Intra-Party Norms and Affective Polarization[*,†]

Samuel Frederick[‡]
Columbia University

This version: September 20, 2024[§]
Click Here for Most Recent Version

Abstract

Antipathy toward members of the opposing political party, or affective polarization, has risen in recent years in the United States. While previous research evaluates the individual-level or intergroup drivers of affective polarization, less work has considered the intragroup social determinants of partisan animosity. Moreover, few studies have explored affective polarization among politicians. In this paper, I examine the extent to which intra-party social norms constrain and shape affective polarization at the mass and politician levels. I focus on norms against the expression of partisan incivility. Using original surveys of elected officials and the American public, I show that there is a strong norm against partisan incivility among politicians, but the norm among the mass public favors partisan incivility. Within my surveys, I embed experiments to evaluate the causal effects of priming this social norm on affective polarization. I find tentative support for the idea that social norms around partisan incivility can influence hostility toward the opposing party. In addition, inparty norms increase warmth toward the inparty, suggesting a potential normative basis for inparty identity. My findings shed light on the social roots of affective polarization, offering a deeper understanding of the causes of partisan animosity in American politics.

Keywords: Social Influence | Social Norms | Affective Polarization | Partisanship

[*]This research was made possible (in part) by a grant from the Rapoport Family Foundation. The statements made and views expressed are solely the responsibility of the author. This paper also benefited from the Partisan Animosity Survey Time Sharing program at the Polarization Research Lab (Iyengar, Lelkes and Westwood, 2024). All studies were reviewed by the Institutional Review Board at Columbia University (AAAU4412 & AAAV1660). The author thanks Donald Green, Shigeo Hirano, Robert Shapiro, Gregory Wawro, and participants at the 2024 APSA Conference and Political Psychology Pre-Conference for their helpful comments.

[†]Pre-registration for the survey of local elected officials can be found at this link:
https://osf.io/bg486/?view_only=6c63aee1f7164f8aa5a43a37cfaacd80
Pre-registration materials for the mass-level survey can be found at the following links:
https://osf.io/3bstm/?view_only=1d58d66d61124fd08043cd1157c18b0e
https://osf.io/s5tre/?view_only=f9155db45d654be1af788a6ef315a6bd

[‡]Ph.D. Candidate, Department of Political Science, Columbia University. sdf2128@columbia.edu.

[§]Please do not circulate without permission from the author.

Introduction

"...when they go low, we go high." - Michelle Obama (2020)

"When they go low, we kick them." - Eric Holder, in Merica (2018)

The quotes above set out competing visions of norms for members of the Democratic Party. Specifically, they define how members of the ingroup behave and *should* behave, and they define the identity of the ingroup (the "we") in *contrast* to the outparty (the "they"). Michelle Obama attempts to define the party norm as one of high-minded civility, even in the face of incivility from the opposing party, while Eric Holder attempts to define the party norm as one that fights fire with fire, incivility with incivility. More broadly, these quotes highlight the extent of animosity between the two parties: the opposing party is presumed to attack, and the inparty is either expected to take the high road or to return the attack.

Partisan animosity is rising in American politics (Iyengar et al., 2019). Those who identify with political parties increasingly hate their political opponents and dislike interacting with them (Huber and Malhotra, 2017; Iyengar, Sood and Lelkes, 2012). Previous research suggests a variety of reasons for this affective polarization. People's identities are becoming more likely to reinforce one another (Mason, 2018). Partisans misperceive members of the outparty (Ahler and Sood, 2018; Stone, 2023). And the lack of contact between opposing partisans does not facilitate empathy (e.g., Kalla and Broockman, 2022; Levendusky and Stecula, 2021; Levendusky, 2023). These existing explanations largely attempt to explain affective polarization using individual psychological factors and intergroup contact. Few studies have explored the role of intragroup norms, like those in the quotes above, in shaping affective polarization. Yet, as Blair et al. (2023) highlight, individuals often have little contact with opposing partisans, meaning that intra-party norms may have more potential to drive individual attitudes than interparty contact. Indeed, with her quote, Michelle Obama explicitly attempts to shape party norms to restrain growing animosity.

Moreover, most research on affective polarization focuses on the mass public. However,

politicians play a crucial role in the American system. Politicians not only create policy, but they also influence public opinion (Broockman and Butler, 2017; Lenz, 2012). Their effect on the public's attitudes extends to affective polarization (Druckman and Levendusky, 2019; Huddy and Yair, 2021; Skyette, 2021) and support for political violence (Kalmoe, 2020; Kalmoe and Mason, 2022). Recent work also finds that politicians are more affectively polarized than the masses (Enders, 2021), but the reasons for affective polarization among politicians are not well understood. Given the centrality of politicians to partisan hostility, a more comprehensive account of affective polarization should take the attitudes of politicians into account.

In this paper, I examine the prevalence of norms that inparty members "go high" among both elites and masses. Using original surveys of local elected officials and the mass public, I measure intra-party norms against partisan incivility. Both politicians and the public believe that norms favor civil but negative partisan rhetoric over partisan incivility, but only politicians in my study believe that an intragroup norm proscribes partisan incivility. In fact, partisan incivility is *supported* by intragroup norms among the mass public. I find tentative support for the hypothesis that a norm against incivility increases expressed warmth toward the opposing party and a norm in favor of incivility decreases expressed warmth toward the opposing party. At the same time, intra-party norms generally appear to increase affective polarization, in part by increasing warmth toward the inparty.

My results contribute first to the literature on affective polarization, suggesting additional avenues for reducing partisan hostility. The focus on intra-party, as opposed to intergroup, social dynamics may prove fruitful for constraining partisan hostility (Blair et al., 2023). I also bring politicians' attitudes into the study of affective polarization, recognizing that politicians have unique motivations and experiences, like re-election, socialization, and interactions with constituents, which can contribute to our understanding of partisan affect. Second, this paper broadens our understanding of partisan identities: particularly among

politicians, group norms appear to play an important role in shaping attitudes toward the inparty. Third, a growing body of work probes the relationship between affective polarization and support for democratic norms (Broockman, Kalla and Westwood, 2022; Finkel et al., 2024; Kingzette et al., 2021; Voelkel et al., 2023), yet, within this area of research, little work has examined whether democratic norms can affect partisan hostility—a question I explore in this paper.

The next section provides an overview of previous literature on social influence, including the ways in which norms can constrain and shape group prejudice. Individuals rely on social groups for information and self-definition (e.g., Festinger, 1954; Turner et al., 1987), and they may behave in specific ways to avoid social disapproval or punishment (Asch, 1956; Bicchieri, 2017). I hypothesize that, much like other social groups, groups of inparty members establish norms which provide informational cues, circumscribe acceptable behavior, and help define individual identities. Through these channels of influence, group norms may play an important role in defining and shaping accepted standards for the expression of partisan affective polarization. The primary focus of my analysis is norms regarding partisan incivility (e.g., a norm that the inparty "goes high"). Previous work has measured approval of incivility but not social norms regarding incivility. Therefore, to better capture norms, I measure individual perceptions of ingroup behavior and ingroup approval of uncivil rhetoric. I hypothesize that norms against incivility influence the expression of partisan animus, reducing expressed hostility toward the outparty. I also expect that people compare their perceptions of inparty and outparty norms, leading to warmer feelings toward the inparty.

To test my hypotheses, I rely on vignette experiments in surveys of the masses and politicians. I randomize whether respondents are asked about a co-partisan who either uses uncivil or civil but negative rhetoric about the opposing party. I then ask how they believe other inparty members would react to the vignette rhetoric, capturing social norms regarding partisan incivility. This design allows me to estimate the effect of incivility on

perceptions of norms, showing that respondents in both samples believe norms favor civil over uncivil rhetoric. Importantly, I also randomize whether respondents are asked about norms around incivility either before or after answering party feeling thermometers. By randomizing question order, I estimate the effect of priming norms regarding partisan incivility on partisan affective polarization, finding that, while norms can influence outparty hostility, they appear to increase warmth toward the inparty and consequently, increase affective polarization.

Social Influence

Traditionally, political scientists have focused on individual psychological factors and intergroup contact as key drivers of affective polarization (e.g., Ahler and Sood, 2018; Frederick, Miller and Green, 2024; Kalla and Broockman, 2022; Mason, 2018; Rossiter and Carlson, 2024), directing much less attention toward *intra*group dynamics. Social psychologists, on the other hand, have long recognized the potential of social groups to shape individual attitudes and behavior. Groups can influence everything from how individuals report perceiving the physical world (Asch, 1955; Sherif, 1965) to their responses to pandemics (Baxter-King et al., 2022; Lipsitz, Pop-Eleches and Robertson, 2023; Wu and Huber, 2021). What's more, individuals often underestimate the impact of social influence on their beliefs and behavior (Nolan et al., 2008), but group influence can nonetheless have significant and long-lasting effects (Ferraro, Miranda and Price, 2011). Both the homogeneity of social networks and the consistent effects of groups on individuals suggest that political scientists should more thoroughly examine the role of social influence in partisan phenomena like affective polarization.

Social groups exert much of their influence through social norms. Bicchieri (2017) defines social norms as beliefs about how others act (empirical or descriptive expectations) and how others expect one to act, possibly including sanctions for violations of these expectations (normative expectations). Three mechanisms have been shown to be important for the influence of norms on behavior: informational, normative, and identity-based mechanisms.

Since the writing of Sherif (1965), scholars have noted that individuals rely on group

information to adapt and orient themselves to new or uncertain situations. Particularly in the realm of subjective opinions and attitudes, individuals strive to be correct but have few external benchmarks for evaluation. As a result, they evaluate the correctness of their opinions through comparison with others in their group (Festinger, 1954). Sherif's (1965) classic study had subjects evaluate the movement of a stationary light in a dark room; in the uncertainty, subjects adapted their perceptions to those of others in the group. Bicchieri and Xiao (2008) manipulate students' empirical expectations by providing information about how others behave in a dictator game; this information about others' behavior influences how students themselves act in the game (see also Sparkman and Walton, 2017). Individuals construct social reality within groups and evaluate the correctness of their beliefs in relation to their group's social reality. Thus, simply learning about how others behave is sufficient to change perceptions and behavior via the information embedded in others' behavior

Much of the influence of groups on behavior comes through normative influence: people believe that certain actions are approved of or disapproved of by their group, and in certain cases, acting contrary to the group's expectations may even result in punishment (Bicchieri, 2017). In one of the original studies of social influence, Asch (1956) placed subjects in groups of confederates and asked them to compare the length of different lines. The confederates were instructed to answer certain questions unanimously and incorrectly. Surprisingly, Asch (1956) found that many subjects answered even objectively verifiable tasks incorrectly to conform with the group. Some subjects who conformed reported not wanting to be perceived in a negative light, suggesting that group influence did not necessarily occur via information about the correct answers but rather through fear of social disapproval. A separate study expanded on the findings of Asch (1956), showing that much of the pressure to conform in Asch's original study came from the physical presence of confederates with the subjects. The study, conducted by Deutsch and Gerard (1955), removed the confederates from the physical presence of the subjects (and thus, removed the element of normative influence and social

disapproval). Without the physical manifestation of normative pressure, subjects were less likely to conform to the incorrect answers of the group. Other studies have confirmed the importance of normative influence for a variety of behaviors (c.f., Bicchieri and Xiao, 2008), including voter turnout (Gerber, Green and Larimer, 2008), partisanship (White and Laird, 2020), policy positions (Levitan and Verhulst, 2016), and energy consumption (Schultz et al., 2007). Alongside the informational elements of social influence, individuals may conform to social norms to avoid drawing social disapproval.

Finally, norms may influence behavior through their relationship to individual identities. According to self-categorization theory, when individuals categorize themselves as part of a group, they assimilate themselves to the group's norms (Turner, 1982; Turner et al., 1987). Adopting the prototype of the group's identity is important both because social identities provide an important basis of self-concept. Social identities contribute to individual self-esteem (Tajfel and Turner, 1986) and help reduce uncertainty through self-definition (Hogg and Mullin, 1999). Consistent with this theory, individuals feel pride when their attitudes match group norms and shame otherwise (Suhay, 2015). Further, individuals appear to conform with group norms more when they identify strongly with the group and when their group identity is more salient (Terry and Hogg, 1996; Toff and Suhay, 2019; White, Hogg and Terry, 2002). Indeed, Goldstein, Cialdini and Griskevicius (2008) present results of a field experiment finding that individuals are most responsive to group norms when the group identity is relevant to the behavior in question: hotel guests are more likely to reuse towels when told that previous guests in their room did so than when told that other members of their gender or other citizens did so.

Overall, then, groups influence individual attitudes and behavior through the establishment of social norms—expectations about the behavior of others and beliefs about how others expect one to act. In this way, groups convey information about social reality; lay out ground rules for expected behavior, including possible sanctions for deviation and rewards

for compliance; and define the behavior inherent in individual social identities. The extent of group influence can range from superficial compliance due to the desire to avoid sanctions to true acceptance of group norms (Kelman, 1961).

Norms and Group Affect

Social influence extends to intergroup affect as well, driving attitudes and behaviors toward outgroups. Intergroup prejudice may not be rooted in any particular conflict *between* groups but is often produced and reproduced *within* groups (Sechrist and Stangor, 2007). In fact, Allport (1954) observed that, in some cases, "it is easier to change group attitudes than individual attitudes" (40) and that changing group norms may, in turn, change individual attitudes. Closely mirroring research on broader social influence, the literature on group norms and group affect finds that groups influence prejudice through informational, normative, and identity-based means.[1]

As noted above, groups provide individuals with information about social reality (e.g., Festinger, 1954). Among the pieces of information provided about the social world is information about outgroups. Sechrist and Stangor (2007) conduct an experiment demonstrating that unfamiliarity with a group leads individuals to rely more on social information about the group. This finding suggests that a reliance on group stereotypes may be dependent, in part, on the need for information. Additionally, simply providing information about group beliefs and attitudes influences students' racial prejudice and discrimination (Sechrist and Stangor, 2001; Stangor, Sechrist and Jost, 2001).

The expression of prejudice is also conditioned by normative influence: people are

[1]It should be noted that there is some disagreement whether social influence on prejudice represents "true" change or merely superficial compliance with norms to avoid sanctions (Kelman, 1961). Some literature suggests that counternormative prejudice still exists—albeit in concealed form (e.g., Stephens-Davidowitz, 2017). Likewise, Paluck (2009) shows that infotainment changes social norms and behavior but not prejudice. Still, other norm treatments affect attitudes expressed in private and at least one week after treatment (Blanchard et al., 1994; Stangor, Sechrist and Jost, 2001; Terry, Hogg and Blackwood, 2001), suggesting that social influence can effect true changes in individuals' prejudicial attitudes. Moreover, even superficially compliant behavior can eventually become accepted and internalized by individuals, meaning social enforcement is no longer necessary to sustain the norm (Crandall, Eshleman and O'Brien, 2002).

more willing to express prejudice when norms are less clear or even support prejudice, and they tend to reduce the expression of prejudice when norms condemn it. As norms against explicit racism have grown stronger in the United States, politicians have turned to the use of more subtle racial code to conceal racist appeals, fearful of electoral penalties (Kinder and Sanders, 1996; Mendelberg, 2001). The mass public, too, has become more reluctant to be openly racist: an experiment shows that subjects are less willing to express overt racism in the presence of Black experimenters (McConahay, Hardee and Batts, 1981). Munger (2017) demonstrates that confronting people who use racial slurs on Twitter reduces the use of racial slurs, revealing that social sanctions are an effective tool for reducing racist expressions. That norms constrain the expression of racism does not mean that racism has disappeared, however: more subtle measures of racism, which are less susceptible to social desirability bias, still detect a great deal of racial animus in the American public (Kinder and Sanders, 1996; McConahay, Hardee and Batts, 1981; Meertens and Pettigrew, 1997; Stephens-Davidowitz, 2017). Just as normative influence can constrain racist expression, so too can it embolden racism and prejudice. Álvarez Benjumea and Winter (2020) create a bespoke online discussion board and randomly assign subjects to view a mixed-norm condition containing hate speech or stronger-norm conditions without hate speech. After a terrorist attack, hate speech increased only in the mixed-norm condition where the normative status of hate was unclear. Similarly, when normative obligations are ambiguous, Gaertner and Dovidio (1977) show that people are more likely to discriminate against Black people, and when normative influence seems to favor racism, people are more willing to express racism (Greenberg and Pyszczynski, 1985). It is clear, then, that the expression of intergroup prejudice is dependent on whether individuals expect that their actions will be attributed to prejudice. If their actions are attributed to prejudice, it could carry social consequences or lead to internalized guilt for transgressing against norms.

A final factor in the relationship between group norms and intergroup prejudice is group

identity. The group norm theory of prejudice holds that intergroup prejudice is a norm that develops through interactions within and between groups (Sherif and Sherif, 1966). Sherif and Sherif (1966) argue that individuals adhere to group norms of prejudice to remain part of the group, to avoid social sanction, and because the group identity is part of individual identity. Other work supports this theory. For example, Munger (2017) finds that ingroup confrontation of racial slurs is a particularly effective deterrent of future racism. Moreover, information about ingroup norms is more influential than information about outgroup norms (Haslam et al., 1996; Stangor, Sechrist and Jost, 2001), and there is some evidence that individuals who identify most strongly with a group appear to internalize group norms against prejudice the most (Crandall, Eshleman and O'Brien, 2002). Thus, individual identities appear to play an important role in shaping the relationship between norms and prejudice.

Norms and Affective Polarization

As affective polarization has increased (Iyengar et al., 2019), scholars have begun to examine potential ways of reducing partisan animosity. Within this broader literature, some studies have looked to intra-party norms to explain polarization. Peters (2021), for example, notes that media discussion of rising polarization likely creates a descriptive norm, which could spur further polarization as people conform (see also Levendusky and Malhotra, 2016). Through several experiments, Connors (2023) demonstrates that affective polarization is seen as socially desirable by partisans, meaning that as social pressure increases, people become more affectively polarized. Despite widespread norms supporting affective polarization, other scholars have attempted to reduce affective polarization using social influence. Reading an article discussing the apolitical norm of the military reduces hostility toward the outparty among military service members (Mullinix and Lythgoe, 2023). A training that instructs attendees how to confront polarization within their own groups can reduce affective polarization while trainings with strategies for intergroup contact have little impact (Blair et al., 2023). Wojcieszak, Winter and Yu (2020) test the influence of norms of

open- and closed-mindedness on bias in information consumption as well as on downstream affective polarization, finding that priming norms can affect the information consumed but have mixed effects on affective polarization. Finally, the work of Dimant (2024) suggests potential limits to the impact of certain norms on affective polarization: descriptive norms can increase positive behavior toward the outgroup, but apparently, the descriptive norms treatments from this study are circumscribed in their ability to reduce intergroup discrimination. Partisans discriminate even in the presence of norms to the contrary because they perceive opposing party norms as contradicting cooperation (Dimant, 2024).

