"Be Nice or Leave Me Alone": An Intergroup Perspective on Affective Polarization in Online Political Discussions

Marchal, Nahema

DOI: https://doi.org/10.1177/00936502211042516

Article

"Be Nice or Leave Me Alone": An Intergroup Perspective on Affective Polarization in Online Political Discussions

Communication Research
1–23

DOI: 10.1177/00936502211042516
journals.sagepub.com/home/crx


Nahema Marchal[1]

Abstract

Affective polarization—growing animosity and hostility between political rivals—has become increasingly characteristic of Western politics. While this phenomenon is well-documented through surveys, few studies investigate whether and how it manifests in the digital context, and what mechanisms underpin it. Drawing on social identity and intergroup theories, this study employs computational methods to explore to what extent political discussions on Reddit's *r/politics* are affectively polarized, and what communicative factors shape these affective biases. Results show that interactions between ideologically opposed users were significantly more negative than like-minded ones. These interactions were also more likely to be cut short than sustained if one user referred negatively to the other's political in-group. Conversely, crosscutting interactions in which one of the users expressed positive sentiment toward the out-group were more likely to attract a positive than a negative response, thus mitigating intergroup affective bias. Implications for the study of online political communication dynamics are discussed.

Keywords

affective polarization, online political discussions, Reddit, intergroup dynamics

As digital spaces have grown as arenas of political communication, scholars have become increasingly concerned with the nature and diversity of online political talk. Digital communication has strong affective dimensions (Papacharissi, 2014; Stark, 2020) and there is a burgeoning literature exploring how emotions and sentiment on

[1]University of Oxford, UK

Corresponding Author:
Nahema Marchal, Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK.
Email: nahema.marchal@oii.ox.ac.uk

social media impact various political outcomes, from political misperceptions to engagement (Eberl et al., 2020; Papacharissi, 2014; Papacharissi & de Fatima Oliveira, 2012; Weeks & Garrett, 2019). Over the last two decades, the question of whether political discussions take place in ideological silos has also emerged as a major source of scholarly inquiry (for reviews of the literature see Barberá, 2020; Tucker et al., 2018). Individuals' tendency to associate with similar others and to seek out content that reinforces their pre-existing beliefs are well-established precepts of communication research (Knobloch-Westerwick et al., 2015). Evidence for them in the online context, however, is mixed. While some degree of ideological homophily is commonly observed in online social networks (Bright, 2018; Halberstam & Knight, 2016), a growing body of studies challenge the "echo chamber" hypothesis (Bruns, 2019; Zuiderveen Borgesius et al., 2016), arguing that Internet and social media are key drivers of exposure to oppositional viewpoints (Bakshy et al., 2015; Barberá et al., 2015; Barnidge, 2017; Dubois & Blank, 2018).

Despite the fact that individuals frequently come into contact with others they disagree with online, however, we still know relatively little about the *affective dimension* of these interactions. In recent years, public opinion scholars have noted the rise of "affective polarization" across the US and Europe: growing animosity and distrust between members of rival parties and ideological groups (Iyengar et al., 2019; Mason, 2018a; Reiljan, 2020; Wagner, 2021). Yet, thus far few studies have explored whether online political discussions are characterized by similar dynamics and what communicative factors might be shaping these patterns. This is especially pertinent, as social media continues to provide exposure to politically-charged content and communications with the potential to distort social attitudes toward others (Settle, 2018; Zhuravskaya et al., 2020). While users may engage in frequent crosscutting talk, the nature of these exchanges (especially if adversarial) might be more consequential for political trust and tolerance than not interacting at all.

Using data from Reddit's *r/politics*—one of the largest and most active political discussion forums online—in this paper, I leverage computational methods to show that ideological rivals are more affectively biased toward one another than they are toward like-minded peers. Results further indicate that interactions between opposed users are more likely to end than be followed up by another exchange if they refer negatively to the other's political in-group. Conversely, expressions of positive sentiment toward the out-group in crosscutting talk are more likely to drive further positive engagement, thus mitigating affective polarization. The contributions of this research are twofold. First, it makes an important methodological contribution by proposing a way to identify the ideological leanings of Reddit users based on their prior engagement with other political groups on the site. Second, these findings expand the current literature on affective polarization by showing that crosscutting online political discussions are characterized by affective biases—with users treating members of their political in-group more favorably than the out-group—and that these are significantly mediated by the direction and the sentiment of the messages exchanged between them.

The following sections are structured as follows. The next section provides a theoretical overview of affective polarization as a process driven by cognitive and affective

responses to other people's political group identity. The second section builds on social identity and intergroup relations theories to generate hypotheses about how intergroup factors might shape affective polarization in political discussions. The third section details the data collection, methodology, and operationalization of key variables and reports results. Findings are then discussed in light of their implications for future research on online polarization.

Polarization and Political Group Identity

Political identity is a powerful predictor of social and political behavior. Decades of research show that partisanship and ideological affiliation are key in determining our perceptions, evaluations, and treatment of others. Partisans are not only less inclined to trust and interact with political opponents, but also overwhelmingly discriminate against them in resource allocation, online labor markets (McConnell et al., 2018), and in their dating preferences (Huber & Malhotra, 2017). For social identity theorists, these biases are the direct corollary of partisanship's role as a social identity. According to this view, most people develop a sense of self through their membership in various social groups, such as their family or football club. This subjective sense of belonging generates powerful allegiances to the group ("us") that, in turn, shape our actions and social attitudes toward perceived out-group members ("them") in predictable ways (Brewer & Kramer, 1985). Individuals tend to preferentially engage with in-group members, for instance, and generally strive to protect their own group's status and standing, while developing negative perceptions and hostile feelings toward perceived rivals: a pattern described as "in-group favoritism" and "out-group hostility" (Brewer & Kramer, 1985). Even the most minimal or arbitrary form of group distinction, such as the color of one's uniform, has been shown to drive these effects (Frank & Gilovich, 1988), and scholars suggest that people can form identities around ideologies and opinion-based groups (Devine, 2015; Malka & Lelkes, 2010; Mason, 2018b) that are just as potent as partisanship in generating intergroup biases (McGarty et al., 2009). Following from this idea, in this paper I focus specifically on the "liberal" and "conservative" ideological identities as drivers of social behavior.

The core mechanism behind the effects of group identity on social behavior is identity salience (Mullen et al., 1992)—how significant an identity is to one's self-concept and to their perception of others—which itself, hinges on other factors such as how frequently one is reminded of their affiliation to the group (Gaertner et al., 1993). In situations of intense intergroup competition, individuals are acutely aware of their group membership and therefore primed to feel greater animosity and aversion toward members of the opposite team. There is mounting evidence that several features of online communication environments make political identities highly salient, with direct consequences for social behavior. According to social identity model of deindividuation effects (SIDE) scholars, the relative absence of personally identifiable information in anonymous digital settings compared to face-to-face settings, coupled with the abundance of partisan language and cues denoting political group membership, including "we" talk and other ways to discursively invoke one's in-group or distance

oneself from out-groups (Carr, 2017; Hinck & Carr, 2020; Scott, 2007), all make political labels more salient, giving them more weight during interactions (Lea et al., 2001). Confirming this pattern, SIDE research has shown that internet users operating under conditions of anonymity tend to form impressions of each other based on social group affiliations rather than individual differences (Li & Zhang, 2021; Postmes et al., 1998; Spears et al., 2002). Recent scholarship also shows that Facebook users readily infer the political affiliation of their interlocutors based on the content they post (Settle, 2018) and modify their online behavior accordingly. Salient political identities have been shown to determine people's word choices (Tamburrini et al., 2015); their propensity to be uncivil in online conversations (Gervais, 2015; Rains et al., 2017); their evaluations of users (Settle, 2018; Suhay et al., 2018), as well as their responses to moral suasion from perceived in-group members (Munger, 2021).

