AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identity

Nicholas Dias and Yphtach Lelkes University of Pennsylvania

Abstract: *Democrats and Republicans clearly dislike one another. Yet, scholars debate whether policy disagreement or partisan identity, per se, drives interparty animus. Past studies suggest the relationship between partisanship and interpersonal affect is spurious, driven by inferred policy preferences. We argue, instead, that policy preferences signal partisan identity when the parties' stances on an issue are well-known. Using a nationally representative survey and four preregistered experiments, we disentangle the effects of policy disagreement and partisan identity on interpersonal affect. Our findings suggest that partisan identity is the principal mechanism of affective polarization, and that policy preferences factor into affective polarization largely by signaling partisan identity. However, our results also affirm that policy disagreement in itself drives interpersonal affect. This provides evidence that partisanship reflects an emotional attachment to a political party, not merely a running tally of rational considerations.*

Verification Materials: The data and materials required to verify the computational reproducibility of the results, procedures, and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/JHJJW0.

Republicans and Democrats loathe each other now more than ever previously measured in surveys (Iyengar et al. 2019). This widening chasm between in- and out-party affect—called *affective polarization*—encourages citizens to judge politicians and political outcomes in emotional and biased ways, compromising their ability to hold elected leaders accountable (Martherus et al. 2021). It leads voters to reward politicians that eschew compromise in favor of grandstanding (Hetherington and Rudolph 2015). Worse still, affective polarization's effects are not limited to politics. It also motivates social segregation along party lines and potentially distorts economic markets (e.g., McConnell et al. 2018; Settle and Carlson 2019).

Proposed causes of affective polarization include social and partisan media (e.g., Levendusky 2013) and a decline in cross-cutting social identities between the parties (e.g., Mason 2018). Yet, many studies have focused on the role (or lack thereof) of Americans' increasingly extreme or sorted policy preferences (Bougher 2017; Webster and Abramowitz 2017). These accounts assume partisanship influences interpersonal affect via one of two mechanisms.[1] The *party-over-policy hypothesis* suggests affective polarization is a powerful manifestation of an evolved tendency to dislike social out-groups, notwithstanding substantive policy disagreements (e.g., Mason 2018). By contrast, the *policy-over-party hypothesis* argues that affective polarization precisely reflects the parties' disagreements about salient policy issues (e.g., Webster and Abramowitz 2017).

Although the concept of affective polarization is relatively new to political science, debates about the importance of identifying with one's party (hereafter called partisan identity) versus policy disagreement have been central to the voting-behavior literature for more than 50 years (Campbell et al. 1960; Fiorina, 1981). A defining limitation of that literature, as well as the literature on affective polarization, is that the effects

Nicholas Dias, Ph.D. student in Communication and Political Science, Annenberg School for Communication, Department of Political Science, School of Arts & Sciences, University of Pennsylvania, 3620 Walnut St, Philadelphia, PA 19104. Yphtach Lelkes, Associate Professor of Communication, Annenberg School for Communication, University of Pennsylvania, 3620 Walnut St, Philadelphia, PA 19104.

We are grateful to Lilla Orr, Gregory Huber, Michael Delli Carpini, Jamie Druckman, Stephen Goggin, Alex Tolkin, and David Broockman for their helpful comments on earlier versions of this article. We also thank our anonymous reviewers and the participants of APSA's 2020 Political Psychology and Elections, Public Opinion, and Voting Behavior webinar.

[1] In line with common usage, we use the term "partisanship" to refer to party affiliation or identification. It does not necessarily denote bias or prejudice in itself.

American Journal of Political Science, Vol. 66, No. 3, July 2022, Pp. 775–790

 DOI: 10.1111/ajps.12628

of partisan identity and policy disagreement are often inseparable (e.g., Goggin, Henderson, and Theodoridis 2019; Milita et al. 2017).

A typical experiment comparing the party-over-policy and policy-over-party hypotheses has participants rate vignette subjects based on their partisanship and preferences about salient policy issues, both randomly assigned (Orr and Huber 2020). Yet, partisanship entails information about policy preferences (Milita et al. 2017; Rothschild et al. 2019). For example, if one learns an individual is Republican, one also learns that person likely supports abortion restrictions. To quote Fiorina (1981), the "issues are *in* party identification" (p. 200).

Conversely, party identification is also *in* the issues. As the parties' have well-known stances on many policy issues—hereafter called *party-branded issues*—learning a person's policy preferences allows one to infer her partisanship or party loyalty (Goggin, Henderson, and Theodoridis 2019; Heit and Nicholson 2016).[2] As such, if an experiment signals that a person is pro-choice, it likely also signals that the person is a Democrat. Alternatively, if an experiment indicates a person is a pro-life Democrat, it potentially signals that the person is a *less-than-loyal* Democrat.

Because party-branded policy preferences and partisanship cue one another, studies relying on party-branded issues cannot disentangle the effects of policy disagreement from partisan identity. They present a compound treatment and, therefore, likely overestimate how much policy disagreement drives affective polarization (Dafoe, Zhang, and Caughey 2018). The goal of this article is to cleanly estimate the effects of both partisan identity and policy disagreement on affective polarization. In doing so, it responds to recent claims that the relationship between partisanship and interpersonal affect is effectively spurious (Lelkes 2021; Orr and Huber 2020).

We begin with a survey of American adults (Study 1), which shows that Republicans and Democrats have easily distinguishable stances on policy issues used in past studies of affective polarization. This study underscores the need to test the effects of policy preferences that do not cue partisan identity. Then, using four preregistered experiments, we disentangle the effects of policy disagreement and partisan identity by experimentally manipulating three features in personal vignettes: partisanship, the explicit statement of policy preferences, and what stated policy preferences signal about partisan identity.[3]

Our first experiment (Study 2) examines whether introducing preferences about comparably salient party-branded and unbranded issues weakens the effect of partisanship on interpersonal affect. We find that party-branded preferences nearly erase the relationship between partisanship and interpersonal affect, whereas unbranded preferences have a much smaller (though substantial) effect. With our second experiment (Study 3), we demonstrate that unbranded preferences behave as party-branded preferences when randomly associated with a party. This confirms that Study 2's results are not driven by differences in the salience of party-branded and unbranded issues. Additionally, to solidify our findings, we conceptually replicate Study 3 with another experiment (Study 4) and, with a final experiment, show that the effects of partisan-identity signals are *not* mediated by inferences about other policy preferences (Study 5).

Ultimately, our results suggest that both partisan identity and policy disagreement independently drive interpersonal affect. Partisanship's effect on interpersonal affect is principally a function of partisan identity. Moreover, policy preferences factor into affective polarization largely by signaling partisan identity. However, policy disagreement, apart from partisan identity, also drives interpersonal affect. These insights underscore the importance of understanding how the effects of party and policy flow through (i.e., are mediated by) one another, and are critical to formulating effective attempts to address affective polarization. They also substantiate that partisanship is not merely a "running tally" of policy preferences and past political experiences, but a visceral attachment to the political parties as social groups (Campbell et al. 1960; Fiorina 1981).

Literature Review

Affective polarization describes the increasing effect of partisanship on interpersonal affect. The average individual-level gap between feeling-thermometer ratings of the Republicans and Democratic parties grew from roughly 23 degrees in 1978 to more than 41 degrees in 2016 (Iyengar et al. 2019). Similarly, only 4% of Democrats and 5% of Republicans in 1960 reported they would be unhappy if their children married someone from the out-party (Almond and Verba 1963). Those

[2]We distinguish partisan branding from party issue ownership, which refers to perceptions of the parties' competency in different policy areas (Petrocik 1996).

[3]Preregistered analysis plans for all studies can be found at https://bit.ly/3nu1X6P.

numbers had risen to 33% and 49%, respectively, by 2010 (Iyengar, Sood, and Lelkes 2012).

Though scholars have myriad explanations for the rise in affective polarization (for a review, see Iyengar et al. 2019), the majority assume partisanship influences interpersonal affect via one of two mechanisms. The first, which we call the party-over-policy hypothesis, posits that affective polarization is a manifestation of humans' innate tendency to dislike social out-groups (e.g., Mason 2018). The second, which we call the policy-over-party hypothesis, argues affective polarization reflects the increasingly deep or broad policy disagreements between the parties (e.g., Webster and Abramowitz 2017).