While these studies suggest that norms may reduce partisan hostility, several of the treatments bundle norms with other factors that could potentially also influence affective polarization. For example, Mullinix and Lythgoe (2023) emphasize norms among military service members, which may increase the salience of the military member identity and reduce the salience of partisanship (Levendusky, 2023). Blair et al. (2023) employ a workshop on strategies to combat both internal and intragroup expressions of polarization, and it is possible that the treatment effect is transmitted by some aspect of the training other than the creation and reinforcement of group norms. Further, some studies rely solely on descriptive norms and do not focus on norms associated with a strong identity (e.g., Dimant, 2024 uses descriptive norms among prior study participants). By using only descriptive norms without an accompanying strongly held identity, previous work may overlook normative and identity-based social influence. Many of these studies manipulate but do not measure extant norms, gaining greater experimental control but potentially losing some insight into how people perceive and interact with norms in their daily life (Blair et al., 2023; Mullinix and Lythgoe, 2023; Wojcieszak, Winter and Yu, 2020). Finally, these studies focus exclusively on the mass public, overlooking the role of political leaders in setting and maintaining democratic norms (Key, 1961).

In this paper, I build on these previous studies and compare norms between the masses

and politicians. I measure perceptions of both inparty descriptive norms and social approval. This design allows me to capture descriptive, normative, and identity-based social influence. I focus on intra-party norms regarding partisan incivility—a norm that "when others go low, we go high." Civility is a key democratic norm. For democracies to thrive, different political camps must treat one another as legitimate opponents, not as enemies or existential threats (Levitsky and Ziblatt, 2018). This implies that there must be some modicum of respect expressed toward opposing partisans (i.e., civility, as in Bybee, 2016; Frimer and Skitka, 2018 or mutual toleration as in Levitsky and Ziblatt, 2018). Incivility, by contrast, places partisan opponents outside the realm of legitimate opposition, denying them even the most basic respect (Gitlin, 2013). In other words, partisan incivility is a behavioral manifestation of partisan animus, or affective polarization, signalling how much the opposing party is detested by the speaker (Allport, 1954).

When studying incivility, it is crucial to consider the role of political leaders given the consequences of politicians' incivility. Incivility among political leaders can drive mass-level affective polarization (Skyette, 2021). Incivility begets further incivility, potentially generating an escalatory cycle (Gervais, 2014, 2015, 2017). This style of "nasty politics" may spiral into violence, as extreme action is legitimized when opponents are painted in stark terms (Kalmoe, 2020; Kalmoe and Mason, 2022; Zeitzoff, 2023). Thus, in evaluating the functioning of the democratic system, norms regarding partisan incivility among elites should be examined. These norms among politicians may form a particularly important democratic backstop in light of previous findings that the mass public is less constrained by democratic norms (McClosky and Brill, 1983; Stouffer, 1955).

A great deal of research indicates that the mass public disapproves of partisan incivility. Incivility by politicians decreases trust in government (Mutz, 2015; Mutz and Reeves, 2005) and decreases approval of politicians (Frimer and Skitka, 2018, 2020). People dislike those who are uncivil: inparty incivility decreases affective polarization, while incivility from

out-partisans increases affective polarization (Druckman et al., 2019). Using conjoint experiments, Costa (2021) shows that voters prefer politicians who do not attack the opposing party. The results of Frimer and Skitka (2020) do reflect some approval for partisan incivility from ordinary people (as opposed to politicians). Still, given the number of studies showing that partisan incivility elicits negative attitudes, I expect that both politicians and the public will disapprove of incivility by their inparty peers more than civil yet negative rhetoric. If they have internalized a norm against incivility, the masses should feel more guilty or ashamed for saying something uncivil compared to something civil but negative.

Few studies have measured norms surrounding incivility as social norms. Instead, previous scholars have generally inferred that incivility transgresses against norms because individuals disapprove—with little consideration of how respondents perceive the broader social acceptability of incivility. In contrast, I explicitly measure perceptions of *social* norms regarding incivility alongside personal opinions. If there is a social norm against incivility, I should find that politicians and the masses believe both that their ingroup peers are less likely to express incivility than civil but negative rhetoric and that their ingroup approves less of incivility than civil communication of partisan disagreements. These factors capture the main elements of Bicchieri's (2017) definition of norms—descriptive or empirical expectations and normative expectations, respectively. Along these lines, politicians and the public should also expect sanctions will be more likely for incivility than civil disagreement. Specifically, inparty peers may "think less of" someone for using incivility (incivility may result in reputational harm). For politicians, voters may choose not to vote for an uncivil candidate, and inparty politicians may refuse to help the uncivil politician or may try to avoid her. For the mass public, sanctions may be conveyed through social discomfort with intense partisan incivility or the expression of disagreement with the offending individual. Overall, then, I expect that there is an inparty social norm against partisan incivility among politicians and the mass public. This norm should be most apparent when comparing uncivil to civil but negative

partisan speech and should encompass elements of descriptive and normative expectations (Bicchieri, 2017).

Building on research demonstrating that social norms influence prejudice (e.g., Blanchard et al., 1994; McConahay, Hardee and Batts, 1981), I hypothesize that priming the norm against partisan incivility will reduce affective polarization. As discussed above, partisan incivility is a clear display of partisan hostility. Consequently, reminding individuals of a context in which the expression of such animus is not socially accepted should reduce willingness to express hostility toward the outparty. This influence may occur for several reasons. First, the norm prime may increase the salience of the norm regarding incivility, making it more influential when individuals evaluate the parties (Cialdini, Reno and Kallgren, 1990). By priming the norm, the treatment should activate the script associated with the norm—a script that discourages the expression of hostility (Bicchieri, 2006). Second, to the extent that the norm proscribes partisan incivility, it may convey information about the outparty (Sechrist and Stangor, 2007). People may reason that, if the inparty condemns the expression of hostility against the opposing party, the outparty may not be as bad as expected. Third, priming a norm against incivility could reduce willingness to express hostility due to the desire to avoid potential sanctions for transgressions or because reminders of sanctions convey that the expression of hostility is undesirable. In either case, people should avoid expressing socially undesirable views. Fourth, by looking specifically at intra-party norms against partisan incivility, I aim to emphasize the "we" in "we go high," highlighting that others who share their identity view the expression of partisan animus unfavorably. Given that partisanship is a social identity (Greene, 1999; Iyengar, Sood and Lelkes, 2012), the norm prime should both heighten the salience of partisanship and generate an image of what that partisanship entails. I expect that partisans should adopt the group norm disfavoring incivility, reducing partisan hostility (Turner et al., 1987). In sum, I expect that norms against incivility should reduce partisan hostility and affective polarization through

information, social desirability, and identity.

Following the results of Dimant (2024), the norm against incivility may also make individuals feel more positively toward their own party. As part of the process of self-categorization into social identities, people compare the ingroup to the relevant outgroup (Turner et al., 1987). Priming ingroup norms surrounding incivility may raise the salience of the ingroup identity and make the norm regarding incivility more central to the interparty comparison. If people believe that the norm is not shared by both parties, the norm may be an additional piece of information which positively differentiates their party from the outparty (Tajfel and Turner, 1986). For these reasons, I expect that individuals who think about the norm against incivility immediately prior to expressing their feelings toward the political party will display (1) more warmth toward the opposing party, (2) less distance between evaluations of the two parties (i.e., less affective polarization), and (3) more warmth toward the inparty.

Research Design

To test these hypotheses, I conducted two original surveys—one of the mass public and one of local elected officials.

Survey of the Mass Public

The survey of the mass public was collected by the Polarization Research Lab via YouGov between July 5th and 16th in 2024 as part of their long-running online tracking poll of American partisan animosity. There were 2,000 survey respondents, of whom 1,669 (83.5%) were party identifiers or leaned toward one of the parties. All of my analyses for the mass sample exclude pure independents but include partisans and party leaners (Iyengar, Sood and Lelkes, 2012). Each respondent was shown a short vignette. The vignette asked respondents to imagine a social situation in which they are discussing either politics or sports with ingroup family members and friends. At this social event, a member of their group says

something uncivil or civil but negative about opposing partisans or a rival sports team. The topic, sports or politics, was randomized with probabilities 0.25 and 0.75, respectively. Within topics, respondents had a probability of 0.75 of seeing an uncivil vignette. I also randomized among several different examples of (in)civility with equal probability to ensure that I was fully capturing the underlying constructs of interest and that my results would generalize beyond a single example of each type of rhetoric.[2]

To measure norms against incivility, I asked respondents if they personally would approve of the rhetoric in the vignette, if they believed others in their group would approve (normative expectations), and if they believed others in their group would use language like the vignette (descriptive expectations). Following Wu and Huber (2021), I also captured norm internalization by determining whether respondents would feel guilty for using rhetoric like the vignette. I included three measures of potential social sanctions for using the rhetoric in the vignette: the likelihood that ingroup members would (1) think less of the respondent for using rhetoric like the vignette, (2) be uncomfortable with the rhetoric, and (3) confront the person in the vignette. Importantly, these questions captured the main elements of social norms highlighted by Bicchieri (2017), including personal attitudes, beliefs about others' attitudes (normative expectations), and beliefs about others' behavior (descriptive expectations). Full details of questionnaires, vignettes, and randomization schemes can be found in Supplementary Information (SI) A.

Using the vignettes about politics, I measure the prevalence of norms against incivility with separate regressions for each norm question with the form

$$Y_i = \alpha + \beta_1 * Uncivil_i + \varepsilon_i$$

where Y_i is the respondent's answer on a given norm question. $Uncivil_i$ is an indicator vari-

[2]Results in SI D.1 indicate that respondents perceive the uncivil and civil vignette examples differently from one another and that uncivil (civil) examples are perceived as similar to other uncivil (civil) examples.

able for whether the vignette contains an example of incivility, and errors are heteroskedasticity robust (HC2).[3]

Crucially, I also randomized whether respondents saw the vignette and answered the accompanying norms questions immediately before or after completing party feeling thermometer ratings—a conventional measure of affective polarization (Iyengar, Sood and Lelkes, 2012). Respondents saw the norms questions before or after with probability 0.5, block randomized within topic and (in)civility. This design allows me to determine whether increasing the salience of norms against the expression of partisan hostility can reduce reported affective polarization.

I estimate the effect of priming norms by regressing each party feeling thermometer measure (Y_i) on an indicator ($NormsFirst_i$) for whether the norms questions came before the feeling thermometer questions with heteroskedasticity-robust (HC2) errors:

$$Y_i = \alpha + \gamma_1 * NormsFirst_i + \varepsilon_i$$

The feeling thermometer measures denoted by Y_i are outparty ratings, inparty ratings, and affective polarization (the difference between inparty and outparty thermometers). Because I want to measure the impact of norms against partisan incivility, I limit the affective polarization analyses to respondents who viewed an uncivil politics vignette, which most directly captures my construct of interest.

Survey of Local Elected Officials

The survey of local elected officials was conducted by CivicPulse between March 20$^{\text{th}}$ and April 10$^{\text{th}}$ of 2024 using their database of officials in localities with populations of at least 1,000 throughout the United States. In all, 513 local government officials responded

[3]In SI E, I show that respondents believe norms similarly favor civility over incivility in discussions about sports; however, respondents believe norms about sports disfavor incivility in absolute terms—in contrast to their perceptions of norms regarding incivility in politics.

to the survey, with 450 (87.7%) respondents identifying with or leaning toward one party on the standard party identification questions. My main analyses rely only on partisans and party leaners. However, in SI C.2, I also present results including feeling thermometer partisans (those who did not identify with or lean toward a party but nonetheless expressed a preference for one party over the other on party feeling thermometers). The survey obtained a final response rate of 7%. Like the mass sample, each respondent was shown a brief vignette which asked respondents to imagine an inparty politician running for office in their state. The politician in the vignette campaigns by either saying something uncivil ($p = 0.75$) or civil but negative ($p = 0.25$) about opposing partisans. Multiple examples of each type of rhetoric were randomized with equal probability within the categories of civility to ensure both construct validity and generalizability.[4] Examples of (in)civility were block randomized within groups determined by party[5] and randomized question order.

As in the mass survey, respondents were asked for their personal approval of the vignette rhetoric, their perception of other inparty politicians' approval (normative expectations), and their belief that other inparty politicians would use similar rhetoric (descriptive expectations). There were three questions about potential sanctions for the vignette rhetoric: (1) the share of voters who would vote for the politician in the vignette, (2) the likelihood that other inparty politicians would avoid or refuse to help the campaign of the vignette politician, and (3) the likelihood that people the respondent cares about would think less of them for campaigning like the vignette politician. These questions were designed to capture personal attitudes, descriptive expectations, and normative expectations (Bicchieri, 2017). Within party groups, the order of questions was block randomized with a probability of 0.5

[4]Results in SI D.2 indicate that civil and uncivil examples were perceived differently from one another by the sample and that uncivil (civil) examples were perceived similarly to other uncivil (civil) examples.

[5]There were five groups corresponding to (1) Democratic identifiers and leaners, (2) Republican identifiers and leaners, as well as those who showed a preference for the (3) Democratic or (4) Republican Party on pre-treatment feeling thermometers, and (5) pure independents. My results in the main text focus only on groups (1) and (2).

to determine the effect of priming the norm against incivility on affective polarization. Respondents viewed the vignette and answered norms questions either before or after answering party feeling thermometer questions. One additional aspect of the elite survey differed from the mass survey: the post-treatment feeling thermometers asked about party "activists," "candidates and elected officials," or "voters." The specific thermometer version was determined by equal-probability block randomization within party and question order groups. I also asked respondents about their feelings toward the parties themselves prior to the norms treatment. Full details of questionnaires, vignettes, and randomization schemes can be found in SI B.

I use linear regressions to determine the extent of norms against incivility among politicians, separately regressing each measure of norms (Y_i) on an indicator ($Uncivil_i$) for whether the vignette contains incivility.

$$Y_i = \alpha + \beta_1 * Uncivil_i + \varepsilon_i$$

These regressions are fitted with heteroskedasticity-robust (HC2) errors. Similar to the mass sample above, I estimate the effect of priming norms against incivility with regressions of the form:

$$Y_{i,t} = \alpha + \gamma_1 * NormsFirst_i + \gamma_2 * Y_{i,t-1} + \gamma_3 * CandidateTherm_i + \gamma_4 * ActivistTherm_i + \varepsilon_i$$

where $Y_{i,t}$ corresponds to the inparty thermometers, outparty thermometers, and affective polarization (the difference between the inparty and outparty thermometers).[6] As above, $NormsFirst_i$ indicates whether the norm is primed prior to the feeling thermometer questions. Both because the sample of politicians is somewhat small and because the post-

[6]Analyses of inparty thermometer ratings were not pre-registered for the sample of politicians, and as such, they are exploratory. I pre-registered regressions using inparty thermometer ratings for the mass sample.

treatment feeling thermometers were directed at party subgroups, I include a measure of pre-treatment party feeling thermometers ($Y_{i,t-1}$) in my regressions. Additionally, I include dummy variables for whether the outcome feeling thermometers asked about "candidates and elected officials" ($CandidateTherm_i$) or "activists" ($ActivistTherm_i$). I limit the analyses regarding affective polarization to the sample of respondents who viewed an uncivil vignette as the uncivil vignettes most closely measured my construct of interest: a norm against partisan incivility.

Results

The Norm Against Incivility

First, I establish the extent to which the masses and politicians believe that a norm against incivility exists. I regress each measure of normative beliefs on an indicator variable for the presence of incivility in the vignette. Figure 1 displays estimates of the treatment effects for both samples. Across all of my measures, I find strong evidence that individuals believe incivility is viewed less favorably than civil rhetoric. Specifically, I find that people disapprove of uncivil political speech more than civil speech. The masses are also more prone to feeling guilty or ashamed for using uncivil speech than civil speech—a key measure of norm internalization.[7] These effects are substantively quite large, accounting for roughly 15% of the entire scale in the mass sample and more than 30% of the scale in the politician sample.

Not only do individuals express more personal distaste for incivility relative to civility, but they believe their peers would as well. Both politicians and masses believe that their ingroups would be less likely to use incivility than civil speech. Substantively, the decrease in the perceived share of the ingroup who would say something like the vignette caused by incivility accounts for between 15% and 20% of the scale. Crucially, both samples also

[7]Due to space constraints on the survey of politicians, I was unable to include this measure of internalization in the politician sample.

believe that their ingroups would be less approving of incivility than civility–shifts that correspond to 15% of the scale among the masses and nearly 25% among politicians. Finally, I show that both samples believe they are more likely to face sanctions for incivility than for civility. Politicians and the mass public are more likely to agree that people they care about would "think less of" them for using uncivil as opposed to civil speech. Given the central role of voters in many of politicians' calculations (Mayhew, 1974), I also asked politicians what share of voters they believe would vote for the candidate in the vignette. My results show that politicians believe a candidate who campaigns using incivility would receive eight percentage points less of the vote than one who campaigns using civil language. Additionally, politicians believe that ingroup members are 13 percentage points more likely to either avoid or refuse to endorse an inparty candidate who uses incivility than one who campaigns against the outparty civilly. Turning to the mass sample, uncivil vignettes caused respondents to report a higher perceived likelihood that their ingroup would be uncomfortable with the vignette speech and that their ingroup would confront the person in the vignette by expressing disagreement with them. In sum, my results indicate that incivility in political speech elicits more personal disapproval from respondents and generates more expected condemnation from other ingroup members than does civility. As expected, norms favor civility over incivility in both samples.

The treatment effects display stronger normative support for civility *relative* to incivility, but they do not tell us the extent of *absolute* support for either civility or incivility. To better assess the strength of the norm against incivility, I turn to predictions from the treatment effect regression models above (see Figure 2). Looking first at predicted norm outcomes for civil vignettes, we can see that both politicians and the mass public approve of civil but negative speech about the outparty on average. The mass sample also disagrees that they would feel guilty or ashamed for using civil but negative speech about the outparty. In other words, neither group appears to have any internalized distaste for negativity toward

the outparty when delivered civilly. Moreover, we can see that both groups believe that similar proportions of the ingroup would speak about the outparty in a civil manner—the average is located approximately halfway between "some" and "most" of the ingroup—and they believe that the ingroup would approve of the civil yet negative speech at similar rates to their own approval. Largely tracking the perceived acceptance of civil rhetoric within the ingroup, neither sample believes that the ingroup would be likely to sanction civil but negative rhetoric. Politicians neither agree nor disagree that people they care about would think less of them for expressing the sentiments in the civil vignette, and the mass sample disagrees with the idea that they would face a sanction for civil speech. Additionally, politicians believe a majority of voters would favor a politician who attacks the opposing party in a civil manner and that it is unlikely that other politicians would avoid them or refuse to help their campaign. Similarly, voters believe it is unlikely that their friends and family members would be uncomfortable with or would express disagreement with civil but negative rhetoric about the opposing party. Notably, with the exception of perceptions of people they care about, politicians and voters express comparable average beliefs about civil rhetoric.

