rather, they are judgments made on the basis of those attitudes (and affects).

The resulting set of items compares reasonably well to previous applications of BNA. Boutyline and Vaisey’s original work included 46 total items, a few less than the number of attitudinal items I’m using here, but almost double the number of affective items. And, indeed, it should be noted that the relative paucity of affective items is certainly a limitation here. However, it must also be noted that items which are available provide decent coverage of the political and affective field and the real question for the method is whether the empirically observed network provides sufficient coverage of the actual belief system. Now, both attitudinal and affective items are dwarfed by the scale of DellaPosta’s (2020) application of BNA, but his argument specifically required coverage of as wide a network as possible (there was no possibility of “irrelevant” beliefs). Mine, on the other hand, is only interested in a more limited system and I believe that 56 attitudinal and 24 affective items, for a total of 80 items in the combined network, is more than sufficient.

4.3.1 Belief Network Measures

I have already said a fair bit about belief network analysis in the course of setting stakes of the analysis, but I will now lay out in more detail the particular measures which will be key to the analysis. First, to reiterate, BNA is a formalization of the idea of belief systems as a network: each survey question is a node, which is connected to each other node with weight equal to the square of their correlation. In comparison of networks which follows, I use a number of different network analytic measure to try and capture different aspects of the empirical expectations laid out above:

- *Density*: the average edge weight, i.e. tie strength. It is a good general measure of how well connected the network is.
- *Centrality*: a rating of how “central” a node is in the network. There are many different definitions, but I am following Boutyline and Vaisey (2017) and using *shortest-path*

Table 4.2: Affective Variables

Short Name	Description
Therm Blacks	GROUP THERMOMETER: Blacks
Therm Whites	GROUP THERMOMETER: Whites
Therm Business	GROUP THERMOMETER: Big Business
Therm Unions	GROUP THERMOMETER: Labor Unions
Therm Lib	GROUP THERMOMETER: Liberals
Therm Con	GROUP THERMOMETER: Conservatives
Therm Dems	GROUP THERMOMETER: Democratic Party
Therm Reps	GROUP THERMOMETER: Republican Party
Therm Poor	GROUP THERMOMETER: Poor People
Therm Gays	GROUP THERMOMETER: Gays and Lesbians
Therm Fem	GROUP THERMOMETER: Feminists
Therm Hisp	GROUP THERMOMETER: Chicanos/Hispanics
Therm Illegal	GROUP THERMOMETER: Illegal Aliens
Pres Angry	CANDIDATE AFFECTS: President- Angry
Pres Proud	CANDIDATE AFFECTS: President- Proud
Dem Cand Angry	CANDIDATE AFFECTS: Democratic Presidential Cand- Angry
Dem Cand Afraid	CANDIDATE AFFECTS: Democratic Presidential Cand- Afraid
Dem Cand Hopeful	CANDIDATE AFFECTS: Democratic Presidential Cand- Hopeful
Dem Cand Proud	CANDIDATE AFFECTS: Democratic Presidential Cand- Proud
Rep Cand Angry	CANDIDATE AFFECTS: Republican Presidential Cand- Angry
Rep Cand Afraid	CANDIDATE AFFECTS: Republican Presidential Cand- Afraid
Rep Cand Hopeful	CANDIDATE AFFECTS: Republican Presidential Cand- Hopeful
Rep Cand Proud	CANDIDATE AFFECTS: Republican Presidential Cand- Proud
Dem Cand Therm	POLITICAL FIGURE THERMOMETER: Democratic Presidential Candidate
Rep Cand Therm	POLITICAL FIGURE THERMOMETER: Republican Presidential Candidate
Pres Therm	POLITICAL FIGURE THERMOMETER: U.S. President
Congress Therm	GROUP THERMOMETER: Congress

betweenness centrality. I first convert tie weights to tie lengths by taking their inverse such that the strongest ties become the longest. Then, a node’s centrality is simply the proportion of shortest paths between other pairs of nodes which run through it. In this understanding, we might think of centrality as being a kind of bottleneck, the node you go through if you are trying to go anywhere else. Given the assumptions of a center-periphery structure, discussed above, this measure is not only a generic indication of structure, but rather locates the node(s) which play an organizing role in the network.

- *Centralization*: a measure of the degree to which the network has a single, well defined center. It is defined as “the sum of pairwise differences between the centrality of the most central node and the centrality of each other node, all normalized by the maximum possible value such a sum could obtain in any network of M nodes” (Boutyline and Vaisey 2017:1383). For our purposes, a more centralized network more strongly matches the center-periphery structure just discussed.
- Several measures based on partitioning the network into “communities” of nodes which are, in some sense, more closely tied to each other than to the rest of the network. Specifically, I follow DellaPosta (2020) in using the “walktrap” community detection algorithm, which identifies communities based on random walks in the network following the intuition that short random walks will tend to remain trapped densely connected communities rather than moving between them. Based on this, I consider:
 - The number of communities, i.e. the number of distinct groups of nodes.
 - The proportion of nodes in the largest and largest two communities, which gives a sense of the consolidation of the network.
 - *Modularity*: a measure of how clustered the network is, defined as the proportion of edges within clusters (communities) compared to the proportion one would expect if the same total edges were distributed at random.
- Enmeshing of attitudinal and affective networks, measured by comparing the average edge weight of connections among items of the same type to those between different types.

None of these measures, however, would be terribly convincing without some kind of check for robustness to sampling variability. To provide this, I follow previous work in using a nonparametric bootstrap (Boutyline and Vaisey 2017; DellaPosta 2020), generating 2500 bootstrapped replications for each network. To do this, I resampled the original ANES respondents with replacement and computed new networks based on the new samples. This allowed me to also recompute all the aforementioned measures, whose distribution I then used to estimate confidence intervals for the analysis below.

4.4 Findings and Discussion

Before getting into the metrics, however, we should first see what the belief networks themselves actually look like. Though there is no good way to establish the statistical reliability of a particular network visualization, these graphs are nonetheless informative. They help to

contextualize the more statistically robust measures discussed below and allow us to make a fuller sense of the results. My aim here is not merely to present a point-by-point comparison of theoretical expectation and metric result, but rather to attempt to provide a more complete empirical story.

With that said, we can now turn to Figures 4.1, 4.2, 4.3 and 4.4. The first two show the 1992 and 2016 networks side-by-side, first for the affective network (Figure 4.1) and then the attitudinal (4.2). The combined networks are larger and are thus shown separately (1992 in Figure 4.3, 2016 in Figure 4.4) for the sake of readability. All six graphs have been laid out using a force-directed algorithm (specifically ForceAtlas2 in Gephi, see Jacomy et al. 2014 for more), in which nodes repulse and edges attract, placing strongly connected nodes closer together and more weakly connected nodes further apart. A few modifications have been made for the sake of readability and ties with weight less than .0225 (i.e. an absolute correlation between items < .15) have been omitted for the same reason.⁷ The thickness of ties is scaled with edge weight, node size is scaled by centrality, and node color shows community membership. Finally, node names are only displayed for the three most central nodes in each community, again for readability.

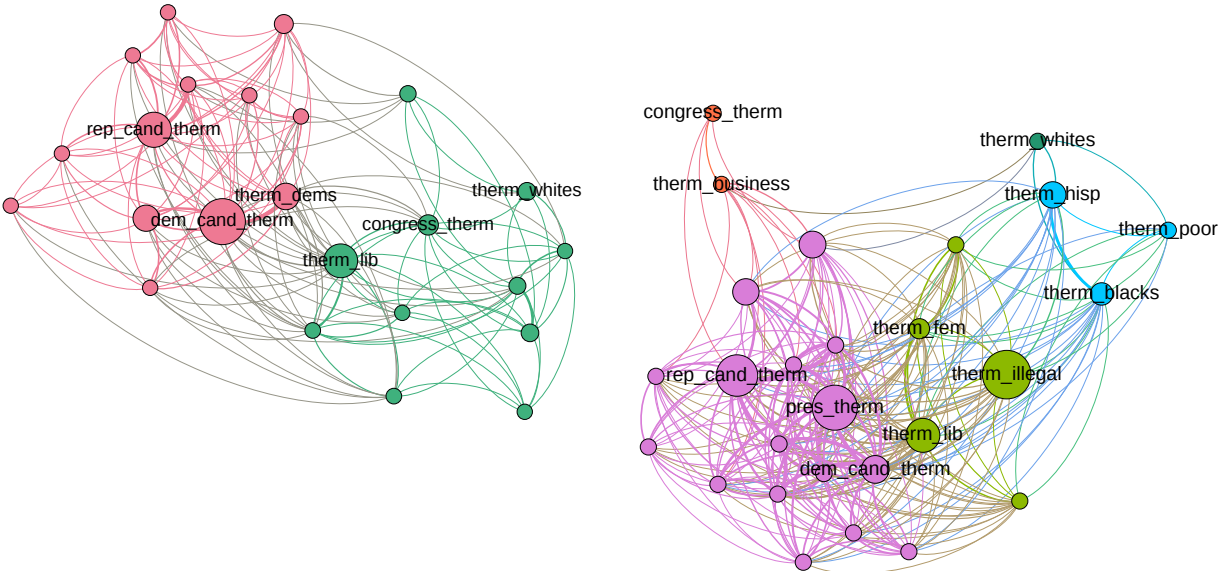


Figure 4.1: The 1992 and 2016 Affective Networks

⁷It should be noted explicitly, however, that all analysis was conducted on the full graphs. The omission of edges was done solely for the generation of these visualizations.

Looking at the networks shows some obvious differences, both across types and over time. In all cases, we see more different communities in 2016 than in 1992, alongside an apparent increase in ties. We also see a universal increase in apparent density, with the 2016 networks consistently having more points closer together, while the 1992 networks show a more even spread. Across types, we see that the affective networks seem simpler than the others, with fewer discernible groups, while the attitudinal and combined networks look more similar.

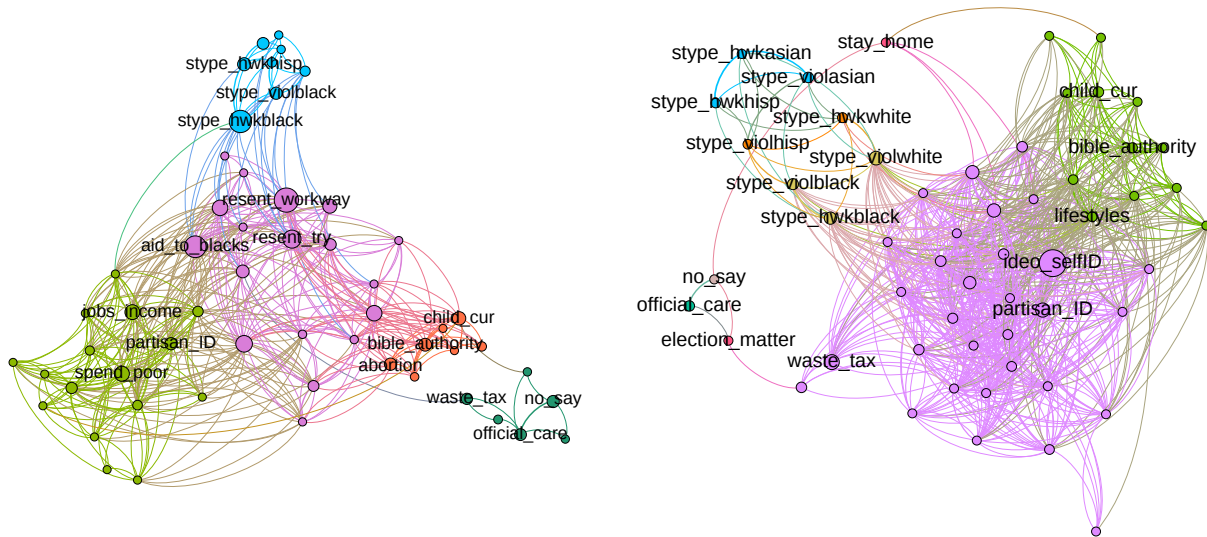


Figure 4.2: The 1992 and 2016 Attitudinal Networks

4.4.1 Network Types over Time

Though they provide useful context, the graphs themselves are not the centerpiece of the analysis presented here. For that, we turn to the network measures described above, which allow us to make statistically robust comparisons and to begin to assemble a meaningful story of these different belief networks and how they have changed over time. The first set of measures are shown in Figure 4.5, including the overall centralization, density, number of communities, and modularity of each network as well as the percentage of nodes in the largest community and the largest two communities. Figure 4.5 plots each of these, including point estimates and 95% confidence intervals for each network,⁸ allowing simple comparisons across time and type.

We start by comparing the attitudinal network and the affective networks, because our first basic question is whether affective items are plausibly similar enough to attitudes to justify

⁸Note that the point estimates displayed here are the estimates generated by taking the mean of the bootstrap replications, rather than the actual value in the sample. The only values where this creates a meaningful difference is for some estimates of the number of communities; however, in no case does it substantively change the analysis. As an aside, this also explains the otherwise strange fractional point estimates for the number of communities.