Affective Dynamics and Intergroup Communication

Emotional expression is a core feature of political discourse on social media (Stark, 2020; Tettegah, 2016). This centrality of emotion is well exemplified by the widespread use of emojis, and the ubiquity of emotional "reaction" features (e.g., like, upvote, or heart buttons) across platforms like Facebook, Twitter, Instagram, or Reddit. Research shows that affective information is easily transferred within online social networks (Ferrara & Yang, 2015, p. 20; Harris & Paradice, 2007; Kramer et al., 2014). Experimental evidence indeed suggests that individuals on the receiving end of a message can easily detect the emotional state of a sender through linguistic markers and emotional cues (Harris & Paradice, 2007). Despite the fact that most Internet and social media users regularly come across individuals they disagree with online (Barnidge, 2017), however, studies that investigate the affective dynamics of communication between politically different users are still scant. This is partly explained by the difficulty of inferring the party or ideological affiliations of large numbers of users (Barberá & Alvarez, 2015) and adequately linking sentiment to different types of interactions (Hillmann & Trier, 2012). There are a handful of notable exceptions, however. Comparing the incidence of out-group hostility in discussion of a political controversy in Israel across different platforms, Yarchi et al. (2021) found that interactions across political camps on Twitter and WhatsApp tend to be more negative than the ones taking place between supporters. Using network-based methods to analyze Twitter discourse on climate change, Tyagi et al. (2020) similarly found climate deniers to be more hostile toward climate believers than toward like-minded users and vice versa. Based on these prior findings, I therefore hypothesize that crosscutting interactions on Reddit will be more negative than like-minded ones:

H1: Interactions between ideologically opposed users will be more negative than like-minded ones.

Beyond efforts to document the phenomenon, little attention has been paid to the communicative factors that could reinforce or mitigate these affective biases between

users. Decades of psychological research have established that individuals respond differently to positive and negative emotional stimuli—a phenomenon termed "negativity bias" (Soroka, 2014)—with negative information typically generating stronger affective and behavioral responses than positive information (for a review, see Skowronski & Carlston, 1989; Soroka et al., 2019). In the digital context, however, studies that explore the role of emotions in shaping online conversation dynamics have thus far yielded conflicting results. Leveraging digital trace data, several large-scale studies of emotional contagion and reaction on social media appear to confirm the "negativity bias" hypothesis. Emotionally-charged posts—especially negative ones—have been shown to travel faster and elicit more feedback and more shares than neutral ones (Hornik et al., 2015; Stieglitz & Dang-Xuan, 2013). Contradicting these findings, however, a number of recent studies indicate that positive content actually attracts more attention from online users (Nave et al., 2018), generating more feedback and reciprocity (Stieglitz & Dang-Xuan, 2013) and spreading wider than negative content (Berger & Milkman, 2012; Ferrara & Yang, 2015; Kramer et al., 2014). Beyond message sentiment, communication scholars have also linked the presence of incivility in online political talk to negative emotions in those who encounter it (Gervais, 2015; Han & Brazeal, 2015), which can trigger copycat aggression (Masullo Chen & Lu, 2017). Incivility is still a contested concept, however, and although uncivil speech in the form of profanities or derogatory comments is mostly negatively-valenced, not all negative utterances amount to incivility (Chen, 2017). It is thus important to distinguish between both concepts. Moreover, several scholars have argued that incivility alone does not preclude meaningful dialogue across political lines and that forms of expressions such as hate speech carry far worse consequences for political tolerance (Bilewicz & Soral, 2020; Rossini, 2020).

Some of the conflicting patterns identified above, I would argue, can be explained by the fact that the role of intergroup dynamics in shaping behavioral responses to message sentiment is rarely considered in studies of social media communication. Yet, how a user experiences and responds to emotionally-charged messages depends on the social context within which they are shared (Mackie & Silver, 2004). Research shows, for example, that partisans are more likely to perceive their political rivals as more impolite than their peers during an exchange, even if the two engage in the same behaviors (Muddiman, 2017; Mutz, 2015). Studying how online political talk impacts users' behavior toward one another thus requires one to look beyond message sentiment in isolation, and examine instead how the context within which a message is shared and who it is aimed at impacts its reception. Here, work from intergroup emotion theory (IET) in particular offers some useful avenues.

First, it is important to note that communication need not be taking place between groups as units to be regarded as "intergroup"; rather, the term applies to situations in which "the transmission or reception of message is influenced by the group memberships of the individuals involved" (Harwood et al., 2005, p. 3). Scholars of intergroup emotions have shown that individuals can experience emotions (such as fear, anger, and pride) on behalf of their in-group (Iyer & Leach, 2008) and so even in the absence of other group members (Hogg et al., 2004). These are triggered by events or

appraisals that can be interpreted as harming the individual's in-group. Exposure to criticism directed at one's in-group, for example, tends to be interpreted as a personal attack—especially among strong identifiers (Mackie et al., 2008). This, in turn, can provoke feelings of anger, humiliation and hate, and motivate a desire for retribution by attacking "the other side" (Lickel, 2012). Extending these findings to the context of crosscutting online discussions between users from opposing ideological inclinations, one would therefore expect that negative sentiment aimed at a users' in-group would spur others to retaliate. This leads to the following hypothesis:

H2a: In crosscutting interactions, a negative reply that mentions the sender's in-group is more likely to be followed by a negative response than a positive one.

Other studies in the literature point in a slightly different direction. Another branch of social psychology concerned with the potential for communication to stoke or mitigate intergroup rivalries is intergroup contact theory (Hewstone et al., 2014). At odds with the first proposition, work in this area has found that negative encounters with members of an out-group can increase individuals' perceptions that the group is threatening, therefore reinforcing intergroup anger (Hayward et al., 2017). In these situations, negative contact may actually hamper intentions to engage in contact with said out-group in the future (Barlow et al., 2012)—a phenomenon called the "avoidance generalization effect" (Meleady & Forder, 2019). These contrasting findings lead me to formulate the following competing hypothesis:

H2b: In crosscutting interactions, a negative reply that mentions the sender's in-group is more likely to end the discussion than be followed up by another crosscutting response of any type.

While individuals are prone to come across members of political communities they are opposed to online, conflict scholars often underline that the mere existence of incompatibility "does not always result in a confrontation" (Knapp & Daly, 2011, p. 480). To the contrary, it is reasonable to conceive that positive intergroup sentiment in exchanges between dissimilar users could have positive effects on future contact between those groups, thus decreasing affective bias. The "intergroup contact hypothesis" (Allport, 1954) suggests that, under certain conditions, positive and meaningful interactions between members of different social groups will have positive effects on reducing intergroup prejudice (Hewstone et al., 2014). Online communication has expanded possibilities for contact and scholars suggest that the specific affordances of computer-mediated communication—disembodied, text-based exchanges coupled with a lack of interpersonal cues—make digital discussion environments particularly fitting venues for effective intergroup interactions (Amichai-Hamburger, 2008; Kim & Wojcieszak, 2018). Positive direct online contact with out-group members through online comments improves user attitudes toward them and predicted intentions to interact again in the future (Kim & Wojcieszak, 2018). In the context of this study, I

therefore expect that positive references of the out-group will be followed by a positive response.