On the left side of Figure 2, we can see that differences between the politician and mass samples seem to occur primarily in response to uncivil speech against out-partisans. First, on average, politicians disapprove of uncivil attacks against the opposing party, while the masses neither approve nor disapprove of such rhetoric. Additionally, while the masses are more likely to agree that they would feel guilty or ashamed for using uncivil language to criticize the opposing party than civil language, they disagree that they would feel guilty on average. Thus, it appears that the masses do not have any internal compunctions about attacking the opposing party uncivilly. Second, the politicians and masses also differ in their perceptions of their reference groups. Politicians and the mass public believe that few of their ingroup members would use uncivil attacks against the opposing party (descriptive expecta-

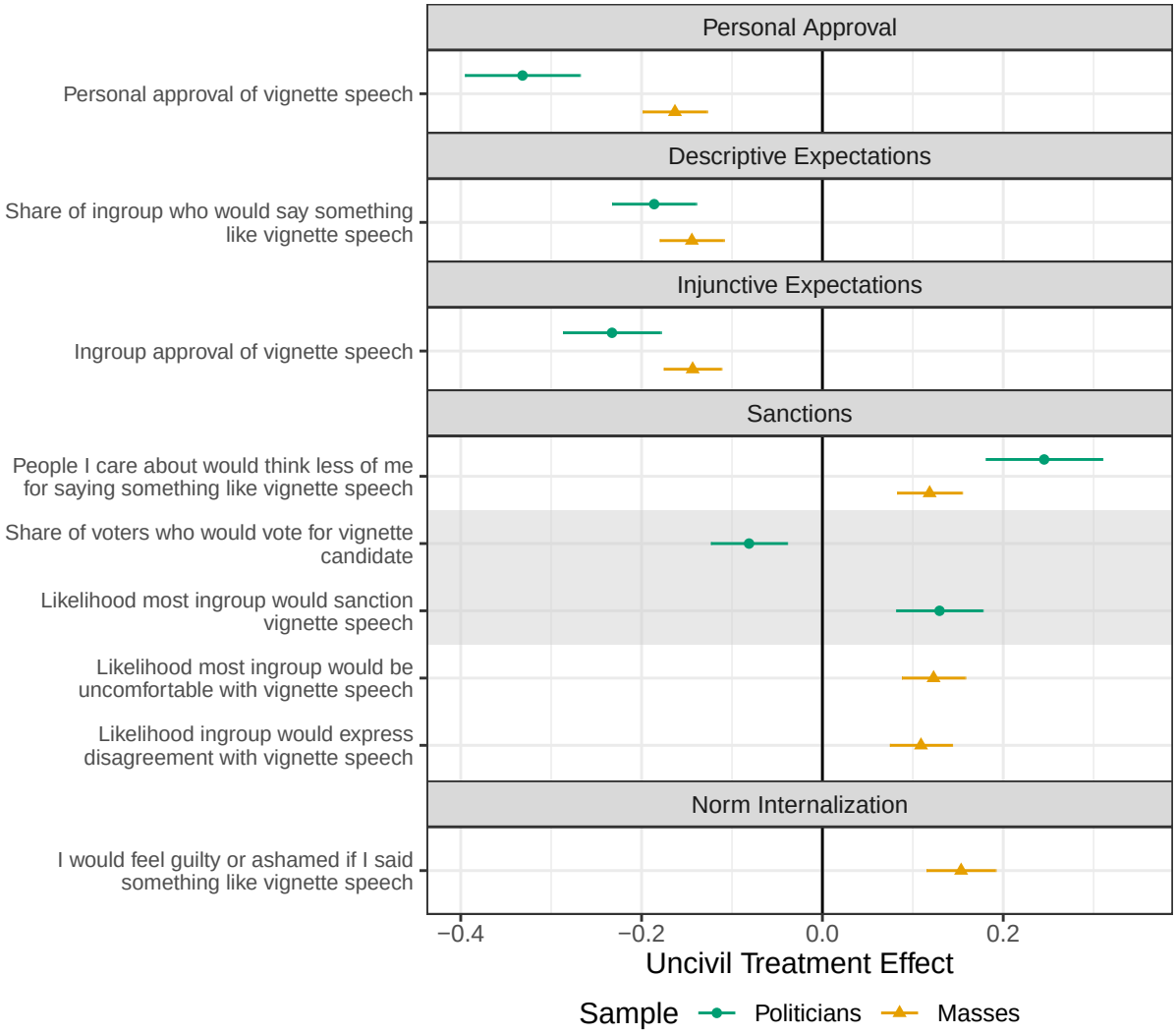
tions). Politicians believe that the ingroup would disapprove of uncivil rhetoric (normative expectations), yet in the mass sample, respondents believe that their ingroup would actually approve of partisan incivility. Though the perceived ingroup approval is significantly lower than for civil rhetoric, the mass public still believes that incivility is socially acceptable. This pattern of mass approval and elite condemnation is also reflected in the anticipation of possible sanctions for incivility. Politicians agree that people they care about would think less of them for engaging in incivility, while the mass public disagrees somewhat with that proposition. Politicians also believe that an uncivil politician would receive less than 50% of the vote in an election, and that most other inparty politicians would be neither likely nor unlikely to avoid or refuse to help the campaign of an uncivil politician. On the other hand, the mass sample views it as unlikely that the ingroup would be uncomfortable with or confront the uncivil attacks on the opposing party.

The results of the previous analyses reveal that both in relative and absolute terms politicians believe that a norm against political incivility exists: they disapprove of uncivil political attacks, believe their ingroup disapproves as well, and expect that such behavior would lead to at least some social and electoral sanctions. In contrast, the mass sample believes that their ingroup norms favor civility over incivility, but in absolute terms, mass respondents do not believe that a norm against incivility exists. Members of the mass public do not personally disapprove of uncivil speech. While mass partisans do not believe many of their peers engage in incivility in political dialogue, they believe their reference group approves of uncivil rhetoric and would not sanction such rhetoric when used. If anything, this indicates that the norm among the mass public supports partisan incivility. It is worth noting that there is no descriptive norm favoring incivility among the masses—a result somewhat at odds with work contending there is a descriptive norm favoring polarization (Peters, 2021). Instead, it is normative pressures that favor incivility, consistent with Connors's (2023) argument that affective polarization is socially desirable. These findings are also interesting

in light of other work that suggests the mass public dislikes incivility (e.g., Druckman et al., 2019; Frimer and Skitka, 2018; Mutz, 2015).

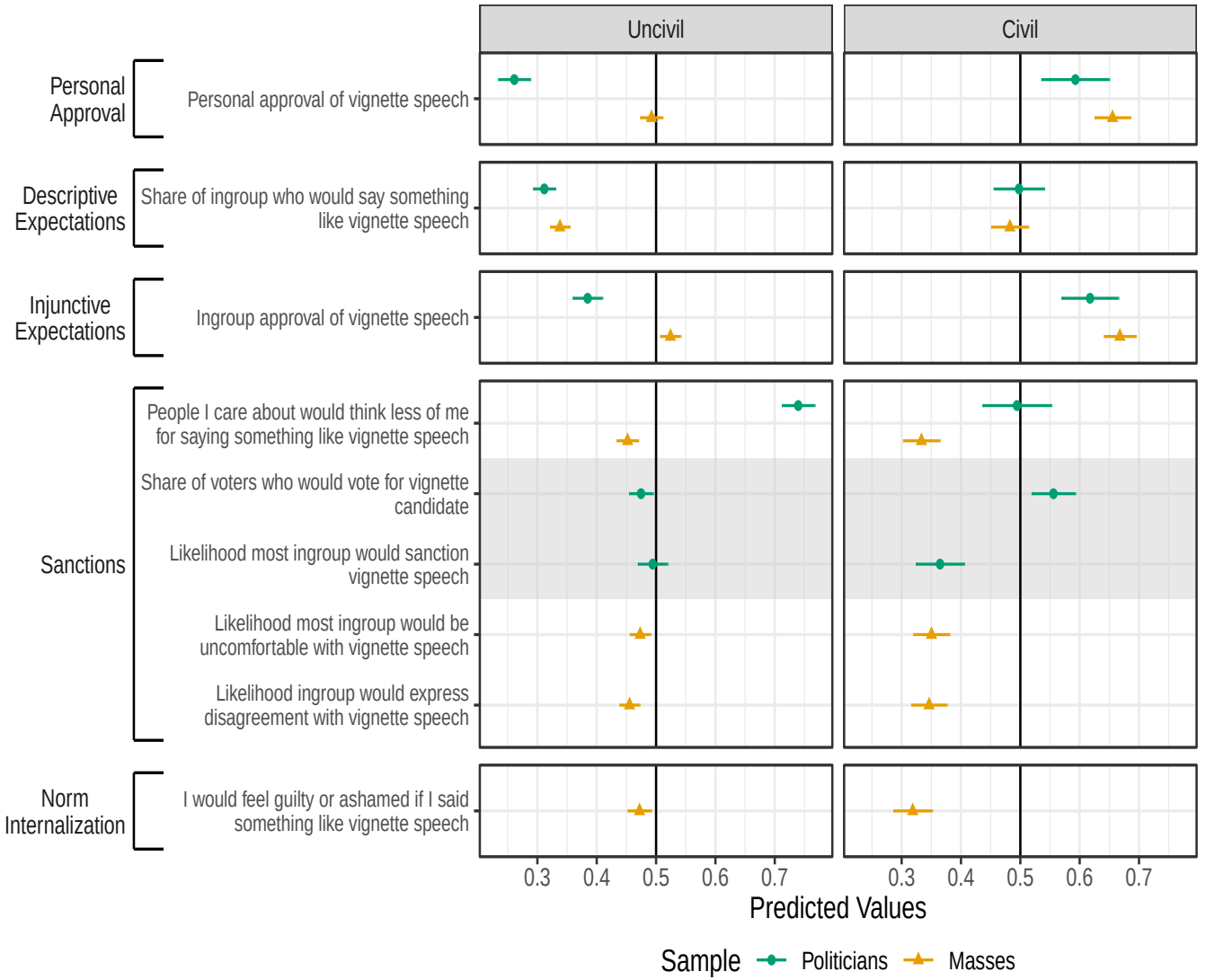
Differences in norms around incivility between the mass and politician levels could occur for two reasons. First, it may be the case that norms are shaped by the context contained in the vignette: politician vignettes were written about campaigning for office, but mass vignettes concerned conversations with family and friends at a social gathering. It is possible that norms among politicians similarly favor partisan incivility in social settings. Second, it may be that politicians start with stronger norms against incivility and select into government office, or they are socialized into new norms when they enter government (McClosky and Brill, 1983; Stouffer, 1955). In either case, my vignette design captured a social norm against partisan incivility among politicians but failed to do so at the mass level.

Figure 1: Average Treatment Effects of Vignette Incivility on Norm Outcomes



Note: Average treatment effects from OLS regressions of norm variables (rescaled to 0 to 1) on an indicator for vignette incivility in political vignettes. Regression models fit with HC2 standard errors. Bars display 95% confidence intervals. Full regression output is in SI C.1.1 for the mass public and SI C.2.1 for politicians.

Figure 2: Vignette Incivility and Predicted Norm Outcomes



Note: Predicted values for norm outcomes from OLS regressions of norm variables (rescaled to 0 to 1) on an indicator for vignette incivility in political vignettes. Regression models fit with HC2 standard errors. Bars display 95% confidence intervals. Full regression output is in SI C.1.1 for the mass public and SI C.2.1 for politicians.

Norms and Affective Polarization

Having examined the prevalence of norms against incivility, I next examine the effects of priming this norm on affective polarization. If people are reminded that the expression of outparty hostility is socially unacceptable in their reference group, we might expect them to express less hostility toward the outparty when asked how they feel about the outparty. By randomizing whether respondents answer questions about norms either before or after party feeling thermometers, we can determine whether making the norm against partisan incivility salient reduces affective polarization. In Figure 3, we can see estimates of the average treatment effects for the norm prime on party feeling thermometer ratings. The leftmost pane displays the treatment effects of priming on outparty feeling thermometer ratings. While none of the estimates are statistically significant, the politicians' estimate is positive, as expected. The small, positive coefficient for the sample of politicians indicates that priming the norm against incivility increases reported warmth toward the opposing party slightly (i.e., it decreases the willingness of respondents to report cold feelings toward the outparty). Given the fairly small sample of politicians, we cannot rule out small effects of the treatment; further research is necessary to determine whether a smaller effect of norms on outparty thermometer ratings exists among politicians.

Contrary to expectations, the treatment effect estimate for the mass sample is small and negative, indicating that priming perceived norms regarding incivility actually slightly increases respondents' willingness to report cold feelings toward the opposing party, albeit insignificantly. While unanticipated, this is likely due to the weak evidence of a norm against incivility found in the mass sample, discussed in the previous section. Since the mass sample appears to believe that their reference groups do not sanction incivility and even approve of it, it makes sense that the treatment does not successfully reduce outparty animosity. In fact, if anything, the norm in the mass sample *supported* incivility against opposing partisans, meaning that the observed decrease in feeling thermometer ratings of the opposing

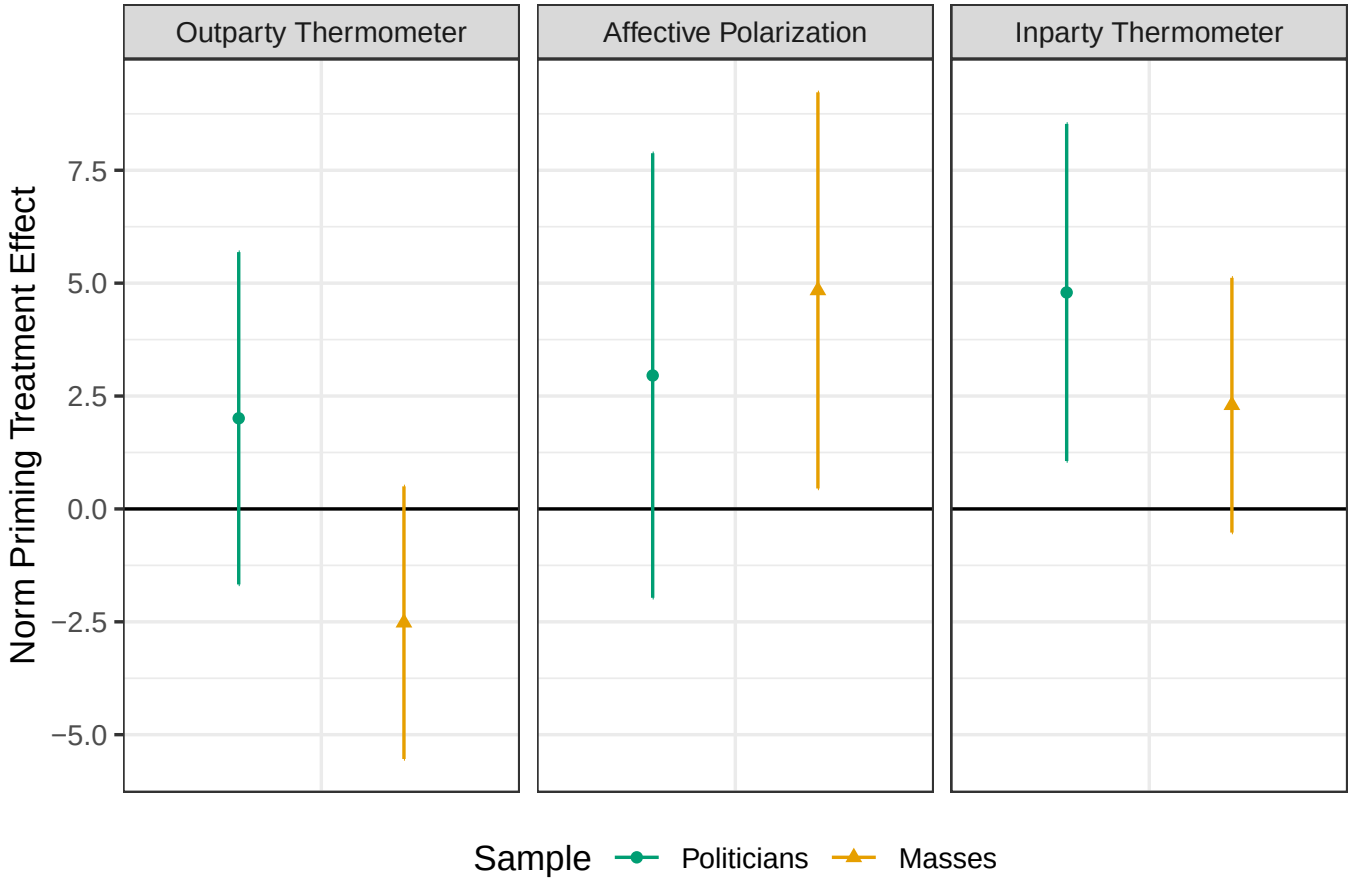
party would actually be normatively consistent behavior for the mass public (Connors, 2023; Peters, 2021). Further, to the extent that norms affect mass attitudes, the countervailing forces of descriptive and normative expectations could reduce the effect.

The middle pane of the figure contains treatment effect results for the difference between the inparty and outparty feeling thermometer ratings (affective polarization). Again, contrary to expectations, my results show that priming the norm against incivility increases affective polarization for both the mass and politician samples, though the effect does not reach conventional levels of significance for the politician sample. Substantively, the increase in affective polarization observed for the mass sample amounts to roughly 2.5% of the entire range of affective polarization. To better understand why priming the norm against incivility seems to increase affective polarization, I decided to also examine the effects of priming on inparty thermometer ratings. This analysis was not pre-registered for the politician sample and should therefore be regarded as exploratory. I did however pre-register this analysis for the mass sample. Both the mass and politician samples show a positive effect of priming the norm on inparty thermometer ratings, though this effect is insignificant for the mass sample. Among politicians, the observed effect on the inparty thermometer ratings is quite large—approximately 5% of the entire range of the inparty thermometer. The small (or negative) effects of priming the norm against incivility on outparty thermometers combined with the positive effects on inparty thermometers to produce increases in affective polarization.

Ultimately, these results indicate that reminding individuals how their reference group feels about partisan incivility can increase warmth toward the inparty, suggesting that ingroup norms may play an important role in defining a positive inparty identity (as opposed to a negative partisan identity as in Abramowitz and Webster, 2016; Zhong, Galinsky and Unzueta, 2008; Zhong et al., 2008). This is particularly notable given that politicians and the masses perceived different ingroup norms regarding incivility—mass reference groups approved of incivility, while politician reference groups disapproved. Additionally, while

positive effects on inparty thermometers and consequently affective polarization were unexpected, my results do offer some support for a norms-based understanding of outparty hostility. Among politicians, whose reference group was perceived to condemn partisan incivility, the priming treatment slightly increased reported warmth toward the opposing party. Among the mass public, where the reference group was perceived to promote partisan incivility, the priming treatment slightly decreased reported warmth toward the opposing party. Both treatment effects are consistent with movement in the direction of perceived norms.

Figure 3: The Effect of Norms on Affective Polarization



Note: Average treatment effects of priming the norm against incivility on affective polarization in uncivil political vignette condition. Regression models fit with HC2 standard errors. Bars display 95% confidence intervals. Full regression output is in SI C.1.2 for the mass public and SI C.2.2 for politicians.

Exploratory Analysis: Decomposing the Compound Norm Prime

The priming treatment is a compound treatment: treated subjects are reminded of their own personal views about partisan incivility as well as the practices and attitudes of their reference group regarding uncivil attacks on the opposing party. The bundled nature of the treatment can disguise important heterogeneity in the treatment effect that depends on individual perceptions of norms. For example, the perceived support for incivility among the masses means that priming reference group attitudes may have reminded the average respondent of normative support for partisan hostility, thereby generating the observed increase in affective polarization. However, the treatment effect may be different for respondents who perceived that partisan incivility was not supported by their reference group. In this section, I assess whether the treatment effects of the norms-based prime vary depending on perceptions of the norm. The analyses in this section were not pre-registered but were undertaken to better understand the precise mechanisms for the observed effects in light of the unanticipated support for incivility in the mass sample. It is important to determine whether the observed increase in affective polarization is the result of perceived approval of incivility; in other words, if stronger norms against incivility exist, can they reduce hostility toward the opposing party and affective polarization?

