Mediatized Campaign Attacks Fuel Affective Polarization if Perceived as Negative: Experimental Evidence with American Voters

ALESSANDRO NAI
The University of Amsterdam, The Netherlands

JÜRGEN MAIER
The University of Koblenz-Landau, Germany

There is mounting evidence in the United States and worldwide that highlights a widespread and deepening “principled dislike” between partisan groups. Stemming from group identity dynamics, such as “affective polarization,” it is likely to be triggered by exposure to intra-elite conflicts, such as campaign negativity and incivility. However, empirical evidence for this effect is scarce, and it rests only on survey data; causal evidence linking campaign attacks and affective polarization is still missing. In this article, we advance the hypothesis that the effects of exposure to mediatized political attacks are likely mediated by how negative such attacks are perceived. To test our expectations, we leverage new evidence from an online experiment with convenience sample of American voters (N = 1,081). Our results suggest that exposure to intra-elites’ political attacks can drive affective polarization, but this unfolds mostly as a function of perceived negativity of those messages, and only for respondents that are ideologically affiliated with the target of the attack. Negativity is in the eye of the beholder, especially when one is being attacked.

Keywords: affective polarization, negative campaigning, incivility, experiment, USA

Many have observed that the political discourse in the United States is becoming increasingly “negative” (e.g., Geer, 2012; West, 2018), and this development could be associated with various detrimental political and societal outcomes. Decreasing turnout and political mobilization, lower political efficacy and trust, higher apathy, and a generally “gloomier” public mood have all been linked, directly or indirectly, to exposure to political attacks (e.g., Ansolabehere, Iyengar, Simon, & Valentino, 1994; Yoon, Pinkleton, & Ko, 2005). Campaign negativity might also have a potentially positive role to play—for instance, by making new relevant information salient in the mind of voters (Polborn & Yi, 2006). However, the risk of nefarious systemic effects cannot be underestimated. Importantly, because of exposing rifts between irreconcilable political positions, attacks between political elites can cause ideological polarization of the masses (Ansolabehere & Iyengar, 1995) and deepen political disagreements. Recent research points to a...
more fundamental degree of disagreement and polarization, built on affective elements (affective polarization; e.g., Iyengar & Krupenkin, 2018; Iyengar, Leikes, Levendusky, Malhotra, & Westwood, 2019; Iyengar, Sood, & Leikes, 2012; Sood & Iyengar, 2016). While ideological polarization describes a dissent with regard to political issues, affective polarization contains a fundamental and “principled” dislike of political opponents. The implications of affective dislike for opponents are potentially far-reaching. While ideological differences can generally be addressed through informational exchanges, negotiation, and mutual give and take, compromises are less likely if political opponents are perceived as hostile, despicable, and deserving contempt. In this sense, affective polarization could cement political and societal divisions that are beyond the reach of educational and civic initiatives.

To what extent are political attacks between elites associated with the deepening affective rift among voters? Recent evidence suggests that voters experience negative emotions when exposed to political attacks (e.g., Roseman, Mattes, Redlawsk, & Katz, 2020; Walter & Ridout, 2021), likely creating incentives for the targeted in-group to (affectively) move away even further from the sponsor of the attack. Yet, only little evidence links exposure to attacks with out-group dislike and affective polarization more in general. The little existing evidence furthermore rests on observational data (e.g., Iyengar et al., 2012), unable to ultimately disentangle the underlying causal mechanisms (Bougher, 2017). Since affective polarization could also be hypothesized to foster attack politics—either, strategically, to mobilize an increasingly polarized base by cueing affective dislike for the opponents in election campaigns, or simply because conflict calls for conflict—empirical evidence is still needed to assess the causal association between (exposure to) political attacks and affective polarization.

In this article, we investigate one specific mechanism linking exposure to political attacks and increased affective polarization, focusing on how negative messages are perceived. On average, political attacks tend to suffer from a normatively unfavorable image, as the public at large tends to dislike them (Fridkin & Kenney, 2011, 2019, pp. 45–46). However, some segments of the electorate seem to have more nuanced opinions about political attacks (Mattes & Redlawsk, 2015) or even appreciate them (Nai & Maier, 2021b). It is generally expected that evaluative perception of (negative) campaign messages links exposure to such attacks with affective polarization. Furthermore, in line with evidence suggesting the existence of a partisan bias in the perception of incivility (Liang & Zhang, 2021), we expect this effect to be contingent on the partisan affiliation of the attacker and the affiliation of those exposed to such attacks. In doing so, our article contributes to the existing evidence on negative campaigning and affective polarization in a threefold way: (i) by presenting causal evidence linking the two phenomena; (ii) by showing that this link is contingent on how negative political attacks are perceived; and (iii) by exploring the social identity dynamics of such effect, most notably in terms of in- and out-party proximity with the attacker and the target.

