PNAS Nexus, 2024, 3, pgae395
https://doi.org/10.1093/pnasnexus/pgae395
Advance access publication 15 October 2024
Research Report

Attraction to politically extreme users on social media

Federico Zimmerman[a,b], David D. Bailey[a,b], Goran Muric[c], Emilio Ferrara[c,d,e], Jonas Schöne[f], Robb Willer[f], Eran Halperin[g], Joaquín Navajas[h,i,j], James J. Gross[k] and Amit Goldenberg[a,b,l,*]

[a]Harvard Business School, Harvard University, Cambridge, MA 02163, USA
[b]Digital, Data, & Design Institute, Harvard University, Cambridge, MA 02163, USA
[c]Information Sciences Institute, University of Southern California, Los Angeles, CA 90089, USA
[d]Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA
[e]Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, CA 90089, USA
[f]Department of Sociology, Stanford University, Stanford, CA 94305, USA
[g]Department of Psychology, Hebrew University of Jerusalem, 9190501, Jerusalem, Israel
[h]Laboratorio de Neurociencia, Universidad Torcuato Di Tella, C1428BCW, Buenos Aires, Argentina
[i]Escuela de Negocios, Universidad Torcuato Di Tella, C1428BCW, Buenos Aires, Argentina
[j]National Scientific and Technical Research Council (CONICET), C1425FQB, Buenos Aires, Argentina
[k]Department of Psychology, Stanford University, Stanford, CA 94305, USA
[l]Department of Psychology, Harvard University, Cambridge, MA 02138, USA
*To whom correspondence should be addressed: Email: agoldenberg@hbs.edu
Edited By Erik Kimbrough

Abstract

Political segregation is a pressing issue, particularly on social media platforms. Recent research suggests that one driver of segregation is *political acrophily*—people's preference for others in their political group who have more extreme (rather than more moderate) political views. However, acrophily has been found in lab experiments, where people choose to interact with others based on little information. Furthermore, these studies have not examined whether acrophily is associated with animosity toward one's political out-group. Using a combination of a survey experiment ($N = 388$) and an analysis of the retweet network on Twitter (3,898,327 unique ties), we find evidence for users' tendency for acrophily in the context of social media. We observe that this tendency is more pronounced among conservatives on Twitter and that acrophily is associated with higher levels of out-group animosity. These findings provide important in- and out-of-the-lab evidence for understanding acrophily on social media.

Significance Statement

Political segregation is a central problem contributing to intergroup conflict and prejudice. Traditionally, it has been thought that segregation is driven by political homophily, the tendency to affiliate with similar others. Recently, it has been suggested that an additional driver of segregation is political acrophily, the tendency to affiliate with others who are more politically extreme. However, acrophily has only been examined in tightly controlled lab experiments. Therefore, testing it in natural interactions is critical to establishing its role in real-world segregation. In two studies, a controlled experiment and an observational study of digital trace data, we show that social media users are more likely to connect with more politically extreme users. Findings provide important out-of-the-lab evidence for acrophily.

Introduction

Political segregation can be seen in almost any social interaction, from the choice of social partners (1), the structure of online social networks (2, 3), to the composition of neighborhoods, and cities (4, 5). Segregation represents a major challenge in the United States and elsewhere. It is associated with attitudinal polarization, intergroup hostility, and increased spread of misinformation (6–8).

Prior work has typically proposed that a key driver of political segregation is *political homophily*, the tendency to affiliate with others who have similar political views (9–12). Political homophily is a pervasive and enduring propensity (13), and it seems to be increasing. For example, according to Iyengar *et al.* (14), in the past 50 years there has been an increase of about 35% in the percentage of Americans who would be somewhat or very unhappy if their child married someone of the opposite party. Many have argued that homophilous political preferences are especially salient in the digital era, where social ties can be formed and dissolved quickly, resulting in the formation of echo chambers of like-minded people who rarely interact across political lines (2, 3, 15–17).