To parse the role of perceptions of the norm in driving effects on affective polarization, I fit causal forests to the data, examining treatment effect heterogeneity depending on answers to the norms questions (Wager and Athey, 2018). Causal forests utilize the random forest framework to split the data to estimate conditional average treatment effects, allowing for nonlinear relationships between variables and treatment effects as well as complex interactions among variables (Wager and Athey, 2018). Specifically, I estimate the conditional average treatment effects (CATE) of priming norms against incivility, depending on individuals' support for incivility and their beliefs about support for incivility within their reference group. If the norms prime is affecting party feeling thermometers through group norms, we

should expect positive effects on outparty feeling thermometers for individuals who believe their group would disapprove of and sanction incivility. The results in this section should be interpreted cautiously: they are exploratory analyses, and the politician sample is fairly small for a reliable analysis of heterogeneous treatment effects. Additionally, it is possible that answering party feeling thermometer before the norms question influenced the reported perceptions of norms—though I find no evidence that answers to the norms questions differ depending on the question order. With these caveats in mind, these analyses are nonetheless potentially informative about causal mechanisms at play in the data.

Figure 4 displays locally weighted smoothing (LOESS) lines for predicted CATE estimates against each norm question. Interestingly, there is very little variation in the predicted CATEs in the politician sample; this could be, in part, a reflection of the relatively small sample. Moreover, the LOESS lines for politicians' CATEs are almost always positive, indicating that the norm prime is predicted to increase warmth toward both the inparty and outparty, irrespective of the perceived norms. The mass sample, however, shows more variation. As my main focus in this paper is the impact of norms on reported partisan hostility, I look primarily at the predicted CATE estimates for outparty thermometer ratings. The norm prime is predicted to have a positive impact on outparty thermometer ratings when individuals disapprove most strongly of partisan incivility, but the effect decreases and becomes negative as individual approval increases. Likewise, respondents who say they would feel guilty or ashamed for expressing partisan incivility experience a positive effect of the norm prime. This effect is negative for respondents who have not internalized a norm against incivility. With respect to perceptions of other members of the ingroup, there is a positive treatment effect when respondents perceive that the reference group strongly disapproves of partisan incivility and a less negative effect when the respondent believes that no one in the reference group would use partisan incivility. Meanwhile, the norm prime decreases ratings of the opposing party most when respondents believe that their ingroup approves of partisan

incivility or that many in their ingroup would use incivility about the opposing party.

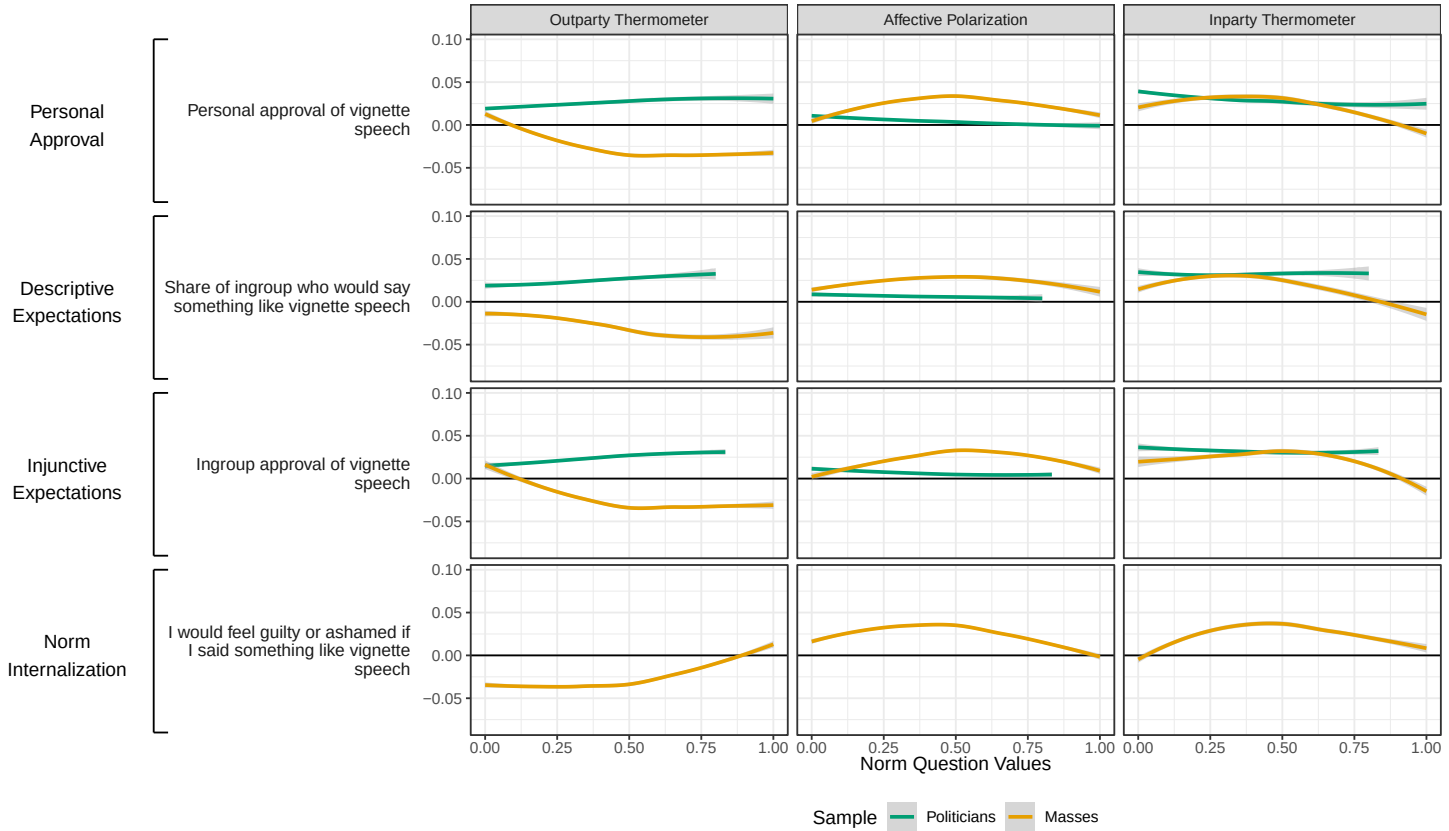
In Figure 5, I plot LOESS lines to predicted CATEs against measures of perceived sanctions for partisan incivility. In the mass sample, I find a similar pattern as above, consistent with my norms-based hypothesis about outparty hostility: respondents who believe their reference group is most likely to sanction partisan incivility have the most positive effects of the norm prime on outparty warmth. On the other hand, those who believe their reference group is unlikely to sanction incivility display negative conditional treatment effects. My results at the mass level suggest, therefore, that when the intended norm is primed (i.e., when respondents believe there is a strong norm against partisan incivility), it decreases expressed hostility toward the opposing party. However, since most members of the mass public do not hold such a norm, the question-order prime appears to invoke the perceived approval of partisan incivility, generating even more hostility toward the opposing party.

I highlight two additional findings from Figure 4 and Figure 5. First, the strongest trends in the politician CATEs appear to be for measures of voter and politician sanctions with outparty thermometer ratings (see Figure 5). These trends are in the opposite direction of expectations. Politicians who believe more voters would support an uncivil candidate report warmer feelings toward the outparty when the norm is primed. Similarly, politicians who believe more inparty politicians would sanction uncivil partisan rhetoric have a smaller CATE than those who believe incivility is less likely to be sanctioned. These trends suggest that politicians may be less susceptible to sanctions in norm enforcement. Second, across all norms measures for the mass sample, there is a curvilinear relationship between norm perceptions and affective polarization and inparty thermometer ratings. The largest predicted CATES are observed for people who disapprove slightly of uncivil rhetoric and who disagree somewhat with the statement measuring norm internalization. They are people who believe not many in their reference group would use partisan incivility and that their reference group disapproves slightly of incivility. Finally, the largest predicted CATEs are for those

who believe ingroup sanctions for incivility are neither likely nor unlikely. This consistent curvilinear relationship suggests that people who believe their ingroup norms are somewhat negative toward partisan incivility experience the strongest boost in inparty warmth from treatment. This offers suggestive evidence that the effect of norms on inparty affect may not be due to pride in strong inparty norms relating to incivility.

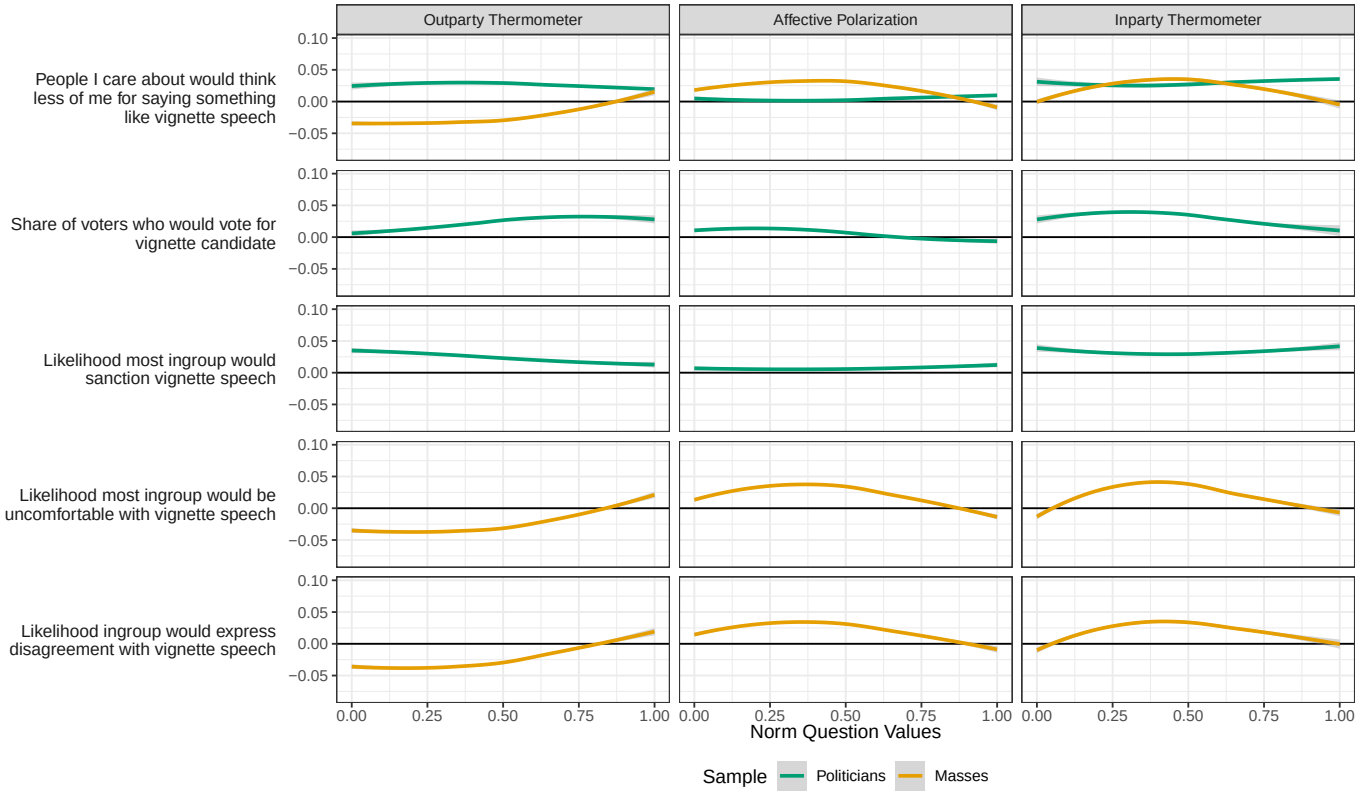
The exploratory results in this section suggest that feeling thermometer ratings of the opposing party are related to intra-party norms. For members of the public who believe there is a strong norm against incivility, priming this norm decreases their willingness to report negative feelings toward the opposing party. On the other hand, for the members of the public who believe there is a strong norm in favor of partisan incivility, priming this norm increases their willingness to report negative feelings toward the opposing party. The effects of the norm prime on inparty feeling thermometers are greatest when the ingroup norm is neutral to negative toward partisan incivility. While the sample of politicians is small, there appears to be much less variation in treatment effects for politicians across the norm measures.

Figure 4: Predicted CATE Estimates of Norm Prime and Perceived Norms



Note: LOESS with span of 1 fitted to predicted CATE estimates from causal forests with norm variables (rescaled to 0 to 1). Causal forests fit using uncivil political vignettes.

Figure 5: Predicted CATE Estimates of Norm Prime and Perceived Sanctions



Note: LOESS with span of 1 fitted to predicted CATE estimates from causal forests with norm variables (rescaled to 0 to 1). Causal forests fit using uncivil political vignettes.

Discussion

In this paper, I examine the prevalence of a norm against partisan incivility, offering new insights into the dynamics of civility by measuring incivility as a social norm. Original surveys of the mass public and politicians show that both groups believe that incivility is frowned upon by their ingroups relative to civil disagreement. However, in absolute terms, the mass public believes that social norms *favor* partisan incivility: they believe their ingroup peers would approve, and they believe that sanctions from their ingroup are unlikely. On the other hand, politicians strongly disapprove of partisan incivility and believe their inparty peers do as well.

I also connect social norms to the expression of partisan affective polarization. Priming the norm against incivility among politicians slightly increased warmth toward the outparty. At the same time, this norm significantly increased warmth toward the inparty and consequently, increased affective polarization. In the mass sample, the norm favoring incivility appears, if anything, to have slightly increased negative feelings toward the opposing party, while simultaneously increasing inparty warmth and affective polarization. The signs of the effects on outparty thermometer ratings are consistent with norms, indicating that norms surrounding incivility can influence partisan hostility; however, the unexpected favorable norm toward incivility may have led results from the mass sample in the opposite direction of expectations. Conditional average treatment effect estimates from causal forests ultimately corroborated the impact of norms on partisan hostility in the mass sample: the priming treatment increased warmth toward the opposing party among those who believed that the norm proscribed partisan incivility, while the treatment made respondents feel more negatively toward the outparty among those who believed that the norm supported partisan incivility.

It is possible that my results, particularly among politicians, are driven by social desirability: politicians may believe that it sounds desirable to say they and their inparty

peers disapprove of partisan incivility. That said, this seems unlikely for several reasons. First, politicians were taking part in an anonymous online survey, reducing potential social desirability incentives. Second, in other work, I find little evidence to suggest that politicians in this survey were answering in a manner consistent with social desirability bias (Frederick, 2024). Third, to the extent that politicians perceive incivility as socially undesirable and answer accordingly, this would be consistent with my argument: there is a norm against partisan incivility among politicians.

Another potential limitation of this study is the generalizability of the politician study. I relied on a sample of local politicians who may behave differently from politicians at other levels of government (e.g., state or federal). Local elected officials may have weaker motivations toward strong partisanship: many local governments are at least nominally nonpartisan and may have little electoral competition. Future work should examine whether strong norms hold among politicians at other levels of government. Still, there are some reasons to suspect that my findings would generalize to other politicians. Local politicians often run for higher office, meaning politicians at other levels of government could be drawn from the pool of local politicians with strong norms. Additionally, my survey questions about norms asked specifically about an inparty congressional candidate and defined the ingroup as other inparty politicians in their state. Thus, local elected officials believe that these norms exist among other inparty politicians in their state, presumably including politicians at higher levels of government.

We should be careful not to overinterpret differences between mass and elite conceptions of norms in this study. In their personal lives, politicians may have norms favoring partisan incivility like the masses. The survey of politicians focused specifically on norms for campaigns, whereas the mass survey centered around norms in informal social settings. Other studies could probe the sensitivity of these results to the setting: do politicians view partisan incivility similarly in non-electoral contexts? One piece of relevant evidence from

my survey worth highlighting is that politicians view people they care about as especially likely to sanction uncivil campaign rhetoric. If their closest circle disapproves of incivility in campaigning, it is conceivable that their closest circle would disapprove in other settings.

Two other elements of this study design bear mentioning. First, because I set out to measure and prime norms simultaneously, I sacrificed some experimental control over which norms, specifically, were primed. At the mass level, most people believed that few of their friends and family would use incivility (the descriptive norm disfavored incivility), while at the same time, they believed incivility was approved of socially (their normative expectations favored incivility). Thus, the pressure from normative and descriptive expectations ran in opposing directions. Respondents had heterogeneous perceptions of norms regarding incivility. As such, the norm prime meant different things to different people. One clear example of heterogeneous norm perceptions is the exploratory CATE analysis: some individuals believed in a norm against partisan incivility, and others believed the norm favored incivility. These two groups experienced different treatment effects. While allowing me to measure perceptions of norms, this aspect of the design induced large potential treatment effect heterogeneity, likely diluting the impact of the treatment on partisan hostility. In future work, I plan to manipulate perceptions of norms directly to determine whether stronger treatments are more effective. Second, due to space constraints in my surveys, I was unable to include longer batteries of items measuring post-treatment affective polarization. Rather, I relied solely on party feeling thermometer ratings as my outcome measures. Though party feeling thermometers are the workhorse of much affective polarization research, it seems probable that their connection to norms regarding incivility was less clear to respondents. Other measures like party trait ratings (Druckman and Levendusky, 2019), partisan antipathy (Finkel et al., 2024), or party dehumanization (Martherus et al., 2021) may be more directly related to the type of animosity conveyed by incivility. Future studies should examine whether norms regarding incivility have a stronger impact on other measures of partisan hostility.

Conclusion

Overall, my original surveys of the mass public and local politicians indicate that inparty norms play a nuanced but overlooked role in shaping affective polarization. Inparty norms that "when others go low, we go high" may decrease coldness toward the opposing party, while norms that "we kick them" may increase coldness toward the opposing party. This paper represents an additional step toward understanding the influence of social norms on political attitudes and behavior. In particular, my results contribute to the study of affective polarization, suggesting that future work on interparty relations should look inward to intra-party social dynamics. Indeed, those hoping to reduce partisan animosity may want to consider the role of social norms in structuring hostility. Interventions designed to change norms and reduce partisan hostility may, at first, lead only to superficial compliance, but over time, individuals often internalize these norms (Crandall, Eshleman and O'Brien, 2002; Kelman, 1961). This process of internalization could durably reduce affective polarization.

Further, this study points to the importance of including politicians in research on affective polarization and democratic norms. Politicians as a group may have different norms than the mass public, developed through their unique socialization process (McClosky and Brill, 1983; Stouffer, 1955). Indeed, norms among politicians may be singularly crucial to the upkeep of democratic institutions (Key, 1961). Though not directly comparable to the mass survey in my study, my survey of local politicians reveals strong social norms proscribing uncivil partisan campaigning. Different norms and socialization processes among politicians mean that the inclusion of politicians in affective polarization research could yield new insights into partisan animosity. Moreover, measuring democratic norms among politicians as *social* norms can generate a deeper understanding of democratic functioning as well as how democratic attitudes are sustained and undermined.

This research opens several avenues for exploration. Future research should analyze the precise mechanisms by which priming ingroup norms may increase inparty warmth. Does

this occur because inparty norms prompt a comparison with outparty norms (e.g., Dimant, 2024; Turner et al., 1987)? Or does inparty warmth increase because the norm prime raises the salience of inparty identity? An additional question raised by this study is the precise mechanism of normative influence. Norms can impact behavior through information provision, social approval, and identity definition. The treatment in this paper attempted to capture all three, and as such, I am unable to parse which mechanisms were most effective in influencing outcomes. Exploratory CATE analyses offer suggestive evidence for both normative and informational influence, but further work is needed to separate these factors. Finally, while this paper measures norms as individuals perceive them, this work does not explore how these individual perceptions of norms develop. Especially if norms can reduce partisan animosity, scholars should examine the dynamics of inparty norms to better understand affective polarization.