By clarifying how partisanship drives interpersonal affect, studies of affective polarization inform a broader debate about the nature of partisanship, providing insight into whether partisanship is a psychological attachment or merely a rational summary of individuals' policy preferences and past political experiences (Campbell et al. 1960; Fiorina 1981). This debate is a decade older than *The American Voter*, which cemented into the study of American politics the notion that citizens have attachments to their parties that transcend rational considerations (Campbell et al. 1960). Yet, despite its long history, this debate is far from resolved (Fowler 2020a, 2020b; Rogers 2020).

Evidence for the Party-Over-Policy Hypothesis

The party-over-policy hypothesis, as first explicated by Iyengar, Sood, and Lelkes (2012), emerges from social identity theory. Social identity theory posits that humans evolved to divide the world into in-groups and out-groups, allies and competitors (Tajfel 1981). Peoples' identification with their social in-groups ties their self-esteem to these in-groups' reputations. Consequently, people elevate their in-groups' standings relative to competing out-groups, as reflected in both in-group favoritism and out-group animosity (Lelkes and Westwood 2017; Martherus et al. 2021). According to the social-identity view of partisanship, citizens' political parties are one of these cherished social in-groups (e.g., Huddy, Mason, and Aarøe 2015).

Several forms of evidence suggest affective polarization reflects the increasing salience of partisan identities, not increasingly stark policy disagreements. Affective polarization and partisan behavior (e.g., campaign activity) are better predicted by partisan identity strength than policy preferences (Huddy, Mason, and Aarøe 2015; Mason 2018). Indeed, Americans' voting behavior can increasingly be predicted with their partisanship alone (Bartels 2000). The growth of affective polarization tracks with the alignment of various social identities in the United States—namely, partisan, ideological, racial, and religious (Mason 2018). As a result of this alignment, Americans have starkly different conceptions of who makes up each political party (Ahler and Sood 2018), fewer cross-cutting identities bind Republicans and Democrats together, and partisans have become militant about their partisan identities (Mason 2018).

Additionally, partisanship appears to determine policy preferences more so than the other way around. Although policy preferences fluctuate over time, partisanship is remarkably stable (e.g., Kinder and Kalmoe 2017). Panel analyses indicate partisanship affects subsequent political judgments (e.g., of policies) to a greater extent than vice versa (Bartels 2002). For example, Lelkes (2018) finds that lagged ideological sorting only predicts affective polarization among the politically knowledgeable, but lagged affective polarization predicts ideological sorting across all levels of political knowledge. Furthermore, partisans readily change their policy preferences when told which policies the in- or out-party supports (Barber and Pope 2019; Cohen 2003).

Yet, the evidence for the party-over-policy hypothesis is far from conclusive. Writing about voting behavior, Fowler (2020a, 2020b) describes how the effects of "partisan intoxication" and policy preferences are often observationally equivalent. Partisan identity may predict interpersonal affect better than individual policy preferences, because partisanship reflects a complex, weighted average of *many* policy preferences (Fiorina 1981). Measurement error could account for the temporal instability of policy preferences and create the false impression that partisanship causes policy preferences (Ansolabehere, Rodden, and Snyder 2008; Fowler 2020b). Moreover, mass partisans may adopt party elites' policy views because they believe those elites represent their interests, not because they blindly follow social-group leaders (Bullock 2011). Ultimately, Fowler (2020b) concludes that there is "no compelling evidence" for the influence of partisan attachments on voting behavior.

Evidence for the Policy-Over-Party Hypothesis

The critique of observational equivalence is core to the policy-over-party hypothesis of affective polarization. According to this hypothesis, partisanship's influence on interpersonal affect is explained—that is, confounded—by policy disagreement. Affect between the parties has

polarized because elites and mass partisans are increasingly extreme or sorted in their views about policy issues (e.g., Bougher 2017; Webster and Abramowitz 2017). Individuals only feel animus toward out-party members because belonging to the out-party implies disagreement about salient policy issues (Orr and Huber 2020).

In support of the policy-over-party hypothesis, some studies have shown that describing political candidates as ideologically "moderate" or "extreme" affects individuals' feelings of warmth toward them (Rogowski and Sutherland 2016; Webster and Abramowitz 2017). Preferences about particular policies predict feelings of coldness toward the out-party and its political candidates (Webster and Abramowitz 2017), and learning a candidate holds extreme policy preferences can lead to larger shifts in feeling-thermometer ratings than learning a candidate's partisanship (Lelkes 2021). Those who disagree with the out-party's stances on multiple policy issues—regardless of the depth of disagreement on any particular issue—feel colder toward the out-party (Bougher 2017). Similarly, citizens who consistently support their in-party's stances on social policy issues are more affectively polarized (Webster and Abramowitz 2017).

Unfortunately, studies supporting the policy-over-party hypothesis are marred by the same equivalence problem that contaminates party-over-policy studies. Policy-over-party studies have overwhelmingly focused on policy preferences that are well-associated with the political parties (e.g., Costa 2021; Lelkes 2021; Orr and Huber 2020). Yet, preferences about party-branded issues may cue partisanship or an individual's loyalty to her party as a social group—that is, her partisan identity. Additionally, the consistency with which an individual endorses her party's policies may itself reflect partisan identity (Barber and Pope 2019; Cohen 2003). Supporting immigration restrictions, abortion restrictions, and gun-owner rights is as much, if not more, a reflection of toeing the Republican Party line than a coherent conservative ideology.

Conflation of Policy Disagreement and Partisan Identity

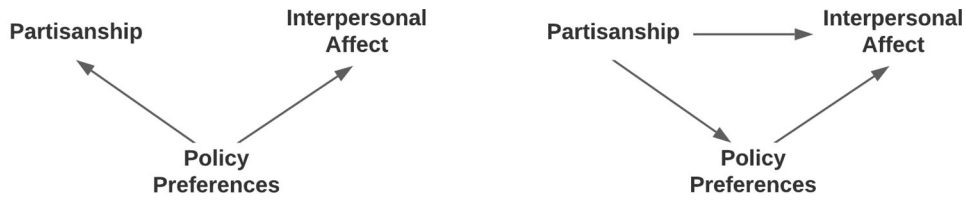
In sum, research supporting both prominent accounts of affective polarization is hampered by reliance on measures of policy preferences that strongly correlate with partisanship (e.g., Goggin, Henderson, and Theodoridis 2019; Milita et al. 2017). Indeed, it is increasingly difficult to find policy issues that do not suggest partisan stances. Since the 1970s, policy preferences and partisanship have become increasingly equivalent. Party elites have become more ideologically polarized, providing "clearer cues" about the parties' issue stances (Levendusky 2010). At the same time, mass partisans from each of the parties have become more ideologically sorted and homogeneous (Levendusky 2009).

The alignment of policy preferences and partisanship is reflected in citizens' ability to readily tie policy preferences to the political parties. In one recent study, Goggin, Henderson, and Theodoridis (2019) found that participants were more likely to infer an unbranded political candidate was Republican when the candidate supported strengthening national defense, addressing the immigration "problem," or reducing the budget deficit. By contrast, participants more readily assumed an unbranded candidate was a Democrat when she supported improving healthcare, assisting the poor, or strengthening Social Security. These effects were consistent across levels of political knowledge. Additionally, Milita et al. (2017) found that participants presume politicians take party-loyal stances on policy issues, unless they explicitly state otherwise (see also Heit and Nicholson 2016; Rothschild et al. 2019). Thus, many policy preferences appear to have become incorporated into Democratic and Republican stereotypes.

If conservative or liberal preferences about issues such as abortion typify Republicans or Democrats, then learning an individual supports abortion restrictions likely prompts people to infer that individual's partisanship, rightly or wrongly (Goggin, Henderson, and Theodoridis 2019). Even when given an individual's partisanship, people may still use policy preferences to infer that individual's attachment or loyalty to her political party as a social group. For example, a pro-choice Democrat may be inferred to be a loyal Democrat, whereas a pro-life Democrat may be inferred to be disloyal.