Recently, however, it has been argued that homophily may not be the only driver of decisions about whom to affiliate with in political contexts. Instead, it has been proposed that tie-selection

Competing Interest: The authors declare no competing interests.
Received: April 12, 2024. **Accepted:** August 27, 2024

decisions in the context of politics are also impacted by *political acrophily*, the tendency to prefer to affiliate with others with more extreme (as opposed to more moderate) political views in the direction of one's political leaning (18). To illustrate, consider a liberal who holds a political stance of a *7* on a *1–10 scale* (five being the center of the political map) and is offered two new potential ties, one of a political stance of an 8 and one of a 6. Based on homophily, the two ties have the same distance from the chooser, and therefore are equally likely to be chosen. Selecting ties based on acrophily would suggest that the 8 would be preferred over the 6. This of course does not mean that homophily—one of the strongest social forces—is not also influential, but rather that acrophily may exist in addition to homophily. Finding evidence for political acrophily is important because it could be a critical catalyst of political segregation in social networks.

So far, one empirical study has directly examined the occurrence of acrophily (18). In a controlled task, participants were asked to rate their emotions or attitudes toward pictures of police brutality against a Black demonstrator. They then saw peers' responses to the same pictures and selected their preferred peers. Based on their selections, they continued being presented with responses from the chosen participants in subsequent trials, while the responses of those not selected were no longer shown. Results showed that participants' peer selections were based on both homophily and acrophily. Furthermore, the tendency for acrophily was associated with individuals' perceptions that more extreme expressions are also considered more prototypical of one's political group, suggesting that norm perceptions may be driving acrophily. Previous analyses have mostly focused on segregation as an outcome and have not examined the potential association between acrophily and processes such as out-group animosity or in-group positivity.

Other findings provide converging—albeit more indirect—support for attraction toward more extreme others. For example, previous research on deviation from group norms suggests that people hold more positive views of those who deviated from the group norm by adopting more extreme views than those who deviated from the norm by adopting more moderate views (19–22). Furthermore, in a recent study of the effects of political beliefs on how well liked someone is, Zimmerman *et al.* (23) found that participants liked peers with more coherent political views more than those with less coherent views, and coherence is positively correlated with extremity.

Previous research also points to potential mechanisms for acrophily that were not directly examined. For example, acrophily may be driven by the fact that extreme partisans are considered more committed to their own party (24) and more authentic (25, 26), which may also mean that they are considered more genuine in-group members. These mechanisms are congruent with sociological research on social categorization (27–29), where extreme partisans might navigate the trade-off between fitting in and standing out by signaling a strong fit with the in-group while differentiating themselves from other partisans. As a result, they might receive more attention on social media without the cost of not being considered copartisans (30). Furthermore, there may be some mechanisms that may make extreme users especially salient and visible on social media. For example, research suggests that extreme users are more active on social media and are promoted by algorithms (31). These factors may make extreme users seem more attractive.

Further insight into the notion of tie-selection and particularly acrophily can be derived from theories on voting behavior. For example, the notion of the "proximity model", in which voters choose candidates based on similarity to their views, aligns with the concept of homophily (32). As for acrophily, it seems to be most similar in some aspects to the concept of "directional voting" (33–35), which suggests that people have a clear preference for representatives of their side, even if the distance in attitudes to these representatives is greater than to representatives from the other side. The two ideas are also different in some aspects, much of the theory of directional voting argues for a preference for a more distant in-group candidate over a more similar out-group candidate (see Rabinowitz and Macdonald (34)). But one extension of directional voting is the idea that people are attracted to representatives who are more extreme than themselves, which is analogous to the concept of acrophily. It is important to note, however, that directional voting has been the focus of a great debate, without clear empirical evidence for its occurrence (36, 37). While directional voting relates to voting decisions and acrophily to the creation of social ties, both types of decisions might be driven by similar mechanisms. Providing real-life evidence of tie-formation on social media could contribute to the development of a single general theory. However, different social and contextual factors could shape these decisions in unique ways. Therefore, exploring both the commonalities and differences of the mechanisms driving social tie-formation and voting decisions should be further explored.