References

Abramowitz, Alan 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.

Ahler, Douglas J. and Gaurav Sood. 2018. "The Parties in Our Heads: Misperceptions about Party Composition and Their Consequences." *Journal of Politics* 80(3):964–981. Publisher: The Southern Political Science Association.

Allport, Gordon. 1954. *The Nature of Prejudice*. Cambridge: Perseus Books.

Asch, Solomon. 1955. "Opinions and Social Pressure." *Scientific American* 193(5):31–35.

Asch, Solomon. 1956. "Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority." *Psychological Monographs: General and Applied* 70(9):1–70.

Baxter-King, Ryan, Jacob Brown, Ryan Enos, Arash Naeim and Lynn Vavreck. 2022. "How Local Partisan Context Conditions Prosocial Behaviors: Mask Wearing During COVID-19." *PNAS* 119(21).

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York: Cambridge University Press.

Bicchieri, Cristina. 2017. *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford: Oxford University Press.

Bicchieri, Cristina and Erte Xiao. 2008. "Do the Right Thing: But Only if Others Do So." *Journal of Behavioral Decision Making* 22:191–208.

Blair, Robert, Jessica Gottlieb, Marie Schenk and Christopher Woods. 2023. "Depolarizing Within the Comfort of Your Party: Experimental Evidence from Online Workshops." *Working Paper* .
URL: *https://osf.io/preprints/osf/eha8r*

Blanchard, Fletcher, Christian Crandall, John Brigham and Leigh Ann Vaughn. 1994. "Condemning and Condoning Racism: A Social Context Approach to Interracial Settings." *Journal of Applied Psychology* 79(6):993–997.

Broockman, David and Daniel Butler. 2017. "The Causal Effects of Elite Position-Taking on Voter Attitudes: Field Experiments with Elite Communication." *American Journal of Political Science* 61(1):208–221.

Broockman, David, Joshua Kalla and Sean Westwood. 2022. "Does Affective Polarization Undermine Democratic Norms or Accountability? Maybe Not." *American Journal of Political Science* 67(3):808–828.

Bybee, Keith. 2016. *How Civility Works*. Stanford: Stanford University Press.

Cialdini, Robert, Raymond Reno and Carl Kallgren. 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology* 58(6):1015–1026.

Connors, Elizabeth. 2023. "Social Desirability and Affective Polarization." *Public Opinion Quarterly* 87(4):911–934.

Costa, Mia. 2021. "Ideology, Not Affect: What Americans Want from Political Representation." *American Journal of Political Science* 65(2):342–358.

Crandall, Christian, Amy Eshleman and Laurie O'Brien. 2002. "Social Norms and The Expression and Suppression of Prejudice: The Struggle for Internalization." *Journal of Personality and Social Psychology* 82(3):359–378.

Deutsch, Morton and Harold Gerard. 1955. "A Study of Normative and Informational Social Influences Upon Individual Judgment." *The Journal of Abnormal and Social Psychology* 51(3):629–636.

Dimant, Eugen. 2024. "Hate Trumps Love: The Impact of Political Polarization on Social Preferences." *Management Science* 70(1):1–31.

Druckman, James N and Matthew S Levendusky. 2019. "What Do We Measure When We Measure Affective Polarization?" *Public Opinion Quarterly* 83(1):114–122.

Druckman, James, S. Gubitz, Matthew Levendusky and Ashley Lloyd. 2019. "How Incivility on Partisan Media (De)Polarizes the Electorate." *Journal of Politics* 81(1):291–295.

Enders, Adam. 2021. "Issues vs. Affect: How Do Elite and Mass Polarization Compare?" *The Journal of Politics* 83(4):1873–1877.

Ferraro, Paul, Juan Jose Miranda and Michael Price. 2011. "The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment." *The American Economic Review: Papers & Proceedings* 101(3):318–322.

Festinger, Leon. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7(2):117–140.

Finkel, Eli, Alexander Landry, James Druckman, Jay Van Bavel and Rick Hoyle. 2024. "Partisan Antipathy and the Erosion of Democratic Norms." *Working Paper* .
URL: *https://www.ipr.northwestern.edu/documents/working-papers/2024/wp-24-01.pdf*

Frederick, Samuel. 2024. "Leading the Leaders: Who Drives Affective Polarization Among Politicians?". Working Paper.
URL: *https://samueldfrederick.com/Who_Polarizes.pdf*

Frederick, Samuel, Michael Miller and Donald Green. 2024. "Does Interaction With Out-Party Elites in a Classroom Setting Diminish Negative Partisanship?" *American Politics Research* .

Frimer, Jeremy and Linda Skitka. 2018. "The Montagu Principle: Incivility Decreases Politicians' Public Approval Even With Their Political Base." *Journal of Personality and Social Psychology: Interpersonal Relations and Group Processes* 115(5):845–866.

Frimer, Jeremy and Linda Skitka. 2020. "Americans Hold Their Political Leaders to a Higher Discursive Standard than Rank-and-File Co-Partisans." *Journal of Experimental Political Psychology* 86:1–19.

Gaertner, Samuel and John Dovidio. 1977. "The Subtlety of White Racism, Arousal, and Helping Behavior." *Journal of Personality and Social Psychology* 35(10):691–707.

Gerber, Alan, Donald Green and Christopher Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102(1):33–48.

Gervais, Brian. 2014. "Following the News? Reception of Uncivil Partisan Media and theUse of Incivility in Political Expression." *Political Communication* 31(4):564–583.

Gervais, Brian. 2015. "Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-Based Experiment." *Journal of Information Technology and Politics* 12(2):167–185.

Gervais, Brian. 2017. "More Than Mimicry? The Role of Anger in Uncivil Reactions to Elite Political Incivility." *International Journal of Public Opinion Research* 29(3):384–405.

Gitlin, Todd. 2013. The Uncivil and the Incendiary. In *Can We Talk? The Rise of Rude, Nasty, Stubborn Politics*, ed. Daniel Shea and Morris Fiorina. New York: Pearson Education, Inc.

Goldstein, Noah, Robert Cialdini and Vladas Griskevicius. 2008. "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels." *Journal of Consumer Research* 35(3):472–482.

Greenberg, Jeff and Tom Pyszczynski. 1985. "The Effect of an Overheard Ethnic Slur on Evaluations of the Target: How to Spread a Social Disease." *Journal of Experimental Social Psychology* 21:61–72.

Greene, Steven. 1999. "Understanding Party Identification: A Social Identity Approach." *Political Psychology* 20(2):393–403.

Haslam, S. Alexander, Penelope Oakes, Craig McGarty, John Turner, Katherine Reynolds and Rachel Eggins. 1996. "Stereotyping and Social Influence: The Mediation of Stereotype Applicability and Sharpness by the Views of In-Group and Out-Group Members." *British Journal of Social Psychology* 35:369–397.

Hogg, Michael and Barbara-A. Mullin. 1999. Joining Groups to Reduce Uncertainty: Subjective Uncertainty Reduction and Group Identification. In *Social Identity and Social Cognition*, ed. Dominic Abrams and Michael Hogg. Oxford: Blackwell Publishers.

Huber, Gregory and Neil Malhotra. 2017. "Political Homophily in Social Relationships: Evidence from Online Dating Behavior." *Journal of Politics* 79(1).

Huddy, Leonie and Omer Yair. 2021. "Reducing Affective Polarization: Warm Group Relations or Policy Compromise?" *Political Psychology* 42(2):291–309.

Iyengar, Shanto, Guarav Sood and Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social-Identity Perspective on Polarization." *Public Opinion Quarterly* 76(3):405–431.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra and Sean Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22:129–146.

Iyengar, Shanto, Yphtach Lelkes and Sean Westwood. 2024. "America's Political Pulse.". **URL:** *https://polarizationresearchlab.org/americas-political-pulse/*

Kalla, Joshua and David Broockman. 2022. "Voter Outreach Campaigns Can Reduce Affective Polarization among Implementing Political Activists: Evidence from Inside Three Campaigns." *American Political Science Review* 116(4):1516–1522.

Kalmoe, Nathan. 2020. *With Ballots and Bullets: Partisanship and Violence in the American Civil War*. New York: Cambridge University Press.

Kalmoe, Nathan and Lilliana Mason. 2022. *Radical American Partisanship: Mapping Violent Hostility, Its Causes, and the Consequences for Democracy.* Chicago: University of Chicago Press.

Kelman, Herbert. 1961. "Processes of Opinion Change." *Public Opinion Quarterly* 25(1):57–78.

Key, V.O. 1961. *Public Opinion and American Democracy.* New York: Alfred A. Knopf, Inc.

Kinder, Donald and Lynn Sanders. 1996. *Divided by Color: Racial Politics and Democratic Ideals.* Chicago: University of Chicago Press.

Kingzette, John, James Druckman, Samara Klar, Yanna Krupnikov, Matthew Levendusky and John Barry Ryan. 2021. "How Affective Polarization Undermines Support for Democratic Norms." *Public Opinion Quarterly* 85(2):663–677.

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

Levendusky, Matthew. 2023. *Our Common Bonds: Using What Americans Share to Bridge the Partisan Divide.* Chicago: University of Chicago Press.

Levendusky, Matthew and Dominik Stecula. 2021. *We Need to Talk: How Cross-Party Dialogue Reduces Affective Polarization.* Cambridge: Cambridge University Press.

Levendusky, Matthew and Neil Malhotra. 2016. "Does Media Coverage of Partisan Polarization Affect Political Attitudes?" *Political Communication* 33(2):283–301.

Levitan, Lindsey and Brad Verhulst. 2016. "Conformity in Groups: The Effects of Others' Views on Expressed Attitudes and Attitude Change." *Political Behavior* 38(2):277–315.

Levitsky, Steven and Daniel Ziblatt. 2018. *How Democracies Die.* New York: Crown Publishing.

Lipsitz, Keena, Grigore Pop-Eleches and Graeme Robertson. 2023. "Local Norms, Political Partisanship, and Pandemic Response: Evidence from the United States." *Perspectives on Politics* 22(1):247–262.

Martherus, James L., Andres G. Martinez, Paul K. Piff and Alexander G. Theodoridis. 2021. "Party Animals? Extreme Partisan Polarization and Dehumanization." *Political Behavior* 43(2):517–540.
URL: *https://doi.org/10.1007/s11109-019-09559-4*

Mason, Liliana. 2018. *Uncivil Agreement: How Politics Became Our Identity.* Chicago: University of Chicago Press.

Mayhew, David. 1974. *Congress: The Electoral Connection.* New Haven: Yale University Press.

McClosky, Herbert and Aida Brill. 1983. *Dimensions of Tolerance: What Americans Believe About Civil Liberties*. New York: Russell Sage Foundation.

McConahay, John, Betty Hardee and Valerie Batts. 1981. "Has Racism Declined in America? It Depends on Who Is Asking and What Is Asked." *Journal of Conflict Resolution* 25(4):563–579.

Meertens, Roel and Thomas Pettigrew. 1997. "Is Subtle Prejudice Really Prejudice?" *Public Opinion Quarterly* 61(1):54–71.

Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton: Princeton University Press.

Merica, Dan. 2018. "Eric Holder on Republicans: 'When They Go Low, We Kick Them'." *CNN* .
URL: *https://www.cnn.com/2018/10/10/politics/eric-holder-republicans-when-they-go-low/index.html*

Mullinix, Kevin and Trent Lythgoe. 2023. "Priming Norms to Combat Affective Polarization." *Political Research Quarterly* 76(1):186–199.

Munger, Kevin. 2017. "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment." *Political Behavior* 39(3):629–649.

Mutz, Diana. 2015. *In-Your-Face Politics: The Consequences of Uncivil Media*. Princeton: Princeton University Press.

Mutz, Diana and Byron Reeves. 2005. "The New Videomalaise: Effects of Televised Incivility on Political Trust." *American Political Science Review* 99(1):1–15.

Nolan, Jessica, Wesley Schultz, Robert Cialdini, Noah Goldstein and Vladas Griskevicius. 2008. "Normative Social Influence Is Underdetected." *Personality and Social Psychology Bulletin* 34(7):913–923.

Obama, Michelle. 2020. "Transcript: Michelle Obama's DNC Speech." *CNN* .
URL: *https://www.cnn.com/2020/08/17/politics/michelle-obama-speech-transcript/index.html*

Paluck, Elizabeth Levy. 2009. "Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda." *Journal of Personality and Social Psychology* 96(3):574–587.

Peters, Uwe. 2021. "How (Many) Descriptive Claims About Political Polarization Exacerbate Polarization." *Journal of Social and Political Psychology* 9(1):24–36.

Rossiter, Erin and Taylor Carlson. 2024. "Cross-Partisan Conversation Reduced Affective Polarization for Republicans and Democrats Even after the Contentious 2020 Election." *Journal of Politics* .

Schultz, P. Wesley, Jessica Nolan, Robert Cialdini, Noah Goldstein and Vladas Griskevicius. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science* 18(5):429–434.

Sechrist, Gretchen and Charles Stangor. 2001. "Perceived Consensus Influences Intergroup Behavior and Stereotype Accessibility." *Journal of Personality and Social Psychology* 80(4):645–654.

Sechrist, Gretchen and Charles Stangor. 2007. "When Are Intergroup Attitudes Based on Perceived Consensus Information? The Role of Group Familiarity." *Social Influence* 2(3):211–235.

Sherif, Muzafer. 1965. *The Psychology of Social Norms.* New York: Octagon Books, Inc. original-date: 1936.

Sherif, Muzafer and Carolyn Sherif. 1966. *Groups in Harmony and Tension: An Integration of Studies on Intergroup Relations.* New York: Octagon Books, Inc.

Skyette, Rasmus. 2021. "Dimensions of Elite Partisan Polarization: Disentangling the Effects of Incivility and Issue Polarization." *British Journal of Political Science* 51(4):1457–1475.

Sparkman, Gregg and Gregory Walton. 2017. "Dynamic Norms Promote Sustainable Behavior, Even if It Is Counternormative." *Psychological Science* 28(11):1663–1674.

Stangor, Charles, Gretchen Sechrist and John Jost. 2001. "Changing Racial Beliefs By Providing Consensus Information." *Personality and Social Psychology Bulletin* 27(4):486–496.

Stephens-Davidowitz, Seth. 2017. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are.* Harper Collins.

Stone, Daniel. 2023. *Undue Hate: A Behavioral Economic Analysis of Hostile Polarization in US Politics and Beyond.* Cambridge: MIT Press.

Stouffer, Samuel. 1955. *Communism, Conformity, and Civil Liberties: A Cross-Section of America Speaks Its Mind.* New York: Doubleday.

Suhay, Elizabeth. 2015. "Explaining Group Influence: The Role of Identity and Emotion in Political Conformity and Polarization." *Political Behavior* 37(1):221–251.

Tajfel, Henri and John Turner. 1986. The Social Identity Theory of Intergroup Behavior. In *The Psychology of Intergroup Relations*, ed. Stephen Worchel and William Austin. Chicago: Hall Publishers.

Terry, Deborah and Michael Hogg. 1996. "Group Norms and the Attitude-Behavior Relationship: A Role for Group Identification." *Personality and Social Psychology Bulletin* 22(8):776–793.

Terry, Deborah, Michael Hogg and Leda Blackwood. 2001. Prejudiced Attitudes, Group

Norms, and Discriminatory Behavior. In *Understanding Prejudice, Racism, and Social Conflict*, ed. Martha Augoustinos and Katherine Reynolds. London: Sage Publications Ltd. pp. 140–155.

Toff, Benjamin and Elizabeth Suhay. 2019. "Partisan Conformity, Social Identity, and the Formation of Policy Preferences." *International Journal of Public Opinion Research* 31(2):349–367.

Turner, John. 1982. Towards a Cognitive Redefinition of the Social Group. In *Social Identity and Intergroup Relations*, ed. Henri Tajfel. Cambridge: Cambridge University Press.

Turner, John, Michael Hogg, Penelope Oakes, Stephen Reicher and Margaret Wetherell. 1987. *Rediscovering the Social Group: A Self-Categorization Theory*. New York: Basil Blackwell Inc.

Voelkel, Jan G., James Chu, Michael N. Stagnaro, Joseph S. Mernyk, Chrystal Redekopp, Sophia L. Pink, James N. Druckman, David G. Rand and Robb Willer. 2023. "Interventions reducing affective polarization do not necessarily improve anti-democratic attitudes." *Nature Human Behaviour* 7(1):55–64.

Wager, Stefan and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association* 113(523):1228–1242.

White, Ismail and Chryl Laird. 2020. *Steadfast Democrats: How Social Forces Shape Black Political Behavior*. Princeton: Princeton University Press.

White, Katherine, Michael Hogg and Deborah Terry. 2002. "Improving Attitude-Behavior Correspondence Through Exposure to Normative Support From a Salient Ingroup." *Basic and Applied Social Psychology* 24(2):91–103.

Wojcieszak, Magdalena, Stephan Winter and Xudong Yu. 2020. "Social Norms and Selectivity: Effects of Norms of Open-Mindedness on Content Selection and Affective Polarization." *Mass Communication and Society* 23(4):455–483.

Wu, Jennifer and Gregory Huber. 2021. "Partisan Differences in Social Distancing May Originate in Norms and Beliefs: Results from Novel Data." *Social Science Quarterly* 102(5):2251–2265.

Zeitzoff, Thomas. 2023. *Nasty Politics: The Logic of Insults, Threats, and Incitement*. New York: Oxford University Press.

Zhong, Chen-Bo, Adam Galinsky and Miguel Unzueta. 2008. "Negational Racial Identity and Presidential Voting Preferences." *Journal of Experimental Social Psychology* 44(6):1563–1566.

Zhong, Chen-Bo, Katherine Phillips, Geoffrey Leonardelli and Adam Galinsky. 2008. "Nega-

tional Categorization and Intergroup Behavior." *Personality and Social Psychology Bulletin* 34(6):793–806.

Álvarez Benjumea, Amalia and Fabian Winter. 2020. "The Breakdown of Antiracist Norms: A Natural Experiment on Hate Speech After Terrorist Attacks." *PNAS* 117(37):22800–22804.