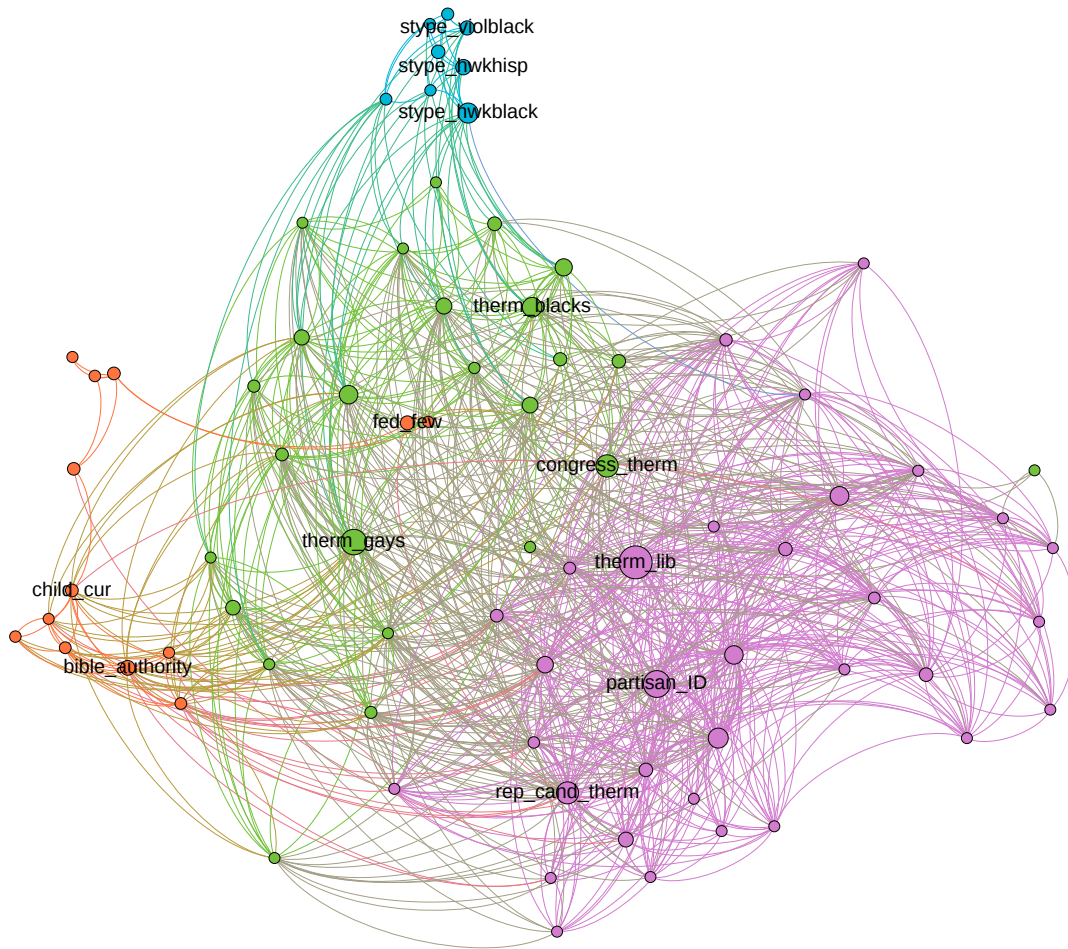


Figure 4.3: The 1992 Combined Network

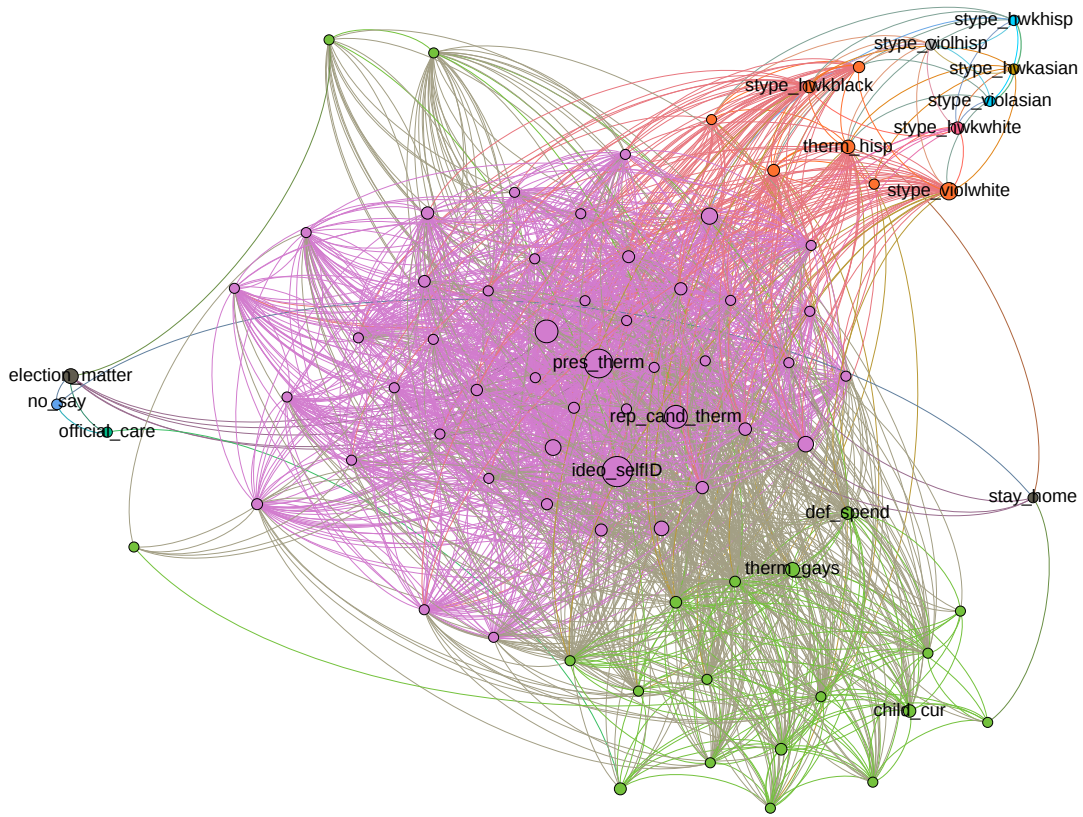


Figure 4.4: The 2016 Combined Network

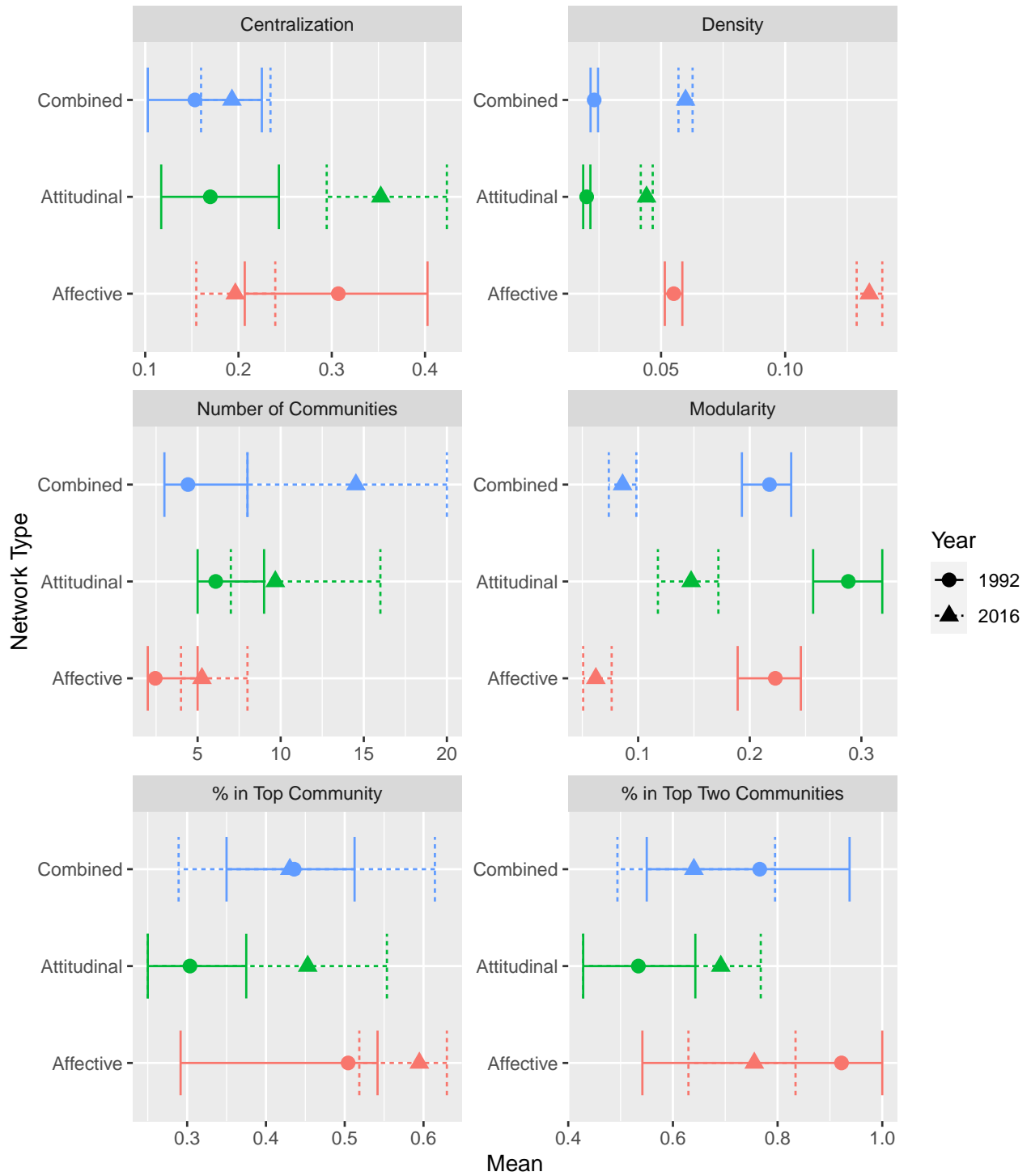


Figure 4.5: Network Type Comparison

their inclusion in belief networks at all. Comparing attitudinal and affective networks then, we see that attitudinal networks in both years have more communities, but a lower percentage of items in their largest communities (though this difference is not always significant). They also have substantially higher modularity and substantially lower density. Perhaps most interestingly, however, the attitudinal network in 1992 has lower centralization than its affective counterpart, but substantially higher centralization in 2016, due both to an increase in centralization in the attitudinal network *and* a decrease in the affective network. We see similar oddities in the direction of change with the percentage of items in two largest communities, where the 1992 affective network starts with a very wide interval which includes 1 and drops in 2016, while the attitudinal network starts lower and rises.

These differences and inconsistencies are interesting, and fit well with our expectation of broad similarity around limited difference. Most of the metrics show some degree of overlap between estimates for the attitudinal and affective networks and, moreover, even those metrics with larger differences (e.g. modularity and density) are not residing in different universes. More importantly, however, the differences (especially those larger differences) are different in the expected direction: the affective network is generally denser than the attitudinal network and has a lower modularity (reflective of overall higher connection across communities).

Furthermore, the differences over time (including the inconsistencies across network types) are clearly explicable in terms of historical processes we know were occurring over the period. Specifically, these comparisons suggest a story of two complementary trends (or, arguably, a single trend with differential impacts). My hypothesis is that these changes represent an overall strengthening of both the attitudinal and affective networks, where the 2016 networks show a public which is both more ideological/ideologically sophisticated and more affectively consistent (i.e. polarized). Basically, in 1992, the attitudinal network is only relatively weakly held together: the combination of low density and relatively low centralization with relatively high modularity (and comparatively lower concentration in larger groups) is indicative of a very diffuse network, where strong connections are relatively rare and, when they do exist, they primarily do so within smaller disconnected groups.⁹ This sort of belief network is consistent with the classic view of the American public's ideological naivete (see, e.g., Converse 1964). By contrast, this has changed significantly by 2016, with the attitudinal network simultaneously becoming much more centralized, doubling in density (though still low in absolute terms), growing in the number of communities, but decreasing in modularity. The decrease in modularity (i.e. groups are getting less dense compared to non-group ties) isn't a sign of a less structured network, but rather of a more strongly connected network overall (see increased density) and a network where the various groups stand in stronger and more meaningful relationship to one another (see increased centralization). Thus, by 2016 we have a belief network more consistent with the findings from DellaPosta (2020) and Kozlowski and Murphy (2021) and likewise consistent with increasing attitudinal polarization in a traditional constraint based sense.

While it might seem odd to argue that the somewhat different changes in the affective

⁹Specifically, these are what we might call "natural" groups, that is, groups where the items are inherently related, e.g., by subject area. So, rather than seeing groups which represent larger coherent ideological structures, the strongest ties are all among items on the same topic. For instance, the ties among questions on the racial resentment scale, or between questions about "lifestyle choices" on the one hand and traditional values on the other. Indeed, 38 of the top 50 ties in the 1992 network are of this kind.

network are consistent with the same basic hypothesis, there are two important things to keep in mind here: the different starting points of the networks in 1992 and, by the same token, the fundamental differences between affect and attitude discussed previously. Specifically, the affective items included here are overall more highly correlated with one another than the attitudinal items, as expected. Based on this, it seems plausible that affects tend more easily towards structure than attitudes.¹⁰

The upshot is that we can almost see the 1992 affective network as being quite similar to the 2016 attitudinal network: it is about as centralized as the 2016 attitude network, with higher modularity and density. In other words, this is a network in which a relatively small number of coherent groups are relatively well connected to one another by some subset of important junctures. By 2016 the network has become radically more dense (a nearly three-fold increase) while its modularity and centralization have declined significantly. Put another way, the overall connected-ness of the graph has increased such that, relatively speaking, the importance of within-group connections and of particular junctions between them has decreased. This doesn't mean that the graph is overall less structured, in the sense that there are fewer groups, that those groups are not meaningfully coherent, or that central nodes no longer matter at all. Indeed, as we'll see momentarily, the 2016 network still has a set of consistently more and consistently less central nodes, as well as perfectly sensible groups of affects. Rather, we might say that the whole affective field has become more coherent, more consistent, such that there is simply less organizational work to be done by these substructures.

Now, to be clear, this explanation of the changes in the attitudinal and affective networks is offered as a plausible and well supported hypothesis. I am not claiming that the network evidence presented here is definitive proof of this narrative. The point, rather, is that this explanation fits into a larger narrative which is supported in other ways; it is consistent with other evidence and, in that consistency, helps both further support this narrative of polarization's development over time and also show the efficacy of these methods.

Turning, then, to the combined networks, we can start to see how attitude and affect fit together. And that really is what we can see in this comparison, because it is an open question. It may be tempting to assume that the combined network should just be an average of the other two: density will be in between the attitudinal and affective networks because density just averages tie strength. But this assumption is, in fact, completely unfounded: the combined networks is not are not just the sum of the affective and attitudinal networks. They are new networks constructed based on the pairwise correlations between all the items in both the attitudinal and affective networks simultaneously. As such, what the networks ultimately look like is entirely dependent on the level of connection *across* the attitude/affect divide. Looking at the combined networks will tell us what that relationship looks like.

With that in mind, we do see an apparent "averaging" with centralization, density, and modularity, in both 1992 and 2016. This sort of "averaging" indicates, at a very basic level, that attitude and affect hang together in at least a minimally coherent system. In other

¹⁰And note that many (though not all) of the attitudinal items have a corresponding affective items (and vice versa). As such, the more general version of the conclusion is perhaps more justified than it at first appears, given that there is at least some degree of equivalence between the attitudinal and affective items and thus some degree to which we can claim to be seeing a comparison between a network of political attitudes towards a set of topics and a network of affects towards those same topics.

words, this is further evidence of the foundational contention that affect and attitude are both constitutive elements in belief systems. The apparent averaging here indicates that attitude and affect interact with each other much as they interact with themselves.

That said, we see something quite different if we look at the number of communities (identified by the community assignment algorithm): in 1992, the combined network divides into a fairly small number of communities (4) which is indeed almost the average of the other two networks ($5 + 2 = 7 / 2 = 3.5$). But in 2016, things look quite different, with the point estimate for the combined network instead being closer to the *sum* of the other two networks (10 communities, the other two have 8 and 5 respectively). In the context of decreasing modularity and increasing density and centralization, this change might at first seem odd, even indicative of decreasing structure and decreasing polarization because of the way it contradicts DellaPosta (2020)’s definition and findings of consolidation in belief networks (which he connected convincingly to his “oil spill model” of polarization).

However, this misses two key points. First, DellaPosta’s understanding of network consolidation and its connection to polarization is contingent on the nature of the network being observed. Specifically, his oil spill model theorizes polarization in terms of the spread of political alignments to more and more beliefs, leading to network consolidation. The networks considered here, due to the nature of the ANES, are composed almost entirely of items which already have (and have had) a political valence. Second, the actual phenomena of interest which consolidation gets at, namely the alignment of more and more conflict around fewer and fewer axes, could also be consistent with the trends observed here, at least in a political network. Comparing the 1992 and 2016 combined networks, what we have, in essence, is more (identifiable) modules, but those tell us relatively less about the structure of the network. On the one hand, this could mean a fragmentation of the network into less meaningful pieces, i.e. less consolidation or alignment of conflict. On the other hand, however, it may be simply a side effect of the overall increasing density: as the whole network comes to be more meaningfully interconnected the identifiable modules, though present, mean less and less in comparison to the baseline connectivity (and alignment) of the network. Indeed, this is one place where the visualizations are helpful: though the 2016 combined network, as shown in 4.4, is divided into ten distinct communities, the visualization also makes clear the wild disparity in the size of these groups. In fact, five of those ten communities have only a single node and two have two. Ultimately, there are fewer communities with more than two nodes in 2016 than in 1992 (three vs. four).

Fortunately, the combined network offers us another metric for evaluating its development which the separate attitudinal and affective networks do not: the degree of integration between item types. That is, comparing levels of connection between items of the same type (e.g. attitudes with attitudes, affects with affects) versus items of different types (e.g. attitudes with affects). In 1992, the average tie strength was .055 between affective items, .02 between attitudinal items, and .019 between items of different types. The average connectivity of the network as a whole increased substantially by 2016 in all three cases, with same-type tie strength averaging .134 for affective items and .044 for attitudinal items. However, the increase for cross-type pairs was proportionally the largest, with average tie strength more than tripling to .058.¹¹

¹¹All these comparisons hold as well if we consider the absolute correlation between items instead of tie

Another way to get a sense of how this enmeshing changed the structure of the network over time is to look at the modularity of the network. However, rather than looking at modularity based on the communities assigned by the walktrap algorithm (as we have thus far), we can treat affective and attitudinal items as communities. Modularity looks at the proportion of edges (or in this case edge weights) within communities versus between communities as compared to the proportion we would expect by chance, in a randomly rewired network with the same distribution of edge weights (Newman 2004). The fact that the modularity decreases from .111 in the 1992 combined network to .06 in the 2016 network demonstrates that the community assignment (in this case of “affect” and “attitude”) tells us much less about the network in 2016, indicating that cross-type ties are playing a bigger role.

What we have, then, is a significant increase in the integration of attitude and affect from 1992 to 2016. Taken together with the other measures, this is strong evidence of a strengthening network over time. What we have is an overall more and more connected network (increasing density), where groups matter less (decreasing modularity), but which is clearly oriented around particular focal points (increasing centralization). And, as we have just shown, this strengthening goes hand in hand with the increasing integration of attitudinal and affective items. This is not only evidence of increasing ideological sophistication and affective consistency (Kozlowski and Murphy (2021)), but of fundamentally strengthening links between those things. In other words, the trends from 1992 to 2016 show a public whose political beliefs are more and more enmeshed in a set of affective responses.

We can get a better sense of what this means qualitatively if we look at what those focal points of organization are, by shifting from the summary network statistics we have so far considered to investigating the centralities of individual items in these networks.

4.4.2 Comparing Centralities

Figures 4.6, 4.7, and 4.8 compare the centrality scores of items in 1992 and 2016 for the affective, attitudinal, and combined networks. These figures show the point estimates and 95% confidence intervals for the *relative* betweenness centrality of each item, or the node’s centrality score divided by the maximum score in the sample. Because the purpose of the comparisons is to see what items were the most reliably central relative to other items, using relative scores in this way produces a more accurate picture than absolute scores would. For instance, if some bootstrap resamples resulted in networks with overall higher centralization, they would widen confidence intervals for absolute scores, even if relative centrality of nodes was the same (see Boutyline and Vaisey 2017 for more). In addition, items with a centrality of .01 or less in both years have been omitted for space reasons.

Looking at the attitudinal and affective networks, we can clearly see the increasing and decreasing (respectively) overall centralization between the 1992 and 2016 networks: in 1992, the attitudinal network shows a confused distribution of centralities with, admittedly, some level of differentiation, but wide confidence intervals with a high degree of overlap. Not only is there no clear center, but, against expectations from Boutyline and Vaisey’s (Boutyline and Vaisey 2017) findings, ideological identity isn’t even particularly close to being that center,

strength (the squared correlation): affect goes from .19 to .31, attitude from .11 to .16, and cross-type from .11 to .20. This is perhaps slightly less disproportionate than the same comparison on tie strength, but it is still disproportionate, with cross-type average correlation nearly doubling.