H3: In crosscutting interactions, a positive reply that mentions the sender's in-group is more likely to be followed by a positive response than a negative one.

Affective polarization is not limited to derogation of and negative affect toward the political outgroup, however. In fact, scholars of intergroup relations have long argued that intergroup conflict is largely driven by "in-group favoritism" rather than "out-group hate" (Halevy et al., 2012). Put differently, when given the choice, people are generally more motivated to support their in-group (e.g., rewarding them in resource allocation) and avoid the company of out-groupers (Iyengar et al., 2019; Mason, 2015) than to directly harm them (Amira et al., 2019; Brewer, 1999; Iyengar & Krupenkin, 2018). Several factors could be reinforcing this tendency toward in-group favoritism. Research from the IET literature demonstrates that, in an intragroup setting, expressing positive feelings such as pride toward the in-group has positive functional effects on strength of identification, cooperation, and motivation for future action, regardless of whether the group is a long standing or fleeting one (Kessler & Hollbach, 2005; Knight & Eisenkraft, 2015). Coupled with findings suggesting that positive discussions among like-minded peers strengthens group identity (Yardi & Boyd, 2010) and encourage continued participation (Joyce & Kraut, 2006), one would therefore expect that positive references to the in-group in a like-minded conversation would drive further contact between group members.

H4: In like-minded interactions, a positive reply that mentions both users' in-group is more likely to be followed by a positive response than a negative one.

Data and Methods

Testing the hypotheses outlined above requires first a careful examination of users' ideological affinities. Moreover, it requires identifying the emotional valence and content of their exchanges, in particular whether these are more positive or negative, and whether they mention specific political entities. To this end, textual and behavioral digital traces can provide valuable insights, provided they are considered as bounded by the social practices and affordances of the medium through which they arose (Jungherr et al., 2017). In this study, I turn to such traces of commenting activity to map out instances of crosscutting and like-minded interactions in my dataset, and extract the sentiment of messages exchanged to determine the nature of these interactions.

Data for this study were collected from Reddit's *r/politics* subreddit between October 15th, 2018 and November 5th, 2018, the day prior to the 2018 US midterm elections, using Jason Baumgartner's Pushshift.io API (Baumgartner et al., 2020). With over 5 million subscribers, *r/politics* is one of the largest and most active political forums on Reddit, itself the third most visited social networking site in the US (Statista,

2021). After filtering out comments that had either been removed, posted by moderator bots or by authors who had deleted their accounts since the data collection, 1.2 million valid comments from 141,171 unique authors nested in 20,623 discussion threads remained in the dataset. Since the unit of interest for this analysis are interactions (a comment followed by at least one response) between pairs of users, only discussion threads that contained more than one interaction were included in the final analysis (12,447 in total). Following all transformations, the final dataset contained a total of 176,339 comment-response dyads.

Estimating Users' Ideological Leaning

Research shows that people tend to engage in online communities that reflect their latent ideological preferences (Barberá & Alvarez, 2015; Shi et al., 2019). Here, I follow Shi et al. (2019) in predicting the ideological position of users in my sample based on their total number of contributions to liberal-leaning and conservative-leaning subreddits outside *r/politics* over a 6-month period ranging from June 2018 to November 2018. For this task, a publicly available list of 200 politically-relevant subreddits, known as the Reddit Politosphere, was obtained from the *r/politics* wiki in August 2019, refined and augmented by searching for "politics" in Reddit search bar and collecting the URLs for all suggested subreddits not already listed, culminating in a final list of 258 subreddits. A team of two coders then proceeded to classify each subreddit by ideology.[1] Each subreddit's titles (e.g., "Progressives"), as well as its description, guidelines, and the content posted in it, were all taken as potential indicators of its users' ideological orientation. The full codebook and confusion matrix for this task are all available in the Supplemental Appendix.

Having collected information about users' commenting activity in liberal-leaning and conservative-leaning subreddits, I proceeded to estimate their ideological alignment using a conservative Bayesian estimation framework (Shi et al., 2019). A "neutral" beta prior Beta(1/3, 1/3) was selected (Kerman, 2011), which assumes that each author is as likely to contribute to a liberal as they are to contribute to a conservative subreddit in order to prevent "lurkers" and occasional commenters from introducing too much uncertainty in the overall estimation of Redditors' ideological alignment.[2] If p is assumed to have a prior distribution $p \sim Beta(a,b)$ before any observations, then the distribution is updated using Bayes' law with observations X (number of comments contributed to liberal subreddits) and Y (comments contributed to both liberal and conservative subreddits), such that: $p \mid X \sim Beta(a+X, b+Y-X)$. Finally, political alignment is defined as the posterior mean of p:

$$E[p|X] = \frac{(X+a)}{(Y+a+b)}$$

Following this method, I was able to infer the ideological leaning of 35,794 authors. To validate these estimations, a random sample of 100 comments posted by authors who had been classified as either "liberal-leaning" or "conservative-leaning" through this method was manually coded following the same coding scheme used to classify subreddits. Comparing the classification generated by the Bayesian framework with

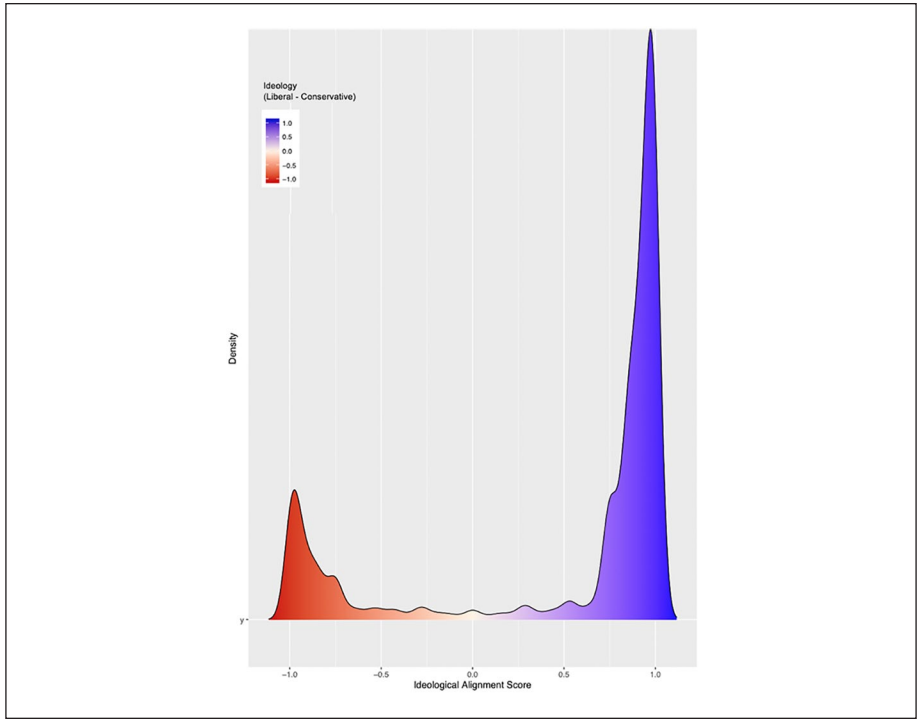

Figure 1. Ideological alignment of Reddit users.
Note. The distribution of ideological alignment of Reddit users, ranging from −1 (most conservative, where users have contributed exclusively and significantly to conservative-leaning subreddits) to +1 (most liberal, where users have contributed exclusively and significantly to liberal-leaning subreddits). $N = 35{,}794$. *Y*-axis represents the proportion of authors who were assigned a specific ideological alignment score.

that obtained through manual coding resulted in a 0.78 accuracy score (95% CI [0.68, 0.86]). Figure 1 below shows the distribution of ideological alignment for *r/politics* users in my sample.