Along these lines, we argue that what policy preferences symbolize in terms of partisan identity is as important as what they entail about substantive policy beliefs. A pro-choice Democrat may dislike a pro-life Democrat—not because they disagree on the issue of abortion, but—because the copartisan has signaled she is not a *loyal* Democrat. Likewise, a pro-choice Democrat may feel endeared toward a pro-choice Republican, simply because she has broken with any Republican norm. The *act* of breaking with the party, not the *substance* of that break, is what matters. Indeed, past studies of social identity have already demonstrated similar patterns. Individuals feel especially warm toward in-group members that are typical on group-relevant traits, but especially cold toward in-group members that are deviant on group-relevant traits (e.g., Marques, Yzerbyt, and Leyens 1988). Typical or deviant out-group members can be similarly punished or rewarded (e.g., Marques et al. 1998).

FIGURE 1 Policy-Over-Party and Party-Over-Policy Hypotheses of Affective Polarization



A Policy-Over-Party **B** Party-Over-Policy

Note: Hypotheses depicted as directed acyclic graphs

Disentangling Policy Disagreement from Partisan Identity

According to the policy-over-party hypothesis, policy preferences confound the relationship between partisanship and interpersonal affect. This implies that, when information about policy disagreement is explicit, the relationship between partisanship and interpersonal affect should disappear (see Figure 1a). Indeed, several studies have leveraged this implication by estimating the effects of partisanship when information about policy preferences is versus is not available (Bullock 2011; Fowler 2020a; Rogers 2020). For example, in a series of experiments, Orr and Huber (2020) showed participants vignettes describing individuals who randomly varied in their partisanship and whether they had a preference about a salient policy issue. When not presented alongside policy preferences, information about partisanship had a large effect on participants' affect toward vignette subjects. However, the effect of partisanship decreased by as much as 55%–62% when information about the subject's policy preferences was also given.

As previously mentioned, however, policy-over-party studies have relied on party-branded policy preferences that likely signal partisan identity. As such, introducing policy preferences may erase the relationship between partisanship and interpersonal affect, because doing so controls for a *mediator*, not a confounder, as suggested by the policy-over-party hypothesis (see Figure 1b). That is, because party-branded preferences give insight into whether a person is a loyal or disloyal partisan, controlling for these preferences when estimating partisanship's effect on interpersonal affect amounts to overcontrol bias (Morgan and Winship 2014). As long as policy preferences convey information about partisan identity, the party-over-policy and policy-over-party hypotheses are empirically indistinguishable.

That said, the opposing causal directions of the party–policy relationship in either hypothesis suggest a critical test. If stripping partisan branding from policy preferences changes the extent to which they diminish the effect of partisanship on interpersonal affect, it would imply that policy preferences inherit part of their significance from partisanship. It would also imply that party-branded preferences act as a mediator between partisanship and interpersonal affect. However, if unbranded policy preferences wipe out the effect of partisanship on interpersonal affect, it would suggest partisanship serves only to cue policy disagreement. Such a test requires identifying policy issues that lack partisan branding, but are nonetheless comparably salient to party-branded issues. We set about this task first.

Study 1

A representative survey was administered to verify whether policy issues typically used in studies of affective polarization are party branded—that is, have well-known party stances. If so, it would suggest preferences about these issues signal partisan identity in addition to policy disagreement, and illustrate the need to study policy issues that lack any such partisan branding. We also sought to identify a handful of unbranded issues that were comparably salient to party-branded issues used in past studies.

Methods

A quota-matched sample of 499 American adults was recruited from Forthright, a high-quality online panel, between February 26 and March 2, 2020 ($M_{\text{Age}} = 45.69$,

s.d.$_{\text{Age}}$ = 16.06, 49.3% female).[4] Participants provided their partisanship using the American National Election Study's (ANES) three-question battery. They then indicated which stance they thought "most Democrats" and "most Republicans" would take on five policy issues. After placing the parties on each issue, participants indicated the issue's salience using a 4-point importance scale ranging from "Not at all important" to "Very important" (Niemi and Bartels 1985).

The five policy issues seen by each participant were randomly selected from a list of 35 issues pulled from the 2016 ANES, Orr and Huber (2020), bills voted upon in Congress or state legislatures,[5] or perennial issues discussed in national news outlets (e.g., requiring a year of national service). For each issue, participants were given two stances (and a "Don't know" option) on which they could place the parties. For example, the two stances given for the issue of gun accessibility were "the federal government should make it more difficult for people to buy a gun" and "the government should make it easier for people to buy a gun."[6] Each policy issue received an average of 61.86 definite (i.e., not "Don't know") ratings (s.d. = 8.45).

Results

Republicans and Democrats were widely perceived to take opposing stances on policy issues typically used in past studies of affective polarization (right side of Figure 2). For instance, participants perceived interparty disagreement about issues used by Lelkes (2021), Orr and Huber (2020), or Costa (2021), including immigration, welfare, gun accessibility, and abortion. Preferences about these issues thus likely signal partisan identity in addition to policy disagreement. However, several issues were not viewed in partisan terms, including the issues of eminent domain, occupational licensing, or whether mobile phone carriers should sell customers' data (left side of Figure 2).

Policy issues varied considerably in terms of salience (M = 2.86, s.d. = 0.95). Party-branded issues tended to be highly salient. For example, the average salience rating for the issues of gun accessibility and naturalizing undocumented immigrant children ("DREAMers") both sat above the "Somewhat important" scale point (M_{Guns} = 3.20, s.d.$_{\text{Guns}}$ = 0.81; M_{DREAMers} = 3.08, s.d.$_{\text{DREAMers}}$ = 0.87). The issue of child adoption by same-sex couples sat just below this point (M = 2.85, s.d. = 0.96). By contrast, nonpartisan (i.e., unbranded) issues tended to be less salient for participants, though several were comparably salient to party-branded issues. For instance, the issue of whether mobile carriers should sell customer data was considered as important as most party-branded issues (M = 3.16; s.d. = 0.78). Likewise, the issues of eminent domain (M = 2.96, s.d. = 0.71) and occupational licensing (M = 2.78, s.d. = 0.90) were comparably salient to that of allowing child adoption by same-sex couples. Indeed, the salience of same-sex couple adoption was not significantly different from that of eminent domain, t (103) = 0.75 or occupational licensing, t (120) = 0.42.

Study 2

Study 1's results suggest that past research on affective polarization has conflated the effects of policy disagreement and partisan identity by relying on policy issues that signal partisan identity. As such, the role of policy disagreement in affective polarization has likely been overestimated, and the notion that policy disagreement confounds the relationship between partisanship and interpersonal affect seems questionable.

Yet, Study 1 also identifies several salient, unbranded issues that can be used to isolate the effects of policy disagreement experimentally. We conducted a critical test of the party-over-policy and policy-over-party hypotheses by replicating one of Orr and Huber's (2020) survey experiments, having dissociated partisan branding from policy disagreement. In a preregistered experiment, participants were randomly assigned to view vignettes that mentioned comparably salient party-branded or unbranded policy issues, as identified in Study 1.