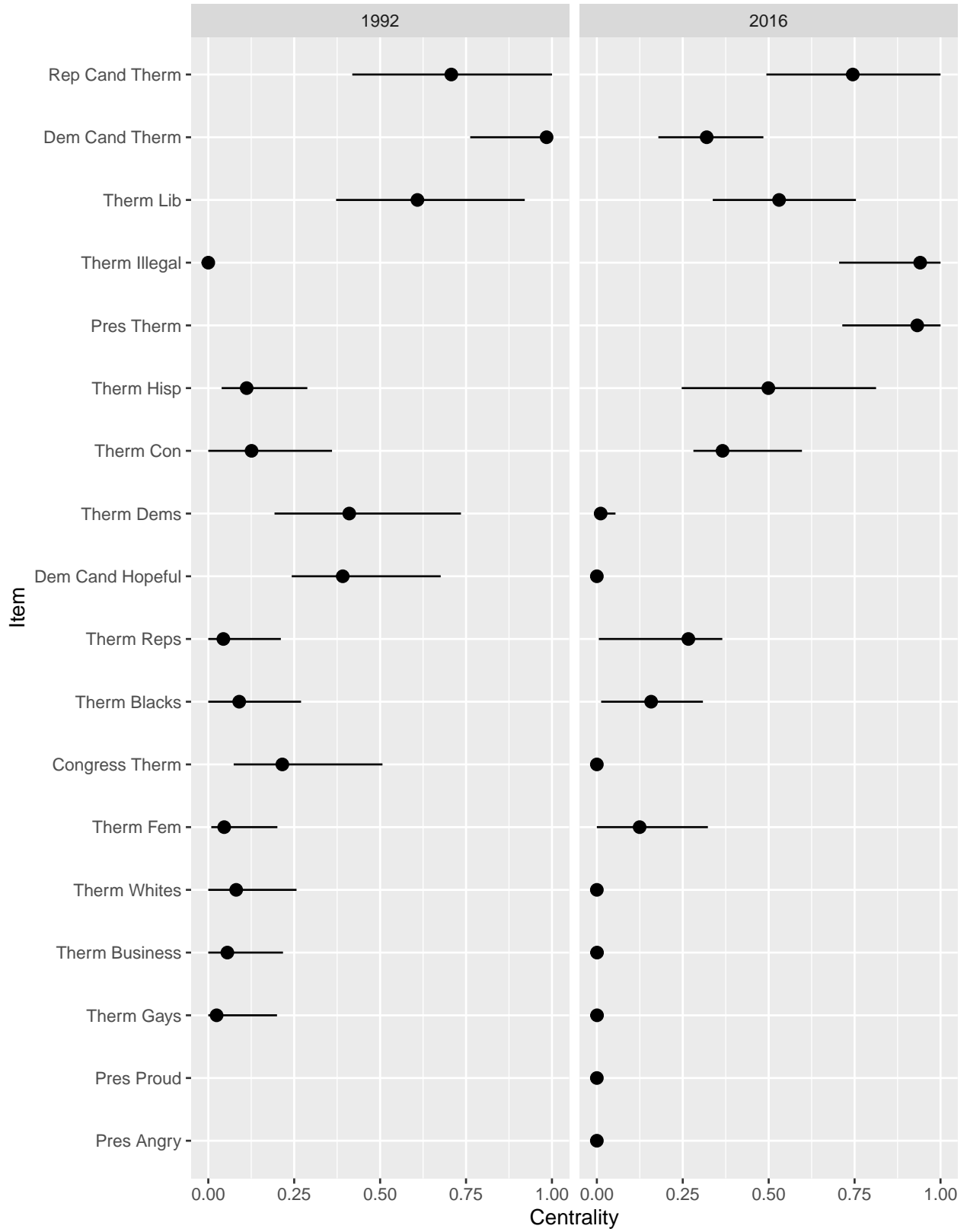


Figure 4.6: Relative Affective Network Centralities

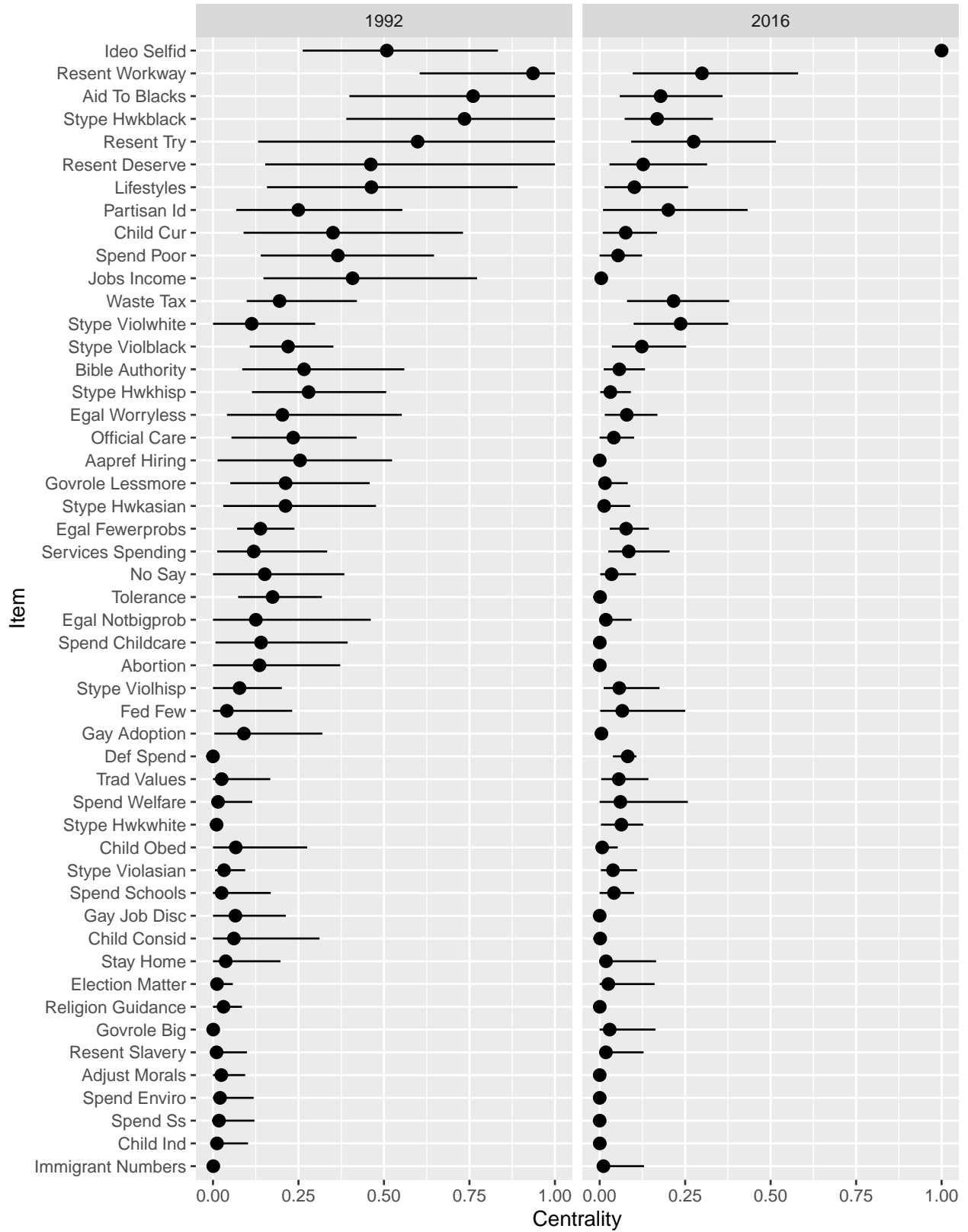


Figure 4.7: Relative Attitudinal Network Centralities

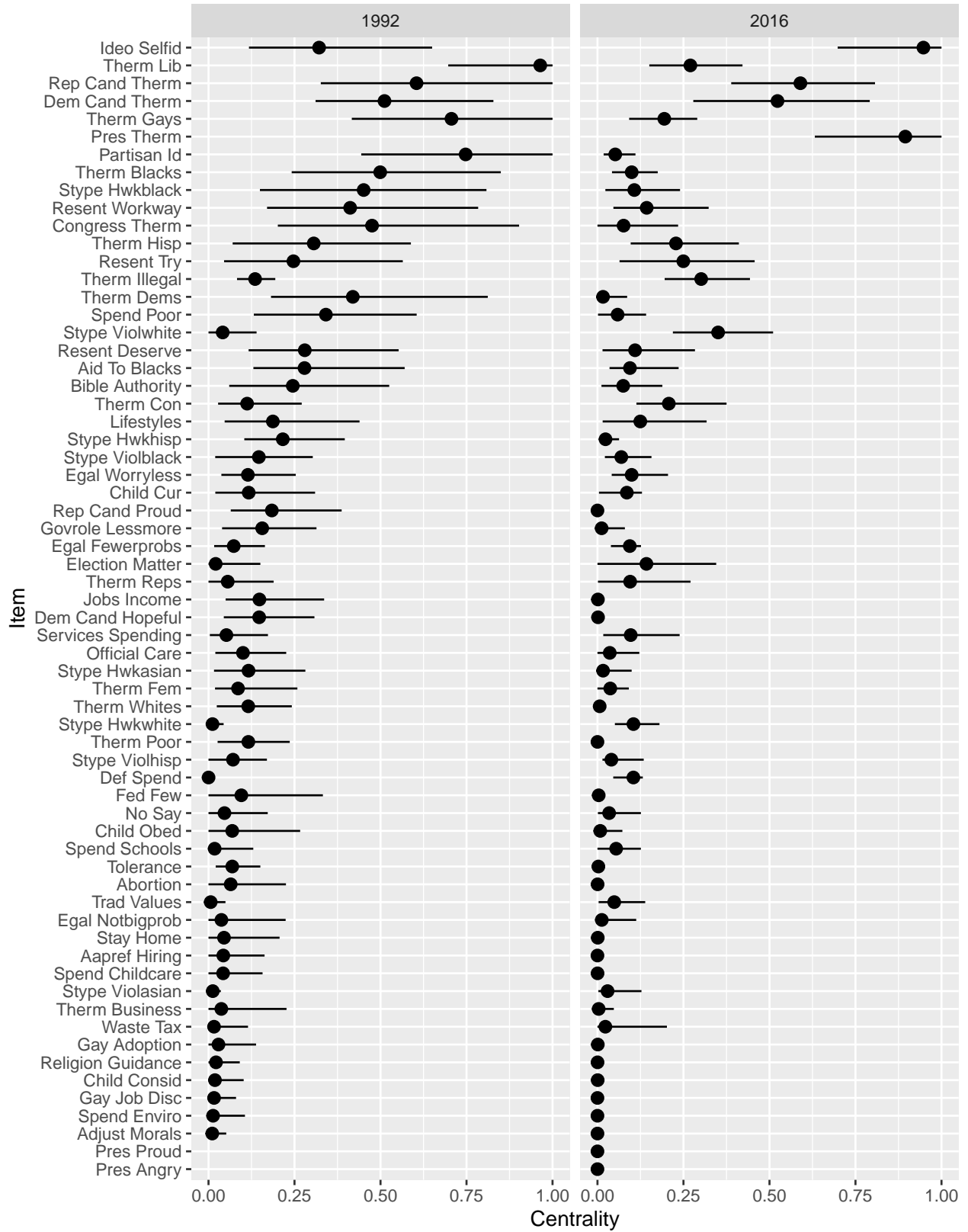


Figure 4.8: Relative Combined Network Centralities

having the highest centrality in just four of the 2500 replications. But the substantial increase in centralization in 2016 comes with an attendant increase in consistency and a tightening of the distribution. There's still a meaningful distribution, but ideological identity has moved decisively to the center of the network. In short, we see here a belief network exactly in line with the network Boutyline and Vaisey observed in the 2000 ANES data. That is, a network fundamentally centered around ideological identity.

Likewise, the affective networks show clear evidence of decreasing centralization. Though in neither year is there a clear and consistent center, the distribution in 1992 shows some level of structure, with a group of candidate, party, and ideology feeling thermometers all appearing at or near the center with some frequency, and with the feeling thermometer for the Democratic candidate a relatively clear favorite for most central. But by 2016, the distribution has widened dramatically, with more items occupying more varied positions across the scale. Most interestingly, perhaps, the set of highest centralities has changed significantly, with feelings towards illegal immigrants as likely as feelings towards the president to occupy the most central position in the network.

That all makes the development of the combined network a bit more surprising. Recall that the combined network saw a small increase in centrality alongside a much more substantial increase in density and decrease in modularity. Given this, we might expect a much more muted change in centralities compared to those we've seen in the attitudinal and affective networks separately. Indeed, this would be even more the case if we imagined the combined network as averaging or addition of the two separate components, each of which experienced opposing changes. However, this is not the case. Rather, we see a transformation almost as substantial as with the attitudinal network alone. In 1992, the combined network shows no real consistent centralized structure, with the centralities of individual nodes varying wildly, point estimates spread heavily across the whole scale and wide confidence intervals. Likewise, though affect towards liberals is clearly the best candidate for the most central node, it has 3 close competitors which overlap substantially both with each other and with five or six further items.

But by 2016, the combined network looks much more like the attitudinal network, with a much tighter distribution of centralities in general and much stronger candidates for the most central node. Specifically, ideological identity and the presidential feeling thermometer stand head and shoulders above any other item, overlapping only with each other (their confidence intervals overlap somewhat with those for the candidate feeling thermometers, but not with any of the point estimates).¹² On its own this result wouldn't necessarily tell us very much: having one attitudinal and one affective option in this way is consistent with a deeply enmeshed combined network, but it could just as easily result from the two separate networks remaining highly distinct, particularly given that the two items were each (one of) the most central in their separate networks.

However, both the previous findings and further investigation clearly show that this is a deeply enmeshed combined attitudinal-affective network. Ideological identity and presidential affect are not two separate centers for two tenuously connected networks, rather they are themselves fairly closely connected and, as is evident in 4.4, play analogous roles within the same central module. Each one is a focal point for connecting one of the other modules (or

¹²This is true regardless of whether we consider the relative or absolute betweenness centralities.

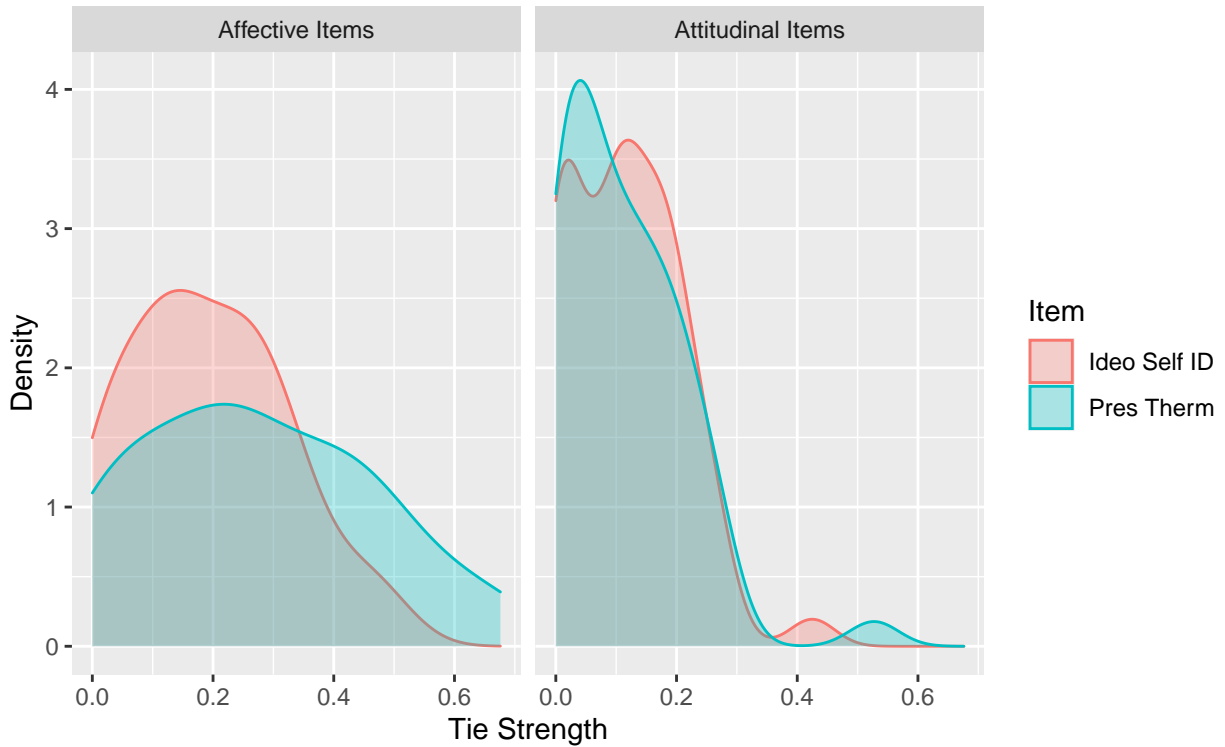


Figure 4.9: Density Plot of Tie Strength, Presidential Feeling Thermometer and Ideological Self Identification

several) to their shared group. And those other modules are not clearly demarcated by item type, rather ideological identity is linked to a “values, religion, spending” module (which, to be clear, includes a number of feeling thermometers) while presidential affect is linked to a “race, racial resentment” module. They are also quite similar in how they are tied to attitudinal and affective items more generally. Looking at Figure 4.9, which shows the distribution of ties for each item, we can see that they are nearly identical with regard to attitudinal items and presidential affect has only a marginal advantage in overall connection to affective items.

4.5 Conclusion

At their most basic, these results show the strengthening of political belief networks from 1992 to 2016, evidence that belief systems became more structured, organized, and consistent, replicating findings from previous work (see, e.g., Boutyline and Vaisey 2017; Kozłowski and Murphy 2021). Previous work, however, has looked only at propositional beliefs or attitudes, treating affect (implicitly and explicitly) as fundamentally different and disconnected from the question of belief systems and their structure. In this paper, I have argued the opposite, that affect must be included in any consideration of belief systems for precisely the same reasons that belief systems are worth considering in the first place. And these results substantiate that argument, not only showing that it is reasonable to include affect in the study of belief systems, but that it is profitable to do so.

In the first place, I have shown that, in both 1992 and 2016, affective networks share fundamental similarities with attitudinal networks. More than that, when attitudes and affects are combined, the resulting networks are not bifurcated messes of topology, but meaningfully combined networks which remain basically comparable to the attitudinal networks examined here (and belief networks examined previously elsewhere) (Boutyline and Vaisey 2017; DellaPosta 2020). In this way, affect and attitude are shown to share a genus, but they still different species, such that the consideration of affect does provide analytical leverage which would not otherwise be available. Specifically, I am able to look at both the degree to and the manner in which the affective and attitudinal networks have become enmeshed with one another.

By doing this, I am able to show that political belief systems not only became more structured, organized, and consistent, but more structured, organized, and consistent in how propositional beliefs and emotions are linked together. Between 1992 and 2016, affect and attitude became more strongly integrated, with disproportionate growth in the strength of cross-type ties and a dramatic reduction in how effectively the division between affect and attitude structured the network. Furthermore, the combined network not only became more centrally organized, it did so in what might be called a “coordinated” fashion, with presidential affect and ideological self identification taking the central position like a pair of binary stars. Ultimately, these results not only show that the political belief system became more ideologically organized, but that that ideological structure was increasingly enmeshed in similarly organized and oriented affective responses.