We do so via new experimental evidence gathered for a convenience sample of American respondents (MTurk; N = 1,081). In the experiment, participants were randomly shown a newspaper article presenting a statement from a fictive Republican candidate either promoting his stance on taxation or attacking his (equally fictive) Democratic opponent on that same issue. The setup allows us to examine the consequences of negativity through the media, complementing research that has shown that media coverage of negative exchanges between political actors can affect political attitudes and behavior (e.g., Min, 2004). Respondents were then asked to evaluate the content of the message they were exposed to (e.g., in terms of negativity) and finally to
describe their feelings about Democrats and Republicans, which we use to measure in- and out-group sentiments. Our results show that exposure to attack messages does indeed increase affective polarization—but predominantly through mediation by the perceived level of negativity. Furthermore, this exists only for respondents that are ideologically affiliated with the target of the attack.

All data, codes, and appendixes are available for replication at the following Open Science Foundation (OSF) repository: https://osf.io/89z5p/.

What Drives Affective Polarization?

Within the American context, two main rationales have been advanced to explain the rise of affective polarization. On the one hand, increased partisan dislike comes from ideological polarization, simply following the idea that "increased ideological differences between political figures produce increasingly polarized affective evaluations" (Rogowski & Sutherland, 2016, p. 485). Ideological polarization has been widely documented in the United States. Data on members of Congress show an increasing divide of the political elites along party lines (e.g., Andris et al., 2015). At the level of the public as well, polls show that supporters of the Democratic and Republican parties have grown increasingly distant in their ideological positions since the 1970s; this trend has accelerated since the beginning of the century (e.g., Webster & Abramowitz, 2017). Growing ideological differences indicate a shrinking of the shared ideological platform across different camps, and such shrinking should naturally come with increasing distrust, if not outright dislike, for those who are seen as not sharing compatible values and political outlooks. Yet, although correlational and experimental data seem to confirm the relationship between ideology and affective polarization (e.g., Rogowski & Sutherland, 2016; Webster & Abramowitz, 2017), the former seems to explain only a small portion of the latter (Lelkes, 2018).

On the other hand, it has been argued that affective polarization is rooted in social identity dynamics. To boost feelings of self-esteem, individuals naturally identify with their "in-group"—be it cultural, social, political, or otherwise—which naturally comes with a devaluation of the "out-group" (Tajfel, 1981). This "in-group bias" is associated with higher levels of political activity and stronger negative emotional reactions when the in-group itself faces a threat (Mason, 2015). In other words, a partisan individual "behaves more like a sports fan than like a banker choosing investment" (Mason, 2015, p. 129; see also Miller & Conover, 2015), and this pattern is reinforced by an intertwining between partisan identities and specific social identities, such as religion or race ("social sorting"; e.g., Mason, 2016). As a result, as only "few political and social identities cut across both parties, partisans no longer see themselves having much in common with the other side" (Lelkes, 2018, p. 68). This likely fosters misperceptions, stereotyping, and hostility toward the out-group (Ahler & Sood, 2018; Iyengar et al., 2012; Mason, 2018).

On top of these two "structural" drivers, a question could be raised as to whether the informational context also leads to changes in affective polarization. Several studies have, directly or indirectly, suggested that the current style of political communication might fuel partisan dislike (e.g., Iyengar et al., 2012, 2019; Lelkes, 2018; Levendusky & Malhorta, 2016; Sood & Iyengar, 2016), as positions expressed by political elites in both camps are increasingly irreconcilable, and explicitly or implicitly surf on out-party hostility.
This trend is, for instance, reflected by the sharp increase of negative campaigning in recent decades (e.g., Fowler, Franz, & Ridout, 2016, p. 53; Geer, 2012). Most attacks simply show a disagreement between the policy positions of the sponsor and the target of the attack (e.g., Benoit, 2007, p. 121), which are likely to reinforce perceptions of ideological polarization among elites. On top of that, political attacks tend to be perceived as rude by most voters (Taylor, 2017)—which is likely to contribute to perceptions of hostility and thus activate subjects’ social identity.

Negative Campaigning and Affective Polarization

The Direct Effect of Campaign Negativity

Iyengar and colleagues (2012) discuss one of the few empirical pieces of evidence suggesting that exposure to negative campaigns fosters affective polarization, based on a triangulation of survey data and content analysis of campaign ads for the 2004 and 2008 U.S. presidential elections (pp. 407–408). In particular, they suggest that the increase of negative campaigning and the enhanced interest of the media in covering confrontational exchanges between the candidates are responsible for the observed affective polarization among partisans (Iyengar et al., 2012, p. 408, 2019). Their results are quite compelling: First, the difference in evaluations for the two parties is significantly larger in battleground states. This difference further magnifies when Election Day approaches—likely because campaigns tend to get more negative at the end (e.g., Fridkin & Kenney, 2019, p. 65). Second, the difference between supporters of the sponsor and supporters of the target of an attack becomes larger with the number of negative ads aired in a state. This aggregate trend is likely enhanced by the tendency of news media to extensively cover negative ads (Geer, 2012). Indeed, exposure to news media covering conflict and partisan polarization tends to increase affective polarization (Levendusky & Malhorta, 2016).