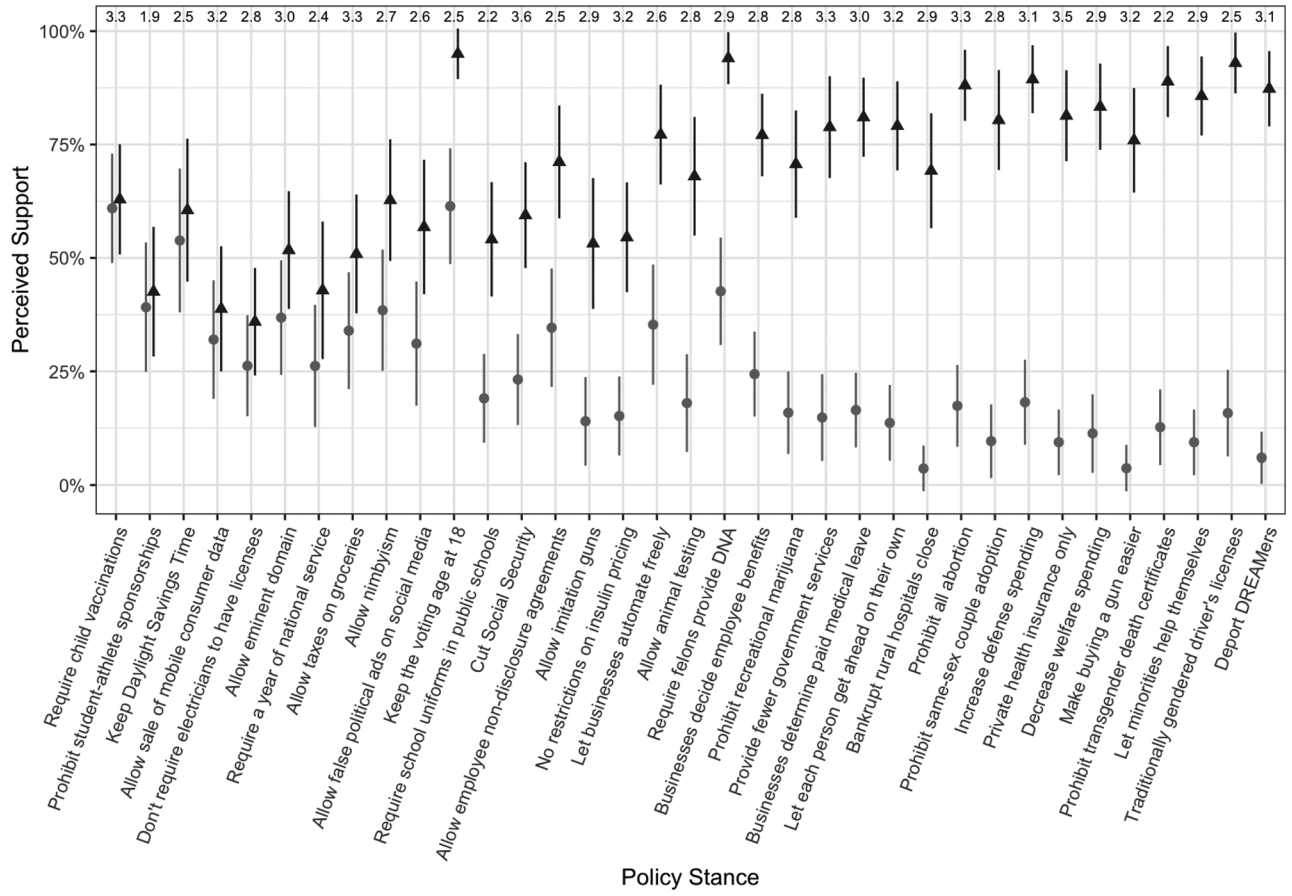
If partisanship strongly predicts interpersonal affect when juxtaposed with unbranded preferences, it would indicate the relationship is not spurious, counter to the policy-over-party hypothesis. It would also suggest that stances on party-branded issues *mediate* the relationship between partisanship and interpersonal affect, thereby supporting the party-over-policy hypothesis. If, however, unbranded preferences wipe out the effect of partisanship on interpersonal affect, the policy-over-party hypothesis would be supported.

[4] Our sample was quota-matched to be representative of American adults with regard to age, gender, education, Census region, and race. The demographics of all study samples are available on p. 5 of our supplemental information (SI). Forthright participants are recruited via mail campaigns based on addressed-based probability sampling, as well as via online ads.

[5] Votes were pulled from the nonpartisan organization Vote Smart (votesmart.org).

[6] "Don't know" responses were coded as missing. The text of each issue stance is available in our replication materials.

FIGURE 2 Perceived Stances of Republicans and Democrats on Unbranded and Party-Branded Policy Issues



Note: Averages from Study 1. Dark triangles represent Republicans' stances. Light circles represent Democrats' stances. Top labels indicate mean salience for each issue. Error bars represent 95% confidence intervals

Methods

Study 2 was split between two waves, separated by 1 week. Adult U.S. residents ($N = 3{,}025$) were first recruited from the Amazon Mechanical Turk (MTurk) on March 9, 2020.[7] All participants who completed Wave 1 were asked to complete Wave 2, with the exception of 324 participants who did not lean toward either the Republican or Democratic parties. In total, 2,266 participants completed both survey waves ($M_{\text{Age}} = 39.31$, s.d.$_{\text{Age}} = 12.63$, 48.6% female).

Wave 1 participants answered demographic questions before indicating their positions on 12 policy issues drawn from Study 1. These included three issues used by Orr and Huber (2020) that were perceived to be highly partisan by Study 1 participants: gun accessibility, DREAMer naturalization, and child adoption by same-sex couples. We also included two unbranded issues found to be comparably salient to these party-branded issues: eminent domain and whether mobile carriers should sell customer data. Other issues were arbitrarily included to obscure the goal of the study (though see footnote 8). After providing their policy preferences, participants indicated their partisanship using the ANES three-question battery. Finally, participants who generally identified with the Republican or Democratic party (i.e., nonleaners) completed Huddy, Mason, and Aarøe's (2015) four-item partisan identity index (Cronbach's alpha = 0.86), which measures how much a participant considers her political party an important social identity.

Wave 2 participants were randomly assigned to read vignettes that described subjects who varied in three key ways: partisanship, the explicit statement of policy preferences, and what stated policy preferences

[7]Although not representative of the American adults, MTurk workers typically replicate findings from experimental studies of political attitudes with national probability samples (Coppock 2019).

FIGURE 3 Example Stimuli from Study 2

Partisanship Only: *Suppose you meet a person who* ***enjoys watching TV****,* ***is from South Carolina****, and is a* ***Democrat****.*

Partisanship + Policy Preference: *Suppose you meet a person who* ***is from Massachusetts****,* ***enjoys reading****, and is a* ***Republican****. They believe that* ***immigrants brought to the U.S. illegally as children should be sent back to where they came from****.*

Policy Preference + Partisanship: *Suppose you meet a person who* ***is from Ohio*** *and* ***works as an electrician****. They believe that* ***the government should never be allowed to seize a citizen's private land, even if the land is needed for a public project and the owner is fairly compensated****. This person is a* ***Republican****.*

Note: Randomly manipulated features are boldfaced

signaled about partisan identity. The partisanship of vignette subjects randomly varied between Democrat and Republican. When a vignette included a policy preference, that preference was randomly selected from one of two possible positions on six policy issues. As such, participants shared the vignette subject's preference roughly half the time. In addition to information about partisanship and policy preferences, all vignettes included two randomly selected social facts—such as the subject's state of residence, occupation, or hobby—to make the descriptions seem completer and more natural. Example vignettes are presented in Figure 3.[8]

The issues referenced in vignettes were evenly split between party-branded and unbranded issues. All issues were asked about in Wave 1, with the exception of occupational licensing.[9] Each participant saw three vignettes. In vignettes that included both partisanship and a policy preference, the order of these details was randomized per participant. Whether a vignette included a policy preference, and whether the issue referenced was party branded, remained constant across the three vignettes each participant read to obscure the goal of the experiment. All other features were randomly assigned per vignette.

We operationalized interpersonal affect in two ways: a feeling-thermometer rating of the vignette subject and an index of two social-closeness measures commonly used in studies of affective polarization (Druckman and Levendusky 2019). The first index item asked participants how comfortable they would be having the vignette subject as a neighbor on their street, on a 4-point scale ranging from "Not at all comfortable" to "Extremely comfortable." The second asked how upset the participant would be if her hypothetical son or daughter married the vignette subject, on a 4-point scale ranging from "Not at all upset" to "Extremely upset." These two items were averaged to produce the index ($r = 0.74$). To maximize statistical power, we concatenated participants' responses to the three vignettes and estimated participant random effects and vignette fixed effects.[10] The following core model was used to test whether varying the partisan branding of policy preferences changed the extent to which they depleted the effect of partisanship on interpersonal affect:

$$Affect_i = \beta_0 + \beta_1 MatchParty + \beta_2 PolicyPresent + \beta_3 MatchParty * PolicyPresent + \epsilon_i.$$

MatchParty indicates whether a vignette subject belonged to the participant's in-party (vs. the out-party),

[8] All vignette features seen can be found on p. 6 of our SI.