This finding of increasing integration of political attitudes and affects is important, not only for it tells us methodologically about the study of belief systems, but for what it shows substantively with regard to mass polarization. Polarization, recall, is about the division of society into an “us” and a “them”, locked in existential conflict (McCoy and Somer 2019). Much previous work has, in various ways, pointed to the integration of political affect and attitudes as central to the development and definition of this “us” and “them” (McCoy and Somer 2019; Simonsen and Bonikowski 2022; Somer and McCoy 2019). However, this centrality has largely remained conceptual rather than definitive or operationalized. Indeed, while some of the theoretical work which functions as the foundation of affective and constraint based approaches to polarization draws on this idea of a growing linkage between affect and emotion, affective and constraint based definitions of polarization do not themselves understand it as definitive (Baldassarri and Gelman 2008; Iyengar et al. 2019, 2012; Lelkes 2018). The approach taken here, the use of belief network analysis in order to assess the integration of attitude and affect in a political belief system, offers an approach to polarization which does. And, in so doing, I offer a meaningful synthesis of affective and constraint based approaches.

More than that, this synthesis has implications for the study of pluralism and constraint more broadly. Constraint based definitions of polarization focus on overlapping versus cross-cutting opinion cleavages, drawing on studies of pluralism to argue that crosscutting cleavages help maintain stable democratic functioning, while aligning or overlapping cleavages are indicative of polarization, with a society being truly polarized when all (political) conflict is oriented around a single dividing cleavage (Baldassarri and Gelman 2008; Somer and McCoy 2019). Overlapping versus crosscutting cleavages matter because they ultimately define the shape and nature of political contestation: work on pluralism has pointed out that pluralistic societies are not made successful by a lack of disagreement or conflict, but rather by a kind

of productive and stable conflict (Blau 1974; Dahl 1961; Durkheim 1984; Lipset and Rokkan 1967). And crosscutting cleavages are the prerequisite for this kind of conflict because they both enable and enforce compromise. For instance, we may disagree about abortion but agree about welfare expansion, so if we want to work together on welfare then we must be willing to compromise enough on abortion that our disagreement is not existential. However, this understanding of political conflict and compromise leaves a major question open: how are the relevant cleavages determined? Not all opinion cleavages are politically relevant; indeed, it feels slightly absurd to label something like “whether the town should plant oak trees or maple trees” an opinion cleavage at all, at least in comparison to something like “in what circumstances should abortion be legal”. Instead, cleavages (opinion or otherwise) must be made relevant by political actors (de Leon et al. 2015). And if we want a way to understand which cleavages are relevant and what the valence of that relevance is, then the integration of affect attitude is one extremely plausible metric at our disposal.

And this point reveals still broader consequences of the findings presented here, ultimately bringing us back to the study of belief systems in general. The upshot of the argument and findings here is not just that the connection between affect and attitudes is critical for our understanding of polarization, nor merely that affect should be considered when we consider belief systems, but rather that the many and varied interconnections among different affects and attitudes are *central* to how we ought to think about belief systems at all. The whole point of a belief systems perspective is that individual beliefs are not fully meaningful in themselves and that it is their relationship to other beliefs which helps determine their full importance. However, some versions of this insight make this a relatively limited matter of functional interdependence or elective affinities between different beliefs (whose dependence and affinity is perhaps to be understood partly by way of their individual meaning). But the results here indicate that we need to go further: belief systems are structures of meaning whose structures *make* their meaning, in a very real sense.

Finally, then, this points back to an expanded synthesis of affect and constraint, a conception of polarization as a kind of worldview, grounded in this relational theory of meaning. Polarization defined neither by a single affect nor an assemblage of propositional attitudes, but by a constellation of affects and attitudes which hold together in a coherent orientation towards politics which dictates strategies of action and defines the nature of the conflict. But it will require much more work to effectively explore this new approach. What I have done here can only be an introduction. On the one hand, the argument here has been limited by the nature of the data available, this historic undervaluing of affect in survey design and its division from attitudes. And on the other, the argument here is necessarily limited in comparison to the substantive importance of presented by the question of polarization and the breadth of the empirical ground opened by this approach.

Chapter 5

Conclusion

Polarization defines modern American politics and, increasingly, seemingly unrelated aspects of everyday life. This dissertation examines the foundational role partisan differences play in this polarization, thereby exploring the co-constitution of identity, meaning making, and political conflict. Drawing on diverse strands of political and cultural sociology, political science, and quantitative methods, this study contributes to our understanding of the specifically interpretive power of partisanship. It also looks at the constitutive partisan asymmetry in polarization, which is grounded in this interpretive power. Indeed, existing quantitative work has had difficulty capturing partisan asymmetry because it misses polarization's interpretive core. Based on the American National Election Studies, the three substantive chapters provide three different responses to these linked problems. Building from diagnosis to solution, this dissertation makes both empirical and theoretical contributions to the study polarization, to wider sociology, and to the substantive understanding of our fractured political world.

The contribution begins with Chapter 2, which used the empirical question of asymmetric polarization to address a long running issue in the study of affective polarization: whether in-party affect plays a constitutive role or if out-party animus alone is all that is necessary. The question, in other words, is whether polarization is just a matter of how much you hate the other side (see, e.g., Abramowitz and Webster 2016; Finkel et al. 2020; Webster and Abramowitz 2017), or if how much you like your own party also matters (see, e.g., Mason 2015; Yarchi et al. 2020). Thus far, the study of affective polarization has been unable to truly adjudicate this question, with both “net affect” and “negative partisanship” approaches receiving theoretical and empirical support (Iyengar et al. 2012; Mason 2015). Yet, the question has relevance beyond the study of polarization, with potential ramifications for social identity theory, on which affective polarization is based (Brewer 1999; Huddy 2001; Tajfel et al. 1971). My analysis of the development of partisan affect from 1978 to 2016 for Democrats and Republicans produced results consistent with *neither* net affect nor negative partisanship understandings of polarization. Instead, although it did find an asymmetry in partisan affect, surprisingly, it was an asymmetry in favor of Democrats rather than Republicans and it was driven almost entirely by in-party affect. Furthermore, the analysis presented in Chapter 2 also demonstrates a relationship between in-party love and out-party hate for Democrats which develops over time, while there is no such connection for Republicans. There is no easy way to fit these results into current understandings of polarization. At best, these would indicate that Democrats are consistently and substantially more polarized than Republicans.

But this makes no sense in the face of the obvious observed asymmetries in every other arena. Something in the current theories is inadequate.

To be clear, both approaches have made significant contributions to our understanding of polarization; indeed, negative partisanship in particular has demonstrated impressive findings with regard to “polarization related” outcomes, e.g. willingness to discriminate against the out-group, assessments of governmental legitimacy, etc. (see, e.g., Druckman and Levendusky 2019; Iyengar et al. 2019, 2012). The findings in Chapter 2 do not imply that we must throw out this previous work; however, we must think about how to approach it in this new context.

The upshot of that context is this: we must think more deeply about what exactly affect *means*. Until now, in both work on polarization and in much other work in the social identity theory paradigm, there has been a fundamental (implicit) claim of a one-to-one correspondence between in- or out-group affect and identity (or, specifically, polarization). But these anomalous findings undermine that assumption. In other words, the meaning of in- and out-party affect can no longer be taken as fixed. This should push us to think more seriously about historical and institutional context, and how this context may play a role in defining what exactly we should call polarization. Put another way, these results are evidence that we can no longer make do with “institutionally ‘thin’” theories of polarization (Pierson and Schickler 2020).

Chapter 3 responded directly to this challenge by proposing a new *worldview model* of polarization and validating it by using it to show exactly the asymmetry in polarization which the affective approaches investigated in Chapter 2 were blind to. This model combines work on worldviews, relational theories of meaning, and affective polarization in order to try and more directly operationalize the basic conceptual definition of polarization as the division of society into meaningful political sides who see each other in a particular (antagonistic, Manichean) way (McCoy et al. 2018). In this understanding, polarization requires not only in- or out-party like/dislike (for instance), but for that dislike to hold together in a persistent, patterned, way with other judgments about the other (and/or one’s own) party.

Defined in this way, the asymmetric development of polarization from 1978 to 2016 becomes clear: while both Democrats and Republicans polarized, Republicans polarized *more*, with their dislike of Democrats more and more strongly connected to perceptions of threat and feelings of fear. This confirms the expected asymmetry at the mass level which is so obvious at the elite level and thereby validates the worldview model. But the worldview model also uncovers evidence of Republican *dissatisfaction* with their own party, lending support to the claim that polarization should lead to intra-party, as well as inter-party, division (McCoy and Somer 2019:246–47).

The worldview model is a powerful synthesis of sociological advances in the study of attitudinal polarization and the relational nature of meaning (Boutyline and Vaisey 2017; DellaPosta 2020) together with the innovation of affective polarization from political science (Iyengar et al. 2012; Mason 2015). The results of this synthesis make its empirical utility clear, but it is also theoretically innovative. My focus on the link between identity and judgment, grounded in a wide range of sociological work (Bourdieu 1984; de Leon et al. 2015; Hennion 2007; Wimmer 2013), goes beyond previous work with social identity theory or motivated reasoning (Druckman et al. 2021; Druckman et al. 2013; Iyengar et al. 2019; Mason 2015) to lay out a new connection between polarization and identity. In so doing, it shifts the grounding assumptions in the debate on the relative importance of attitudes

versus identity in driving affective polarization (Lelkes 2018; Mason 2018; Rogowski and Sutherland 2016; Webster and Abramowitz 2017) and offers a new elaboration of work on partisan differences in public opinion (see, e.g., Brooks and Manza 2013; Morisi et al. 2019). The theory I have introduced here also introduces the concept of identity as a heuristic to the study of polarization (see, e.g., Boutyline and Vaisey 2017). Finally, in highlighting the importance of the specific content of political identities, as opposed to merely the strength of identification (cf. Mason 2015), the worldview model better grasps the interpretive power of polarization and thereby makes a powerful methodological and theoretical contribution to both the study of polarization and the quantitative study of worldviews more generally.

Chapter 4 expands on that contribution by investigating the worldview model with a novel methodological approach. A worldview or belief system is not simply a set of independent propositional opinions, but rather an integrated system in which the structure of beliefs (the way they relate to one another) fundamentally determines the nature of the system (Boutyline and Vaisey 2017; Converse 1964; DiMaggio 1997; Jost et al. 2009; Martin 2002). Thus, this chapter attempts to address the question of polarization by looking at the whole system (so far as possible given the available data) rather than just one part of it. More than that, however, it goes beyond previous quantitative work on belief systems by affects in the system. Although the role of affect is clear in some of the diverse literature on belief systems (or similar concepts, e.g. cultural schemas) (see, e.g., Bourdieu 1984; Hennion 2007; Wimmer 2013), it has generally been ignored in quantitative, public opinion oriented, work (see, e.g., Boutyline and Vaisey 2017; DellaPosta 2020).

Through a novel application of Boutyline and Vaisey's (2017) belief network analysis, this chapter demonstrated the validity and usefulness of bringing affect in. It replicated previous results from belief network analysis, but now with the inclusion of affective items in the network, thereby showing a basic degree of similarity in the way attitudinal and affective items relate to one another. But it also mobilized the additional analytical leverage provided by the inclusion of affect to ground a new approach to the worldview model, defining it in terms of the level of integration between attitudinal and affective items. In this way, this dissertation also goes beyond current work in polarization, which largely seeks to adjudicate the primacy of attitude or affect in defining or driving polarization (see, e.g., Lelkes 2018; Rogowski and Sutherland 2016; Webster and Abramowitz 2017), offering a meaningful synthesis grounded in work on pluralism and constraint (Baldassarri and Gelman 2008; Blau 1974; Dahl 1961; Durkheim 1984; Lipset and Rokkan 1967).

Although the work presented here goes a long way towards exploring both the empirical fact of partisan asymmetry and the interpretive power of polarization which helps create it, it is also limited in important ways and much remains to be done. The worldview model proposed here is powerful and well validated in its foundations, yet it is still new. The methodologies in both Chapters 3 and 4 need empirical expansion: in the first case, other selections of items should be examined, especially drawing on other measures of affective polarization (i.e. social distance measures or trait batteries instead of feeling thermometers) (Druckman and Levendusky 2019). Belief network analysis may be helpful here as a way of identifying closely connected groups of items, but it also calls for further investigation of its own. In particular, due to practical constraints, the application of belief network analysis in Chapter 4 does not deal with partisan differences, which would be a clear next step. More broadly, this study is centrally limited in that it is purely quantitative and based on

observational survey data. As such, although it can provide good evidence for aggregate trends and theoretical concepts, it cannot fully explore the mechanisms by which those trends or concepts operate. The idea of a feedback loop of dissatisfaction, in particular, deserves further focused investigation. This feedback loop may play a central role in explaining the militant and dangerous political mobilizations discussed in the introduction. The story of Republican radicalization looks, at first glance, to fit the story of the feedback loop perfectly, with waves of insurgency (the Tea Party, Donald Trump) rising to counter perceived external threats because the existing Republican Party is understood to be insufficient to the task. But this is not the only possible explanation and, precisely because it would be so substantively powerful if true, it demands further investigation aimed not only at testing it, but at identifying in more detail the mechanisms by which it functions. For these mechanisms may hold the key to contesting this radicalization and thus to protecting, and perhaps even revitalizing, what remains of our democratic institutions.

References

- Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory*.
- Abramowitz, Alan I. 2010. *The Disappearing Center*. New Haven: Yale University Press.
- Abramowitz, Alan I., and Kyle L. Saunders. 2008. "Is Polarization a Myth?" *Journal of Politics* 70(2):542–55.
- 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* 41:12–22.
- Abramowitz, Alan, and Jennifer McCoy. 2019. "United States: Racial Resentment, Negative Partisanship, and Polarization in Trump's America." *Annals of the American Academy of Political and Social Science* 681(1):137–56.
- Almeida, Paul D. 2018. "The Role of Threat in Collective Action." Pp. 43–62 in *The Wiley Blackwell Companion to Social Movements*, edited by D. A. Snow, S. A. Soule, H. Kriesi, and H. J. McCammon. Oxford: Wiley-Blackwell.
- Alter, Alexandra. 2022. "How a Debut Graphic Memoir Became the Most Banned Book in the Country." *The New York Times: Books*, May 1. Retrieved July 23, 2022 (<https://www.nytimes.com/2022/05/01/books/maia-kobabe-gender-queer-book-ban.html>).
- Alwin, Duane F. 2007. *Margins of Error: A Study of Reliability in Survey Measurement*. Hoboken, NJ: Wiley-Interscience.
- Azab, Marian, and Wayne A. Santoro. 2017. "Rethinking Fear and Protest: Racialized Repression of Arab Americans and the Mobilization Benefits of Being Afraid." *Mobilization* 22(4):473–91.
- Baldassarri, Delia, and Andrew Gelman. 2008. "Partisans Without Constraint: Political Polarization and Trends in American Public Opinion." *American Journal of Sociology* 114(2):408–46.
- Baldassarri, Delia, and Amir Goldberg. 2014. "Neither Ideologues nor Agnostics: Alternative Voters' Belief System in an Age of Partisan Politics." *American Journal of Sociology* 120(1):45–95.
- Banda, Kevin K., and John Cluverius. 2018. "Elite Polarization, Party Extremity, and Affective Polarization." *Electoral Studies* 56:90–101.
- Baron, Jonathan, and John T. Jost. 2019. "False Equivalence: Are Liberals and Conservatives in the United States Equally Biased?" *Perspectives on Psychological Science* 14(2):292–303.
- Barry, Dan, Mike McIntire, and Matthew Rosenberg. 2021. "'Our President Wants Us Here': The Mob That Stormed the Capitol." *The New York Times*, January 9. Retrieved June 8, 2021 (<https://www.nytimes.com/2021/01/09/us/capitol-rioters.html>).
- Bermeo, Nancy. 2019. "Reflections: Can American Democracy Still Be Saved?" *Annals of the American Academy of Political and Social Science* 681(1):228–33.

- Billig, Michael, and Henri Tajfel. 1973. "Social Categorization and Similarity in Intergroup Behaviour." *European Journal of Social Psychology* 3(1):27–52.
- Blau, Peter M. 1974. "Parameters of Social Structure." *American Sociological Review* 39(5):615–35.
- Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge: Harvard University Press.
- Boutyline, Andrei, and Stephen Vaisey. 2017. "Belief Network Analysis: A Relational Approach to Understanding the Structure of Attitudes." *American Journal of Sociology* 122(5):1371–1447.
- Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1):63–82.
- Brennan Center for Justice, and Bipartisan Policy Center. 2021. *Election Officials Under Attack: How to Protect Administrators and Safeguard Democracy*. New York: Brennan Center for Justice.
- Brewer, Marilynn B. 1999. "The Psychology of Prejudice: Ingroup Love and Outgroup Hate?" *Journal of Social Issues* 55(3):429–44.
- Brooks, Clem, and Jeff Manza. 2013. "A Broken Public? Americans' Responses to the Great Recession." *American Sociological Review* 78(5):727–48.
- Carr, Nicole. 2022. "White Parents Rallied to Chase a Black Educator Out of Town. Then, They Followed Her to the Next One." June 16, <https://www.propublica.org/article/georgia-dei-crt-schools-parents>. Retrieved July 23, 2022 (<https://www.propublica.org/article/georgia-dei-crt-schools-parents>).
- Clifford, Scott. 2019. "How Emotional Frames Moralize and Polarize Political Attitudes: Moralizing Effects of Persuasive Frames." *Political Psychology* 40(1):75–91.
- Clore, Gerald L., and Alexander J. Schiller. 2018. "New Light on the Affect-Cognition Connection." in *Handbook of emotions*, edited by L. F. Barrett, M. Lewis, and J. M. Haviland-Jones. New York London: The Guilford Press.
- Converse, Philip. 1964. "The Nature of Belief Systems in Mass Publics." Pp. 206–61 in *Ideology and Discontent*, edited by D. E. Apter. New York: Free Press of Glencoe.
- Crimaldi, Laura, and John Hilliard. 2022. "Masked Men Chanting Name of Neo-Nazi Group Demonstrate Outside JP Venue Hosting Drag Queen Story Hour - The Boston Globe." July 23, <https://www.bostonglobe.com/2022/07/23/metro/masked-men-chanting-name-neo-nazi-group-demonstrate-outside-jp-venue-hosting-drag-queen-story-hour/>. Retrieved July 23, 2022 (<https://www.bostonglobe.com/2022/07/23/metro/masked-men-chanting-name-neo-nazi-group-demonstrate-outside-jp-venue-hosting-drag-queen-story-hour/>).
- Dahl, Robert A. 1961. *Who Governs? Democracy and Power in an American City*. New Haven: Yale University Press.
- de Leon, Cedric, Menali Desai, and Cihan Tuğal. 2015. "Introduction." Pp. 1–36 in *Building Blocs: How Parties Organize Society*. Palo Alto: Stanford University Press.
- DellaPosta, Daniel. 2020. "Pluralistic Collapse: The 'Oil Spill' Model of Mass Opinion Polarization." *American Sociological Review* 85(3):507–36.
- Dias, Nicholas, and Yphtach Lelkes. 2021. "The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identity." *American Journal of Political Science* ajps.12628.
- DiMaggio, Paul. 1997. "Culture and Cognition." *Annual Review of Sociology* 23:263–87.