The idea that the link between negativity and affective polarization also exists at the individual level is supported by research showing that exposure to political attacks generates negative emotions. For instance, Roseman and colleagues (2020) show that contempt, a feeling of revulsion “elicited by appraisals that a person has an undesirable trait, such as bad character or incompetence,” (p. 5) is the most important driver of candidate assessments after exposure to attack politics (see also Walter & Ridout, 2021). Gervais (2021) also shows that incivility causes disgust, which occurs when “social or moral boundaries appear to be violated” (Simpson, Carter, Anthony, & Overton, 2006, p. 32). As these negative emotional responses are likely directly associated with the source of the messages, they can be expected to deepen the affective rift between the opposed camps, thus fostering affective polarization. Furthermore, we expect this effect to be a function of the “intensity” of the attacks. Social comparison theory argues that competition influences whether and how individuals protect their identity (Garcia, Tor, & Schiff, 2013; Miller & Conover, 2015), and the extent of such defense mechanisms is a function of the severity of the attacks on the in-group. Harsh messages—character attacks, uncivil attacks—directed against one’s “own” candidate are likely to exacerbate the competitive situation between the in- and the out-group and the subsequent identity threats for the former. Therefore, we expect the direct effect of negativity on affective polarization to be the strongest for uncivil attacks, followed by character attacks.
H1: Exposure to political attacks increases affective polarization.

H2: The positive effect of political attacks on affective polarization is particularly strong for uncivil attacks and character attacks.

The Mediating Role of Message Perception

At a very general level, political attacks make salient political conflicts, and hence, they are likely to activate conflict attitudes in the public. Yet, not all attacks are perceived as such. Growing evidence suggests that different voters perceive political attacks very differently—for instance, voters high in psychopathic tendencies seem more likely to appreciate negativity, whereas voters high in conflict avoidance clearly reject it (Nai & Maier, 2021b). Also, some attacks and critiques can convey interesting and factually relevant information (Fridkin & Kenney, 2011; Polborn & Yi, 2006), and the recipient might not suffer of a normatively negative perception by the public at large (Fridkin & Kenney, 2011, 2019, pp. 45–46). Other attacks, inversely, might particularly resonate for some voters and become perceived as highly negative. With this in mind, the same attack is likely perceived differently by different people. Although humans have a general tendency to physiologically react to negative information (“negativity bias,” see, e.g., Soroka, Fournier, & Nir, 2019), research indicates that “citizen perceptions of negativity differ considerably” (Lipsitz & Geer, 2017, p. 584; see also Sigelman & Kugler, 2003); the same applies for incivility (Liang & Zhang, 2021). What might matter more than simple exposure to an attack is whether such attack was perceived as particularly negative. We thus expect that affective polarization should be higher in situations where an attack message is perceived as strong (i.e., as particularly negative) than in situations where an attack is perceived as weak and therefore not particularly threatening.

H3: Perceiving the attack as more negative increases affective polarization.

Partisan Proximity

Finally, the perception (and effectiveness) of political attacks likely depends on partisan proximity with the sponsor and the target of the attacks. With respect to social identity theory (Tajfel, 1981; Tajfel & Turner, 1986), we can expect that exposure to attack messages will increase the salience of the characteristics of the group to which voters belong. Attacks on the in-group (or one of its standard-bearers) create a competitive situation where the electorate is called upon to decide on rivaling policy stands and personalities. If effective, an attack message can turn the electorate away from the voter’s preferred party. From this perspective, attacks are a threat to the positive evaluation of the in-group—and subsequently to the voter’s self-esteem. This threat will trigger voters to comparatively evaluate the in- and the out-group (Miller & Conover, 2015). To protect or restore their self-concept, voters’ evaluations of in- and out-group will be biased, that is, the in-group will receive positive evaluations, whereas the out-group will receive negative evaluations (Iyengar & Westwood, 2015, p. 691). This bias should increase with the perceived virulence of the attack. In this sense, we should expect attacks to particularly foster affective polarization if targeting the in-group, compared with attacks that come from the in-group. Indeed, research on the perception of negative messages has shown that supporters of the sponsor and the target differ considerably in their evaluations of attack messages. Most notably, Lipsitz and Geer (2017) show that both Democrats and Republicans evaluate an attack leveled by “their” presidential
candidate as fairer compared with supporters of the opposite party. In this sense, we should expect respondents to react more negatively to attacks sponsored by a candidate from the out-group and targeting a candidate from their own party, than the reverse.

\textit{H4: Exposure to political attacks increases affective polarization in particular among those ideologically closer to the target of the attack.}

Research Design

Participants

We report below the results of a between-subjects experiment embedded in an online survey distributed to a convenience sample of U.S. residents. Participants were recruited in December 2019 via the Amazon MTurk online platform for a small compensation ($0.7 in total). MTurk samples cannot be assumed to be representative of the population. Nevertheless, the literature seems to suggest that MTurk offers a cheap and reliable alternative to more traditional surveys (Casler, Bickel, & Hackett 2013; for a more critical take, see Ford, 2017; Harms & DeSimone, 2015; Hauser & Schwarz, 2016).