One may note that the distribution is quite skewed toward liberals (78%), with the peaks at both extrema corresponding to users who have made the highest number of contributions to liberal or conservative subreddits. As shown in Table 1, the vast majority of interactions (69%) over the data collection period appeared to be taking place between like-minded users, especially liberal-leaning ones, while crosscutting interactions were more marginal (31% of total).

Sentiment Scores

Having predicted the ideological leanings of users in my sample, I move on to extracting sentiment scores for each comment in the dataset. Sentiment scores are compiled using NLTK's Vader Sentiment Analyzer—a human-validated sentiment analysis

Table 1. Distribution of Interaction Types.

Type of interaction	*N*	%
Likeminded	121,939	69
Conservative-Conservative	6,463	
Liberal-Liberal	115,476	
Crosscutting	54,400	31
Total	176,339	100

package specifically attuned for sentiment expression on social media and designed to handle emoticons, emojis, as well as emotion-laden internet slang and punctuation (Hutto & Gilbert, 2014). VADER is a rule-based sentiment model that has both a dictionary and associated intensity measures, and returns a compound sentiment score in the [−1, +1] range, where −1 is the most negative sentiment and +1 is the most positive sentiment ($M = -0.07$, $SD = 0.53$). To test the accuracy of VADER's estimations, a random sample of 300 comments from the dataset was manually classified as positive, negative, or neutral, resulting in an overall 0.70 accuracy score (see Supplemental Appendix for complete performance metrics).

Only comments that scored either highly negative (with compound scores in the 25th% percentile) or highly positive (with scores in the 75th% percentile) were classified as respectively "negative" or "positive," while all others were classified as "neutral." While this technique is robust to predict the valence of comments, it should be acknowledged that sentiment analysis is limited in its ability to identify and deal with instances of sarcastic or non-literal comments (Muresan et al., 2016).

Measures

Dependent variables. Affective polarization is operationalized as the tendency for users to engage positively with like-minded peers and negatively with ideological opponents. This is consistent with other studies in this line of work and theories of intergroup behavior that see intergroup prejudice as characterized by both in-group favoritism and out-group hostility (Brewer, 1999). The primary task, therefore, is to determine the affective nature of interactions between crosscutting and like-minded users. Averaging the sentiment scores of both original and response comments in each dyad indicates that like-minded interactions are slightly less negative ($n = 121{,}939$, $M = -0.057$, $SD = 0.39$) than crosscutting ones ($n = 54{,}400$, $M = -0.06$, $SD = 0.39$).

This study also seeks to explore what communicative factors play a role in driving or mitigating affective bias. Here, I am particularly interested in how the tone and direction of the last reply in a conversation dyad impacts future interactions. I distinguish between seven follow-up interaction types, depending on (1) the ideological affinity between past and future commenter and (2) whether a future comment is negative, positive, or neutral, as well as the possibility that an interaction is not followed at all. A transition to a different interaction is recorded when an initial comment-response

dyad elicits another reply, forming another dyad between either the same or a different pair of users, and so on until the exchange ends.

Predictors. To identify the presence of in-group or out-group political entities in the comment corpus, I used spaCy's named entity recognition feature with a custom dictionary of common nouns and abbreviations closely associated with the words "liberal" and "conservative," such as "libs," "dems," and "cons" (see Supplemental Appendix for complete list). Users may, of course, refer to liberal or conservative ideologies in comments without explicitly mentioning these groups by name (e.g., through discussion of specific policies or political leaders). Given that the presence of explicit social group cues accentuates identity salience, however, in the analysis I opt to focus primarily on these identifiers. When a comment references multiple political entities, it is impossible to properly determine the direction in which the sentiment of the comment is expressed—the presence of a liberal ($n = 12{,}582$) or conservative entity ($n = 14{,}678$) was thus only recorded when a reply strictly mentioned one or the other. Following from this, I measure the presence of an in-group or out-group political entity in a response by matching the ideology of the sender and the political entity mentioned. The tone of the last reply in a conversation dyad, finally, is determined based on the sentiment score previously allocated to each comment (i.e., negative or positive).

Controls. Several social factors might influence the tone and nature of users' responses. Here, I first control for a response's *popularity score*, or the number of upvotes minus the number of downvotes a comment has received at the time of data collection, as highest scoring comments have a higher chance of ranking at the top of a discussion thread. Finally, given that the presence of other in-group and out-group members can influence individual behavior in intergroup settings (Coe et al., 2014), I also control for the *proportion of liberal or conservative-leaning users* in a discussion thread.

Models

There are several sources of non-independence in the data: interactions are nested in pairs of users, in that two users can have more than one interaction in the same discussion. Pairs of users are then also separately nested in individual discussions, which form part of larger threads. To determine what proportion of the variance in the sentiment of user interactions can be explained by these levels, I first calculate intra class correlations. Nesting within thread only accounted for 14% of the variance in message sentiment (ICC = 0.14) while nesting within discussion and pair of users explained between 41.7% and 55.1% of the variance in the outcome variable (ICC = 0.42 and 0.55, respectively). To account for these dependencies in the data, I thus adopt a series of multilevel models to address my hypotheses.

The first part of the analysis addresses the question of whether users of different ideological leanings display affective biases in their interactions. To answer this,

Table 2. Effect of Interaction Type on Average Interaction Sentiment.

	Average sentiment of interaction		
	Model 1		
Random effects	Variance	*SD*	
User pair	0.47	0.68	
Discussion	0.01	0.12	
Thread	0.09	0.29	
Residual	0.44	0.66	
Fixed effects	Estimate	*SE*	*t*
Intercept	0.05***	0.007	6.56
Crosscutting (Yes = 1)	−0.02**	0.007	−3.15
Marginal R^2		0.001	
Conditional R^2		0.56	

Note. $N = 176{,}339$.
*$p < .05$. **$p < .01$. ***$p < .001$.

Model 1 examines the effect of ideological affinity between two users on the sentiment of their interactions, including the identification of the user pair, the discussion in which the interaction took place and the thread as random effects. Estimates of statistical significance were computed using the Satterthwaite's degrees of freedom method (Kuznetsova et al., 2017). To calculate marginal and condition R^2, I follow the methods proposed by Nakagawa and Schielzeth (2013).

The second part of the analysis is concerned with how the tone and target of the last comment in a conversation dyad impact follow-up interactions, including the possibility that a conversation ends. To explore the mechanisms outlined in Hypotheses 2a to 4, I run a series of separate multilevel logit models, which allows me to specify the model equation for different binomial contrasts of interest in the outcome variable. Model 2a and Model 3 evaluate the odds of future negative engagement between ideologically opposed users compared to any other types of interactions. Model 2b contrasts the likelihood that a crosscutting interaction ends or is followed by another crosscutting interaction of any type. Model 4, finally, is concerned with the probability of positive future interactions between like-minded users compared to any other types of interactions. There are some limitations associated with this model choice. Notably, individualized models can lead to efficiency loss (Agresti, 2002, p.274). However, in each case this is mitigated by a large sample size in the reference category.