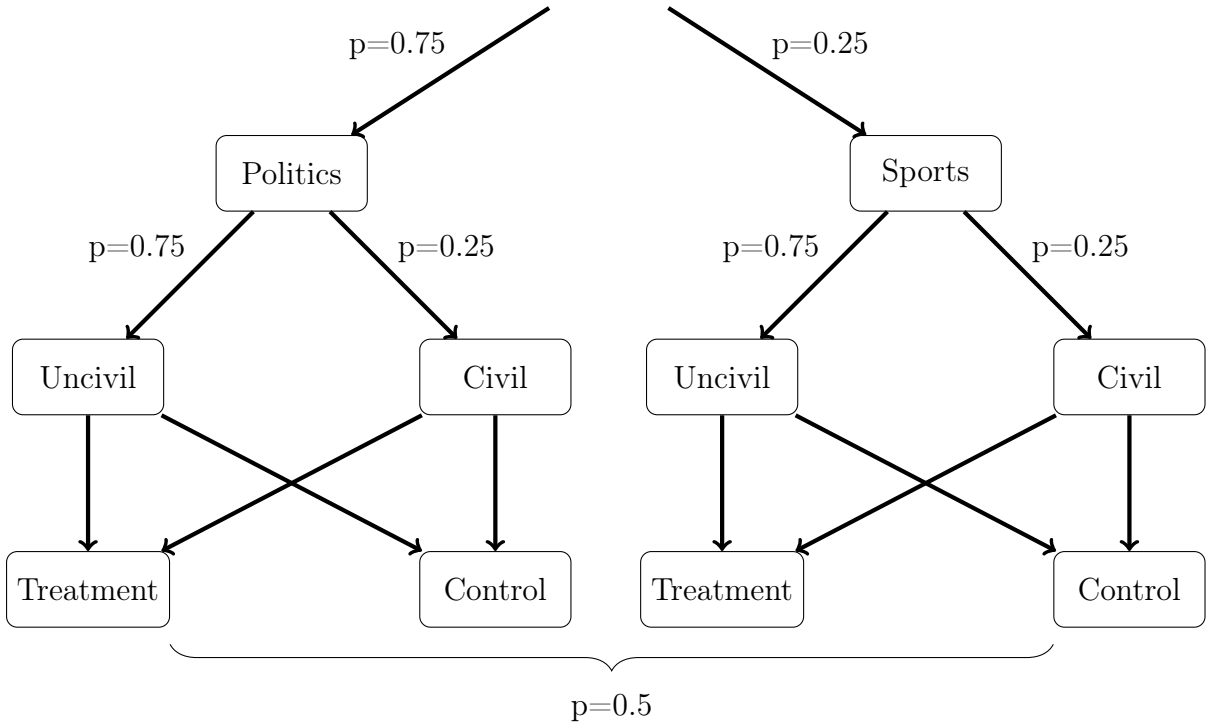
Supplementary Information for "Intra-Party Norms and Affective Polarization"

Samuel Frederick

Contents

A Details of Mass Survey

A.1 Vignette Randomization



A.2 Vignette Content

Imagine you are at a social event with your friends and/or family [PERSON TYPE] discussing [TOPIC]. Someone in your group [VIGNETTE CONTENT].

TOPIC	**PERSON TYPE**	**Civility**	**VIGNETTE CONTENT**
sports	who are fans of a sports team	Uncivil	says that fans of a rival sports team are scum and an embarrassment to the country says that a rival sports team is full of dirty cheaters who can only win by fraud/rigging the game says that a rival sports team is disgraceful and is ruining the sport

TOPIC	PERSON TYPE	Civility	VIGNETTE CONTENT
			says that fans of a rival sports team are classless losers
		Civil	says that fans of a rival sports team are misguided says that they are disappointed by the decisions of a rival sports team says that they disagree with the fans of a rival sports team says that fans of a rival sports team are making flawed decisions
politics	who are [INPARTY]s	Uncivil	says that [OUTPARTY]s are scum and an embarrassment to the country says that the [OUTPARTY] is full of dirty cheaters who can only win by fraud/rigging the game says that [OUTPARTY]s are disgraceful and are ruining the country says that [OUTPARTY]s are classless losers calls [OUTPARTY]s traitors who have betrayed the country
		Civil	says that [OUTPARTY]s are misguided
politics	who are [INPARTY]s	Civil	says that they are disappointed by the decisions of the [OUTPARTY] says that [OUTPARTY]s are making flawed decisions says they disagree with [OUTPARTY]s on policy

A.3 Mass Survey Questionnaire

Party Feeling Thermometers

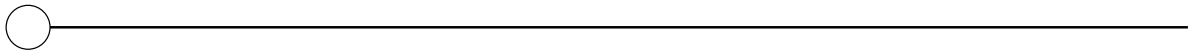
dem_ft. We'd like you to rate how you feel toward some groups on a scale of 0 to 100. Zero means very cold and unfavorable and 100 means very warm and favorable. Fifty means you

do not feel cold or warm.

How would you rate your feelings toward the Democratic Party?

Very cold and negative Neither cold nor warm Very warm and positive

0 100

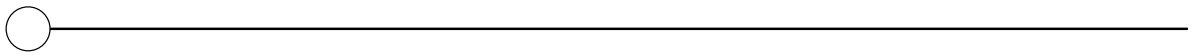


rep_ft. We'd like you to rate how you feel toward some groups on a scale of 0 to 100. Zero means very cold and unfavorable and 100 means very warm and favorable. Fifty means you do not feel cold or warm.

How would you rate your feelings toward the Republican Party?

Very cold and negative Neither cold nor warm Very warm and positive

0 100



Norms Questions

[Norms Questions block comes either before (Treatment) or after (Control) Party Feeling Thermometers (**dem_ft** and **rep_ft**).]

[topic randomized to one of {"sports", "politics"}.]

[person_type matched to **topic**, takes on a value of one of {"who are fans of a sports team", "who are [INPARTY]s"}.]

Personal Approval

norm_personalapprove. Would you personally approve or disapprove of saying something like this while discussing ${topic}?

- ○ Strongly disapprove
- ○ Disapprove
- ○ Somewhat disapprove
- ○ Neither approve nor disapprove
- ○ Somewhat approve
- ○ Approve

- O Strongly approve

Normative Expectations

norm_normative. Would most of your friends and family members ${person_type} approve or disapprove of saying something like this while discussing ${topic}?

- O Strongly disapprove
- O Disapprove
- O Somewhat disapprove
- O Neither approve nor disapprove
- O Somewhat approve
- O Approve
- O Strongly approve

Descriptive Expectations

norm_descriptive. How many of your friends and family members ${person_type} would say something like this while discussing ${topic}?

- O None
- O A few
- O Some
- O Most
- O Almost all
- O All

Reference Group Sanctions

norm_sanction_uncomfortable. How likely is it that most of your friends and family members ${person_type} would feel uncomfortable if someone said something like this while discussing ${topic}?

- O Extremely unlikely
- O Unlikely
- O Somewhat unlikely
- O Neither likely nor unlikely

- O Somewhat likely
- O Likely
- O Extremely likely

norm_peoplecare. Do you agree or disagree with the following statement?

People I care about would think less of me if I said something like this while discussing ${topic}.

- O Strongly disagree
- O Disagree
- O Somewhat disagree
- O Neither agree nor disagree
- O Somewhat agree
- O Agree
- O Strongly agree

norm_sanction_disagree. How likely is it that your friends and family members ${person_type} would express disagreement with the person saying something like this while discussing ${topic}?

- O Extremely unlikely
- O Unlikely
- O Somewhat unlikely
- O Neither likely nor unlikely
- O Somewhat likely
- O Likely
- O Extremely likely

Norm Internalization

norm_internalized. Do you agree or disagree with the following statement?

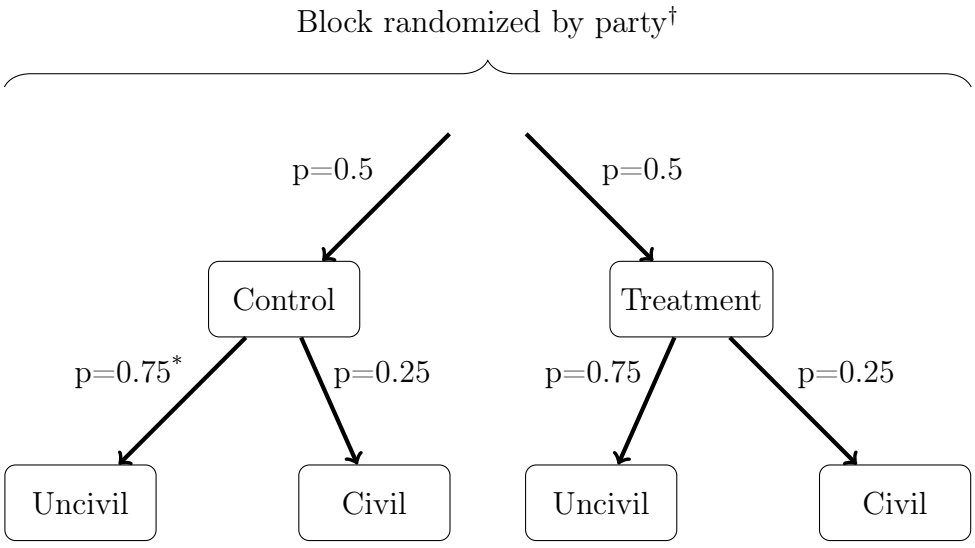
I would feel guilty or ashamed if I said something like this while discussing ${topic}.

- O Strongly disagree
- O Disagree

- O Somewhat disagree
- O Neither agree nor disagree
- O Somewhat agree
- O Agree
- O Strongly agree

B Details of Politician Survey

B.1 Vignette Randomization



† Blocks: Democratic Identifier/Leaner, Republican Identifier/Leaner, Independent, Independent who prefers Democratic Party on feeling thermometer ratings, Independent who prefers Republican Party on feeling thermometer ratings

* Randomization occurred in Qualtrics, giving equal weight to each VIGNETTE CONTENT within civil examples and equal weight to each VIGNETTE CONTENT within uncivil examples with total probabilities equal to 0.75 for uncivil examples and 0.25 for civil examples.

B.2 Vignette Content

Imagine [AN INPARTY] politician in your state who campaigns for U.S. Congress by [VIGNETTE CONTENT].

Civility	VIGNETTE CONTENT
Uncivil	calling [OUTPARTY]s enemies of your area who threaten to destroy the state
	saying [OUTPARTY]s are corrupt and are trying to defraud the citizens of your state
	calling [OUTPARTY]s traitors who have betrayed your state
	calling [OUTPARTY]s thugs and criminals who are a danger to your state
Civil	saying [OUTPARTY]s are not what is best for your state
	saying [OUTPARTY]s would move your state in the wrong direction
	saying they disagree with [OUTPARTY]s on policy

B.3 Politician Survey Questionnaire

First, we'd like to ask your opinions about a range of policy proposals that local elected leaders might face.

Policy Questions

While we recognize that the details of any policy are important, generally speaking, to what extent would you support or oppose each of the following proposals?

Next, we'd like to ask about your political stances and your experience working in local government.

Party Identification Questions

Main Party Question

party. Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what?

- ○ Republican
- ○ Democrat
- ○ Independent
- ○ No preference
- ○ Other, please specify:

	Strongly support	Support	Somewhat support	Neither support nor oppose	Somewhat oppose	Oppose	Strongly Oppose
Expanding voucher options for students to attend private or religious schools.	O	O	O	O	O	O	O
Expanding collective bargaining rights for public sector employees.	O	O	O	O	O	O	O
Increasing the use of cameras, speedbumps, or other traffic tools in residential areas.	O	O	O	O	O	O	O
Increasing government spending on park maintenance.	O	O	O	O	O	O	O
Rezoning residential areas to allow for more commercial use.	O	O	O	O	O	O	O

Party Strength

[Display if **party** is Republican or Democrat]

pty_strength. Would you call yourself a strong ${party} or a not very strong ${party}?

- O Strong ${party}
- O Not very strong ${party}

Party Lean

[Display if **party** is *not* Republican or Democrat]

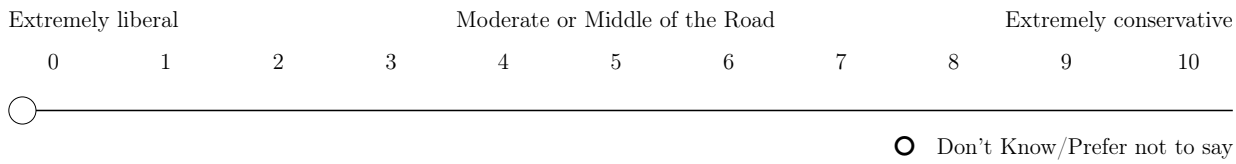
pty_lean. Do you think of yourself as closer to the Republican Party or to the Democratic Party?

- O Closer to the Republican Party
- O Neither

○ Closer to the Democratic Party

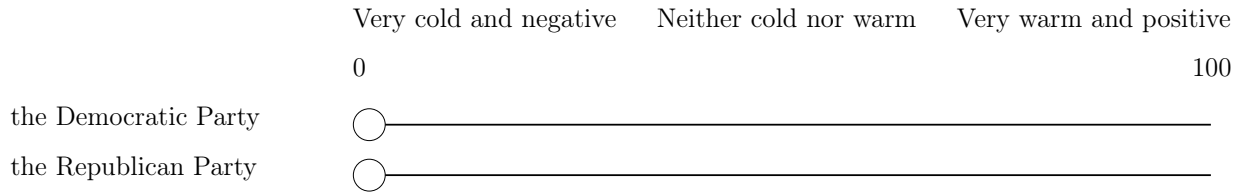
Ideology Question

ideology. Here is an 11-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale? If you don't know or would prefer not to say, please select "Don't know/Prefer not to say."



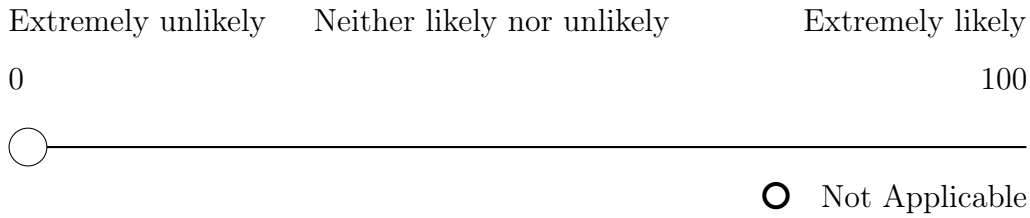
Party Feeling Thermometers

pty_ft. On a scale from 0 (very cold and negative) to 100 (very warm and positive), how do you feel toward...



Perceived Electoral Competition

competition. We know it's challenging to predict, but roughly speaking, how likely do you think it is that you will face a competitive election (e.g., a close race against an opponent) against a ${outparty} candidate in your next campaign?



Years in Government

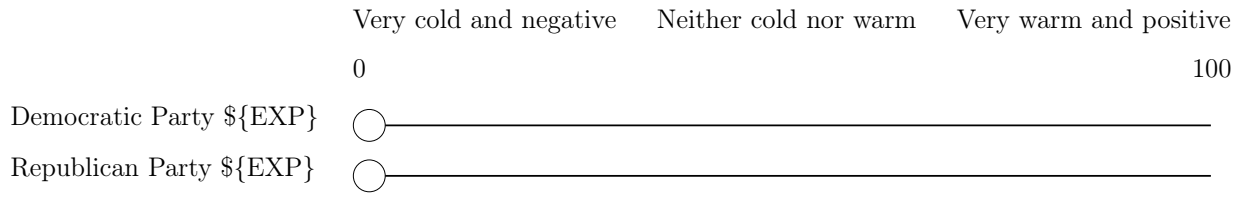
gov_exp. Over your career, how many years have you served in government IN TOTAL?

[Experiment for Other Study]

Party-Group Feeling Thermometers

[EXP block randomized with equal probability by party group and question-order treatment to one of {"activists", "candidates and elected officials", "voters"}]

voter_exp_ft. On a scale from 0 (very cold and negative) to 100 (very warm and positive), how do you feel toward...



Norms Questions

[Norms Questions block comes either before (Treatment) or after (Control) Party-Group Feeling Thermometers (**voter_exp_ft**).]

These next few questions ask your opinions about campaigning for elected office.

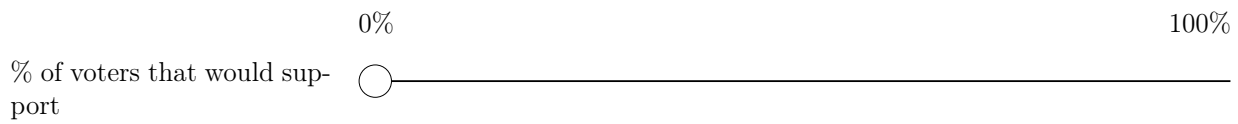
Personal Approval

norm_personalapprove. Would you personally approve or disapprove of campaigning like this?

- ○ Strongly disapprove
- ○ Disapprove
- ○ Somewhat disapprove
- ○ Neither approve nor disapprove
- ○ Somewhat approve
- ○ Approve
- ○ Strongly approve

Voter Sanctions

norms_votersanction. We know it's challenging to predict, but roughly what percent of voters in your state do you think would vote for a politician campaigning like this in a general election?



Descriptive Expectations

norm_descriptive. How many ${inparty} politicians in your state do you think would campaign like this?

- ○ None
- ○ A few
- ○ Some
- ○ Most
- ○ Almost all
- ○ All

Normative Expectations

norm_normative. Do you think most ${inparty} politicians in your state would approve or disapprove of campaigning like this?

- ○ Strongly disapprove
- ○ Disapprove
- ○ Somewhat disapprove
- ○ Neither approve nor disapprove
- ○ Somewhat approve
- ○ Approve
- ○ Strongly approve

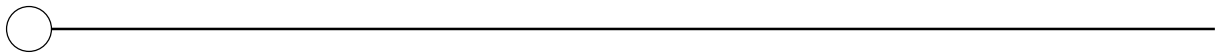
Politician Sanctions

[SANCTION block randomized with equal probability by party group and question-order treatment to one of {"make a point of avoiding the politician campaigning like this in the future", "refuse to help this politician's campaign"}.]

norms_polsantion. We know it's challenging to predict, but roughly how likely do you think it is that most ${inparty} politicians in your state would ${SANCTION}?

Extremely unlikely Neither likely nor unlikely Extremely likely

0 100



Reference Group Sanctions

[Question order randomized with **norms_polsanction**.]

norm_peoplecare. Do you agree or disagree with the following statement?

People I care about would think less of me if I campaigned like this.

- ○ Strongly disagree
- ○ Disagree
- ○ Somewhat disagree
- ○ Neither agree nor disagree
- ○ Somewhat agree
- ○ Agree
- ○ Strongly agree

C Regression Tables

C.1 Mass Regression Models

This section displays the full regression models from the main text for the mass sample in addition to pre-registered models with LASSO covariate selection (e.g., Belloni, Chernozhukov and Hansen, 2014). LASSO-selected covariates are mean-centered. The models from the main text are presented in the columns labeled "(1)." In Section C.1.1, I show that my estimates of the uncivil treatment effects are robust to the inclusion of covariates, changing little depending on specification. My estimates of the effect of the norm prime are somewhat more sensitive to the inclusion of covariates, declining in magnitude, and the ATE estimate of the norm treatment loses statistical significance after including covariates (see Section C.1.2)

C.1.1 Norm Outcomes and Uncivil Treatment

	Personal Approval		Group Descriptive		Group Approval		People Care	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Uncivil Treatment	-0.163***	-0.161***	-0.144***	-0.141***	-0.143***	-0.145***	0.119***	0.121***
	(0.018)	(0.016)	(0.018)	(0.017)	(0.016)	(0.015)	(0.018)	(0.018)
Constant	0.655***	0.654***	0.482***	0.480***	0.668***	0.669***	0.333***	0.331***
	(0.015)	(0.014)	(0.016)	(0.015)	(0.014)	(0.013)	(0.016)	(0.015)
Num. Obs.	1229	1229	1229	1229	1229	1229	1229	1229
Covariates	-	✓	-	✓	-	✓	-	✓
Num. Covariates	-	18	-	25	-	24	-	17

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

	Group Uncomfortable		Group Confront		Internalized	
	(1)	(2)	(1)	(2)	(1)	(2)
Uncivil Treatment	0.123***	0.128***	0.109***	0.119***	0.153***	0.160***
	(0.018)	(0.017)	(0.018)	(0.016)	(0.019)	(0.018)
Constant	0.350***	0.347***	0.346***	0.339***	0.319***	0.313***
	(0.016)	(0.014)	(0.015)	(0.014)	(0.017)	(0.015)
Num. Obs.	1229	1229	1229	1229	1229	1229
Covariates	-	✓	-	✓	-	✓
Num. Covariates	-	21	-	15	-	19

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

C.1.2 Affective Polarization and Norm Treatment

	Affective Polarization		Inparty Thermometer		Outparty Thermometer	
	(1)	(2)	(1)	(2)	(1)	(2)
Norm Treatment	4.839*	3.266+	2.297	1.197	-2.519	-1.977
	(2.235)	(1.886)	(1.437)	(1.242)	(1.538)	(1.354)
Constant	50.054***	50.865***	73.924***	74.491***	23.870***	23.597***
	(1.607)	(1.358)	(1.017)	(0.880)	(1.097)	(0.972)
Num. Obs.	932	932	932	932	933	933
Covariates	-	✓	-	✓	-	✓
Num. Covariates	-	24	-	27	-	25

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

C.2 Politician Regression Models

This section displays the full regression models from the main text for the politician sample in addition to pre-registered models with LASSO covariate selection (e.g., Belloni, Chernozhukov and Hansen, 2014). LASSO-selected covariates are mean-centered. For some outcomes, cross-validated LASSO selected no covariates, meaning the covariate models are equivalent to the base model from the main text. The models from the main text are presented in the columns corresponding to "Party Identifiers" labeled with "(1)." The columns labeled "Thermometer Partisans" include politicians who did not identify with or lean toward one of the parties on the standard party identification questions but who did express a preference for one of the parties on pre-treatment feeling thermometers.