The questionnaires included an “attention check” (Berinsky, Huber, & Lenz, 2012; Oppenheimer, Meyvis, & Davidenko, 2009), where specific instructions—select the option “other” and write a keyword in the entry box—were embedded within a long and digressing question. Respondents that failed such attention check were assumed to skim only through the questions and the treatment and were excluded (\(N = 25\)). The final sample, after exclusion of these respondents, includes \(N = 1,081\) respondents. The final sample is composed of 50.8% female respondents, with an average age of 39.7 years (\(SD = 12.4\)); 55.1% of respondents are Democrats, 34.3% are Republicans, and 10.6% are Independents. See Appendix C (Tables C1 and C2) for more details on sample composition, including the distribution of observations per U.S. state. Comparing the sample composition in terms of gender and age brackets strata with official census data from 2018 (United States Census Bureau, n.d.), it appears that our sample clearly oversamples young males between 20 and 34 (ratio population-sample is only .62), and strongly undersamples older males (ratio 3.95) and even more older females of 65+ (ratio 4.76; 11.9% in the population, and only 2.5% in our sample). Yet, the nonrepresentativeness of the sample is not particularly problematic, as we do not use it to produce point estimates but to compare trends across experimental groups in which observations were fully randomized (including in terms of gender and age; see below).

Procedure

Participants answered several batteries related to their individual profiles. Afterward, they were exposed to the experimental treatment. In the first step, participants received some information on two fictitious male candidates, a Democrat and a Republican. By exposing subjects to media coverage of candidates’ attack behavior, we tested the consequences of mediatized negativity, or negativity through the media, which is conceptually distinct from negativity by the media (i.e., media criticism of politics; for a similar distinction on the role of the media in populist communication, see Esser, Stępińska, & Hopmann,
This is also different from direct exposure to candidates’ negative campaign communication, which is tested in the existing evidence linking negativity and affective polarization (Iyengar et al., 2012). All treatments were designed as a newspaper article including a quote from the Republican candidate on taxation. Although social media has become a major platform for information about politics (Reuters Institute for the Study of Journalism, 2020, p. 88), the traditional news media is still the most important source of information for an overwhelming majority of U.S. adults. In 2019, about 80% said that TV, radio, and the press (including news websites or apps) are “the most common way they get political and election news” (Pew Research Center, 2021, p. 34). Media coverage of campaigns can amplify communication events that are targeted to smaller audiences (e.g., campaign rallies), as well as alternative mass communication channels such as campaign ads (Geer, 2012) or social media posts (e.g., von Nordheim, Boczek, & Koppers, 2018). In addition, information provided by classical news sources can be expected to be particularly persuasive as most outlets are considered as trustworthy in stark contrast to news in social media (Reuters Institute for the Study of Journalism, 2020, p. 88).

Subjects were told that both candidates were competing for a seat in the next U.S. House of Representatives for Minnesota’s 9th Congressional district (Minnesota only has 8). Inspired by the experimental design used by Brooks and Geer (2007), respondents were then randomly exposed to one of 11 treatments. Randomization checks indicate a successful random distribution of respondents across all treatments according to their gender, age, education, and party identification (\( p > .05 \) throughout).

In two of the treatments, the sponsor of the message praised his policy stands. In all other treatments, he attacked his opponent. Attacks were systematically varied with respect to focus (i.e., policy vs. character), and character attacks could be either civil or uncivil. The core statement, i.e., the aspect that mattered in the respective version of the treatment, was highlighted. Within the framework of research on negative campaign communication, candidate messages were distinguished along three dimensions: tone, focus, and (in)civility (e.g., Brooks & Geer, 2007, Hopmann, Vliegenthart, & Maier, 2018). The tone determines whether a message is positive or negative. Negativity is defined as “any criticism leveled by one candidate against another” (Geer, 2006, p. 23). The focus of the message describes whether an attack is on the opponent’s policy positions or his personal characteristics. Finally, a message can be civil or uncivil (regardless of its focus). Although it is not yet clear exactly what incivility is (for a theoretical framework see Bormann, Tranow, Vowe, & Ziegele, 2022), it seems to be clear that uncivil messages “violate some agreed upon standard of society” (Maisel, 2013, p. 204). Hence, negative campaign messages can have very different designs depending on their focus and their level of (in)civility. With this in mind, our analyses will compare the effects of exposure to positive (\( N = 192 \)) versus negative (\( N = 889 \)) messages; policy (\( N = 100 \)) versus character (\( N = 789 \)) attacks; and, within character attacks, civil (\( N = 395 \)) versus uncivil (\( N = 394 \)) attacks. See Appendix B for more details about treatments.

1 For positive treatments, only the focus of the message varied. Character attacks varied in two additional dimensions: the use of fear appeals and the use of populist rhetoric; these two additional factors are not investigated in this article.
Manipulation checks were successful. Compared with respondents exposed to a positive message, respondents who read a negative message were significantly more likely to evaluate it as “negative,” \( t(1,079) = -21.65, p < .001 \). Respondents who were exposed to a character attack were significantly more likely to evaluate it as “harsh” than participants exposed to a policy attack, \( t(887) = -5.52, p < .001 \). Even within respondents exposed to a character attack, those who were exposed to an uncivil attack were significantly more likely to perceive it as “harsh” than respondents exposed to a civil attack, \( t(787) = -12.25, p < .001 \).