- DiMaggio, Paul, John H. Evans, and Bethany Bryson. 1996. "Have Americans' Social Attitudes Become More Polarized?" *American Journal of Sociology* 102(3):690–755.
- DiMaggio, Paul, Ramina Sotoudeh, Amir Goldberg, and Hana Shepherd. 2018. "Culture Out of Attitudes: Relationality, Population Heterogeneity and Attitudes Toward Science and Religion in the U.S." *Poetics* 68:31–51.
- Ditto, Peter H., Brittany S. Liu, Cory J. Clark, Sean P. Wojcik, Eric E. Chen, Rebecca H. Grady, Jared B. Celniker, and Joanne F. Zinger. 2019. "At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives." *Perspectives on Psychological Science* 14(2):273–91.
- Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. 2021. "Affective Polarization, Local Contexts and Public Opinion in America." *Nature Human Behaviour* 5(1):28–38.
- Druckman, James N., and Matthew S. Levendusky. 2019. "What Do We Measure When We Measure Affective Polarization?" *Public Opinion Quarterly* 83(1):114–22.
- Druckman, James N., Erik Peterson, and Rune Slothuus. 2013. "How Elite Partisan Polarization Affects Public Opinion Formation." *American Political Science Review* 107(1):57–79.
- Durkheim, Emile. 1984. *The Division of Labor in Society*. Houndmills: MacMillan Press.
- Epstein, Reid J., and Lisa Lerer. 2021. "Rejecting Biden's Win, Rising Republicans Attack Legitimacy of Elections." *The New York Times*, June 7. Retrieved June 21, 2021 (<https://www.nytimes.com/2021/06/07/us/politics/republican-election-fraud-claims.html>).
- Evans, John H. 2003. "Have Americans' Attitudes Become More Polarized? - An Update." *Social Science Quarterly* 84(1):71–90.
- Finkel, Eli J., Christopher A. Bail, Mina Cikara, Peter H. Ditto, Shanto Iyengar, Samara Klar, Lilliana Mason, Mary C. McGrath, Brendan Nyhan, David G. Rand, Linda J. Skitka, Joshua A. Tucker, Jay J. Van Bavel, Cynthia S. Wang, and James N. Druckman. 2020. "Political Sectarianism in America." *Science* 370(6516):533–36.
- Fiorina, Morris P., and Samuel J. Abrams. 2009. *Disconnect: The Breakdown of Representation in American Politics*. Norman: University of Oklahoma Press.
- Fiorina, Morris P., Samuel J. Abrams, and Jeremy C. Pope. 2011. *Culture War? The Myth of a Polarized America*. 3rd ed. Boston: Longman.
- Fischer, Claude S., and Michael Hout. 2006. *Century of Difference: How America Changed in the Last One Hundred Years*. New York: Russel Sage Foundation.
- Fischer, Claude S., and Greggor Mattson. 2009. "Is America Fragmenting?" *Annual Review of Sociology* 35:435–55.
- Gaertner, Samuel L., John F. Dovidio, Phyllis A. Anastasio, Betty A. Bachman, and Mary C. Rust. 1993. "The Common Ingroup Identity Model: Recategorization and the Reduction of Intergroup Bias." *European Review of Social Psychology* 4(1):1–26.
- Garrett, R. Kelly, and Robert M. Bond. 2021. "Conservatives' Susceptibility to Political Misperceptions." *Science Advances* 7(23):eabf1234.
- Goldberg, Amir. 2011. "Mapping Shared Understandings Using Relational Class Analysis: The Case of the Cultural Omnivore Reexamined." *American Journal of Sociology* 116(5):1397–1436.
- Goldberg, Amir, and Sarah K. Stein. 2018. "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation." *American Sociological Review* 83(5):897–932.
- Goldstone, Jack A., and Charles Tilly. 2001. "Threat (and Opportunity): Popular Action

- and State Response in the Dynamics of Contentious Action.” Pp. 179–94 in *Silence and Voice in the Study of Contentious Politics*, edited by R. R. Aminzade, J. A. Goldstone, D. McAdam, E. J. Perry, W. H. Sewell, S. Tarrow, and C. Tilly. Cambridge: Cambridge University Press.
- Goodwin, Jeff. 2011. “Are Protesters Opportunists? Fifty Tests.” Pp. 277–300 in *Contention in Context*, edited by J. Goodwin and J. M. Jasper. Palo Alto: Stanford University Press.
- Gorman, Brandon, and Charles Seguin. 2018. “World Citizens on the Periphery: Threat and Identification with Global Society.” *American Journal of Sociology* 124(3):705–61.
- Grossmann, Matt, and David A. Hopkins. 2015. “Ideological Republicans and Group Interest Democrats: The Asymmetry of American Party Politics.” *Perspectives on Politics* 13(1):119–39.
- Hacker, Jacob S., and Paul Pierson. 2005. *Off Center: The Republican Revolution And the Erosion of American Democracy*. New Haven: Yale University Press.
- Hacker, Jacob S., and Paul Pierson. 2014. “After the “Master Theory”: Downs, Schattschneider, and the Rebirth of Policy-Focused Analysis.” *Perspectives on Politics* 12(3):643–62.
- Hacker, Jacob S., and Paul Pierson. 2020. *Let Them Eat Tweets: How the Right Rules in an Age of Extreme Inequality*. New York: Liveright Publishing Corp.
- Hanmer, Michael J., and Kerem Ozan Kalkan. 2013. “Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models.” *American Journal of Political Science* 57(1):263–77.
- Heinrichs, Audra. 2022. “A Group of Sex Ed Teachers Are Afraid for Their Lives Amid Right-Wing Harassment Campaign.” March 15, <https://jezebel.com/a-group-of-sex-ed-teachers-are-afraid-for-their-lives-a-1848651425>. Retrieved July 23, 2022 (<https://jezebel.com/a-group-of-sex-ed-teachers-are-afraid-for-their-lives-a-1848651425>).
- Helmore, Edward. 2022. “US Drag Queens Stand Their Ground Amid Intimidation by the Far Right.” *The Guardian: World News*, July 4. Retrieved July 23, 2022 (<https://www.theguardian.com/world/2022/jul/04/drag-queens-targeted-us-far-right>).
- Hennion, Antoine. 2007. “Those Things That Hold Us Together: Taste and Sociology.” *Cultural Sociology* 1(1):97–114.
- Hetherington, Marc J. 2009. “Review Article: Putting Polarization in Perspective.” *British Journal of Political Science* 39(2):413–48.
- Hibbing, John R., Kevin B. Smith, and John R. Alford. 2014. “Differences in Negativity Bias Underlie Variations in Political Ideology.” *Behavioral and Brain Sciences* 37(3):297–307.
- Hlavac, Marek. 2018. “Stargazer: Well-Formatted Regression and Summary Statistics Tables.” <https://CRAN.R-project.org/package=stargazer>. Retrieved (<https://CRAN.R-project.org/package=stargazer>).
- Hochschild, Arlie Russel. 2016. *Strangers in Their Own Land: Anger and Mourning on the American Right*. New York: The New Press.
- Huddy, Leonie. 2001. “From Social to Political Identity: A Critical Examination of Social Identity Theory.” *Political Psychology* 22(1):127–56.
- Iyengar, Shanto, and Masha Krupenkin. 2018. “The Strengthening of Partisan Affect.” *Political Psychology* 39:201–18.
- Iyengar, Shanto, Yphtach Lelkes, Matthew S. 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(7):129–46.

- Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76(3):405–31.
- 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.
- Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software" edited by M. R. Muldoon. *PLoS ONE* 9(6):e98679.
- Jardina, Ashley. 2021. "In-Group Love and Out-Group Hate: White Racial Attitudes in Contemporary U.S. Elections." *Political Behavior* 43(4):1535–59.
- Jasper, James M. 2011. "Emotions and Social Movements: Twenty Years of Theory and Research." *Annual Review of Sociology* 37:285–303.
- Jordan, Soren, Clayton Mc Laughlin Webb, and B. Dan Wood. 2014. "The President, Polarization and the Party Platforms, 1944-2012." *The Forum* 12(1):169–89.
- Jost, John T. 2017. "Ideological Asymmetries and the Essence of Political Psychology." *Political Psychology* 38(2):167–208.
- Jost, John T., Christopher M. Federico, and Jaime L. Napier. 2009. "Political Ideology: Its Structure, Functions, and Elective Affinities." *Annual Review of Psychology* 60(1):307–37.
- Jost, John T., Jaime L. Napier, Hulda Thorisdottir, Samuel D. Gosling, Tibor P. Palfai, and Brian Ostafin. 2007. "Are Needs to Manage Uncertainty and Threat Associated With Political Conservatism or Ideological Extremity?" *Personality and Social Psychology Bulletin* 33(7):989–1007.
- Jurecic, Quinta. 2022. "The Great Rage." *The Atlantic*, May 5. Retrieved July 23, 2022 (<https://www.theatlantic.com/ideas/archive/2022/05/us-politics-threats-violence-harassment/629739/>).
- Kahan, Dan M., David A. Hoffman, Donald Braman, Danieli Evans, and Jeffrey J. Rachlinski. 2012. "'They Saw a Protest': Cognitive Illiberalism and the Speech-conduct Distinction." *Stanford Law Review* 64(4):851–906.
- Kozlowski, Austin C., and James P. Murphy. 2021. "Issue Alignment and Partisanship in the American Public: Revisiting the 'Partisans Without Constraint' Thesis." *Social Science Research* 94:102498.
- Ladd, Jonathan McDonald, and Gabriel S. Lenz. 2008. "Reassessing the Role of Anxiety in Vote Choice." *Political Psychology* 29(2):275–96.
- Lauka, Alban, Jennifer McCoy, and Rengin B. Firat. 2018. "Mass Partisan Polarization: Measuring a Relational Concept." *American Behavioral Scientist* 62(1):107–26.
- Lawton, Sophie. 2022. "Right-Wing Media Pushed Alarmist Claims That Title IX Protections for LGBTQ Students Will Mean the End of 'Traditional Christian Values' in Schools." June 15, <https://www.mediamatters.org/right-wing-media-furious-over-biden-administration-effort-to-protect-lgbtq-students>. Retrieved July 23, 2022 (<https://www.mediamatters.org/right-wing-media-furious-over-biden-administration-effort-to-protect-lgbtq-students>).
- Leeper, Thomas J. 2018. "Interpreting Regression Results Using Average Marginal Effects with R's Margins."
- Lelkes, Yphtach. 2016. "Mass Polarization: Manifestations and Measurements." *Public Opinion Quarterly* 80:392–410.
- Lelkes, Yphtach. 2018. "Affective Polarization and Ideological Sorting: A Reciprocal, Albeit Weak, Relationship." *Forum* 16(1):67–79.

- Lewis, Jeffery B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. 2020. "Voteview: Congressional Roll-Call Votes Database."
- Lipset, Seymour M., and Stein Rokkan. 1967. "Cleavage Structures, Party System, and Voter Alignments: An Introduction." Pp. 1–64 in *Party System and Voter Alignments*, edited by S. M. Lipset and S. Rokkan. New York: Free Press.
- Long, J. Scott, and Sarah A. Mustillo. 2018. "Using Predictions and Marginal Effects to Compare Groups in Regression Models for Binary Outcomes." *Sociological Methods & Research* 004912411879937.
- Lu, Yanqin, and Jae Kook Lee. 2019. "Partisan Information Sources and Affective Polarization: Panel Analysis of the Mediating Role of Anger and Fear." *Journalism & Mass Communication Quarterly* 96(3):767–83.
- Mackie, Diane M., Thierry Devos, and Eliot R. Smith. 2000. "Intergroup Emotions: Explaining Offensive Action Tendencies in an Intergroup Context." *Journal of Personality and Social Psychology* 79(4):602–16.
- Mann, Thomas E., and Norman J. Ornstein. 2016. *It's Even Worse Than It Looks: How the American Constitutional System Collided With the New Politics of Extremism*. New and Ex. New York: Basic Books.
- Martin, John Levi. 2002. "Power, Authority, and the Constraint of Belief Systems." *American Journal of Sociology* 107(4):861–904.
- Mason, Lilliana. 2015. "'I Disrespectfully Agree': The Differential Effects of Partisan Sorting on Social and Issue Polarization." *American Journal of Political Science* 59(1):128–45.
- Mason, Lilliana. 2018. "Ideologues Without Issues: The Polarizing Consequences of Ideological Identities." *Public Opinion Quarterly* 82:866–87.
- Mason, Lilliana, Julie Wronski, and John V. Kane. 2021. "Activating Animus: The Uniquely Social Roots of Trump Support." *American Political Science Review* 1–9.
- McCarty, Nolan M. 2019. *Polarization: What Everyone Needs to Know*. New York, NY: Oxford University Press.
- McCoy, Jennifer, Tahmina Rahman, and Murat Somer. 2018. "Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities." *American Behavioral Scientist* 62(1):16–42.
- McCoy, Jennifer, and Murat Somer. 2019. "Toward a Theory of Pernicious Polarization and How It Harms Democracies: Comparative Evidence and Possible Remedies." *The Annals of the American Academy of Political and Social Science* 681(1):234–71.
- Mohr, John W. 1998. "Measuring Meaning Structures." *Annual Review of Sociology* 24(1):345–70.
- Mora, Maria De Jesus, Rodolfo Rodriguez, Alejandro Zermeño, and Paul Almeida. 2018. "Immigrant Rights and Social Movements." *Sociology Compass* 12(8):1–20.
- Morisi, Davide, John T. Jost, and Vishal Singh. 2019. "An Asymmetrical President-in-Power Effect." *American Political Science Review* 113(2):614–20.
- Munro, Geoffrey D., Carrie Weih, and Jeffrey Tsai. 2010. "Motivated Suspicion: Asymmetrical Attributions of the Behavior of Political Ingroup and Outgroup Members." *Basic and Applied Social Psychology* 32(2):173–84.
- Munson, Ziad W. 2008. *The Making of Pro-Life Activists: How Social Movement Mobilization Works*. Chicago: University of Chicago Press.
- Mutz, Diana C. 2018. "Status Threat, Not Economic Hardship, Explains the 2016 Presidential

- Vote.” *Proceedings of the National Academy of Sciences* 115(19):E4330–39.
- Newman, M. E. J. 2004. “Analysis of Weighted Networks.” *Physical Review E* 70(5):056131.
- Pierson, Paul, and Eric Schickler. 2020. “Madison’s Constitution Under Stress: A Developmental Analysis of Political Polarization.” *Annual Review of Political Science* 23:37–58.
- Polletta, Francesca, and Jessica Callahan. 2017. “Deep Stories, Nostalgia Narratives, and Fake News: Storytelling in the Trump Era.” *American Journal of Cultural Sociology* 5(3):392–408.
- Ring, Trudy. 2022. “Drag Queen Story Hour Canceled in N.C. Town Due to Violent Threats.” June 7, <https://www.advocate.com/pride/2022/6/07/drag-queen-story-hour-canceled-nc-town-due-violent-threats>. Retrieved July 23, 2022 (<https://www.advocate.com/pride/2022/6/07/drag-queen-story-hour-canceled-nc-town-due-violent-threats>).
- Rogowski, Jon C., and Joseph L. Sutherland. 2016. “How Ideology Fuels Affective Polarization.” *Political Behavior* 38(2):485–508.
- Salgado, Beck Andrew. 2022. “Followers of Right-Wing Forum Libs of TikTok Harass a Tosa First-Grade Educator for Teaching Pronouns at Past Job.” *Journal Sentinel*, July 20. Retrieved July 23, 2022 (<https://www.jsonline.com/story/communities/west/news/2022/07/20/followers-right-wing-forum-libs-tiktok-harass-new-tosa-teacher/10096346002/>).
- Schwartz, Jeremy. 2022. “A Push to Remove LGBTQ Books in One County Could Signal Rising Partisanship on School Boards.” February 7, <https://www.propublica.org/article/a-push-to-remove-lgbtq-books-in-one-county-could-signal-rising-partisanship-on-school-boards>. Retrieved July 23, 2022 (<https://www.propublica.org/article/a-push-to-remove-lgbtq-books-in-one-county-could-signal-rising-partisanship-on-school-boards>).
- Simonsen, Kristina B., and Bart Bonikowski. 2022. “Moralizing Immigration: Political Framing, Moral Conviction, and Polarization in the United States and Denmark.” *Comparative Political Studies* 001041402110602.
- Sinclair, Barbara. 2006. *Party Wars: Polarization and the Politics of National Policy Making*. Norman: University of Oklahoma Press.
- Skocpol, Theda. 2020. “The Elite and Popular Roots of Contemporary Republican Extremism.” Pp. 3–28 in *Upending American Politics: Polarizing Parties, Ideological Elites, and Citizen Activists from the Tea Party to the Anti-Trump Resistance*, edited by T. Skocpol and C. Tervo. Oxford University Press.
- Smith, Eliot R., Charles R. Seger, and Diane M. Mackie. 2007. “Can Emotions Be Truly Group Level? Evidence Regarding Four Conceptual Criteria.” *Journal of Personality and Social Psychology* 93(3):431–46.
- So, Linda. 2021. “Trump-Inspired Death Threats Are Terrorizing Election Workers.” *Reuters*, June 11. Retrieved July 10, 2021 (<https://www.reuters.com/investigates/special-report/usa-trump-georgia-threats/>).
- Somer, Murat, and Jennifer McCoy. 2019. “Transformations Through Polarizations and Global Threats to Democracy.” *The Annals of the American Academy of Political and Social Science* 681(1):8–22.
- Stavrakakis, Yannis. 2018. “Paradoxes of Polarization: Democracy’s Inherent Division and the (Anti-)Populist Challenge.” *American Behavioral Scientist* 62(1):43–58.
- Tajfel, Henri, M. G. Billig, R. P. Bundy, and Claude Flament. 1971. “Social Categorization and Intergroup Behaviour.” *European Journal of Social Psychology* 1(2):149–78.
- Tajfel, Henri, and John Turner. 1979. “An Integrative Theory of Intergroup Conflict.” in