The current set of studies was designed to address a few important gaps in our understanding of acrophily. First, while these initial studies are a promising first step, there is yet no direct evidence for the occurrence of acrophily in natural social interactions. Such evidence is essential because the only direct test of acrophily was conducted in a tightly controlled laboratory setting in which people were equally exposed to co- and counter-partisan peers. In real interactions, however, networks are already segregated, and people are mostly surrounded by similar others, which may also eliminate the tendency for acrophily. Second, participants' choices of social ties in the previous task were solely based on others' responses to a specific political response. But in natural social interactions, social ties may be chosen for a variety of reasons, which may dilute or eliminate the effect of acrophily. Third, looking at acrophily online could help us examine its association with other behaviors on social media, and specifically the association between acrophily and individuals' feelings toward the opposite political party and their own. Such association has not been examined but can be inferred from the fact that affective polarization, which is the tendency to feel negative feelings toward one's out-group and positive feelings toward one's in-group, is driven by people's attraction to their group identity (8, 38). It is therefore likely that attraction to more extreme members of one's own political group will also be related to out-group animosity and in-group positivity. More specifically, we expect this preference for political extremes to be associated with out-group animosity, as it is a stronger predictor of political behaviors such as voting and the dissemination of fake news (39, 40).

The present research

The current research aims to test two main hypotheses. The first hypothesis (H1) is that acrophily is present in real-life interactions. The second hypothesis (H2) is that acrophily is associated with out-group animosity. While acrophily has been found in lab settings, it could be more prevalent on social media, where political norms are more extreme. Conversely, it might not be observed in real-life networks that are already strongly segregated.

Moreover, the individual characteristics associated with higher levels of acrophily have not been thoroughly studied. We hypothesize that a preference for political extremes is associated with out-group animosity, as it is a strong predictor of political identity and behavior.

This research consists of two studies, each designed to examine both hypotheses across different contexts. In study 1, we test our hypotheses in an online controlled study. We design a lab experiment in which participants are presented with different user profiles and are asked to rate both the likelihood of following these profiles back on social media and to evaluate each one across multiple dimensions (e.g. intelligence and confidence).

In study 2, we test our hypotheses in real-life settings using a different set of data and methodologies compared to those used in study 1. H1: we examine tendencies for acrophily in actual tie connections using a large Twitter dataset (41) of tweets on political issues. We examine retweet connections between people using a simulation network approach to evaluate acrophily (18). H2: in addition to looking at general tendencies of acrophily across political groups, we also examine the association between acrophily at the individual level and out-group animosity. To do this, we implement sentiment and categorization analysis of users' tweets (42, 43). We capture out-group animosity by analyzing the interaction between the use of negative emotions and the frequency of third-person plural pronouns in users' tweets.

Results

Study 1: Evidence for acrophily in reported intentions on social media

The goal of study 1 was to test the two main preregistered hypotheses (https://aspredicted.org/8WL_7KT). H1: participants would prefer extreme copartisans over moderate ones. We predicted that this preference would be reflected in participants being more likely to follow back extreme copartisans compared to moderate ones. Moreover, we expected to observe direct evidence for the occurrence of acrophily by looking at the distance between participants' self-reported political affiliation and their estimation of the profiles' political affiliation. H2: the preference for extreme copartisans would be associated with higher levels of out-group animosity, measured by the feeling thermometer. This study was performed by $N = 388$ US citizens, recruited via Connect by Cloud Research, who frequently use Twitter (see "Materials and methods"). The sample was balanced in terms of party affiliation (Democrats/Republicans) and gender (male/female).