[9] The issue of compelled child vaccinations was asked about in Wave 1, and initially was to be included in the vignette experiment. However, the surge of COVID-19 in the United States coincided with the launch of Wave 2, creating concerns that the partisan branding and salience of compelled vaccinations had changed since Wave 1. As such, we replaced this issue with that of occupational licensing for electricians.

[10] Vignette fixed effects were not preregistered. However, including these controls made no qualitative difference to our models.

and *PolicyPresent* indicates whether a vignette included a policy preference. Thus, the interaction term (β_3) represents the change in the effect of sharing the vignette subject's partisanship when information about a policy preference is added to the vignette. Party-branded issues and unbranded issues were analyzed separately.[11]

Results

Participants felt approximately 21 degrees warmer toward copartisan vignette subjects, as compared with out-partisan vignette subjects, when no information about the subject's policy preferences was given ($SE_{Unbranded} = 1.25$, $SE_{Party\text{-}Branded} = 1.44$, $p < 0.001$). The effect of shared partisanship shrunk when vignettes included the subject's preference about a policy issue. However, as expected, the size of this change varied dramatically depending upon whether that preference referred to a party-branded or unbranded issue.

As shown on the left side of panel A in Figure 4, the effect of shared partisanship decreased by 57% (21 to 9 degrees) when preferences about party-branded issues were introduced to vignettes (SE = 1.80, $p < 0.001$). This difference is roughly consistent with the findings of Orr and Huber (2020), who relied on party-branded issues in their study. However, the effect of shared partisanship shrunk by only 29% (21 to 15 degrees) when vignettes introduced the subject's preference on an unbranded issue (SE = 1.55, $p < 0.001$).

A post hoc linear mixed-effects model indicated that the difference between effect of shared partisanship, given party-branded versus unbranded policy preferences, was statistically significant (Difference = 6.03, SE = 1.47, $p < 0.001$). Analyses of each individual policy issue also indicated the difference between these effects was unlikely the result of issue idiosyncrasies: Every unbranded issue depleted the effect of shared partisanship less than every party-branded issue (see replication materials). As such, policy disagreement did not confound the relationship between partisanship and thermometer ratings. This relationship only weakened substantially when policy preferences provided cues about partisan identity. By signaling partisan identity, party-branded policy preferences appear to have mediated the relationship between shared partisanship and interpersonal affect.

[11] As in Orr and Huber (2020), and in line with our preregistration document, participants who did not see the vignette subject's partisanship were dropped from these analyses. Analyses examining the effects of shared policy preferences can be found on p. 12 of our SI.

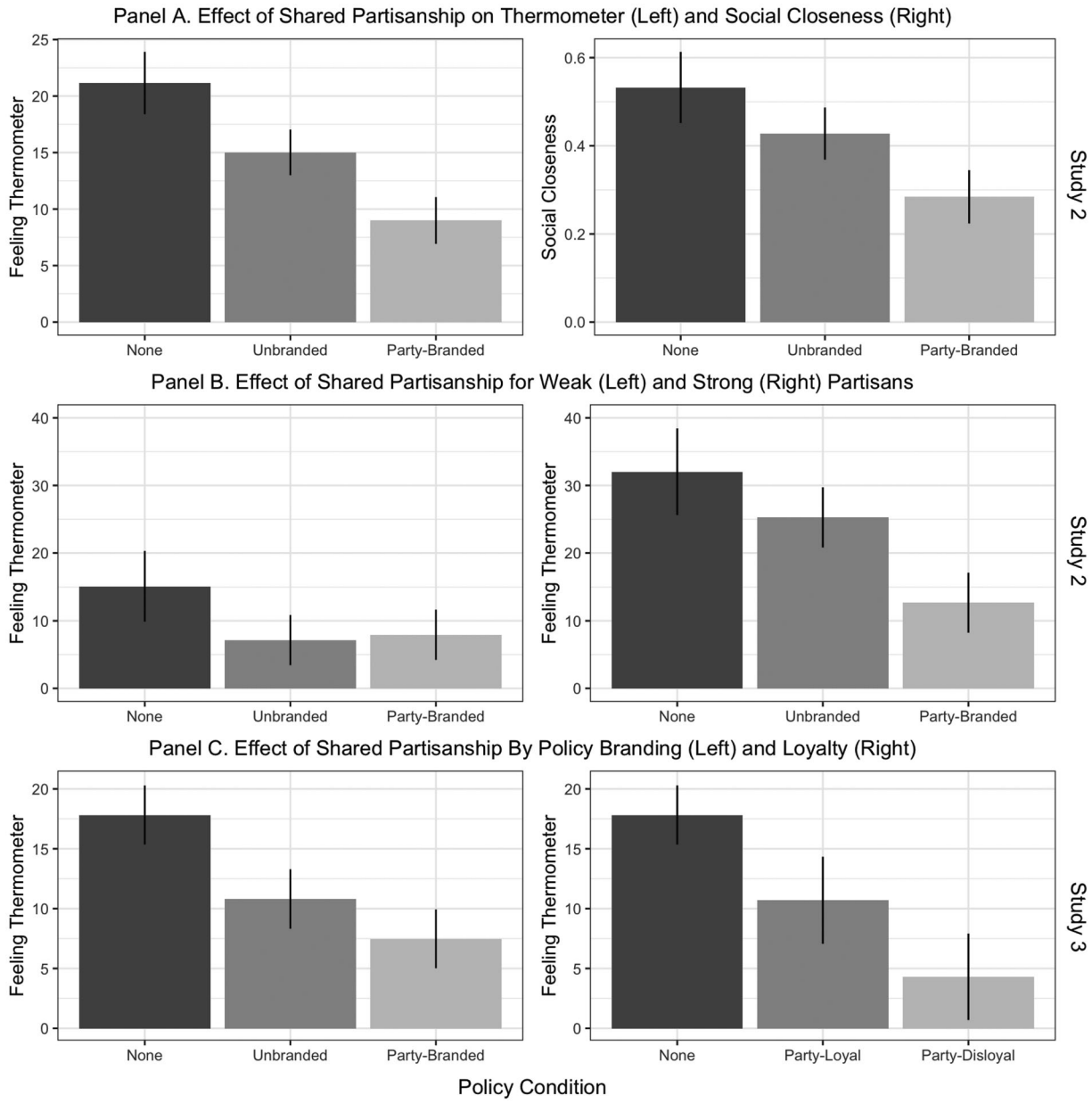
The effects of shared partisanship on social closeness mirrored those on thermometer ratings (right side of panel A in Figure 4). Shared partisanship mattered considerably when participants were given no information about a vignette subject's policy preferences: Participants were more willing to have close social ties with copartisan subjects than out-partisan subjects, by about a half-scale point on a 4-point scale ($\beta = 0.53$, SE = 0.04, $p < 0.001$). As seen on the right side of panel A in Figure 4, the effect of shared partisanship decreased by 47% when vignettes introduced the subject's preference about a party-branded issue ($\beta = -0.25$, SE = 0.05, $p < 0.001$). However, the decline in the effect of shared partisanship was much less dramatic (20%) when unbranded preferences were introduced to vignettes ($\beta = -0.10$, SE = 0.05, $p < 0.05$).

A post hoc linear mixed-effects model indicated that the difference between the effect of shared partisanship on social closeness, given party-branded versus unbranded preferences, was statistically significant (Difference = 0.14, SE = 0.04, $p < 0.001$). Moreover, analyses of individual issues revealed that every unbranded issue depleted the effect of shared partisanship less than every party-branded issue (see replication materials). Thus, our findings show a consistent pattern: The effect of partisanship on interpersonal affect remains strong in the presence of unbranded policy preferences, but not party-branded preferences. The relationship between partisanship and interpersonal affect is not confounded by policy disagreement, but mediated by the identity cues provided by party-branded preferences.