**Measures**

**Affective Polarization**

Following Iyengar and colleagues (2012), we measured affective polarization as the extent to which respondents hold strong negative opinions for the out-party while holding strong positive opinions about the in-party when it comes to traits that are stereotypically attributed to party supporters (p. 412). In two batteries (asked after the experiment), respondents had to declare whether they agreed or disagreed with the fact that Republicans and Democrats could be described as “patriotic,” “closed-minded,” “intelligent,” “hypocritical,” “selfish,” “honest,” “open-minded,” “generous,” or “mean” (from 1 “disagree strongly” to 7 “agree strongly”). After reversing the coding of some items and accounting for respondent’s party ID,\(^2\) we calculated the mean score across these adjectives to obtain a measure of positive sentiments for the in-party. The same procedure was used to obtain a measure of negative sentiments for the out-party.\(^3\) Unsurprisingly, positive sentiments for the in-party are significantly correlated with negative sentiments for the out-party, \( r(965) = .32, p < .001 \). The average score across these two variables yields a measure of affective polarization that reflects positive opinions of the in-group and negative opinions of the out-group. The obtained variable ranges from 1 “very low” to 7 “very high” (\( M = 4.92, SD = .88 \)).

**Perceived Negativity of the Message**

To measure perceived negativity of the treatment, respondents were asked the extent to which they agreed or disagreed that the excerpt they just read was “negative.” Responses were measured on a 7-point scale from 1 “disagree strongly” to 7 “agree strongly” (\( M = 4.99, SD = 1.91 \)).

---

\(^2\) Partisan identification was measured following the protocol used in the American National Election Study. Respondents were first asked whether they usually think of themselves as a Democrat, a Republican, or an independent; in the first two cases, respondents were then asked if they would strongly call themselves a strong Democrat or Republican (yes, no); respondents initially identifying as independents were asked whether they would think of themselves as closer to the Republican Party or to the Democratic Party (or neither). The combination of these variables yielded a simplified binary variable (0 Democrat, 1 Republican), plus a third category of independents (\( N = 114, 10.6\% \)). Because a clear out-party could not be identified for them, which was central for the measure of affective polarization, respondents classified as independents were excluded (\( N = 114, 10.6\% \)).

\(^3\) Reliability of the index was high (\( \alpha = 0.91 \)) for both sentiments about Republicans and Democrats.
Results

Campaign Negativity and Affective Polarization

Table 1 estimates affective polarization for all respondents in our sample as a function of exposure to different types of political messages. Model 1 (M1) estimates the effect of exposure to a political attack (vs. a positive message); M2 estimates the effect of exposure to a character (vs. policy) attack; and within character attacks, M3 estimates the effect of exposure to uncivil (vs. civil) attacks. Neither a negative message in general nor the harshest form of attacks (uncivil ones) significantly impacts affective polarization on its own. This being said, respondents that were exposed to a character attack scored significantly higher on the affective polarization scale than respondents exposed to a policy attack. Because this effect remains relatively marginal (the difference is of .2 points on 1–7 scale), and because only one dimension of negativity has a significant effect, we prefer to generally reject H1 and H2. The simple exposure to campaign negativity is unlikely to move respondents substantially when it comes to their partisan perceptions.

Table 1. Exposure to Campaign Messages and Affective Polarization; Direct Effects.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coef</td>
<td>SE</td>
<td>Sig</td>
<td>Coef</td>
</tr>
<tr>
<td>Negative&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.03</td>
<td>(.07)</td>
<td></td>
</tr>
<tr>
<td>Character attack&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>.21</td>
<td>(.10)</td>
</tr>
<tr>
<td>Uncivil&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>Constant</td>
<td>4.90</td>
<td>(.07)</td>
<td>***</td>
</tr>
<tr>
<td>N</td>
<td>967</td>
<td>792</td>
<td>698</td>
</tr>
<tr>
<td>R²</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
</tr>
</tbody>
</table>

Note. All models are OLS regressions. Dependent variable in all models is affective polarization, measured as the existence of positive in-party stereotypes and negative out-party stereotypes (1–7).

<sup>a</sup> Reference category is “Positive campaigning.”
<sup>b</sup> Reference category is “Policy attack” (nonattacks excluded).
<sup>c</sup> Reference category is “Civil attack” (nonattacks and policy attacks excluded).