Section C.2.1 shows that my estimates of the effects of the uncivil treatment on norm outcomes are robust to both the inclusion of covariates and the data subset. Including feeling-thermometer partisans does not alter the statistical or substantive significance of the coefficient estimates. Unlike in the mass sample, the estimates of the effects of the norm prime mostly do not change in substantive or statistical significance after adjusting for pre-treatment covariates or including feeling-thermometer partisans. In particular, the significant inparty thermometer effect estimates remain positive and statistically significant even after adjusting for covariates. The notable exception to the broader robustness of the effects among politicians are the estimates of the norm prime's ATE for outparty thermometer ratings: the ATE estimate declines somewhat after including covariates but declines much more when feeling-thermometer partisans are included. This decline is consistent with an identity-based explanation for norms results because feeling-thermometer partisans feel much less warmly toward their "inparty" and, by definition, do not identify with either party. Thus,

their behavior may be less responsive to "inparty" norms, leading to the observed decrease in the estimate of the norm treatment for outparty thermometers.

C.2.1 Norm Outcomes and Uncivil Treatment

	Personal Approval				Group Descriptive			
	Party Identifiers		Thermometer Partisans		Party Identifiers		Thermometer Partisans	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Uncivil Treatment	-0.332***	-0.313***	-0.327***	-0.311***	-0.186***	-0.183***	-0.190***	-0.188***
	(0.032)	(0.032)	(0.030)	(0.029)	(0.024)	(0.024)	(0.022)	(0.022)
Constant	0.593***	0.327***	0.581***	0.320***	0.498***	0.349***	0.499***	0.348***
	(0.029)	(0.012)	(0.027)	(0.011)	(0.022)	(0.009)	(0.020)	(0.008)
Num. Obs.	448	431	487	471	448	445	487	484
Covariates	-	✓	-	✓	-	✓	-	✓
Num. Covariates	-	14	-	12	-	3	-	3

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

	Group Approval				People Care			
	Party Identifiers		Thermometer Partisans		Party Identifiers		Thermometer Partisans	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Uncivil Treatment	-0.233***	-0.232***	-0.237***	-0.236***	0.245***	0.237***	0.231***	0.231***
	(0.028)	(0.028)	(0.026)	(0.026)	(0.033)	(0.033)	(0.031)	(0.031)
Constant	0.617***	0.431***	0.618***	0.430***	0.494***	0.692***	0.515***	0.515***
	(0.024)	(0.011)	(0.022)	(0.011)	(0.030)	(0.013)	(0.028)	(0.028)
Num. Obs.	448	446	487	485	449	434	488	488
Covariates	-	✓	-	✓	-	✓	-	✓
Num. Covariates	-	1	-	1	-	6	-	0

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

	Voter Support				Ingroup Sanctions			
	Party Identifiers		Thermometer Partisans		Party Identifiers		Thermometer Partisans	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Uncivil Treatment	-0.081***	-0.081***	-0.067**	-0.067**	0.130***	0.128***	0.126***	0.126***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.024)	(0.024)	(0.023)	(0.023)
Constant	0.556***	0.556***	0.541***	0.541***	0.365***	0.467***	0.374***	0.473***
	(0.019)	(0.019)	(0.018)	(0.018)	(0.021)	(0.011)	(0.020)	(0.010)
Num. Obs.	441	441	479	479	443	440	481	478
Covariates	-	✓	-	✓	-	✓	-	✓
Num. Covariates	-	0	-	0	-	4	-	4

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

C.2.2 Affective Polarization and Norm Treatment

	Affective Polarization			
	Party Identifiers		Thermometer Partisans	
	(1)	(2)	(1)	(2)
Norm Treatment	2.955	3.460	4.004+	4.025+
	(2.503)	(2.444)	(2.378)	(2.281)
Affective Polarization (Pre)	0.707***	0.624***	0.708***	0.622***
	(0.043)	(0.047)	(0.041)	(0.046)
Activist Thermometer	0.014	-1.405	0.595	-1.235
	(3.204)	(3.053)	(3.041)	(2.886)
Candidate Thermometer	6.105*	5.367+	6.778*	6.071*
	(2.958)	(2.996)	(2.852)	(2.810)
Constant	1.015	36.137***	-0.449	34.609***
	(3.055)	(1.713)	(2.860)	(1.583)
Num. Obs.	353	346	381	373
Covariates	-	✓	-	✓
Num. Covariates	-	11	-	10

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

	Inparty Thermometer			
	Party Identifiers		Thermometer Partisans	
	(1)	(2)	(1)	(2)
Norm Treatment	4.794*	4.763*	5.040**	4.680*
	(1.898)	(1.901)	(1.923)	(1.892)
Inparty Thermometer (Pre)	0.559***	0.512***	0.589***	0.539***
	(0.065)	(0.066)	(0.062)	(0.067)
Activist Thermometer	-18.069***	-13.311***	-18.271***	-14.371***
	(2.719)	(2.472)	(2.692)	(2.425)
Candidate Thermometer	-4.061*		-3.814+	
	(2.037)		(2.083)	
Voter Thermometer		4.526*		4.485*
		(2.000)		(2.044)
Constant	32.939***	62.566***	29.978***	61.109***
	(4.598)	(1.289)	(4.473)	(1.279)
Num. Obs.	356	349	384	376
Covariates	-	✓	-	✓
Num. Covariates	-	7	-	11

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

	Outparty Thermometer			
	Party Identifiers		Thermometer Partisans	
	(1)	(2)	(1)	(2)
Norm Treatment	2.007	1.512	0.939	0.243
	(1.871)	(1.864)	(1.808)	(1.812)
Outparty Thermometer (Pre)	0.717***	0.636***	0.702***	0.617***
	(0.048)	(0.055)	(0.048)	(0.056)
Activist Thermometer	-18.190***	-5.732**	-19.074***	-6.639***
	(2.453)	(2.073)	(2.387)	(1.948)
Candidate Thermometer	-11.669***		-11.968***	
	(2.298)		(2.264)	
Voter Thermometer		12.047***		12.075***
		(2.273)		(2.230)
Constant	22.468***	26.657***	23.394***	26.977***
	(2.639)	(1.388)	(2.659)	(1.384)
Num. Obs.	353	345	381	372
Covariates	-	✓	-	✓
Num. Covariates	-	18	-	13

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

C.2.3 Affective Polarization and Norm Treatment, Uncivil Interaction

	Affective Polarization			
	Party Identifiers		Thermometer Partisans	
	(1)	(2)	(1)	(2)
Norm Treatment	-1.298	-1.705	-1.925	-3.060
	(4.275)	(4.277)	(4.050)	(4.091)
Uncivil Treatment	-5.554+	-6.989*	-6.738*	-7.497*
	(3.159)	(3.227)	(2.999)	(3.012)
Treatment*Uncivil	4.210	4.868	5.909	6.751
	(4.986)	(4.878)	(4.715)	(4.673)
Affective Polarization (Pre)	0.705***	0.630***	0.707***	0.633***
	(0.038)	(0.041)	(0.036)	(0.039)
Activist Thermometer	-0.623	-1.541	-0.003	-1.059
	(2.828)	(2.669)	(2.673)	(2.524)
Candidate Thermometer	5.220*	4.643+	5.901*	5.360*
	(2.605)	(2.616)	(2.509)	(2.477)
Constant	7.179*	43.119***	6.818*	42.278***
	(3.627)	(2.771)	(3.408)	(2.596)
Num. Obs.	441	432	480	470
Covariates	-	✓	-	✓
Num. Covariates	-	12	-	13

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

	Inparty Thermometer			
	Party Identifiers		Thermometer Partisans	
	(1)	(2)	(1)	(2)
Norm Treatment	0.540	-0.172	0.490	-0.021
	(3.487)	(3.477)	(3.295)	(3.328)
Uncivil Treatment	-3.066	-3.959+	-3.966+	-4.091+
	(2.249)	(2.339)	(2.178)	(2.260)
Treatment*Uncivil	4.142	4.188	4.462	4.382
	(4.057)	(3.966)	(3.874)	(3.880)
Inparty Thermometer (Pre)	0.542***	0.478***	0.570***	0.535***
	(0.056)	(0.058)	(0.054)	(0.056)
Activist Thermometer	-18.315***	-13.189***	-18.177***	-13.870***
	(2.377)	(2.110)	(2.327)	(2.111)
Candidate Thermometer	-4.459*		-3.809*	
	(1.821)		(1.844)	
Voter Thermometer		5.042**		4.418*
		(1.793)		(1.792)
Constant	37.351***	66.320***	35.155***	64.921***
	(4.427)	(1.959)	(4.254)	(1.861)
Num. Obs.	444	433	483	473
Covariates	-	✓	-	✓
Num. Covariates	-	13	-	9

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

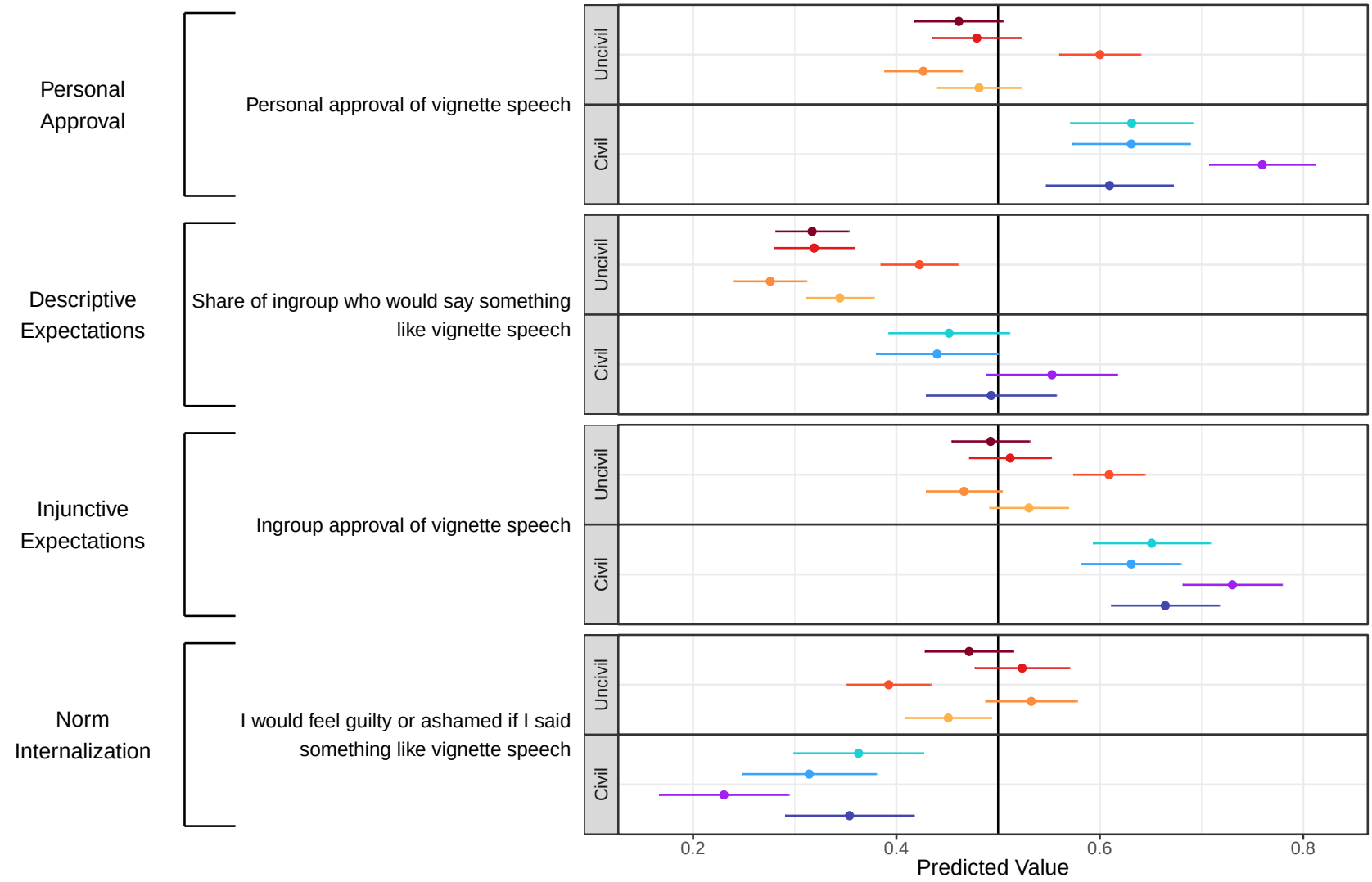
	Outparty Thermometer			
	Party Identifiers		Thermometer Partisans	
	(1)	(2)	(1)	(2)
Norm Treatment	2.648	2.362	2.897	3.450
	(3.022)	(3.081)	(2.916)	(2.916)
Uncivil Treatment	2.735	3.048	2.992	3.741
	(2.506)	(2.606)	(2.419)	(2.418)
Treatment*Uncivil	-0.575	-0.703	-1.907	-3.005
	(3.557)	(3.573)	(3.424)	(3.447)
Outparty Thermometer (Pre)	0.695***	0.625***	0.689***	0.631***
	(0.043)	(0.048)	(0.042)	(0.047)
Activist Thermometer	-17.860***	-6.227***	-18.354***	-6.956***
	(2.164)	(1.746)	(2.091)	(1.685)
Candidate Thermometer	-10.915***		-10.878***	
	(2.058)		(2.023)	
Voter Thermometer		10.965***		10.691***
		(2.041)		(2.016)
Constant	19.756***	23.026***	20.037***	22.605***
	(2.936)	(2.242)	(2.864)	(2.024)
Num. Obs.	441	431	480	470
Covariates	-	✓	-	✓
Num. Covariates	-	20	-	12

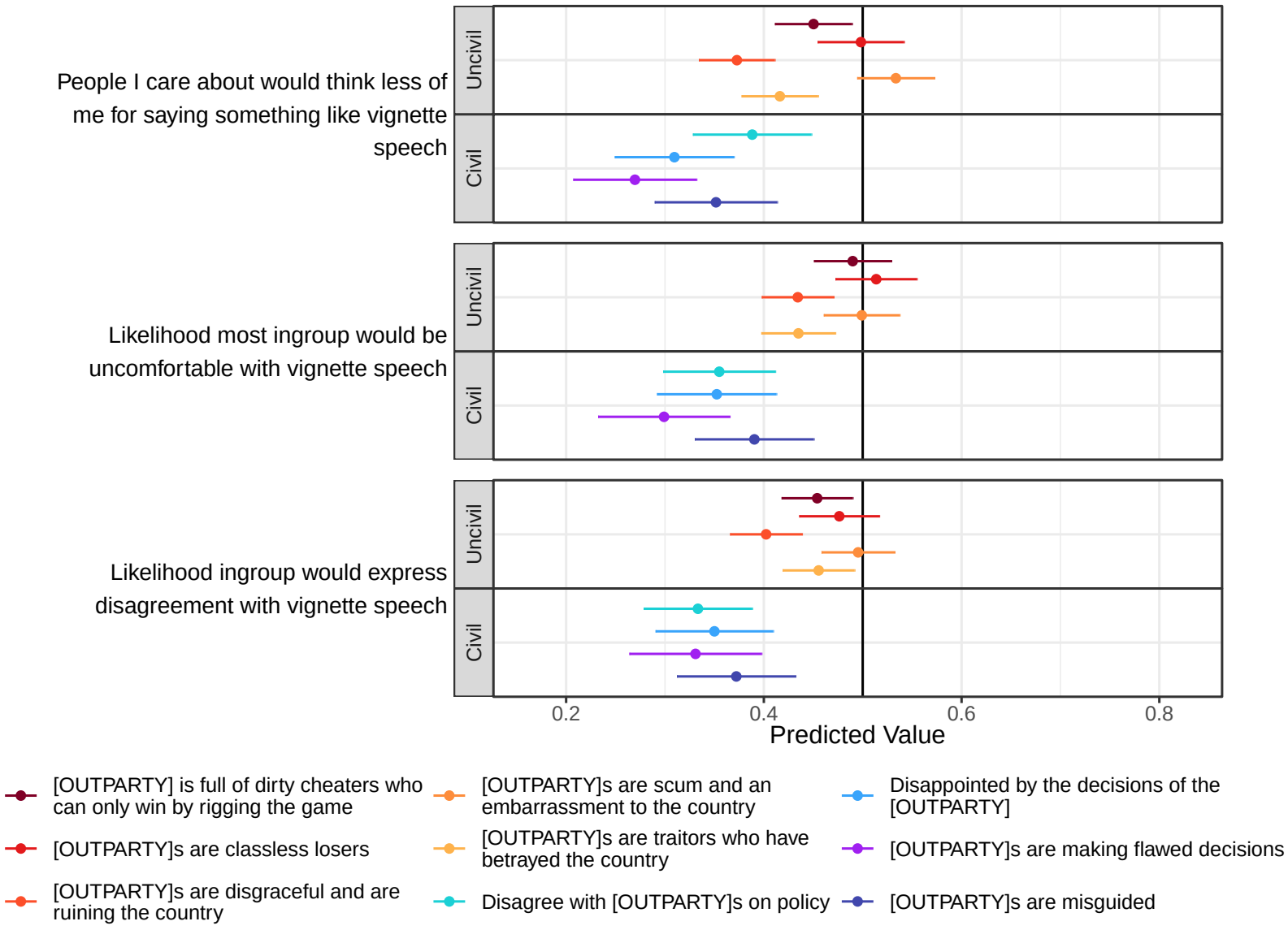
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D Construct Validity of Vignette Examples