Moderation Analyses. If the gap in the effects of party-branded and unbranded policy preferences is truly a function of these preferences' differing capacity to signal partisan identity, then this gap should be larger for those participants with strong partisan identities (Huddy, Mason, and Aarøe 2015). Simply put, signals about partisan identity should matter more to participants who are themselves strongly identified with a political party. To test this possibility post hoc, we repeated our analysis with those participants in the top and bottom tertiles of partisan identity strength, respectively (Hainmueller, Mummolo, and Xu 2019).

Strength of partisan identity dramatically moderated the conditional effects of shared partisanship on thermometer ratings. As shown on the right side of panel B in Figure 4, the difference between the effect of shared partisanship, given party-branded preferences versus unbranded preferences, was particularly large for strong party identifiers (Difference = 12.58, SE = 3.21, $p < 0.001$). However, as seen on the left side of panel B in

FIGURE 4 Conditional Effects of Shared Partisanship (Studies 2 and 3)



Note: Linear mixed-effects model estimates from Studies 2 and 3. All models include participant random effects and vignette fixed effects. Policies matched on salience in Study 2. Policies held constant in Study 3. Weak (strong) partisans represent participants in the bottom (top) tertile of partisan identity strength. Partisan identity measured using Huddy, Mason, and Aarøe's (2015) index. Error bars represent 95% confidence intervals

Figure 4, this gap was statistically insignificant for weak party identifiers (Difference = 0.79, SE = 2.67).[12] An identical pattern was observed for social closeness: The difference between the effects of party-branded and unbranded policy preferences was large for strong party identifiers (Difference = 0.26, SE = 0.10, $p < 0.01$) and insignificant for weak party identifiers (Difference = 0.06, SE = 0.08).[13]

[12]The three-way interaction between shared partisanship, issue type, and partisan identity strength was also significant ($\beta = -13.50$ SE = 4.17, $p < 0.01$).

[13]The three-way interaction between shared partisanship, issue type, and partisan identity strength was significant ($\beta = -0.32$, SE = 0.12, $p < 0.01$).

Comparative Effects of Shared Policy Preferences. Although our primary goal is to understand whether partisanship's effect on interpersonal affect—that is, affective polarization—is confounded or mediated by policy preferences, policy preferences may influence interpersonal affect for reasons unrelated to partisanship. It is thus important to distinguish testing the mechanisms of affective polarization from comparing the *total* effects of policy preferences and partisanship on interpersonal affect. However, the total effect of shared policy preferences on interpersonal affect does provide useful context to our analysis.

Subsetting to those participants who saw both the partisanship and policy preferences of vignette subjects allowed us to simultaneously estimate the effects of shared partisanship and shared policy preferences. In a post hoc test, we ran the following core model with participant random effects and vignette fixed effects:

$$Affect_i = \beta_0 + \beta_1 MatchParty + \beta_2 MatchPolicy + \epsilon_i.$$

Participants felt 16 degrees warmer toward vignette subjects when they shared a political party, but 24 degrees warmer toward vignette subjects when they shared an *unbranded* policy preference ($SE_{MatchParty} = 1.10$, $SE_{MatchPolicy} = 1.16$, $p < 0.001$). However, the effects of shared partisanship and shared policy preferences on social closeness were comparable, increasing social closeness by roughly a half-scale point ($SE_{MatchParty} = 0.03$, $SE_{MatchPolicy} = 0.04$, $p < 0.001$). As with the effects of shared partisanship, the effects of shared policy preferences were larger in isolation ($\beta_{Thermometer} = 32.92$, $\beta_{Social\ Closeness} = 0.76$, see p. 14 of SI). Altogether, these findings suggest that, although partisan identity is the principal mechanism of the partisanship–affect relationship, the total effect of policy on interpersonal affect may be larger than that of party. However, the effect of partisanship is still substantial, even when controlling for policy preferences. We return to this point in our "Discussion" section.

Study 3

The results of Study 2 firmly support the party-over-policy hypothesis. Per Study 1, the unbranded issues and party-branded issues used in Study 2 were comparably salient. As such, these unbranded preferences signal substantive disagreement about salient policy issues in the same way as the party-branded preferences, without signaling partisan identity. Though the unbranded preferences were not previously linked to the political parties, these preferences should have confounded the effect of partisanship to the same extent as the party-branded issues—if the policy-over-party hypothesis were correct. They did not.

Though we relied on a widely used measure of issue importance (Niemi and Bartels 1985), Study 2's issues may not have been effectively matched on salience. In other words, perhaps unbranded preferences mattered less in Study 2, because these policies were simply irrelevant (or different in some other way). To address these concerns, we sought to make our unbranded issues *behave* as party-branded issues by explicitly stating where Republicans and Democrats stood on these issues. By holding issue substance constant, this approach offers a purer test of the policy-over-party and party-over-policy hypotheses. If unbranded policies substantially shrink the effect of partisanship once linked to parties, it would indicate that policy preferences are important precisely *because* they signal partisan identity.

In another preregistered experiment, Democrats and Republicans were randomly attributed stances on unbranded issues later mentioned in vignettes. This approach depends on our ability to brand issues as partisan within an experiment. Moreover, the importance of the policy issues to the political parties was not emphasized to participants. Thus, although the effects of partisan branding will likely be smaller than those in Study 2, significant findings under this paradigm would provide strong support for the idea that party-branded preferences mediate the relationship between partisanship and interpersonal affect.

We also considered whether the effects of party-branded preferences varied according to whether they signaled a vignette subject was *loyal* or *disloyal* to her party. Given that partisanship is arguably synonymous with loyalty to one's political party (e.g., Milita et al. 2017), informing participants that a partisan endorses her party's policies may convey little additional information to participants. Thus, our preregistered hypothesis was that the effects of party-branded preferences would be especially pronounced when branding indicated a vignette subject was *disloyal* to her party.

Methods

U.S. residents ($N = 2{,}004$) were recruited from MTurk on March 21, 2020, and a total of 1,799 Republicans and Democrats (including party leaners) completed the study

FIGURE 5 Example Stimuli from Study 3

Introduction without Partisan Branding: ***Some*** *argue that the government should be able to seize private land if it is needed for a public project and the owner is fairly compensated.* ***Critics*** *disagree and argue that the government should never be allowed to seize a citizen's private land, even if the land is needed for a public project and the owner is fairly compensated.*

Introduction with Partisan Branding: ***Republicans*** *argue that electricians should be required to have licenses, because poor electrical work can create serious damage and even danger.* ***Democrats*** *disagree and argue that requiring electricians to have licenses does not improve the quality of their work, and prevents people from joining the profession.*

Note: Randomly manipulated features are boldfaced

($M_{Age} = 38.33$, s.d.$_{Age} = 12.87$, 52.9% female).[14] Within a single survey, participants indicated their partisanship using the ANES three-question battery, answered demographic questions, and then completed a modified version of the experiment from Study 2.

In this version of the experiment, policy preferences were drawn exclusively from the unbranded issues of occupational licensing, eminent domain, and allowing mobile carriers to sell customer data. Instead of randomizing vignettes with party-branded or unbranded issues, participants whose vignettes included policy preferences were randomly assigned to view one of two prevignette introductions to the referenced policy issues. One indicated that Democrats and Republicans had taken opposing stances about the issue, whereas the other merely referenced that the issue had been debated (see Figure 5). Republicans' and Democrats' stances on the issues were randomized. Participants saw a total of two vignettes and only provided thermometer ratings for the vignette subjects. The resulting data were analyzed in a manner identical to Study 2.

Results

Similar to Study 2, participants felt approximately 18 degrees warmer toward copartisan vignette subjects, as compared with out-partisan vignette subjects, when given no information about the subject's policy preferences ($SE_{Unbranded} = 1.21$, $SE_{Party\text{-}Branded} = 1.18$, $p < 0.001$). As expected, adding policy preferences to vignettes diminished the effect of shared partisanship on thermometer ratings, and the size of this change depended on whether these preferences had been indicated to be partisan.