We created fictional social media users and varied their degree of political extremity. We then conducted a pilot study to validate that these fictional social media users were indeed perceived as having different levels of political extremity (see "Materials and methods"). In the actual experiment, each participant was shown three fictional Twitter profiles appearing at a random order: a neutral profile, a moderate in-group, and an extreme in-group. Each profile included the user's names, picture, and biography (Fig. 1A). The profiles' political extremity was manipulated by incorporating a party icon in the profile picture and by modifying the user bio to indicate stronger partisanship. After viewing each profile, participants were asked to imagine that the profile had followed them on Twitter and were asked if they would follow the profile back. We used two measures to evaluate participants' likelihood of following each profile back. The first was a binary scale (yes/no), and the second was a continuous scale in which we asked participants to estimate the likelihood of following back on a six-point scale. In addition to these measures, to further validate that attraction to extremes was indeed driven by acrophily—which is defined as the preference or motivation to affiliate with others from one's groups who are more extreme than oneself—we asked participants to place both themselves and each profile on a seven-point political affiliation scale ranging from "strongly liberal" to "strongly conservative."

Participants next filled out an interpersonal attraction questionnaire asking them to indicate the extent to which the person in the profile would likely be a friend, would fit in their circle of friends, and could be depended on to get things done (see Supplementary material for analysis). Participants also rated the individuals portrayed in the profiles on eight attributes, such as intelligence and self-confidence (see "Materials and methods" for full list). Finally, to examine the association between the preference for extreme profiles and out-group animosity, participants completed a five-point Likert feeling thermometer scale evaluating both political parties, ranging from "very cold" to "very warm."

Evaluating the tendency to follow back extreme profiles

We examined whether participants were more likely to follow back extreme profiles compared to moderate ones (Fig. 1A). Following the pre-registration, we performed one mixed-effects logistic regression considering the yes/no question as a dependent variable (DV) and one linear mixed model including the six-point Likert scale question as the DV. The regression predictors included profiles' extremity, and participants' ideology and extremity. These three variables were categorized as dummy variables under the following classifications: neutral, moderate, or strong in-group profile; Democrat or Republican participant; and moderate or strong partisan participant. We also considered the interaction between profiles' extremity and participants' partisanship to determine if the preference for extreme in-groups is specific to one political party, and we added a random intercept for participant id.

First, we confirmed that participants preferred copartisans over neutral profiles [Binomial generalized linear mixed model considering the binary DV: $b = 0.99$, $z(1,158) = 3.92$, $P < 0.001$, $R^2 = 0.06$, 95% CIs = (0.49, 1.48), Fig. 1B; linear mixed-effects model considering the discrete DV: $b = 0.58$, $t(11,157) = 4.76$, $P < 0.001$, $R^2 = 0.04$, 95% CIs = (0.34, 0.82), Fig. 1C]. Then, we focused the analysis only on in-group profiles and found that people were more likely to follow back an extreme in-group over a moderate one [Binomial generalized linear mixed model: $b = 0.63$, $z(770) = 2.07$, $P = 0.04$, $R^2 = 0.05$, 95% CIs = (0.03, 1.23), Fig. 1B; linear mixed-effects model: $b = 0.37$, $t(769) = 3.04$, $P < 0.001$, $R^2 = 0.03$, 95% CIs = (0.13, 0.60), Fig. 1C]. Republicans were more likely than Democrats to follow back others in general [main effect: $b = 1.46$, $z(770) = 3.04$, $P = 0.002$, 95% CIs = (0.52, 2.39); $b = 0.55$, $t(769) = 2.59$, $P = 0.01$, 95% CIs = [0.14, 0.97)], with no interaction between political affiliation and attraction to extreme copartisans [interaction: $b = -0.10$, $z(770) = -0.23$, $P = 0.82$, 95% CIs = (−0.95, 0.75); $b = -0.02$, $t(769) = -0.09$, $P = 0.93$, 95% CIs = (−0.35, 0.32); see Supplementary material for details]. According to the next analysis in the pre-registration, to verify the robustness of our findings, we considered participants' perceptions of profile extremity as predictors in the regression models and their intentions to follow them back as the DV. When taking participants' perceptions of profiles' extremity into account, the effect remains significant, suggesting that the extremity of the profiles is influencing participants' preferences (see Supplementary material for details).