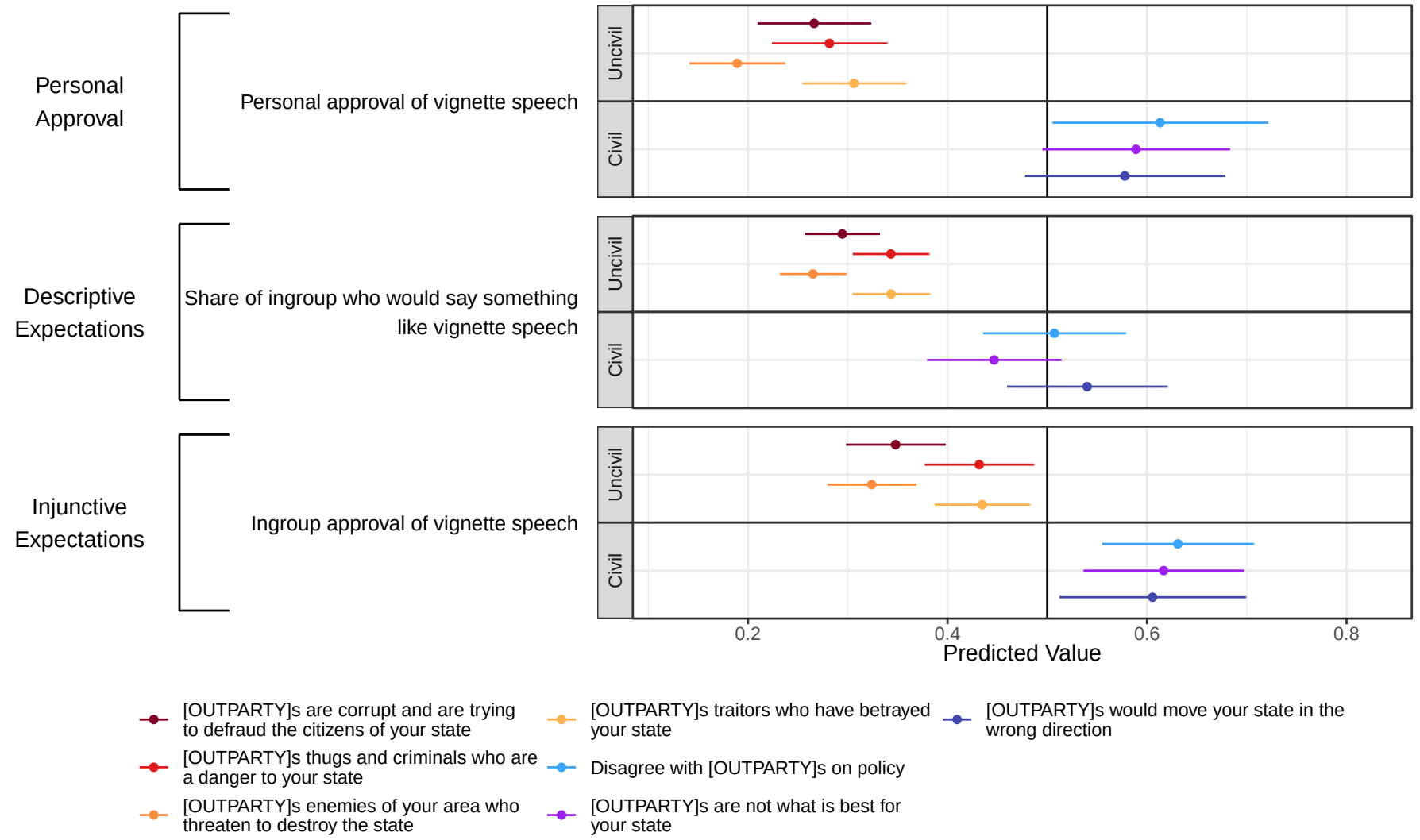
In this section, I display predicted values from regressions of norm outcomes on dummy variables for each individual vignette example. My results indicate that members of the mass public and politicians largely do perceive the uncivil examples differently from the civil but negative examples. There is some variation within the civil and uncivil clusters, but it appears that uncivil (civil) examples are mostly perceived as more similar to other uncivil (civil) examples than to civil (uncivil) examples. In other words, people do tend to approve less of, believe others would approve less of, and believe others might sanction the uncivil examples more than the civil examples. Further, people believe that their ingroup would be less likely to engage in uncivil examples as opposed to civil examples. The main exception to this clustering of civil and uncivil examples is that the mass public seems to perceive the vignette example calling the outparty "disgraceful" as similar to the civil examples on some norm measures. The masses do, however, see the "disgraceful" example similarly to the other uncivil examples when it comes to the likelihood of sanctions from their ingroup. These results suggest that my vignette examples are successfully capturing the intended construct: uncivil examples are almost uniformly perceived as less socially acceptable than civil examples.

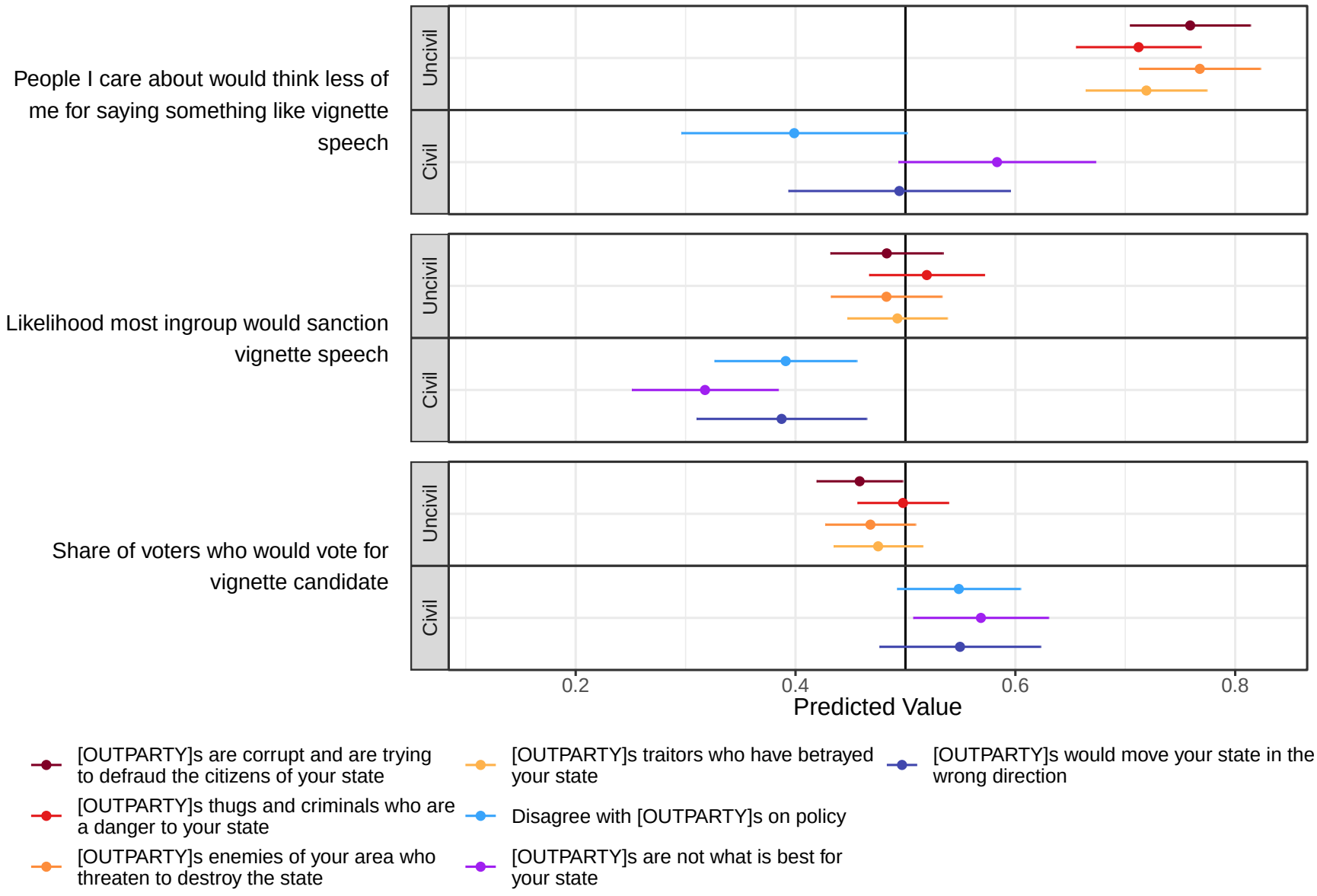
D.1 Mass Sample Vignette Examples




People I care about would think less of me for saying something like vignette speech
Likelihood most ingroup would be uncomfortable with vignette speech
Likelihood ingroup would express disagreement with vignette speech
Uncivil
Civil
0.2
0.4
0.6
0.8
Predicted Value
[OUTPARTY] is full of dirty cheaters who can only win by rigging the game
[OUTPARTY]s are classless losers
[OUTPARTY]s are disgraceful and are ruining the country
[OUTPARTY]s are scum and an embarrassment to the country
[OUTPARTY]s are traitors who have betrayed the country
Disagree with [OUTPARTY]s on policy
Disappointed by the decisions of the [OUTPARTY]
[OUTPARTY]s are making flawed decisions
[OUTPARTY]s are misguided

D.2 Politician Sample Vignette Examples




People I care about would think less of me for saying something like vignette speech
Likelihood most ingroup would sanction vignette speech
Share of voters who would vote for vignette candidate
Uncivil
Civil
0.2
0.4
0.6
0.8
Predicted Value
[OUTPARTY]s are corrupt and are trying to defraud the citizens of your state
[OUTPARTY]s thugs and criminals who are a danger to your state
[OUTPARTY]s enemies of your area who threaten to destroy the state
[OUTPARTY]s traitors who have betrayed your state
Disagree with [OUTPARTY]s on policy
[OUTPARTY]s are not what is best for your state
[OUTPARTY]s would move your state in the wrong direction

E Norms in Politics and Sports

This section examines how individuals in the mass sample evaluate incivility in discussions about sports as opposed to politics. To do so, I regress norm measures on the uncivil treatment, an indicator for the topic (politics), and the interaction between the two. Figure E.0.1 displays ATE estimates of the uncivil treatment on norm outcomes for sports and politics. The ATE estimates indicate that norms favor civility over incivility in both sports and politics. Moreover, the uncivil treatment effects are very similar for both topics—with two exceptions. First, the treatment effect for descriptive expectations is smaller for sports than for politics: uncivil and civil rhetoric seem less differentially prevalent for sports. Second, individuals report a larger gap in feelings of guilt or shame between civil and uncivil for sports as opposed to politics, consistent with more internalization of a norm against incivility for sports. Next, in Figure E.0.2, I display predicted values for norm outcomes from the interacted models. While I found that the masses view incivility similarly relative to civility across sports and politics, Figure E.0.2 shows that, in absolute terms, there is a fairly strong norm against incivility in discussions about sports. Conversely, there is no such norm against incivility in politics, as reported in the main text, and in fact, the norm favors incivility in politics among the mass public.

These results suggest there may be a general norm against incivility that is not applied to the outparty: there appear to be limits on acceptable criticism even for groups it is presumably socially acceptable to criticize (e.g., rival sports teams). However, these limits appear to not apply to political outgroups. This could be due to the unique dislike of the opposing party which may make members of the public less tolerant of the outparty (Iyengar, Sood and Lelkes, 2012; Mason, 2018; Sullivan, Piereson and Marcus, 1982).

Figure E.0.1: Average Treatment Effect Estimates of Incivility for Sports and Politics

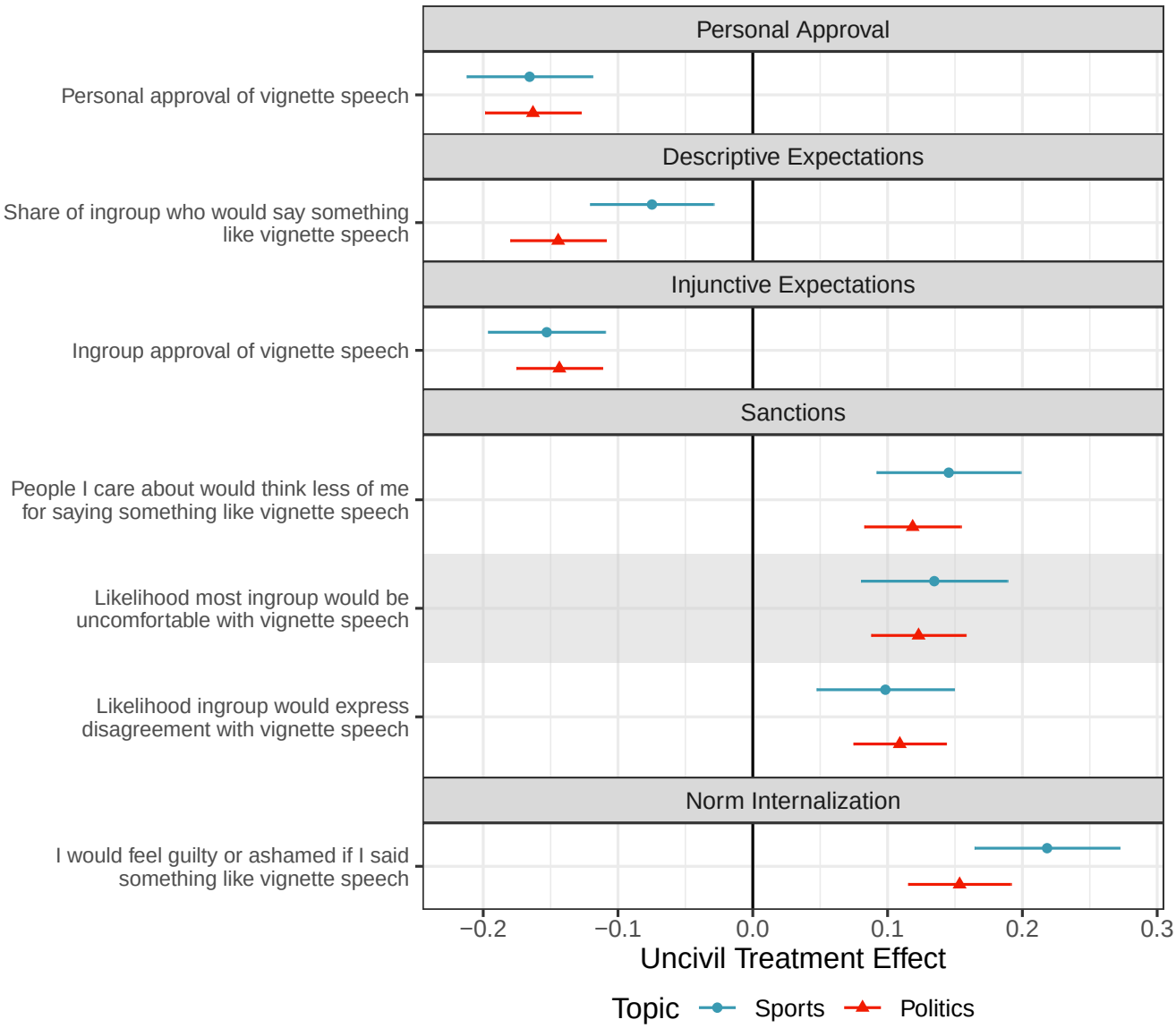
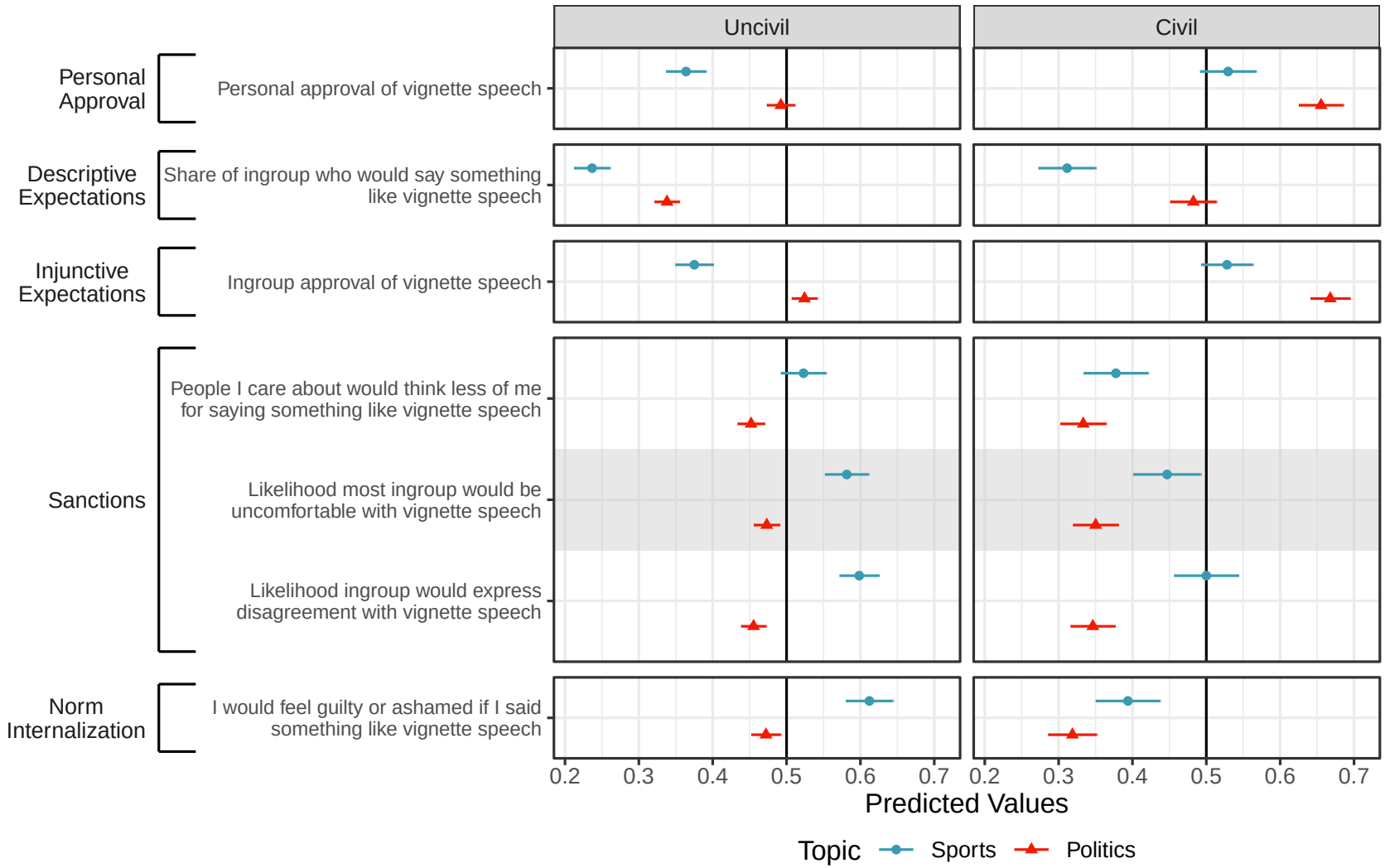


Figure E.0.2: Predicted Norm Values for Civil and Uncivil Sports and Politics Examples



References

Belloni, Alexandre, Victor Chernozhukov and Christian Hansen. 2014. "Inference on Treatment Effects After Selection Among High-Dimensional Controls." *The Review of Economic Studies* 81(2):608–650.

Iyengar, Shanto, Guarav Sood and Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social-Identity Perspective on Polarization." *Public Opinion Quarterly* 76(3):405–431.

Mason, Liliana. 2018. *Uncivil Agreement: How Politics Became Our Identity.* Chicago: University of Chicago Press.

Sullivan, John, James Piereson and George Marcus. 1982. *Political Tolerance and American Democracy.* Chicago: University of Chicago Press.