When branded as partisan, introducing policy preferences decreased the effect of shared partisanship by 58% (left side of panel C in Figure 4, $\beta = -10.31$, SE $= 1.66$, $p < 0.001$). This effect is nearly identical to the effect of party-branded preferences in Study 2 (57%). However, when preferences were presented without reference to the political parties, the effect of shared partisanship decreased by only 39% ($\beta = -7.01$, SE $= 1.72$, $p < 0.001$). Thus, policy preferences diminished the effect of partisanship on interpersonal affect considerably more when they provided signals about partisan identity, but the effect of policy disagreement alone was still substantial. A post hoc linear mixed-effects model indicated that the difference between the effects of shared partisanship, given policy preferences that had versus had not been branded as partisan, was statistically significant (Difference $= 3.33$, SE $= 1.78$, $p < 0.05$, one-tailed).

Yet, the main effect of party-branded preferences also masked considerable heterogeneity between the effects of party-loyal and party-disloyal preferences. When a vignette subject took a party-loyal policy position, the effect of shared partisanship declined by only 40%, an effect nearly identical to that of unbranded preferences (right side of panel C in Figure 4, $\beta = -7.13$, SE $= 2.24$, $p < 0.01$). However, when the vignette subject's position was disloyal to the subject's party, the effect of shared partisanship declined by 76% ($\beta = -13.52$, SE $= 2.23$, $p < 0.001$). A post hoc linear mixed-effects model indicated

[14] As with Study 2, we limited our analyses to self-identified partisans and those participants who leaned toward either the Republican or Democratic party.

that the effect of shared partisanship, given party-loyal versus party-disloyal preferences, was statistically significant (Difference = 6.39, SE = 2.71, $p < 0.05$). Thus, these results confirm that party-branded preferences mediate the effects of partisanship on interpersonal affect.

Tests of Robustness and Alternative Hypotheses

By holding policy substance constant, Study 3 more directly confirms that policy preferences drive affective polarization by signaling partisan identity. However, in introducing policy issues, Study 3 may have also made policies particularly salient to participants' decisions. As a robustness test, on p. 9 of our SI, we present a preregistered, conceptual replication of Study 3 that used a shorter, within-vignette manipulation of partisan branding. This study (Study 4) was less successful in branding nonpartisan issues, and the main effect of partisan branding was insignificant. However, the gap between the effects of party-loyal and party-disloyal preferences was significant, as before ($\beta = 3.84$, SE = 1.43, $p < 0.01$).

An alternative interpretation of Study 3's and Study 4's results might argue that party loyalty and disloyalty signals allow participants to infer *other* policy preferences a vignette subject might hold. In this way, simply stating that a partisan is loyal or disloyal may call to mind additional policy preferences (e.g., "If the subject toed the Republican line on this issue, she may also be pro-life, progun, and anti-welfare.") As such, the difference between the effects of party-branded and unbranded preferences may still be a function of policy disagreement. On p. 17 of our SI, we present the results of another preregistered experiment wherein we manipulated participants' capacity to make policy inferences (Study 5).

In this alternative-hypothesis test, some participants responded to vignettes under time pressure and were encouraged to rely on emotion when making judgments, whereas others were forced to wait to respond and encouraged to rely on logic when making judgments (Levine et al. 2018). Control participants completed a facsimile of Study 4. Consistent with the party-over-policy hypothesis, party-disloyal preferences had identical effects, regardless of whether participants were able to make additional policy inferences. Moreover, predictions about vignette subjects' stances on party-branded issues did not mediate the effects of party-disloyal preferences on interpersonal affect (p. 17 of SI).

Another potential complication for our results is that seemingly innocuous social facts (e.g., pop-culture preferences) can signal partisanship or identities connected to partisanship (e.g., race) and thus mediate partisanship's effects (Hetherington and Weiler 2018). To account for this possibility, we sought to identify social facts that differently affected Republicans and Democrats (e.g., endeared Republicans but repelled Democrats) and thus might mediate the effects of partisanship. As a final robustness test, we reran all of our main models while controlling for the interaction between participant partisanship and each social fact. As shown on p. 15 of our SI, these controls made no qualitative difference to our findings. Moreover, social facts did not have consistent effects across our studies, though our statistical power was likely insufficient to estimate these effects.

Discussion

Our findings demonstrate that the relationship between partisanship and interpersonal affect is not confounded by policy disagreement. Rather, policy preferences drive affective polarization, in large part, by signaling partisan identity. Past studies have apparently overestimated policy disagreement's role in affective polarization by relying on issues that signal partisan identity (Study 1). By contrast, we find that policy preferences only erase the effect of partisanship on interpersonal affect when preferences are party branded—that is, when preferences signal partisan identity (Study 2).

Linking the parties with unbranded policy preferences, we confirm that party-branded preferences mediate the partisanship–affect relationship by signaling partisan identity, and particularly party disloyalty (Studies 3 and 4). These effects are not a function of inferences about other policy preferences (Study 5). Altogether, our results suggest that partisan identity is the principal mechanism of affective polarization.

Our results underscore that, when arbitrating between the party-over-policy and policy-over-party hypotheses, scholars must consider how the effects of policy and party flow through (i.e., are mediated by) one another. The effects of policy preferences on interpersonal affect appears to be less mediated by partisanship (Path A of Figure 6a; p. 12 of SI). That is, policy preferences' effects on affect are primarily direct. As we have shown, partisanship also has a large effect on interpersonal affect. But this effect is largely mediated by party-branded policy preferences (Paths B' and C' of Figure 6b). When studies rely on party-branded preferences, these features can create the impression that party does not matter relative to policy. These revised models might

FIGURE 6 Policy-Through-Party and Party-Through-Policy Hypotheses of Affective Polarization

A Policy-Through-Party **B** Party-Through-Policy

Note: Hypotheses depicted as directed acyclic graphs. Letters indicate distinct paths

be more accurately termed the *policy-through-party* and *party-through-policy* hypotheses.

Though our studies suggest policy disagreement's role in affective polarization has been exaggerated, policy disagreement still has a substantial effect on interpersonal affect. Preferences about unbranded policy issues diminished the effect of partisanship on thermometer ratings by around a third—29% in Study 2 and 39% in Study 3. As unbranded issues are orthogonal to partisanship, these effects can only be a function of policy disagreement. Thus, both partisan identity and policy disagreement contribute to affective polarization. Moreover, the *total* effect of shared policy preferences (all paths of Figure 6a) on thermometer ratings exceeds that of shared partisanship (all paths of Figure 6b). However, the total effects of partisanship and policy preferences on social closeness are comparable.

We believe our dependent variables—thermometer ratings of, and social closeness toward, vignette subjects—capture feelings toward mass partisans and not party elites, though we cannot be certain. The gap between in- versus out-party affect in our studies is comparable to that observed by Orr and Huber (2020), who focused on affect toward mass partisans (18–23 degrees), and smaller than that in the ANES (41 degrees), which Druckman and Levendusky (2019) argue captures attitudes toward party elites. Though smaller than the party–elite gap in affect, affective polarization between mass partisans is still substantial, and has troubling consequences for social cohesion and fairness. For instance, Iyengar and Westwood (2015) find that citizens discriminate between job candidates more on the basis of partisanship than race. We also cannot judge whether Study 1 participants had party elites in mind when deciding where "most Republicans" or "most Democrats" stood on policy issues. Nonetheless, citizens are willing to generalize the stances of party elites on party-branded issues to mass partisans (Orr and Huber 2020; Rothschild et al. 2019).

Our findings have both methodological and substantive implications. On the former, they underscore the importance of design-based approaches to overcoming information equivalence in survey experiments (Dafoe, Zhang, and Caughey 2018). On the latter, our findings contradict the idea that affective polarization is primarily a function of Republicans' and Democrats' increasingly extreme or sorted policy preferences (e.g., Bougher 2017; Webster and Abramowitz 2017). However, our results cannot say why partisanship has become an increasingly important social identity for Americans.