Results

Hypothesis 1 predicted that interactions between crosscutting users would be more negative than those between like-minded users. Results from this model (Model 1) are summarized in Table 2. Model 1 clearly shows that ideologically opposed users tend to engage in significantly more negative interactions than like-minded ones, though the effect size is modest ($\beta = -0.02$, $p < .01$). Crosscutting interactions have, on

average, a 0.02 lower sentiment score than like-minded ones. Hypothesis 1 is thus supported.

Before addressing the second part of the analysis, I ran an initial model with the same sets of random effects to assess the baseline likelihood of transitioning from crosscutting to like-minded interactions (see Model 0 in Supplemental Appendix). This indicates that users rarely transition between these two states: interactions between opposed users are significantly more likely to be followed by another crosscutting exchange than a like-minded one (OR = 61.05, $p < .001$).

Moving on to the hypotheses of interest, Hypotheses 2a and 2b suggested that negative crosscutting interactions that referenced the sender's in-group (e.g., if a liberal user made a negative comment about conservatives as a response to something a conservative user wrote) would be more likely to prompt another negative crosscutting exchange (H2a) or to halt discussion altogether (H2b). Model 2a and 2b in Table 3 address this set of hypotheses.

The results in Table 3 show that negative interactions between opposed users are slightly less likely to be followed up by a negative crosscutting response than a positive one (OR = 0.12, $p < .001$). Examining the interaction term, negative exchanges in which a respondent negatively references another user's political in-group are not significantly associated with a higher chance of negative response, compared to all other types of replies. Hypothesis 2a is therefore not supported. Interestingly, in this scenario, large proportions of political in-groupers in the thread also significantly drive the odds of a negative crosscutting response, and so to a high degree (OR = 40.43, $p < .001$), while the negative effect of the presence of out-group members is small (OR = 0.83, $p < .001$). In the same type of interactions, a high score comment is also slightly more conducive to a negative follow-up (OR = 1.08, $p < .001$).

Turning now to Model 2b, there is clear evidence that the mention of a user's political identity moderates the relationship between previous and future crosscutting interactions. Negative crosscutting interactions in which the reply negatively references the sender's political in-group are indeed significantly more likely to end than to be followed up by another crosscutting exchange of any kind (OR = 1.37, $p < .001$), compared to all other types of replies. Hypothesis 2b is thus confirmed. Among control variables, the effect of high proportions of in-group and out-group members in the thread are both significant, though their effect is small (OR = 0.08 and 1.12 respectively).

The third hypothesis dealt with the possibility that positive intergroup contact would enhance the likelihood of further positive engagement between ideologically opposed users. As is evident from Model 3 in Table 3, interactions between different users in which the respondent positively references the sender's in-group in a comment (e.g., a conservative expressing positive sentiment toward liberals in an exchange with a liberal user) are significantly more likely to be followed up by another positive exchange than a negative one (OR = 2.06, $p < .001$) compared to all other types of comments. This confirms Hypothesis 3. Among control variables, the presence of fellow in-group members is also a significant driver of crosscutting positive engagement (OR = 53.55, $p < .001$). In contrast, high proportions of out-group members in the thread only have a small impact on the outcome variable (OR = 1.07, $p < .05$).

Table 3. Effect of Interaction Features on Follow-Up Response.

	Crosscutting negative	End	Crosscutting positive	Like-minded positive
	Model 2a	Model 2b	Model 3	Model 4
Fixed effects	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]
Intercept	0.27*** [0.25, 0.28]	0.70*** [0.67, 0.73]	0.05*** [0.05, 0.06]	0.05*** [0.05, 0.06]
Crosscutting negative (1 = Yes)	0.12*** [0.11, 0.12]	0.92*** [0.89, 0.96]		
Crosscutting positive (1 = Yes)			0.19*** [0.15, 0.23]	
Like-minded positive (1 = Yes)				0.23*** [0.20, 0.27]
Refers sender's ingroup (1 = Yes)	1.16** [1.06, 1.27]	0.69*** [0.63, 0.74]	0.99 [0.86, 1.13]	1.01 [0.90, 1.13]
Comment score	1.08*** [1.05, 1.11]	0.13*** [0.12, 0.15]	0.97 [0.92, 1.02]	0.99*** [0.96, 1.03]
Proportion of ingroup members in thread	40.43*** [27.10, 60.33]	0.08*** [0.07, 0.09]	53.55*** [27.71, 103.50]	1.06** [1.02, 1.09]
Proportion of outgroup members in thread	0.83*** [0.79, 0.86]	1.12*** [1.08, 1.15]	1.07* [1.01, 1.13]	8.24*** [6.07, 11.19]
Crosscutting negative × refers sender's ingroup	0.93 [0.70, 1.23]	1.37*** [1.18, 1.59]		
Crosscutting positive × refers sender's ingroup			2.06* [1.13, 3.75]	
Likeminded positive × refers sender's ingroup				1.29 [0.75, 2.21]
Marginal R^2	0.16	0.25	0.11	0.12
Conditional R^2	0.95	0.83	0.96	0.72
AIC	70,323.7	100,758.0	34,119.8	61,176.4

Note. $N = 176{,}339$ for Model 2a, 3, and 4 and $N = 98{,}886$ for Model 2b. All models summarize fixed effects. OR = odd ratio; CI = confidence interval.
*$p < .05$. **$p < .01$. ***$p < .001$.

Having reviewed the drivers of intergroup affective bias, I now turn to intragroup dynamics. Hypothesis 4 made predictions about the effect of positive references to the in-group in like-minded exchanges. Model 4 in Table 3 shows that, controlling for other factors, positive interactions between similar users are actually less likely to be followed up by a positive like-minded response (OR = 0.23, $p < .001$) than other interaction types. Furthermore, references to the political in-group do not moderate that relationship. Hypothesis 4 is therefore not supported.

Discussion

The present research set out to determine whether discussions between liberal and conservative users on Reddit's *r/politics* were affectively polarized, and what communicative factors shaped these affective biases—the first study to do so. Drawing on theories of intergroup relations and leveraging computational methods to infer the political leanings of forum members, I find that interactions between ideologically-opposed users are indeed polarized along affective lines. In other words, *r/politics* users express more negative sentiment when interacting with people they disagree with than with like-minded peers. This finding joins a nascent body of studies that have identified similar patterns on other platforms (Yarchi et al., 2021), suggesting the need to expand current definitions of online polarization to include affective dimensions.

The analyses presented here also indicate that the contents of these conversations and the group context within which they take place are significant determinants of affective polarization. Confirming initial expectations about "generalized avoidance," negativity aimed at the opposition in conversation between two opposed users had the effect of discouraging further interaction of the same kind. This held true even after controlling for comment popularity and proportion of in-group members in the conversation, underscoring that, even in the face of threats, retaliation is often a less preferable option than opting out from a chat. This outcome may also reflect the fact that individuals are quite adverse to uncivil and excessively partisan behavior (Klar et al., 2018; Shafranek, 2020) and generally prefer cooperating with their political in-group than engaging in overt intergroup conflict (Halevy et al., 2012).