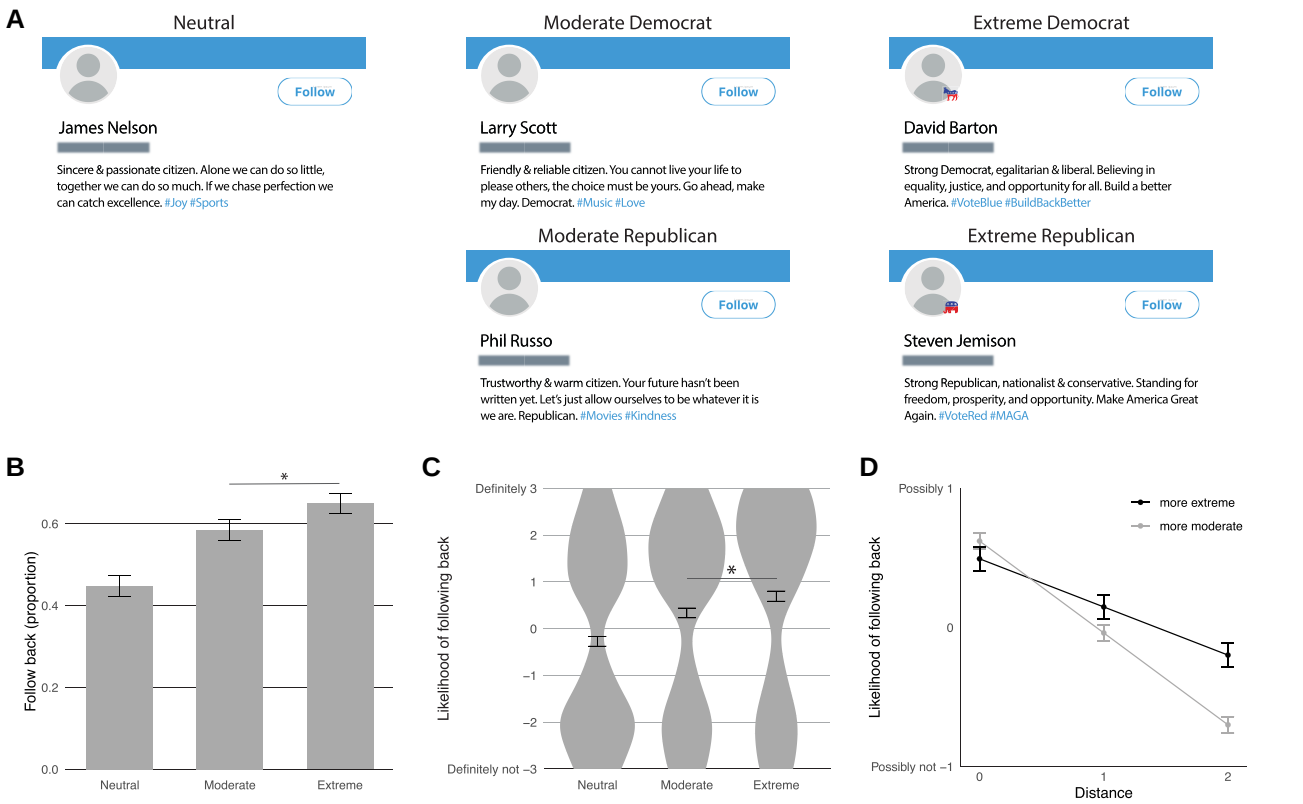

Fig. 1. Results from study 1. A) We present examples of the stimuli showing a neutral profile, a moderate Democrat, a moderate Republican, an extreme Democrat, and an extreme Republican profile. The stimuli were validated in a previous pilot study. The faces (not shown here) were created through the website of Generated Media Inc. (https://generated.photos/). B and C) Participants were more likely to follow an extreme partisan over a moderate one. B) Results for the yes/no follow-back question. The figure shows the proportion of participants following back each profile and its SE of proportion (*$P < 0.05$). C) Results for the likelihood of following back each profile (six-point Likert scale). Response distribution is shown in gray and the mean value and SEM in black (*$P < 0.05$). D) Participants were more likely to follow back in-group profiles that are more extreme than themselves, compared to more moderate, consistent with what acrophily predicts. The figure shows the estimates and SE from the linear regression model, depicting the likelihood of following back in-group profiles based on their distance on the political orientation scale. It compares this likelihood for profiles that are more extreme (black) and more moderate (gray) than the participant.

So far, results suggested that participants were attracted to extreme profiles more than to moderate ones, but these findings may not necessarily be driven by acrophily. In an extreme case, where all the participants are extreme partisans, the results above could be all driven by attraction to similarity, i.e. homophily. We therefore wanted to validate that indeed attraction to extreme was driven by acrophily. To address this concern, we considered the distance between the political positions of each in-group profile, evaluated by each participant, and the participant's own positions on a seven-point political orientation scale, ranging from "strongly liberal" to "strongly conservative." We performed one mixed-effects logistic regression with the yes/no question as the DV and one linear mixed model including the six-point Likert scale question as the DV. The predictors in the regressions were the political absolute distance between participants and profiles, a dummy variable indicating whether the profile is more extreme or moderate than the participant, and the interaction between these variables. We also included participants' political position as a control variable and a random intercept for participant id. We checked that any increase in distance