*** p <.001, ** p < .01, * p < .05, † p < .1

Table 2 replicates the same models but this time includes two key covariates central for our remaining hypotheses: message perception and respondent ideological stance. The models also control for age, gender, and education of respondents. The key result in Table 2 is that, for all three types of attacks, respondents that perceive them as more negative are significantly more likely to score higher on affective polarization. This provides a first important piece of information in support of H3, according to which the effects of exposure to negative and harsh campaigns is mediated by perception of such messages.
Table 2. Exposure to Campaign Messages and Affective Polarization; Profile Controls.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th></th>
<th></th>
<th>M2</th>
<th></th>
<th></th>
<th>M3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef</td>
<td>SE</td>
<td>Sig</td>
<td>Coef</td>
<td>SE</td>
<td>Sig</td>
<td>Coef</td>
<td>SE</td>
</tr>
<tr>
<td>Negative</td>
<td>-.21</td>
<td>(.08)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>-.02</td>
<td>(.07)</td>
</tr>
<tr>
<td>Character attack</td>
<td>.18</td>
<td>(.09)</td>
<td>†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncivil</td>
<td></td>
<td></td>
<td></td>
<td>-.02</td>
<td>(.07)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived negativity</td>
<td>.08</td>
<td>(.02)</td>
<td>***</td>
<td>.12</td>
<td>(.02)</td>
<td>***</td>
<td>.12</td>
<td>(.02)</td>
</tr>
<tr>
<td>Republican</td>
<td>-.28</td>
<td>(.06)</td>
<td>***</td>
<td>-.21</td>
<td>(.06)</td>
<td>***</td>
<td>-.18</td>
<td>(.07)</td>
</tr>
<tr>
<td>Female</td>
<td>.08</td>
<td>(.05)</td>
<td></td>
<td>.05</td>
<td>(.06)</td>
<td></td>
<td>.09</td>
<td>(.06)</td>
</tr>
<tr>
<td>Age</td>
<td>.01</td>
<td>(.00)</td>
<td>***</td>
<td>.01</td>
<td>(.00)</td>
<td>***</td>
<td>.01</td>
<td>(.00)</td>
</tr>
<tr>
<td>Education</td>
<td>-.03</td>
<td>(.02)</td>
<td>†</td>
<td>-.03</td>
<td>(.02)</td>
<td>†</td>
<td>-.05</td>
<td>(.02)</td>
</tr>
<tr>
<td>Constant</td>
<td>4.82</td>
<td>(.24)</td>
<td>***</td>
<td>4.24</td>
<td>(.29)</td>
<td>***</td>
<td>4.47</td>
<td>(.29)</td>
</tr>
<tr>
<td>N</td>
<td>966</td>
<td></td>
<td></td>
<td>792</td>
<td></td>
<td></td>
<td>698</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.10</td>
<td></td>
<td></td>
<td>.12</td>
<td></td>
<td></td>
<td>.12</td>
<td></td>
</tr>
</tbody>
</table>

Note. All models are OLS regressions. Dependent variable in all models is affective polarization, measured as the existence of positive in-party stereotypes and negative out-party stereotypes (1–7).

a Reference category is “Positive campaigning.”
b Reference category is “Policy attack” (nonattacks excluded).
c Reference category is “Civil attack” (nonattacks and policy attacks excluded).
d Reference category is “Democrat” (independent and unaffiliated voters excluded).

*** p < .001, ** p < .01, * p < .05, † p < .1

All three models in Table 2 show, additionally, that Republicans are significantly less likely to score high on affective polarization, all things considered. While this is outside of the scope of our investigation, such effect was possibly driven by the political circumstances in the United States at the time of data collection, late 2019, deeply into the (first) Trump’s presidency. While evidence exists that asymmetric polarization is mostly due to the Republican party moving further away from the political center than the Democratic party (e.g., Hacker & Pierson, 2015), the profound loathing of democrats and liberals toward President Trump is well documented (e.g., Jacobson, 2020; Nai & Maier, 2021a). We believe that a case can be made that such loathing likely exacerbated affective polarization, especially among Democrats in this specific moment in time.

Table 2 also presents a somewhat puzzling result: Controlling for the new set of covariates, the effect of negativity on affective polarization becomes significant and negative. In other terms, accounting for how negative the message is seen by respondents (and their profile), exposure to negative messages decreases affective polarization. While the magnitude of the effect is not particularly strong, its direction is noteworthy. We can only speculate about the reasons of such effect, which in our opinion could be driven by two factors. On the one hand, this result suggests that exposure to positive campaigning increases affective polarization, perhaps by beefing up positive feelings for the in-group. Although the effect of the message perception remains strong, the fact that this effect is no longer significant in models using only the negative affect for the out-party as alternative measure of affective polarization (Appendix A, Table A2) seems to provide support for this hypothesis. On the other hand, this could simply reflect the idea that once the negative perceptions of negativity (message perception) are filtered out from the equation, negativity
in itself can be seen as a “normal” element of electoral democracy, perhaps even providing useful and relevant information for citizens to make up their mind (e.g., Polborn & Yi, 2006). Relevant information increases the likelihood that voters tolerate negativity and has less detrimental effects on, for example, candidate evaluation (Fridkin & Kenney, 2019).

**Mediated Effects**

Our second expectation is that the effect of campaign messages on affective polarization is mediated by their perception as being negative. Figure 1 presents the results of three mediation models, where affective polarization is estimated directly via exposure to the three different campaign messages (negative, character attack, and uncivil), plus indirectly via the perception of such messages as negative.

![Figure 1. Mediated effects by perceived negativity.](image)

*Note.* Coefficients are bootstrapped unstandardized regression coefficients (1,000 iterations). The dashed arrow represents the indirect effect. Dependent variable in all models is affective polarization, measured as the existence of positive in-party stereotypes and negative out-party stereotypes (1–7).