Conversely, I find that positive crosscutting engagement have a *depolarizing* effect: when users made positive references to the opposition, they were more likely to engage in further positive crosscutting interactions. This finding enriches current research on the effects of online intergroup contact by showing that positive encounters with ideologically different "others" not only improve individuals' feelings toward that outgroup (Kim & Wojcieszak, 2018), but also their disposition toward them. In line with findings from Wojcieszak and Warner (2020), Huddy and Yair (2021), and Skytte (2020), it also provides support for warm group relations and positive contact as effective strategies for reducing partisan prejudice in online discussions.

Finally, and contrary to initial expectations, in-group praise did not have a significant impact on affective polarization among like-minded users. One possible explanation for this has to do with the fact that group norms tend to vary with situational

factors: the extent to which groups will engage in discriminatory behavior, for example, depends heavily on group members' perception of threat from rivals (Hogg et al., 1984). In situations when the power differential between dominant and minority groups is strong, as is the case on *r/politics*—a predominantly liberal enclave—the need for positive distinctiveness is reduced (Rains et al., 2017). In-group praise might have therefore not carried the same normative weight in this particular context as it would in other situations.

This study presents several limitations that must be acknowledged. However, each presents clear opportunities for future research. Given the disproportionately high number of liberal users in my sample, the opportunity for crosscutting interactions in these discussions was inherently restricted. Moreover, as I highlight above, the power imbalance between liberals and conservative users may have affected the way in which the two groups interacted with one another. It is entirely possible, for example, that the minority of conservative users engaged on the forum would behave differently in a space where they had a dominant presence. It will therefore be important to follow up this study with analyses of other subreddits that are either more ideologically segregated or where the power dynamic is reversed to see if they reveal similar patterns.

Second, the present analysis only focused on a restricted number of communicative and group factors, namely the presence of other group members, tone, and salience of political identities in exchanges between users. The strategy proposed for the coding of political entities, in particular, might have excluded comments that involved clear intergroup dynamics without making explicit references to "liberals" or "conservatives."[3] Future studies could thus usefully expand on these findings by analyzing the effects of more granular communicative features and constructs on affective polarization, such as discussion of policy issues or the presence of incivility, name-calling or dismissal.

A third limitation is the potential loss of relevant data through community moderation. Though Reddit is characterized by a strong free speech ethos, in practice subreddit moderators are empowered to remove or delete messages that violate their community guidelines if they contain personal attacks, hate speech, or flaming. While only a small proportion of comments had been deleted by moderator bots at the time of data collection, some of these omissions might have constituted the most extreme displays of negative sentiment on the forum.

Finally, it will be important for future studies to test whether the findings presented here hold with alternative sentiment analysis techniques, including entity-level sentiment analysis and supervised approaches. Previous work in this area has shown group identification to take place primarily through textual cues and language use. However, many social media platforms are also rich in visual cues that denote one's political affiliation and views. Beyond visually anonymous environments, it could thus be interesting to explore how political cues contained in users' real names (including emojis), self-descriptions, and even profile pictures affect the dynamics outlined above.

Acknowledgments

The author would like to thank Eliott Ash, Jonathan Bright, Philip Howard, Yptach Lelkes, Victoria Nash, Bertie Vidgen, Taha Yasseri, and participants of the 2019 POLTEXT conference for their valuable comments and engagement with earlier versions of this manuscript.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author gratefully acknowledges support from the Economic and Social Research Council (ESRC) and the British Federation of Women Graduates (BFWG) for this research.

ORCID iD


Nahema Marchal https://orcid.org/0000-0002-8518-3840

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Ideological orientation is, of course, not unidimensional and may be denoted in a number of ways, from stated social values and positions on policy issues to one's attitudes toward political foes (Hanson et al., 2019). Linguistic choices in political conversations are also regarded as a reliable indicator of an individual's ideological leaning: what liberals refer to an "estate tax," for example, is typically called the "death tax" by a conservative, with no ideologically neutral alternative (Lakoff, 2002). With this in mind, a holistic approach was taken to the classification scheme.
2. Authors were included if they had at least contributed one comment to either type of subreddit.
3. As one reviewer pointed out, only about 27,000 comments in the sample mentioned either a liberal or a conservative entity (and not both at the same time), meaning that the great majority of comments were not given an in-group or out-group label.

References

Agresti, A. (2002). *Categorical data analysis*. John Wiley & Sons, Ltd.

Allport, G. (1954). *The nature of prejudice*. Addison-Wesley.

Amichai-Hamburger, Y. (2008). The contact hypothesis reconsidered: Interacting via internet: Theoretical and practical aspects. In A. Barak (Ed.), *Psychological aspects of cyberspace: Theory, research, applications* (pp. 209–227). Cambridge University Press.

Amira, K., Wright, J. C., & Goya-Tocchetto, D. (2019). In-group love versus out-group hate: Which is more important to partisans and when? *Political Behavior*, *43*, 473–494. https://doi.org/10.1007/s11109-019-09557-6

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Political science. Exposure to ideologically diverse news and opinion on facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Barberá, P. (2020). Social media, echo chambers, and political polarization. In N. Persily & J. A. Tucker (Eds.), *Social media and democracy: The state of the field, prospects for reform* (pp. 34–56). Cambridge University Press.

Barberá, P., & Alvarez, R. M. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, *23*(1), 76–91. https://doi.org/10.1093/pan/mpu011

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*(10), 1531–1542. https://doi.org/10.1177/0956797615594620

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R., Harwood, J., Rubin, M., & Sibley, C. G. (2012). The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, *38*(12), 1629–1643. https://doi.org/10.1177/0146167212457953

Barnidge, M. (2017). Exposure to political disagreement in social media versus face-to-face and anonymous online settings. *Political Communication*, *34*(2), 302–321. https://doi.org/10.1080/10584609.2016.1235639

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). *The Pushshift Reddit dataset*. Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020), Atlanta, Georgia, USA, *14*(1), 830–839.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *JMR, Journal of Marketing Research*, *49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, *41*(S1), 3–33. https://doi.org/10.1111/pops.12670

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brewer, M. B., & Kramer, R. M. (1985). The psychology of intergroup attitudes and behavior. *Annual Review of Psychology*, *36*(1), 219–243. https://doi.org/10.1146/annurev.ps.36.020185.001251

Bright, J. (2018). Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. *Journal of Computer-Mediated Communication*, *23*(1), 17–33. https://doi.org/10.1093/jcmc/zmx002

Bruns, A. (2019). *Are filter bubbles real?* Polity.

Carr, C. T. (2017). Social media and intergroup communication. In J. Nussbaum (Ed.) *Oxford research encyclopedia of communication*. Oxford University Press.

Chen, G. M. (2017). *Online incivility and public debate*. Palgrave Macmillan.