- The Social psychology of intergroup relations*, edited by W. G. Austin and S. Worchel. Monterey, Calif: Brooks/Cole Pub. Co.
- Tan, Shelly, Youjin Shin, and Danielle Rindler. 2021. “How One of America’s Ugliest Days Unraveled Inside and Outside the Capitol.” *Washington Post*, January 9. Retrieved June 8, 2021 (<https://www.washingtonpost.com/nation/interactive/2021/capitol-insurrection-visual-timeline/>).
- Tensley, Brandon. 2022. “Analysis: Proud Boys Crashed Drag Queen Story Hour at a Local Library. It Was Part of a Wider Movement.” July 21, <https://www.cnn.com/2022/07/21/us/drag-lgbtq-rights-race-deconstructed-newsletter-reaj/index.html>. Retrieved July 23, 2022 (<https://www.cnn.com/2022/07/21/us/drag-lgbtq-rights-race-deconstructed-newsletter-reaj/index.html>).
- Tushnet, Mark V. 2004. “Constitutional Hardball.” *J. Marshall L. Rev* 37:523–53.
- Vaisey, Stephen. 2009. “Motivation and Justification: A Dual-Process Model of Culture in Action.” *American Journal of Sociology* 114(6):1675–715.
- Wagner, John, Felicia Sonmez, Mike DeBonis, Karoun Demirjian, Amy B. Wang, Colby Itkowitz, and Paulina Firozi. 2021. “Pence Declares Biden Winner of the Presidential Election After Congress Finally Counts Electoral Votes.” *Washington Post*, January 7. Retrieved June 8, 2021 (<https://www.washingtonpost.com/politics/2021/01/06/congress-electoral-college-vote-live-updates/>).
- Ward, Dalston G., and Margit Tavits. 2019. “How Partisan Affect Shapes Citizens’ Perception of the Political World.” *Electoral Studies* 60:102045.
- Webster, Steven W., and Alan I. Abramowitz. 2017. “The Ideological Foundations of Affective Polarization in the U.S. Electorate.” *American Politics Research* 45(4):621–47.
- Webster, Steven W., and Bethany Albertson. 2022. “Emotion and Politics: Noncognitive Psychological Biases in Public Opinion.” *Annual Review of Political Science* 25(1):annurev-polisci-051120-105353.
- Willer, Robb, Matthew Feinberg, and Rachel Wetts. 2016. *Threats to Racial Status Promote Tea Party Support Among White Americans*. SSRN Scholarly Paper. ID 2770186. Rochester, NY: Social Science Research Network.
- Williams, Richard. 2012. “Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects.” *The Stata Journal: Promoting Communications on Statistics and Stata* 12(2):308–31.
- Wilson-Mendenhall, Christine D., and Lawrence W. Barsalou. 2018. “A Fundamental Role for Conceptual Processing in Emotion.” in *Handbook of emotions*, edited by L. F. Barrett, M. Lewis, and J. M. Haviland-Jones. New York London: The Guilford Press.
- Wimmer, Andreas. 2013. *Ethnic Boundary Making: Institutions, Power, Networks*. New York: Oxford University Press.
- Wines, Michael. 2020. “Here Are the Threats Terrorizing Election Workers.” *The New York Times*, December 3. Retrieved July 10, 2021 (<https://www.nytimes.com/2020/12/03/us/election-officials-threats-trump.html>).
- Yarchi, Moran, Christian Baden, and Neta Kligler-Vilenchik. 2020. “Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media.” *Political Communication* 1–42.
- Zmigrod, Leor. 2020. “The Role of Cognitive Rigidity in Political Ideologies: Theory, Evidence, and Future Directions.” *Current Opinion in Behavioral Sciences* 34:34–39.

Zmigrod, Leor, Peter Jason Rentfrow, and Trevor W. Robbins. 2020. "The Partisan Mind: Is Extreme Political Partisanship Related to Cognitive Inflexibility?" *Journal of Experimental Psychology: General* 149(3):407–18.

Appendix A

Face Validity Checks for Chapter 2

In addition to the regression strategy deployed in the body of the paper, I also looked at plots of average net affect, in-party affect, and out-party affect over time, by party, in order to assess the basic shape of the trends. These trends are shown in Figures A.1, A.2, and A.3, below. These figures alone do not show strong trends, but they do seem to indicate an overall asymmetry in net affect towards *Democrats*, an asymmetry which is far stronger on in-party affect than on out-party affect, and which arguably seems to grow over time. Again, I am not claiming that these figures alone are strong results; however, they do provide a basic degree of face validity and give context to the regressions in the body of the paper. They lend some credence to the idea that the similar positive results in the regressions are a sharpening of true associations, rather than uncovering trends which exist *only* net of other covariates.

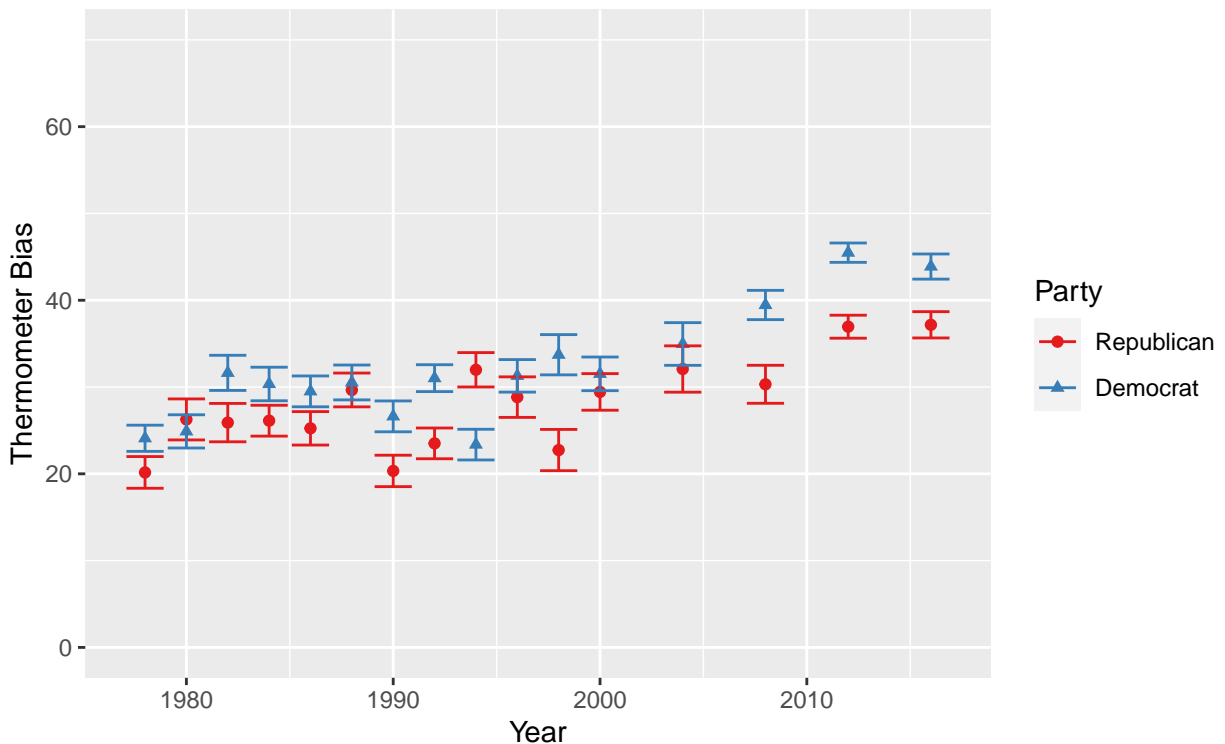


Figure A.1: Mean Thermometer Bias by Party

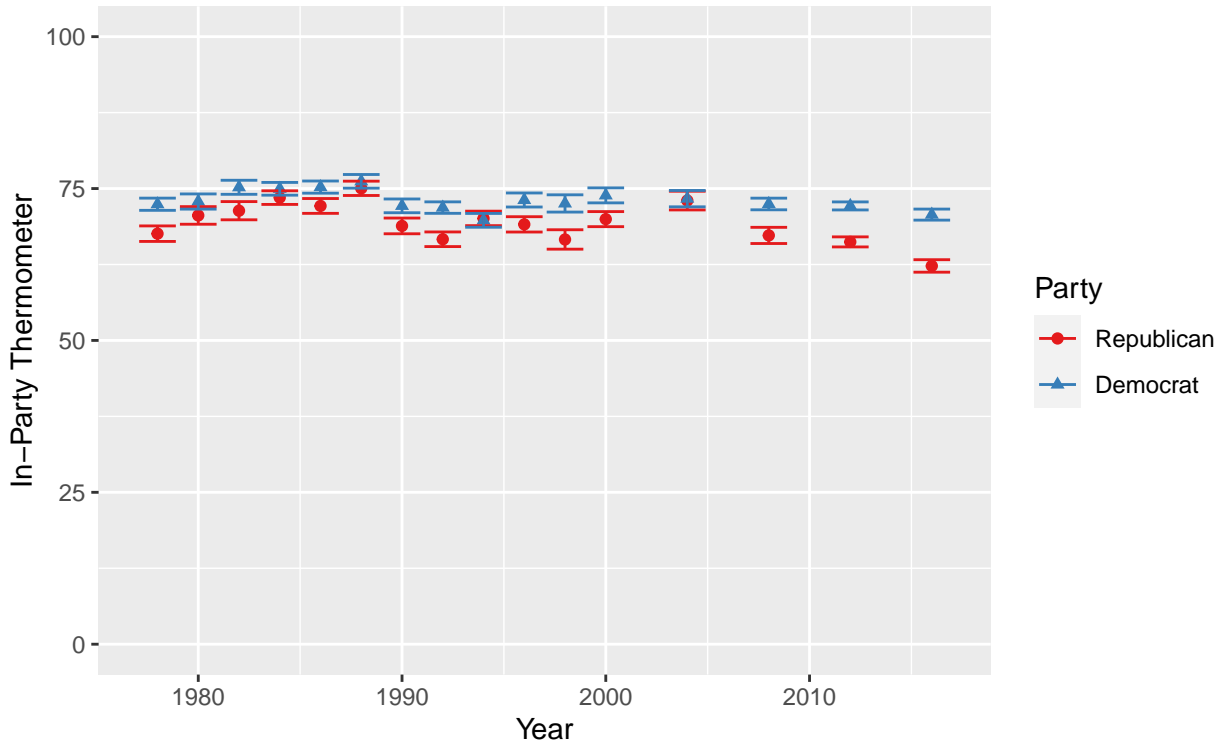


Figure A.2: Mean In-Party Affect by Party

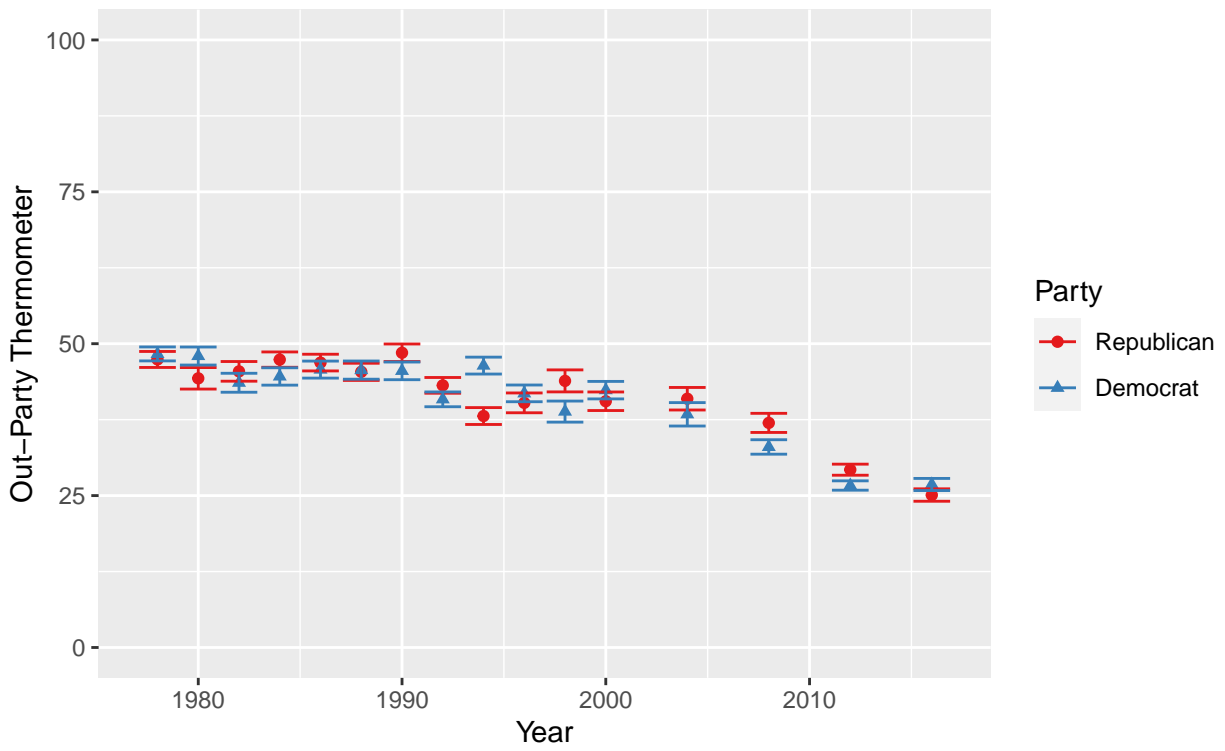


Figure A.3: Mean Out-Party Affect by Party

Appendix B

Putting Trends in Context

The feeling thermometers I use as my dependent variables were first included in 1978, which obviously limits the chronological scope of my analysis. Given what we know about the development of polarization in the US, this not a major issue: 1978 is around when the modern period of increasing polarization kicked off, at least on a mass level. I am able to go beyond that as well, however, through the use of alternative dependent variables which allow me to look at affective polarization from 1952 to 2004. In this way I can put the developments in the more recent period in context and show that they are not continuations of longer term trends with other explanations.

For my alternate measure of affect, I turn to use party likes/dislikes: respondents were asked to mention things they like and dislike about each party, and distinct mentions were counted (up to five).¹ To turn these into measures of in- and out-party affect, the number of dislikes for each party is subtracted from the number of likes, creating an affect score towards each party; for a measure of net affect, the affect towards the out-party is then subtracted from the affect towards the in-party, resulting in a value between -10 and 10.²

First, Figure B.1 shows average like bias over time and Figure B.2 shows average like bias by party. The two things to note are, first, that like bias actually tended downward from 1952 to 1978, at which point it began to increase again. This perfectly lines up with the start of the main dependent variable and its upward trend. Second, as with the thermometer affect by party figure (Figure A.1 in Appendix A), while the simple display of average like bias by party doesn't perfectly correspond to an increasing asymmetry towards Democrats from 1978, it does provide some amount of face validity and also a clear pre/post 1978 difference.

In the same vein, Figures B.3 and B.4 show the marginal effects of Democratic identification on like bias, in-party like affect, and out-party like affect from regressions similar to those in the body of the paper, with the following differences: due variable availability, the models do not include ideological extremity and must use a recoded version of the church attendance variable. In addition, and more importantly, because the expectation is for no particular linear trend in the pre-1978 period, these models use year as an indicator variable in all cases,

¹Although less common than the feeling thermometers I use as my primary dependent variables (due to their chronological limitations), party likes/dislikes have also been used elsewhere as measures of affective polarization (see, e.g., Mason 2015).

²As with the feeling thermometers, these measures are undefined for true independents, as they have no in- or out-party.