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .1$

As the figure clearly shows, negative messages, character attacks, and incivility are all significantly perceived as more negative than their counterparts (positive messages, policy attacks, civil attacks). Such
perception, as discussed earlier, drives affective polarization upward in all three cases. Importantly, exposure to negative messages, character attacks, and uncivil attacks significantly increases affective polarization via the perception of such messages as negative (indirect path).

The figure shows again the negative direct effect of campaign negativity on affective polarization (top-left panel), which, as discussed above, exists only when controlling for how negative the message was perceived. The fact that exposure to negative messages increases affective polarization via such a perception (indirect path) seems to confirm the argument we advanced above: Negativity could be perceived positively (or as relevant) and as such decrease polarization. The more a message is explicitly perceived as negative, though, the more it likely increases polarization. In this sense, it is the unfavorable normative evaluation of the message that drives its detrimental effects (increased affective polarization), not the message itself. Negativity is in the eye of the beholder (Lipsitz & Geer, 2017; Sigelman & Kugler, 2003), and indeed our results confirm all in all that message perceptions—above and beyond simple exposure to attacks—drive affective polarization. H3 is generally supported.

We also expected different effects of respondents depending on their in- or out-groups with regard to the attacker. Figure 2 shows the results of a series of mediation analyses that replicate the previous ones, but separately for different party IDs. Left-hand panels report results of analyses only run on Democrats (in-group with the target and out-group with the attacker), whereas right-hand panels are for analyses run on Republicans (in-group with the attacker and out-group with the target). In other terms, panels on the left are for respondents that should have particularly loathed the attack, as it targeted their own party tribe, whereas respondents on the right were on the side of the attacker.
Figure 2. Mediated effects by perceived negativity, by party affiliation.

Note. Coefficients are bootstrapped unstandardized regression coefficients (1,000 iterations). The dashed arrow represents the indirect effect. Dependent variable in all models is affective polarization, measured as the existence of positive in-party stereotypes and negative out-party stereotypes (1–7).

Left-hand panels are run only for respondents declaring a partisan affiliation with the Democrats (out-group with the attacker, in-group with the target of the attacks). Right-hand panels are run only for respondents declaring a partisan affiliation with the Republicans (in-group with the attacker, out-group with the target of the attacks).

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .1$
The mediation models show, first, that in general, negative messages, character attacks, and incivility are perceived as significantly more negative than their counterparts (positive messages, policy attacks, civil attacks) for both groups of respondents, with one exception: Republicans (in-party with the attacker) are not significantly more likely to find character attack as more negative than policy attacks—that is, they might have a greater tolerance for harsher attacks. Importantly, for Republicans, perceiving the messages as more negative is, in all three cases, unrelated to affective polarization. The main effects driven by this difference between Democrats and Republicans are highly consistent with our expectations (H4). For respondents exposed to an out-group attack (Democrats, left-hand panels), exposure to the attacks indirectly rather strongly impacts affective polarization (via perceived negativity); this is the case for all three types of messages (negative, character attacks, uncivil attacks). Inversely, for respondents exposed to an in-group attack (Republicans), exposure to the attack does not impact affective polarization, neither directly nor indirectly via message perception.

Interestingly, the negative direct effect of negativity on affective polarization discussed above for all respondents, once controlling for message perception, exists only among Democrats (top-left panel). Again, this effect is likely driven by the fact that simple negativity can be perceived as relevant once its negative evaluation is removed from the picture. Once such negative evaluation is controlled for, being exposed to critiques from the out-group might be perceived as a relevant new piece of information which can lead respondents to reevaluate their ideological priors. Unsurprisingly, this effect does not exist for character attacks and incivility.

All in all, our results suggest that exposure to media coverage of negative campaign messages can indeed drive affective polarization, but this unfolds mostly as a function of perceived negativity of those messages and only for respondents that are ideologically affiliated with the target of the attack. Negativity is in the eye of the beholder, especially when being attacked.

Robustness Checks

We have replicated our models using an alternative dependent variable, measuring affective polarization simply as the extent to which people dislike the out-party (reversed feeling thermometer, 0–100). The two variables are conceptually related, and empirically rather strongly (but not perfectly) correlated, $r(965) = .63$, $p < .001$. Results with this alternative dependent variable are robust—all effects described above for our main measure of affective polarization exist also for out-party dislike (Appendix A, Tables A1 and A2 and Figures A1 and A2). In one case, these alternative models even pick up an additional effect, absent in our main results but perfectly in line with the general trend described: For respondents exposed to an in-group attack (Republicans), exposure to an attack reduces affective polarization if that message is perceived as negative (Figure A2, top-right panel).