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10.1111/jcom.12104

Devine, C. J. (2015). Ideological social identity: Psychological attachment to ideological ingroups as a political phenomenon and a behavioral influence. *Political Behavior*, *37*(3), 509–535. https://doi.org/10.1007/s11109-014-9280-6

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information Communication & Society*, *21*(5), 729–745. https://doi.org/10.1080/1369118x.2018.1428656

Eberl, J.-M., Tolochko, P., Jost, P., Heidenreich, T., & Boomgaarden, H. G. (2020). What's in a post? How sentiment and issue salience affect users' emotional reactions on facebook. *Journal of Information Technology & Politics*, *17*(1), 48–65. https://doi.org/10.1080/19331681.2019.1710318

Ferrara, E., & Yang, Z. (2015). Measuring emotional contagion in social media. *PLoS One*, *10*(11), e0142390. https://doi.org/10.1371/journal.pone.0142390

Frank, M. G., & Gilovich, T. (1988). The dark side of self- and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology*, *54*(1), 74–85. https://doi.org/10.1037/0022-3514.54.1.74

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common in-group identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology*, *4*(1), 1–26. https://doi.org/10.1080/14792779343000004

Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, *12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416

Halberstam, Y., & Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. *Journal of Public Economics*, *143*(1), 73–88. https://doi.org/10.1016/j.jpubeco.2016.08.011

Halevy, N., Weisel, O., & Bornstein, G. (2012). "In-group love" and "out-group hate" in repeated interaction between groups. *Journal of Behavioral Decision Making*, *25*(2), 188–195. https://doi.org/10.1002/bdm.726

Han, S. H., & Brazeal, L. M. (2015). Playing nice: Modeling civility in online political discussions. *Communication Research Reports*, *32*(1), 20–28. https://doi.org/10.1080/08824096.2014.989971

Hanson, K., O'Dwyer, E., & Lyons, E. (2019). The individual and the nation: A qualitative analysis of US liberal and conservative identity content. *Journal of Social and Political Psychology*, *7*(1), 378–401. https://doi.org/10.5964/jspp.v7i1.1062

Harris, R. B., & Paradice, D. (2007). An investigation of the computer-mediated communication of emotions. *Research Journal of Applied Sciences*, *3*(12), 2081–2090.

Harwood, J., Giles, H., & Palomares, N. (2005). Intergroup theory and communication processes. In H. Giles & J. Harwood (Eds.), *Language as social action. Intergroup communication: Multiple perspectives* (pp. 1–17). Peter Lang Publishing.

Hayward, L. E., Tropp, L. R., Hornsey, M. J., & Barlow, F. K. (2017). Toward a comprehensive understanding of intergroup contact. *Personality and Social Psychology Bulletin*, *43*(3), 347–364. https://doi.org/10.1177/0146167216685291

Hewstone, M., Lolliot, S., Swart, H., Myers, E., Voci, A., Al Ramiah, A., & Cairns, E. (2014). Intergroup contact and intergroup conflict. *Peace and Conflict: Journal of Peace Psychology*, *20*(1), 39–53. https://doi.org/10.1037/a0035582

Hillmann, R., & Trier, M. (2012). Sentiment polarization and balance among users in online social networks. *AMCIS 2012 Proceedings*, *10*, 3228–3237.

Hinck, A. S., & Carr, C. T. (2020). Advancing a dual-process model to explain interpersonal versus intergroup communication in social media. *Communication Theory*, qtaa012. https://academic.oup.com/ct/advance-article-abstract/doi/10.1093/ct/qtaa012/5898399 redirectedFrom=fulltext

Hogg, M. A., Abrams, D., Otten, S., & Hinkle, S. (2004). The social identity perspective. *Small Group Research*, *35*(3), 246–276. https://doi.org/10.1177/1046496404263424

Hogg, M. A., Turner, P. J., & Smith, P. M. (1984). Failure and defeat as determinants of group cohesiveness. *British Journal of Social Psychology*, *23*(2), 97–111. https://doi.org/10.1111/j.2044-8309.1984.tb00619.x

Hornik, J., Shaanan Satchi, R., Cesareo, L., & Pastore, A. (2015). Information dissemination via electronic word-of-mouth: Good news travels fast, bad news travels faster! *Computers in Human Behavior*, *45*, 273–280. https://doi.org/10.1016/j.chb.2014.11.008

Huber, G. A., & Malhotra, N. (2017). Political homophily in social relationships: Evidence from online dating behavior. *The Journal of Politics*, *79*(1), 269–283. https://doi.org/10.1086/687533

Huddy, L., & Yair, O. (2021). Reducing affective polarization: Warm group relations or policy compromise? *Political Psychology*, *42*(2), 291–309. https://doi.org/10.1111/pops.12699

Hutto, C. J., & Gilbert, E. E. (2014). *VADER: a parsimonious rule-based model for sentiment analysis of social media text*. Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor, Michigan, USA.

Iyengar, S., & Krupenkin, M. (2018). The strengthening of partisan affect. *Political Psychology*, *39*(S1), 201–218. https://doi.org/10.1111/pops.12487

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Iyer, A., & Leach, C. W. (2008). Emotion in inter-group relations. *European Review of Social Psychology*, *19*(1), 86–125. https://doi.org/10.1080/10463280802079738

Joyce, E., & Kraut, R. E. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, *11*(3), 723–747. https://doi.org/10.1111/j.1083-6101.2006.00033.x

Jungherr, A., Schoen, H., Posegga, O., & Jürgens, P. (2017). Digital trace data in the study of public opinion: An indicator of attention toward politics rather than political support. *Social Science Computer Review*, *35*(3), 336–356. https://doi.org/10.1177/0894439316631043

Kerman, J. (2011). Neutral noninformative and informative conjugate beta and gamma prior distributions. *Electronic Journal of Statistics*, *5*, 1450–1470. https://doi.org/10.1214/11-ejs648

Kessler, T., & Hollbach, S. (2005). Group-based emotions as determinants of in-group identification. *Journal of Experimental Social Psychology*, *41*(6), 677–685. https://doi.org/10.1016/j.jesp.2005.01.001

Kim, N., & Wojcieszak, M. (2018). Intergroup contact through online comments: Effects of direct and extended contact on outgroup attitudes. *Computers in Human Behavior*, *81*, 63–72. https://doi.org/10.1016/j.chb.2017.11.013

Klar, S., Krupnikov, Y., & Ryan, J. B. (2018). Affective polarization or partisan disdain? Untangling a dislike for the opposing party from a dislike of partisanship. *Public Opinion Quarterly*, *82*(2), 379–390. https://doi.org/10.1093/poq/nfy014

Knapp, M. L., & Daly, J. A. (2011). *The SAGE handbook of interpersonal communication*. SAGE Publications.

Knight, A. P., & Eisenkraft, N. (2015). Positive is usually good, negative is not always bad: The effects of group affect on social integration and task performance. *Journal of Applied Psychology*, *100*(4), 1214–1227. https://doi.org/10.1037/apl0000006

Knobloch-Westerwick, S., Westerwick, A., & Johnson, B. K. (2015). Selective exposure in the communication technology context. In S. Shyam Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 405–424). Wiley.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lakoff, G. (2002). Moral politics. *University of Chicago Press*. https://doi.org/10.7208/chicago/9780226471006.001.0001

Lea, M., Spears, R., & de Groot, D. (2001). Knowing me, knowing you: Anonymity effects on social identity processes within groups. *Personality and Social Psychology Bulletin*, *27*(5), 526–537. https://doi.org/10.1177/0146167201275002

Lickel, B. (2012). Retribution and revenge. In L. Troop (Ed.), *The Oxford Handbook of intergroup conflict* (pp. 89–105). Oxford University Press.

Li, S., & Zhang, G. (2021). Intergroup communication in online forums: The effect of group identification on online support provision. *Communication Research*, *48*, 874–894. https://doi.org/10.1177/0093650218807041

Mackie, D. M., & Silver, L. (2004). Intergroup emotions. In L. Z. Tiedens & C. W. Leach (Eds.), *The social life of emotions* (pp. 227–245). Cambridge University Press.