of the profile from participants' own political affiliation led to a reduced tendency to follow back. Indeed, we found that increasing the distance between participants' self-reported political affiliation and their estimation of the profiles' political affiliation was a significant predictor of the likelihood of following back a specific profile, as predicted by homophily [Binomial generalized linear mixed model: main effect: $b = -4.79$, $z(1,070) = -7.73$, $P < 0.001$, $R^2 = 0.06$, 95% CIs $= (-6.01, -3.58)$; linear mixed-effects model: main effect: $b = -0.68$, $t(1,069) = -14.56$, $P < 0.001$, $R^2 = 0.08$, 95% CIs $= (-0.77, -0.49)$]. Moreover, the interaction between the absolute distance and whether the in-group profile is more extreme or more moderate than the participant was significant [interaction: $b = 3.12$, $z(1,070) = 4.36$, $P < 0.001$, 95% CIs $= (1.72, 4.52)$; interaction: $b = 0.46$, $t(1,069) = 5.54$, $P < 0.001$, 95% CIs $= (0.30, 0.63)$; Fig. 1D]. This interaction indicates that participants were more likely to follow back in-group profiles that are more extreme than themselves, compared to more moderate ones, consistent with what acrophily predicts. These results can be seen in Fig. 1D. Regarding the main effect of the distance between participants' self-reported political affiliation and their estimation of the profiles' political affiliation, we see that for peers in both directions, more moderate and more extreme, as the distance increases, the likelihood of following back a target decreases. Moreover, the interaction effect with whether the profile is more extreme or more moderate can be seen in the fact that the decrease in the likelihood of following back a target is steeper for more moderate others compared to more extreme ones. This difference in the slopes signifies a preference for extreme partisans over moderate ones.

In addition to these analyses, we examined participants' perceptions of the profiles on a variety of dimensions as a function

of whether the profiles were extreme or moderate. Consistent with previous studies (18), we found that participants evaluated extreme profiles as more aligned with the mainstream view of their party supporters [paired t test: $t(387) = 9.08$, $P < 0.001$]. Participants also evaluated extreme users to be more confident [$t(387) = 7.02$, $P < 0.001$], interesting [$t(387) = 5.89$, $P < 0.001$], and intelligent [$t(387) = 5.65$, $P < 0.001$], and more likely to provide solid arguments in discussions [$t(387) = 5.59$, $P < 0.001$], see Supplementary material for full analysis.