Results are also robust when including respondents who failed the attention test, and when excluding respondents from Minnesota ($N = 19$), where the fictive candidates in the experiment come from; see materials for replication in the OSF repository.
Summary and Conclusion

Does negative campaigning foster affective polarization? Existing research triangulating survey data and content analysis suggests that this is the case (Iyengar et al., 2012), but at the aggregate level and only in observational terms. Using data from a between-subjects online experiment with U.S. residents, we provide causal evidence to this link. Yet, the relationship between negativity and affective polarization is more complicated than expected. Attack messages do not directly impact affective polarization (or do so very weakly in some cases), leading to the rejection of H1 and H2, but do so through mediation by the perceived level of negativity (supporting H3). In other terms, whether and to what degree candidates’ attack behavior increases affective polarization by and large depends on how negative these attacks are perceived, in line with recent research on the effects of incivility (Liang & Zhang, 2021). Additionally, confirming H4, our results suggest that the mediated effect of negativity on affective polarization exists particularly for respondents that are exposed to attacks that target their ideological in-group (a candidate from their preferred party, in our case Democrats) and is completely absent for respondents that have an ideological affinity with the attacker (in our case, Republicans).

The potential consequences of our findings are manifold: While ideological differences can generally be solved through negotiation and mutual give and take, compromises are notably harder to achieve among people who loathe each other, inevitably leading to political blockades. In the United States, a divided government has become the default after World War II (Ansolabehere, Palmer, & Schneer, 2018), and affective polarization is unlikely to help solving this sclerotic situation. It is easy to imagine that those performance deficits will create disaffection among the public and cause heated debates on who is responsible for the situation. Our results suggest that political attacks, albeit indirectly, play an important role here. Moreover, our results could indicate the existence of a toxic circle of rhetorical discord, fundamental mutual dislike, and political discontent. Negativity remains a fundamental component of the public sphere; its “newsworthiness” is well-known; and indeed increasing evidence suggests that candidates that go negative are more likely to be picked up by news coverage of political events (e.g., Maier & Nai, 2020). At the same time, it is likely that affective polarization and partisan discord fuel an increased use of attack politics—either driven by increased ideological differences (Geer, 2012) or simply to placate increasingly tribal political bases. Negativity fosters polarization; polarization fosters negativity; and the preferential media coverage of political conflict adds fuel to the fire. Breaking this cycle is extremely difficult; it requires either that political actors rethink their behavior, or that the media resist the temptation to provide a platform to—and, thus, amplify the effect of—campaign messages that are “below the belt.” Neither is likely to happen anytime soon.

Our results come with some limitations. First, they are based on a single negative treatment. It would be very interesting to understand in more detail how exposure to repeated attacks works with respect to affective polarization, especially in light of research investigating the differential effects of sequences of negative and positive messages (e.g., Nai & Seeberg, 2018; on serial position effects, see Cowan, Saults, Elliott, & Moreno, 2002; Li & Epley, 2009). Of course, results appearing after one single exposure are likely to be much more conservative than results existing for multiple, repeated, constant exposure to attack politics. Similarly, the fact that the treatments in our study were presented as coming from a neutral source (the journalist covering the campaign in the fictive newspaper article) should yield weaker results when compared with partisan sources such as campaign ads or speeches. Further studies could investigate...
message perceptions since in this study, they are measured via a single statement. Most notably, this could be achieved by providing a more nuanced measurement of voters’ perceptions of negativity and incivility.

Second, the fact that (mediated) effects of negativity on affective polarization exist particularly when attacks are targeted toward respondents’ ideological in-group could potentially be spurious because of the nature of our experimental setup—and, notably, the fact that the attacker is always a Republican and the target is always a Democrat. Evidence in the United States (and abroad) seems to suggest that conservatives are more likely to showcase a preference for attack politics (Lau & Pomper, 2001; Nai & Sciarini, 2018), and even conservative campaign strategists tend to predilect attack politics (Theilman & Whilhite, 1998). Perhaps, what drives the absence of mediated results for Republicans is not because they are part of the in-group with the attacker, but simply because they have a higher tolerance for negativity. Further research that randomizes the partisan affiliation of attacker and target is necessary to disentangle the effects of ideology and group affiliation.

Third, our findings are based on one study and are highly contingent on the setting of the investigation. A replication of our findings, ideally varying the setting, issues, and actors involved, is necessary. The fact that our study focuses on U.S. voters, a particularly extreme outlier when it comes to dynamics of negative politics (and politics in general, really), suggests prudence when translating the trends discussed here to other political and cultural contexts. Similarly, the fact that the attacker in our study was a White male politician is unlikely to yield generalizable effects to candidates with different socio-demographic profiles. Given that negativity can have very different effects when uttered by female politicians (e.g., Herrnson & Lucas, 2006), future studies should manipulate the characteristics of the sender to better understand the influence of candidate characteristics on affective polarization.

In this sense, echoing the provisio too often missing from highly mediatized new findings in medical or neurological research (“. . . in mice;” Chakradhar, 2019), we ought to remind our readers that our findings suggest that negativity can affect affective polarization . . . in a single experiment with fictive White male politicians in the United States. This caveat notwithstanding, our results seem to add another item to the increasingly long list of potential detrimental effects of negative campaigning—and, as such, hopefully pave the way for further research on the topic and independent replications.4

References


Data and replication materials are available in the OSF repository: https://osf.io/89z5p/


Chakradhar, S. (2019, April 15). It’s just in mice! This scientist is calling out hype in science reporting. STAT. Retrieved from https://www.statnews.com/2019/04/15/in-mice-twitter-account-hype-science-reporting/