Mackie, D. M., Smith, E. R., & Ray, D. G. (2008). Intergroup emotions and intergroup relations. *Social and Personality Psychology Compass*, *2*(5), 1866–1880. https://doi.org/10.1111/j.1751-9004.2008.00130.x

Malka, A., & Lelkes, Y. (2010). More than ideology: conservative–liberal identity and receptivity to political cues. *Social Justice Research*, *23*(2-3), 156–188. https://doi.org/10.1007/s11211-010-0114-3

Mason, L. (2015). "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, *59*(1), 128–145. https://doi.org/10.1111/ajps.12089

Mason, L. (2018a). *Uncivil agreement: How politics became our identity*. University of Chicago Press.

Mason, L. (2018b). Ideologues without issues: The polarizing consequences of ideological identities. *Public Opinion Quarterly*, *82*(S1), 866–887. https://doi.org/10.1093/poq/nfy005

Masullo Chen, G., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media*, *61*(1), 108–125. https://doi.org/10.1080/08838151.2016.1273922

McConnell, C., Margalit, Y., Malhotra, N., & Levendusky, M. (2018). The economic consequences of partisanship in a polarized era. *American Journal of Political Science*, *62*(1), 5–18. https://doi.org/10.1111/ajps.12330

McGarty, C., Bliuc, A. M., Thomas, E. F., & Bongiorno, R. (2009). Collective action as the material expression of opinion-based group membership. *Journal of Social Issues*, *65*(4), 839–857. https://doi.org/10.1111/j.1540-4560.2009.01627.x

Meleady, R., & Forder, L. (2019). When contact goes wrong: Negative intergroup contact promotes generalized outgroup avoidance. *Group Processes & Intergroup Relations*, *22*(5), 688–707. https://doi.org/10.1177/1368430218761568

Muddiman, A. (2017). Personal and public levels of political incivility. *Journal of International Communication*, *11*, 3182–3202. https://ijoc.org/index.php/ijoc/article/view/6137

Mullen, B., Brown, R., & Smith, C. (1992). In-group bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*(2), 103–122. https://doi.org/10.1002/ejsp.2420220202

Munger, K. (2021). Don't @ me: Experimentally reducing partisan incivility on twitter. *Journal of Experimental Political Science*, *8*, 102–116. https://doi.org/10.1017/xps.2020.14

Muresan, S., Gonzalez-Ibanez, R., Ghosh, D., & Wacholder, N. (2016). Identification of nonliteral language in social media: A case study on sarcasm. *Journal of the Association for Information Science and Technology*, *67*(11), 2725–2737. https://doi.org/10.1002/asi.23624

Mutz, D. C. (2015). *In-your-face politics: The consequences of uncivil media*. Princeton University Press.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtainingR2from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nave, N. N., Shifman, L., & Tenenboim-Weinblatt, K. (2018). Talking it personally: Features of successful political posts on facebook. *Social Media + Society*, *4*, 1–12. https://doi.org/10.1177/2056305118784771

Papacharissi, Z. (2014). *Affective publics: Sentiment, technology, and publics*. Oxford University Press.

Papacharissi, Z., & de Fatima Oliveira, M. (2012). Affective news and networked publics: The rhythms of news storytelling on #Egypt. *Journal of Communication*, *62*(2), 266–282. https://doi.org/10.1111/j.1460-2466.2012.01630.x

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? *Communication Research*, *25*(6), 689–715. https://doi.org/10.1177/009365098025006006

Rains, S. A., Kenski, K., Coe, K., & Harwood, J. (2017). Incivility and political identity on the internet: Intergroup factors as predictors of incivility in discussions of news online. *Journal of Computer-Mediated Communication*, *22*(4), 163–178. https://doi.org/10.1111/jcc4.12191

Reiljan, A. (2020). 'Fear and loathing across party lines' (also) in Europe: Affective polarisation in European party systems. *European Journal of Political Research*, *59*(2), 376–396. https://doi.org/10.1111/1475-6765.12351

Rossini, P. (2020). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research*. 1–27. https://doi.org/10.1177/0093650220921314

Scott, C. R. (2007). Communication and social identity theory: Existing and potential connections in organizational identification research. *Communication Studies*, *58*(2), 123–138. https://doi.org/10.1080/10510970701341063

Settle, J. E. (2018). *Frenemies: How social media polarizes America*. Cambridge University Press.

Shafranek, R. M. (2020). Political consequences of partisan prejudice. *Political Psychology*, *41*(1), 35–51. https://doi.org/10.1111/pops.12596

Shi, F., Teplitskiy, M., Duede, E., & Evans, J. A. (2019). The wisdom of polarized crowds. *Nature Human Behaviour*, *3*(4), 329–336. https://doi.org/10.1038/s41562-019-0541-6

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131–142. https://doi.org/10.1037/0033-2909.105.1.131

Skytte, R. (2020). Dimensions of elite partisan polarization: Disentangling the effects of incivility and issue polarization. *British Journal of Political Science*. 1–19. https://doi.org/10.1017/s0007123419000760

Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(38), 18888–18892. https://doi.org/10.1073/pnas.1908369116

Soroka, S. N. (2014). *Negativity in democratic politics: Causes and consequences*. Cambridge University Press.

Spears, R., Lea, M., Corneliussen, R. A., Postmes, T., & Haar, W. T. (2002). Computer-mediated communication as a channel for social resistance. *Small Group Research*, *33*(5), 555–574. https://doi.org/10.1177/104649602237170

Stark, L. (2020). Empires of feeling. In M. Boler & E. Davis (Eds.), *Affective politics of digital media* (pp. 298–313). Routledge.

Statista. (2021). *Leading social media websites in the United States*. Statista. https://www.statista.com/statistics/265773/market-share-of-the-most-popular-social-media-websites-in-the-us/

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, *29*(4), 217–248. https://doi.org/10.2753/mis0742-1222290408

Suhay, E., Bello-Pardo, E., & Maurer, B. (2018). The polarizing effects of online partisan criticism: Evidence from two experiments. *International Journal of Press/Politics*, *23*(1), 95–115. https://doi.org/10.1177/1940161217740697

Tamburrini, N., Cinnirella, M., Jansen, V. A. A., & Bryden, J. (2015). Twitter users change word usage according to conversation-partner social identity. *Social Networks*, *40*, 84–89. https://doi.org/10.1016/j.socnet.2014.07.004

Tettegah, S. (2016). *Emotions, technology, and social media*. Academic Press Inc.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature*. Hewlett Foundation. https://www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf

Tyagi, A., Uyheng, J., & Carley, K. M. (2020). *Affective polarization in online climate change discourse on Twitter*. Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, The Hague, Netherlands, pp. 443–447.

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, *69*, 102199. https://doi.org/10.1016/j.electstud.2020.102199

Weeks, B. E., & Garrett, K. (2019). Emotional characteristics of social media and political misperceptions. In J. E. Katz & K. K. Mays (Eds.), *Journalism and truth in an age of social media* (pp. 236–251). Oxford University Press.

Wojcieszak, M., & Warner, B. R. (2020). Can interparty contact reduce affective polarization? A systematic test of different forms of intergroup contact. *Political Communication*, *37*(6), 789–811. https://doi.org/10.1080/10584609.2020.1760406

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, *38*, 98–139. https://doi.org/10.1080/10584609.2020.1785067

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of Science Technology & Society*, *30*(5), 316–327. https://doi.org/10.1177/0270467610380011

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, *12*(1), 415–438. https://doi.org/10.1146/annurev-economics-081919-050239

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1), https://doi.org/10.14763/2016.1.401

Author Biography

Nahema Marchal is a doctoral candidate at the Oxford Internet Institute, University of Oxford.