Out-group animosity

We examined whether attraction to extreme profiles was associated with out-group animosity. Participants reported their party affiliation and completed a feeling thermometer, where we asked them to report their feelings toward both the Democratic and Republican parties. In line with the pre-registration, we quantified participants' preference toward extreme partisans as the difference in the likelihood of following back extreme vs. moderate profiles for each participant, and out-group animosity was measured by the reported feelings toward the opposite political party on a five-point Likert scale ranging from "very cold" to "very hot." As expected, these variables were significantly correlated [Spearman's correlation: $r = -0.19$, $P < 0.001$, 95% CIs = (−0.28, −0.09)]. To control for other individual variables, we performed a linear regression model incorporating participants' party affiliation and extremity, their feelings toward both parties, gender, and age as predictors, and the preference for extreme profiles as the DV. While we found that in-group positivity is also associated with the preference for extreme in-groups [linear model: $b = 0.26$, $t(381) = 2.47$, $P = 0.01$, 95% CIs = (0.05, 0.47)], negative feelings toward the out-group party emerged as the strongest predictor among the variables considered [$b = -0.30$, $t(381) = -2.94$, $P = 0.003$, $R^2 = 0.05$, 95% CIs = (−0.50, −0.10); see Supplementary material for full analysis]. Alternatively, rather than analyzing feelings toward the in-group and out-group separately, we computed the absolute difference between these two variables, i.e. affective polarization, and found that, consistent with the previous approach, affective polarization is significantly correlated with the attraction to extreme profiles (see Supplementary material for analysis).

Study 2: Observational evidence for acrophily in social media

The goal of study 2 was to test the hypotheses that acrophily is present in real-life interactions, and it is associated with out-group animosity. We aimed to assess the ecological validity of the findings from study 1 by examining real-life social media interactions, which encompass the influence of social media algorithms and user behavior patterns. We believe that providing evidence for acrophily on social media is important not only because it reveals the effect outside of the lab, but because social media networks are already segregated, which makes the real-life analysis a conservative test compared to the empirical study.

We examined the occurrence of acrophily by looking at a retweet network derived from a large dataset of Twitter interactions (N = 1,865,559) collected from 2019 May to 2020 December (41). When analyzing the data, we were expecting to find evidence for the occurrence of acrophily, such that users would be more likely to retweet content produced by users who are more politically extreme than themselves. We used a simulation analysis approach because retweet behavior alone could not reveal acrophily in a natural environment, as we cannot control the distribution of political affiliations in the network. For instance, if the distribution is skewed toward the extreme, even random retweeting would appear to show the opposite of acrophily. By simulating different tie-selection strategies, we were able to compare users' behavior to their hypothetical behavior given the existing networks, which allowed us to account for the existing distribution of political affiliation. The current dataset also provided the opportunity of testing which user attributes are associated with acrophily. More specifically, we were looking to find evidence for out-group animosity based on the content that users produce. We chose to focus our analysis by looking at retweet networks because retweet connections provide direct evidence that users have seen content produced by each other and therefore represent an indication for actual connection (44, 45). Retweet networks also reflect how information actually spreads between users and allow an assessment of connection strength—represented by the number of retweets between users (46).