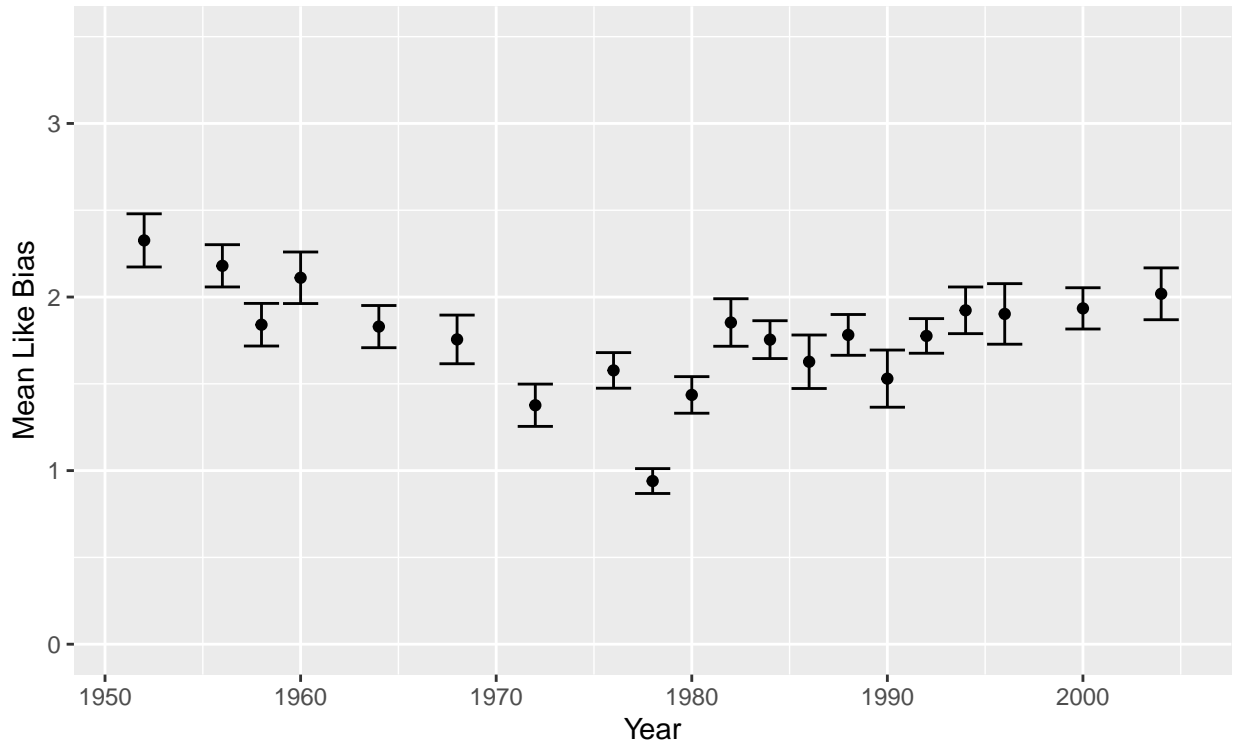


Figure B.1: Mean Like Bias Over Time

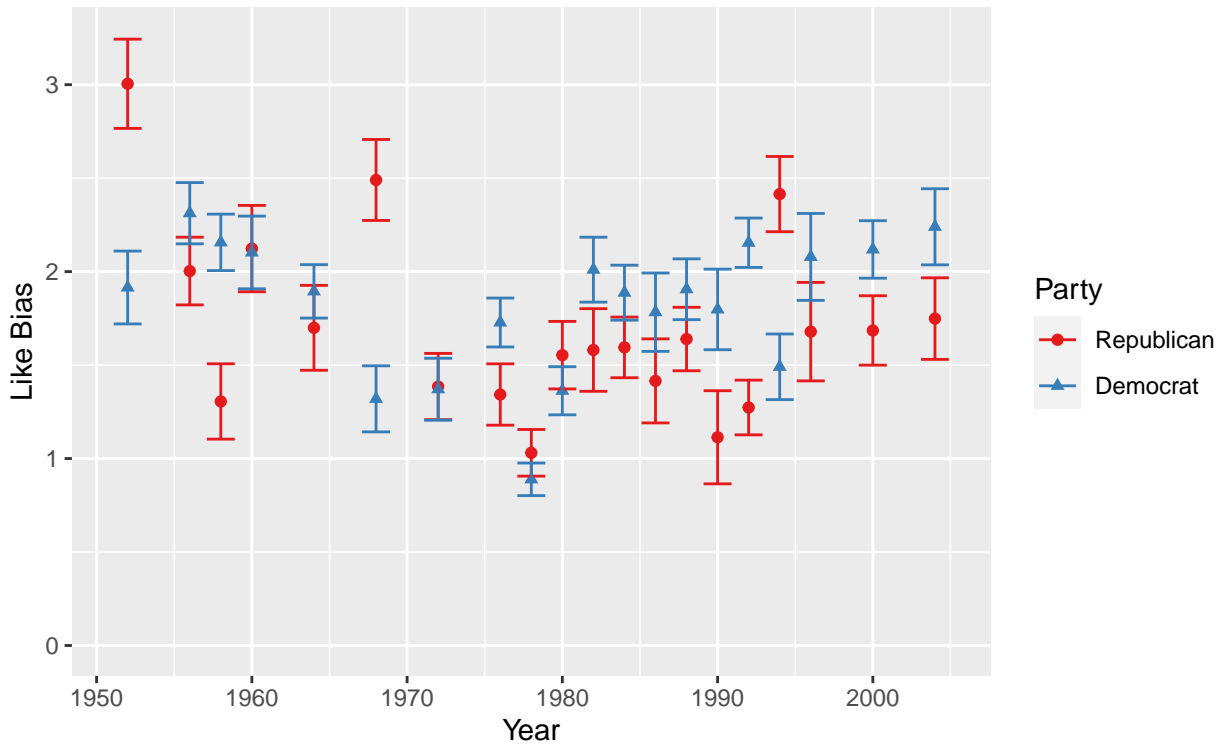


Figure B.2: Mean Like Bias by Party

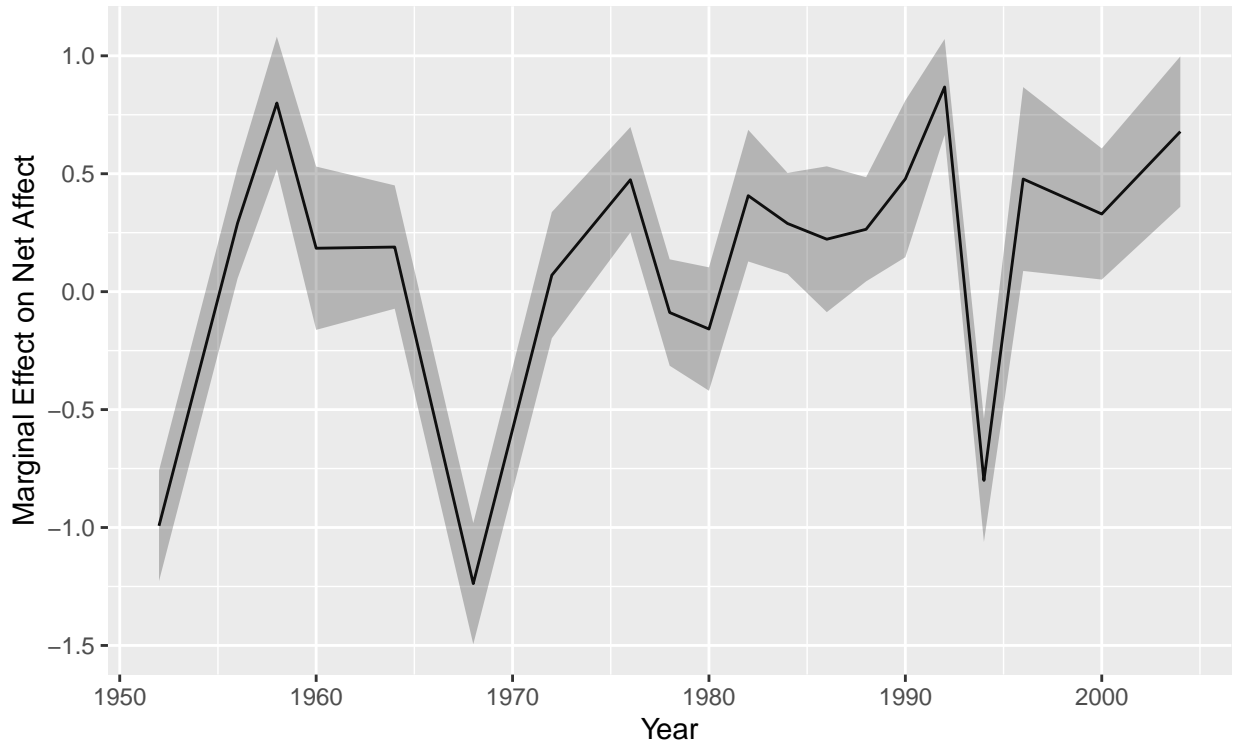


Figure B.3: Marginal Effects of Democratic ID on Like Bias, by Year

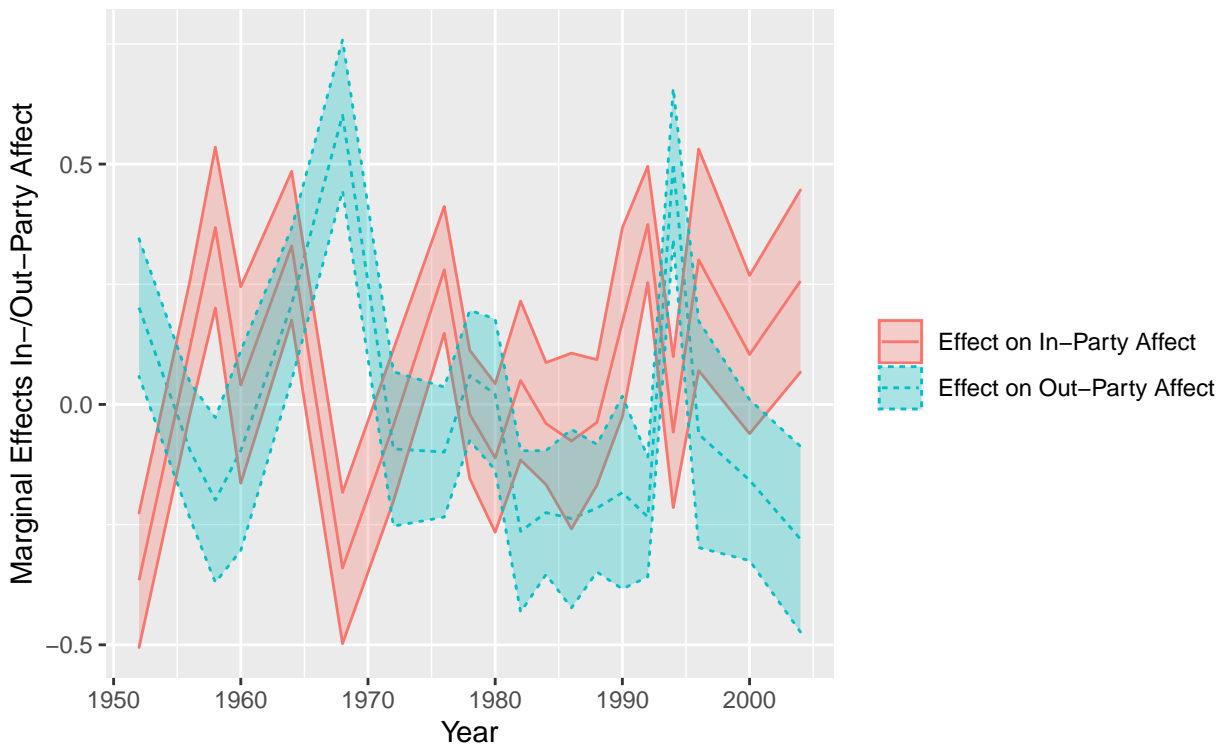


Figure B.4: Marginal Effects of Democratic ID on In- and Out-Party Like Affect, by Year

including in interactions. As such, they allow wild variation in marginal effects from year to year. Accordingly, there are no obvious trends; however, despite lacking the most recent data, we can see the suggestion of trends towards Democrats from around 1978 (with the exception of 1994).

Appendix C

Model Building for Chapter 2

The results in this paper rely upon interaction effects, but a full display and discussion of the comparison with the baseline main effects models proved unwieldy in the body of the paper. Hence, I include the comparison here. Tables C.1, C.2, and C.3 show the baseline main effects models next to the full models with interactions from the body of the paper for each of the three dependent variables (net affect, in-party affect, and out-party affect).

Table C.1: Net Affect Main Effects vs. Full Model Comparison

	<i>Dependent variable:</i>	
	Net Affect	
	(1)	(2)
Democrat	3.387*** (0.339)	-389.147*** (53.378)
Dem x Year		0.197*** (0.027)
Out Therm	13.241*** (0.209)	-90.066** (33.385)
Dem x Out		0.052** (0.017)
Out x Year	5.105*** (0.179)	-183.822*** (28.247)
Dem x Out x Year		0.095*** (0.014)
In Therm	5.876*** (0.526)	5.816*** (0.526)
Dem x In	-0.480** (0.156)	-0.487** (0.156)
In x Year	0.046*** (0.010)	0.046*** (0.010)
Dem x In x Year	1.298*** (0.323)	1.163*** (0.323)
Partisan Strength	0.073 (0.199)	0.047 (0.199)
Partisan Strength x Year	-0.280 (0.351)	-0.376 (0.351)
Ideological Strength	0.586*** (0.105)	0.560*** (0.105)
Observations	26,738	26,738
R ²	0.230	0.233

Note:

*p<0.05; **p<0.01; ***p<0.001
Year fixed effects not displayed.

In all cases, the full models provide a better fit to the data in terms of a higher adjusted R-squared. Furthermore, each of the interactions was shown to significantly contribute to the

Table C.2: In-Party Affect Main Effects vs. Full Model Comparison

	<i>Dependent variable:</i>	
	In-Party Affect	
	(1)	(2)
Democrat	2.313*** (0.211)	-445.394*** (65.856)
Dem x Year		0.225*** (0.033)
Out Therm	-0.011* (0.005)	1.487 (1.241)
Dem x Out		4.854** (1.517)
Out x Year		-0.001 (0.001)
Dem x Out x Year		-0.002** (0.001)
In Therm	7.840*** (0.132)	-42.827* (21.124)
Dem x In		0.025* (0.011)
In x Year	1.003*** (0.113)	-14.257 (17.973)
Dem x In x Year		0.008 (0.009)
Partisan Strength	4.259*** (0.326)	4.127*** (0.327)
Partisan Strength x Year	-0.624*** (0.097)	-0.630*** (0.097)
Ideological Strength	0.047*** (0.006)	0.048*** (0.006)
Ideological Strength x Year	1.402*** (0.200)	1.361*** (0.200)
Black	-1.221*** (0.124)	-1.264*** (0.124)
Income	-1.641*** (0.218)	-1.741*** (0.218)
Age	-0.283*** (0.065)	-0.317*** (0.065)
Observations	26,738	26,738
R ²	0.199	0.204

Note:

*p<0.05; **p<0.01; ***p<0.001
Year fixed effects not displayed.

Table C.3: Out-Party Affect Main Effects vs. Full Model Comparison

	<i>Dependent variable:</i>	
	Out-Party Affect	
	(1)	(2)
Democrat	-1.023*** (0.264)	-41.695 (158.644)
Dem x Year		0.022 (0.079)
Out Therm	-0.017* (0.008)	5.205** (1.715)
Dem x Out		1.640 (2.205)
Out x Year		-0.003** (0.001)
Dem x Out x Year		-0.001 (0.001)
In Therm	-5.209*** (0.173)	-31.185 (28.160)
Dem x In		0.013 (0.014)
In x Year	-4.040*** (0.139)	148.677*** (21.970)
Dem x In x Year		-0.077*** (0.011)
Partisan Strength	-1.527*** (0.409)	-1.444*** (0.412)
Partisan Strength x Year	-0.153 (0.121)	-0.156 (0.121)
Ideological Strength	0.002 (0.007)	0.002 (0.007)
Ideological Strength x Year	0.127 (0.251)	0.195 (0.251)
Black	-1.301*** (0.155)	-1.308*** (0.155)
Income	-1.374*** (0.272)	-1.384*** (0.273)
Age	-0.864*** (0.081)	-0.858*** (0.081)
Observations	26,738	26,738
R ²	0.198	0.201

Note: *p<0.05; **p<0.01; ***p<0.001
Year fixed effects not displayed.

fit of the model based on a series of Wald tests.

Appendix D

Variable Descriptives for Chapter 3

In order to present a clear and efficient summary of descriptive trends for the main variables of interest, I include six figures here. Figures D.1, D.3, and D.5 display mean values for in-/out-party affect,¹ fear of the out-party candidate, and perceived out-party extremity over time by party, while Figures D.2, D.4, and D.6 show the corresponding standard deviations. There are two analytical points worth mentioning here: first, as stated above, the trends here (especially with mean values) are all consistent with polarization, but none provide any evidence of asymmetry. Second, though we see changes in variability over time (sometimes increasing, sometimes decreasing), these trends are *relatively similar* for Democrats and Republicans, providing evidence that the differences in marginal effects are not simply artifacts of different levels of variability in the distributions.

¹This first figure replicates the content of Figures A.2 and A.3 for convenience.