Consistent with the argument of *The American Voter* more than 60 years ago, our results suggest partisanship is more than a "running tally" of policy preferences and past political experiences (Campbell et al. 1960). Partisanship entails an emotional attachment to, and a desire to protect the integrity of, one's political party as a social group. Citizens care about whether others conform to their parties' policy stances, regardless of the content of those stances. This pattern mirrors those found in studies of other social identities (Marques, Yzerbyt, and Leyens 1988; Marques et al. 1998).

Our results are also instructive for attempts to mitigate affective polarization. Insomuch as affective polarization reflects policy disagreement, addressing polarization would likely require reversing decades of elite polarization or mass ideological sorting between the parties (Levendusky 2009, 2010). By contrast, insomuch as affective polarization reflects partisan–identity salience, shifting citizens' focus to shared identities (Levendusky 2018) or correcting party stereotypes (Ahler and Sood 2018) may be useful in reducing polarization. Both are tall orders. However, identity-based interventions have shown promise.

In sum, we believe our results offer compelling evidence that partisan identity fuels affective polarization, independently of policy disagreement. Affective polarization threatens citizens' ability to come together, fairly weigh political outcomes, hold elected leaders accountable, and ultimately self-govern. Nothing could strike closer to the core of the democratic enterprise, or be more essential for the health of American democracy.

References

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

Abraham Almond, Gabriel, and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations.* Princeton, NJ: Princeton University Press.

Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102(2): 215–32.

Barber, Michael, and Jeremy C. Pope. 2019. "Does Party Trump Ideology? Disentangling Party and Ideology in America." *American Political Science Review* 113(1): 38–54.

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

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

Bougher, Lori D. 2017. "The Correlates of Discord: Identity, Issue Alignment, and Political Hostility in Polarized America." *Political Behavior* 39(3): 731–62.

Bullock, John G. 2011. "Elite Influence on Public Opinion in an Informed Electorate." *American Political Science Review* 105(3): 496–515.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter.* Chicago, IL: University of Chicago Press.

Cohen, Geoffrey L. 2003. "Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs." *Journal of Personality and Social Psychology* 85(5): 808–22.

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

Costa, Mia. 2021. "Ideology, Not Affect: What Americans Want from Political Representation." *American Journal of Political Science* 65(2): 342–58.

Dafoe, Allan, Baobao Zhang, and Devin Caughey. 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26(4): 399–416.

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

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections.* New Haven, CT: Yale University Press.

Fowler, Anthony. 2020a. "Partisan Intoxication or Policy Voting?." *Quarterly Journal of Political Science* 15(2): 141–79.

Fowler, Anthony. 2020b. "Defending Sober Voters against Sensationalist Scholars: A Reply to Rogers." *Quarterly Journal of Political Science* 15(2): 213–19.

Goggin, Stephen N., John A. Henderson, and Alexander G. Theodoridis. 2019. "What Goes with Red and Blue? Mapping Partisan and Ideological Associations in the Minds of Voters." *Political Behavior* 42(1): 985–1013.

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27(2): 163–92.

Heit, Evan, and Stephen P. Nicholson. 2016. "Missing the Party: Political Categorization and Reasoning in the Absence of Party Label Cues." *Topics in Cognitive Science* 8(3): 697–714.

Hetherington, Marc J., and Thomas J. Rudolph. 2015. *Why Washington Won't Work: Polarization, Political Trust, and the Governing Crisis.* Chicago, IL: University of Chicago Press.

Hetherington, Marc, and Jonathan Weiler. 2018. *Prius or Pickup?: How the Answers to Four Simple Questions Explain America's Great Divide.* Boston, MA: Houghton Mifflin Harcourt.

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

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

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

Iyengar, Shanto, Yphtach Lelkes, Matthew S. Levendusky, Neil Malhotra, and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22(1): 129–46.

Kinder, Donald R., and Nathan P. Kalmoe. 2017. *Neither Liberal Nor Conservative: Ideological Innocence in the American Public.* Chicago, IL: University of Chicago Press.

Lelkes, Yphtach. 2018. "Affective Polarization and Ideological Sorting: A Reciprocal, Albeit Weak, Relationship." *Forum (Germany)* 16(1): 67–79.

Lelkes, Yphtach. 2021. "Policy Over Party: Comparing the Effects of Candidate Ideology and Party on Affective Polarization." *Political Science Research and Methods* 9(1): 189–96.

Lelkes, Yphtach, and Sean J. Westwood. 2017. "The Limits of Partisan Prejudice." *Journal of Politics* 79(2): 485–501.

Levendusky, Matthew S. 2009. *The Partisan Sort.* Chicago, IL: University of Chicago Press.

Levendusky, Matthew S. 2010. "Clearer Cues, More Consistent Voters: A Benefit of Elite Polarization." *Political Behavior* 32(1): 111–31.

Levendusky, Matthew S. 2013. *How Partisan Media Polarize America.* Chicago, IL: University of Chicago Press.

Levendusky, Matthew S. 2018. "Americans, Not Partisans: Can Priming American National Identity Reduce Affective Polarization?." *Journal of Politics* 80(1): 59–70.

Levine, Emma E., Alixandra Barasch, David Rand, Jonathan Z. Berman, and Deborah A. Small. 2018. "Signaling Emotion and Reason in Cooperation." *Journal of Experimental Psychology: General* 147(5): 702–19.

Marques, José M., Dominic Abrams, Dario Paez, and Cristina Martinez-Taboada. 1998. "The Role of Categorization and In-Group Norms in Judgments of Groups and Their Members." *Journal of Personality and Social Psychology* 75(4): 976–88.

Marques, José M., Vincent Y. Yzerbyt, and Jacques-Philippe Leyens. 1988. "The 'Black Sheep Effect': Extremity of Judgments Towards Ingroup Members as a Function of Group Identification." *European Journal of Social Psychology* 18(1): 1–16.

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): 517–40.

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

McConnell, Christopher, Yotam Margalit, Neil Malhotra, and Matthew S. Levendusky. 2018. "The Economic Consequences of Partisanship in a Polarized Era." *American Journal of Political Science* 62(1): 5–18.

Milita, Kerri, Elizabeth N. Simas, John Barry Ryan, and Yanna Krupnikov. 2017. "The Effects of Ambiguous Rhetoric in Congressional Elections." *Electoral Studies* 46(1): 48–63.

Morgan, Stephen L., and Christopher Winship. 2014. "Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge: Cambridge University Press.

Niemi, Richard G., and Larry M. Bartels. 1985. "New Measures of Issue Salience: An Evaluation." *The Journal of Politics* 47(4): 1212–20.

Orr, Lilla V., and Gregory A. Huber. 2020. "The Policy Basis of Measured Partisan Animosity in the United States." *American Journal of Political Science* 64(3): 569–86.

Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40(3): 825.

Rogers, Steven. 2020. "Sobering Up after 'Partisan Intoxication or Policy Voting?'." *Quarterly Journal of Political Science* 15(2): 181–212.

Rogowski, Jon C., and Joseph L. Sutherland. 2016. "How Ideology Fuels Affective Polarization." *Political Behavior* 38(2): 485–508.

Rothschild, Jacob E., Adam J. Howat, Richard M. Shafranek, and Ethan C. Busby. 2019. "Pigeonholing Partisans: Stereotypes of Party Supporters and Partisan Polarization." *Political Behavior* 41(2): 423–43.

Settle, Jaime E., and Taylor N. Carlson. 2019. "Opting Out of Political Discussions." *Political Communication* 36(3): 476–96.

Tajfel, Henri. 1981. *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge: Cambridge University Press.

Webster, Steven W., and Alan I. Abramowitz. 2017. "The Ideological Foundations of Affective Polarization in the U.S. Electorate." *American Politics Research* 45(4): 621–47.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A: Models in Tabular Form
Appendix B: Participant Demographics for Studies 1–5
Appendix C: Stimuli / Question Phrasing
Appendix D: Conceptual Replication of Study 3 (Study 4)
Appendix E: Comparative Effects of Shared Policy Preferences
Appendix F: Controlling for Social Facts
Appendix G: Party Disloyalty Triggers Policy Inferences? (Study 5)
Appendix H: Ethics Statement