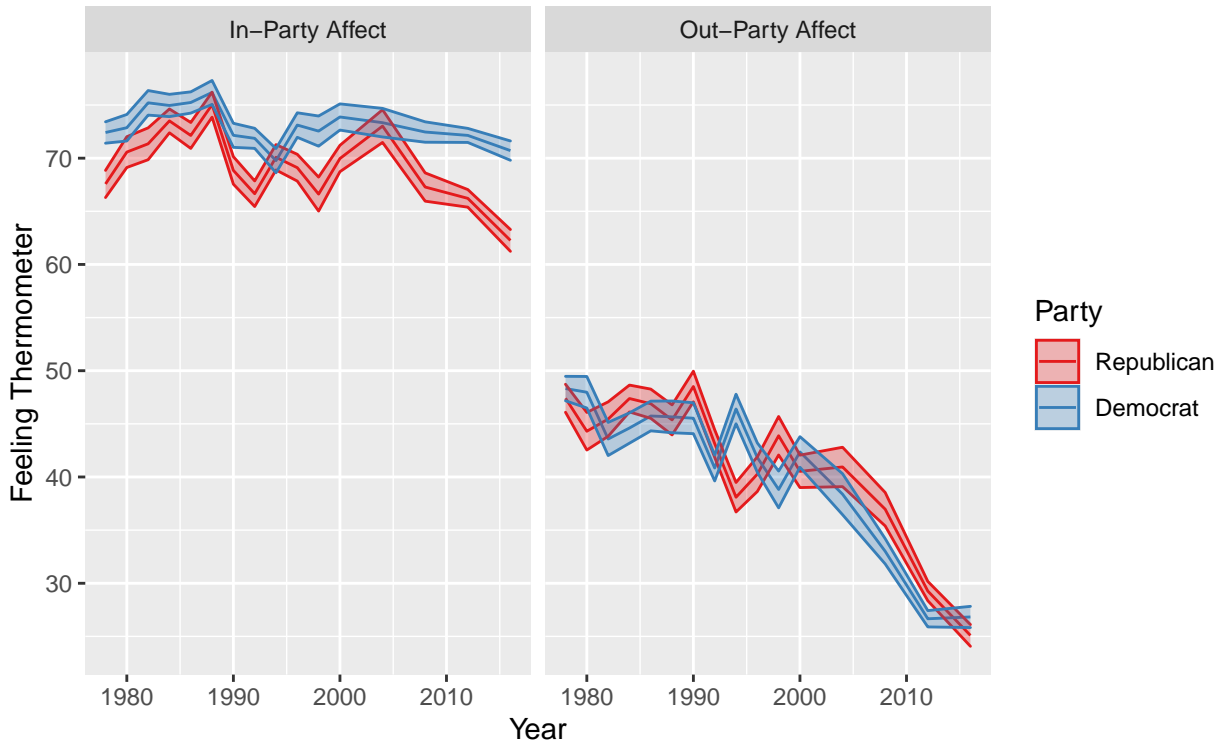


Figure D.1: Mean Feeling Thermometer by Year and Party

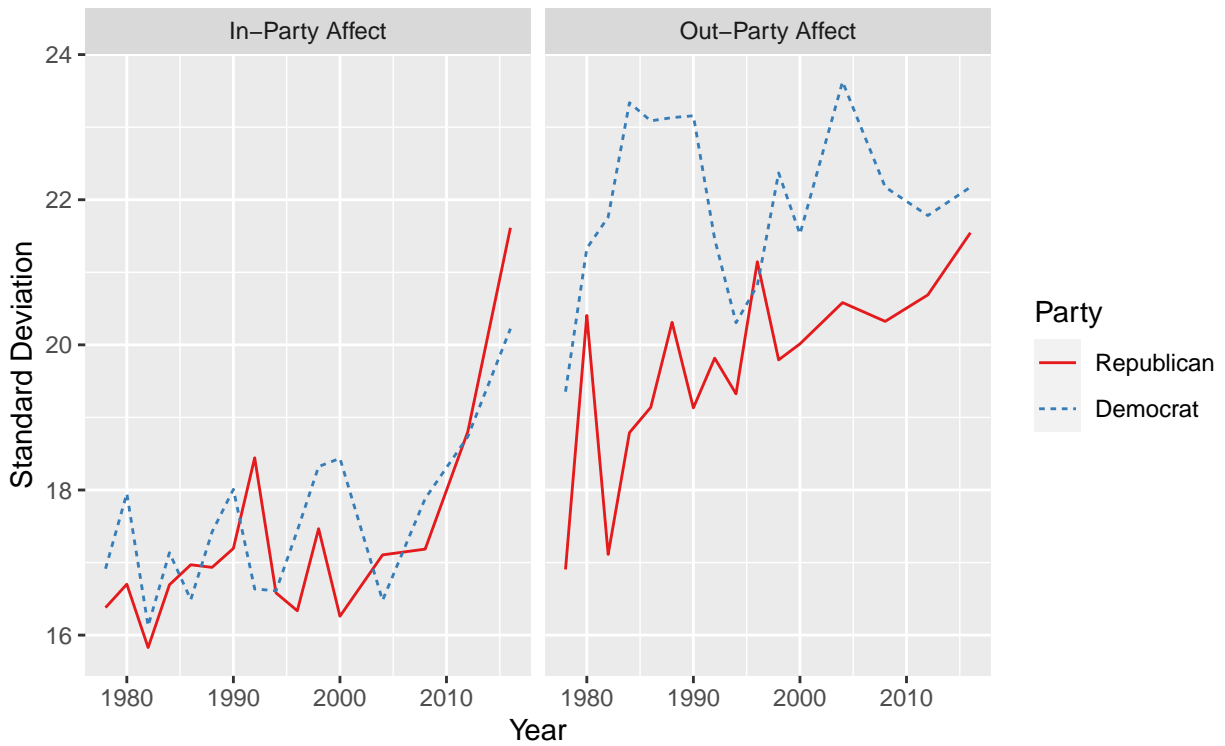


Figure D.2: Feeling Thermometer Standard Deviation by Year and Party



Figure D.3: Proportion Afraid of Out-Party Candidate by Year and Party

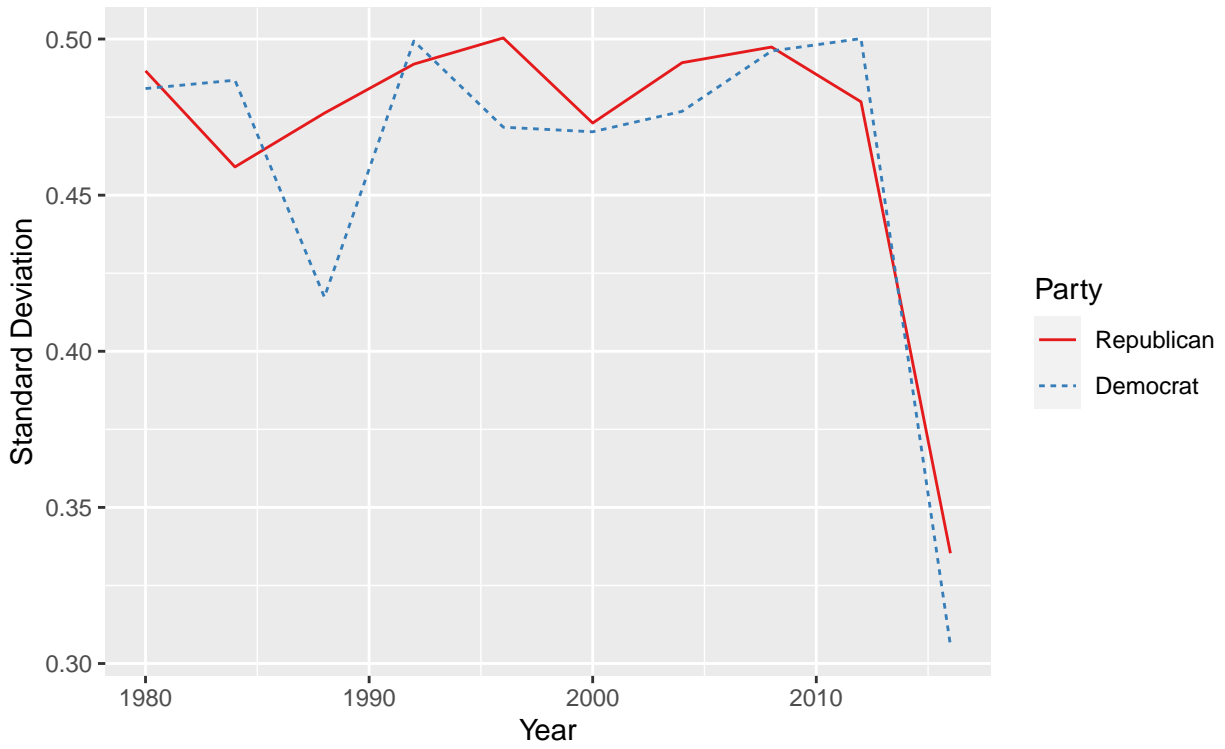


Figure D.4: Standard Deviation of Fear of Out-Party Candidate by Year and Party

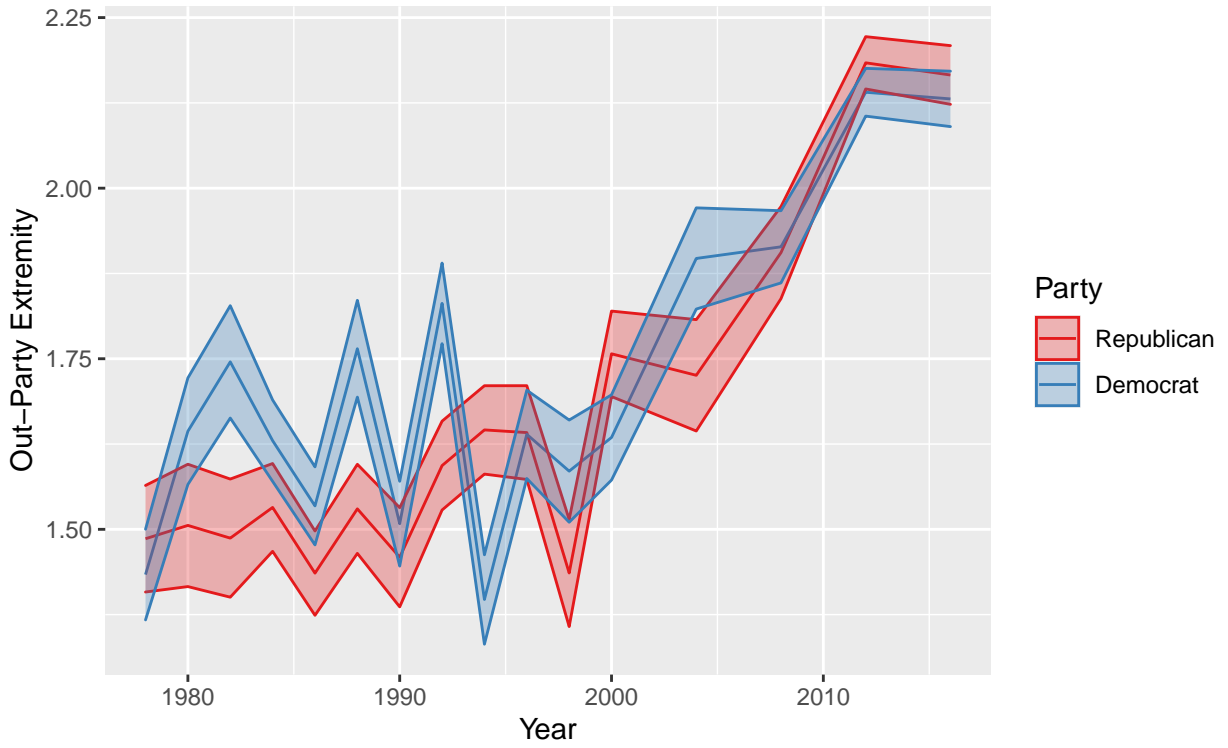


Figure D.5: Mean Perceived Out-Party Extremity by Year and Party

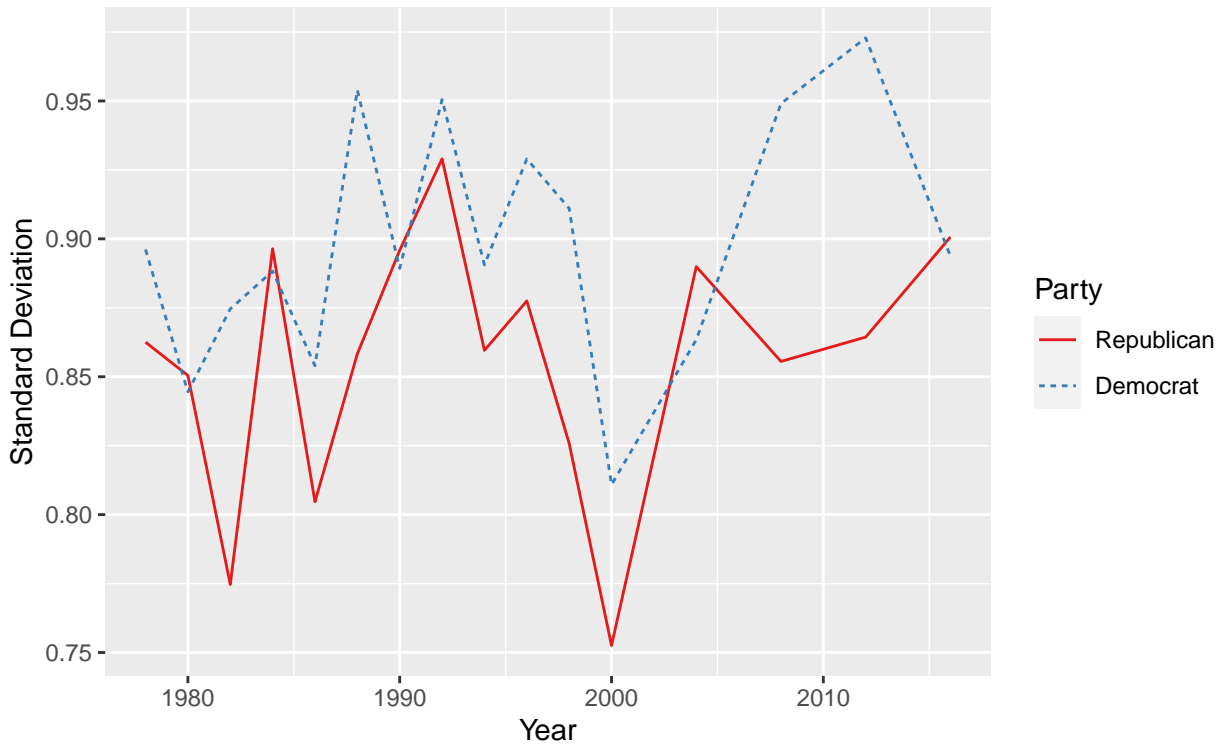


Figure D.6: Standard Deviation of Perceived Out-Party Extremity by Year and Party

Appendix E

Model Building for Chapter 3

The analytical strategy in this paper is based on interaction effects, but including a comparison with the baseline main effects models in the main body of the paper proved unwieldy. Instead, I am including the comparisons here. Tables E.1 and E.2 show the baseline main effects models next to the full models with interactions for each of the four regressions from the body of the paper. Then, Table E.3 shows a summary comparison for all models, including the adjusted R^2 and the results for Wald tests on each set of interaction terms.

In all cases, the models are very similar in terms of explained variance, but, as shown in Table E.3, in every case the full model offers some improvement on adjusted R^2 and the interaction terms are shown to significantly contribute to the fit of the model. The improvements on adjusted R^2 are universally small, but are always present, and the Wald tests clearly show that the addition of the interaction terms does contribute to the model.

Table E.1: Fear Main Effects vs. Full Model Comparison

	<i>Dependent variable:</i>			
	Out-Party Affect		In-Party Affect	
	(1)	(2)	(3)	(4)
Fear	-8.83*** (0.32)	337.16*** (76.10)	2.98*** (0.26)	-124.13* (62.68)
Democrat	-1.31*** (0.31)	393.22*** (72.94)	2.70*** (0.26)	-354.95*** (60.08)
Dem x Year		-0.20*** (0.04)		0.18*** (0.03)
Fear x Dem		-399.75*** (98.86)		-163.96* (81.42)
Fear x Year		-0.17*** (0.04)		0.06* (0.03)
Fear x Dem x Year		0.20*** (0.05)		0.08* (0.04)
Black	-1.64*** (0.49)	-1.62*** (0.49)	4.20*** (0.40)	4.06*** (0.40)
Income	-0.11 (0.14)	-0.11 (0.14)	-0.66*** (0.12)	-0.67*** (0.12)
Age	-0.00 (0.01)	0.00 (0.01)	0.05*** (0.01)	0.05*** (0.01)
Gender	0.73* (0.30)	0.75* (0.30)	1.25*** (0.24)	1.22*** (0.24)
Education	-0.87*** (0.18)	-0.91*** (0.18)	-1.52*** (0.15)	-1.53*** (0.15)
South	-0.82* (0.32)	-0.80* (0.32)	-1.78*** (0.26)	-1.90*** (0.26)
Church				
Attendance	-0.79*** (0.10)	-0.80*** (0.10)	-0.23** (0.08)	-0.25** (0.08)
Partisan				
Strength	-4.96*** (0.19)	-4.95*** (0.19)	7.92*** (0.16)	7.90*** (0.16)
Ideological				
Strength	-3.83*** (0.16)	-3.81*** (0.16)	0.81*** (0.13)	0.82*** (0.13)
Observations	18,592	18,592	18,592	18,592
Adjusted R ²	0.25	0.25	0.21	0.22

Note:

*p<0.05; **p<0.01; ***p<0.001
Year fixed effects not displayed.

Table E.2: Out-Party Perceived Extremity Main Effects vs. Full Model Comparison

	<i>Dependent variable:</i>			
	Out-Party Affect		In-Party Affect	
	(1)	(2)	(3)	(4)
Perceived Extremity	-5.62*** (0.15)	207.38*** (34.85)	1.22*** (0.12)	24.29 (28.74)
Democrat	-0.31 (0.27)	457.72*** (94.79)	2.02*** (0.22)	-240.53** (78.15)
Dem x Year		-0.23*** (0.05)		0.12** (0.04)
Perc Extr x Dem		-159.86*** (45.60)		-80.61* (37.59)
Perc Extr x Year		-0.11*** (0.02)		-0.01 (0.01)
Perc Extr x Dem x Year		0.08*** (0.02)		0.04* (0.02)
Black	-0.85* (0.43)	-0.90* (0.43)	4.31*** (0.35)	4.19*** (0.35)
Income	-0.08 (0.12)	-0.05 (0.12)	-0.74*** (0.10)	-0.76*** (0.10)
Age	0.02* (0.01)	0.02* (0.01)	0.04*** (0.01)	0.04*** (0.01)
Gender	0.42 (0.25)	0.43 (0.25)	1.45*** (0.21)	1.43*** (0.21)
Education	-0.99*** (0.16)	-0.99*** (0.16)	-1.39*** (0.13)	-1.44*** (0.13)
South	-1.44*** (0.28)	-1.43*** (0.28)	-1.78*** (0.23)	-1.89*** (0.23)
Church				
Attendance	-0.81*** (0.08)	-0.82*** (0.08)	-0.28*** (0.07)	-0.31*** (0.07)
Partisan				
Strength	-4.64*** (0.16)	-4.63*** (0.16)	7.68*** (0.14)	7.66*** (0.14)
Ideological				
Strength	-3.53*** (0.14)	-3.47*** (0.14)	0.91*** (0.12)	0.93*** (0.12)
Observations	23,448	23,448	23,448	23,448
Adjusted R ²	0.26	0.26	0.20	0.21

Note:

*p<0.05; **p<0.01; ***p<0.001
Year fixed effects not displayed.

Table E.3: Summary Comparison

IV	DV	Model	Adjusted R ²	Wald Test	
				df	Prob > F
Fear	Out-Party Affect	Main Effects	0.247	20	
		Full Model	0.248	24	0.000
	In-Party Affect	Main Effects	0.209	20	
		Full Model	0.215	24	0.000
Perc Extrem	Out-Party Affect	Main Effects	0.260	26	
		Full Model	0.262	30	0.000
	In-Party Affect	Main Effects	0.204	26	
		Full Model	0.209	30	0.000