We evaluated users' political affiliation based on their media consumption (see "Materials and methods" for full details). The idea behind this method is that the political affiliation of the media outlets that users choose to retweet reflects their own political views. This method has been validated in many previous studies, including ones that evaluated users' political affiliation and compared them to actual political beliefs (47–49). After mapping all the political affiliation ratings for our participants, we rescaled the variable to range from −2 to 2, liberal to conservative. We also removed participants who were located at the middle of the scale on average (see "Data reduction") and converted liberal ratings to positive values so that both political groups would be on a scale of 0 = moderate to 2 = extreme. As can be seen in Fig. 2, the conservative distribution of political affiliation leans more extreme, which may be driven by the number of extreme media outlets on the conservative side. It is important to note that this distribution should make acrophily even harder to detect as these extreme users have fewer users who are more extreme than them to retweet.

Data reduction

Data were reduced in four ways. First, we only included users who produced at least five original tweets, so that we could focus on active accounts. Second, we included only retweet connections between copartisans and therefore excluded all cross-partisan interactions or retweets of users whose political affiliation was equal to zero (center). Third, because we were interested in peer selection, we removed any instance where users retweeted themselves. Fourth, we trimmed the network based on the number of aggregated retweets between users (from now on called egos for the sake of simplicity) and peers (from now on called alters). Because our simulations (described below) included swapping existing retweet connections with others, and because these simulations were all done within a certain retweet strength, we wanted to make sure that we had enough connections available for the simulations. Therefore, we focused on retweet connections of up to five retweets, as connections with six retweets or more each represent <2% of the total data (see Supplementary material for details). These four data reduction efforts resulted in a dataset including 25,639 conservatives and 41,054 liberals (for a total of 66,639 egos) with 1–5 retweet connections. There were 3,898,327 unique ego-alter connections (1,553,814 conservative and 2,344,513 liberal) and 5,722,470 total retweets among all connections (2,335,428 conservative and 3,387,042 liberal).

Visualizing acrophily

One challenge in examining acrophily with social media data is that users' choice of interaction very much depends on the

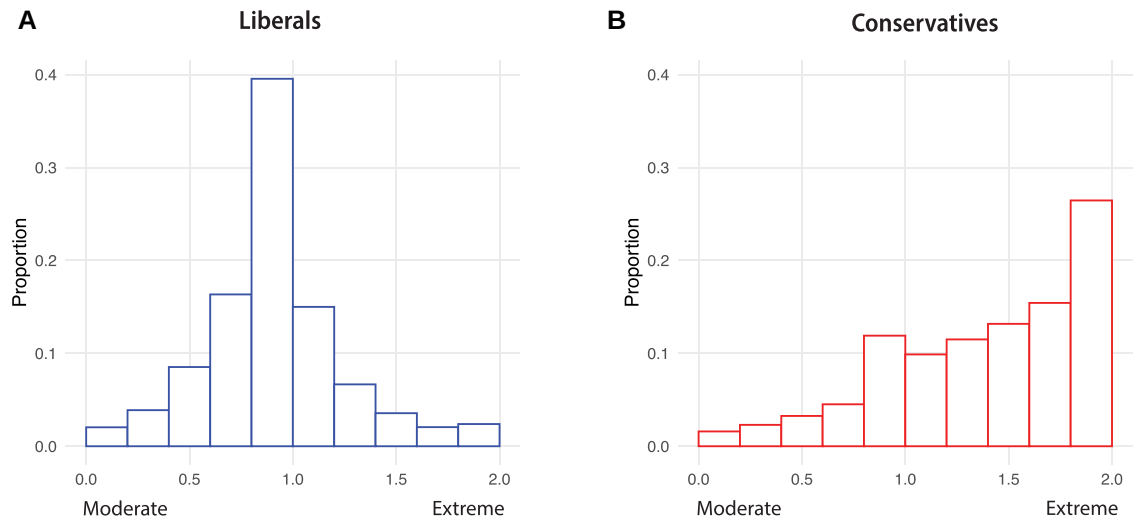

Fig. 2. Users' political affiliation from study 2, ranging from moderate (0) to extreme (2) for each political group. A) shows the distribution for liberals and B) shows the distribution for conservatives. We removed users with political affiliation = 0 as part of our data reduction effort to only examine copartisan interactions.

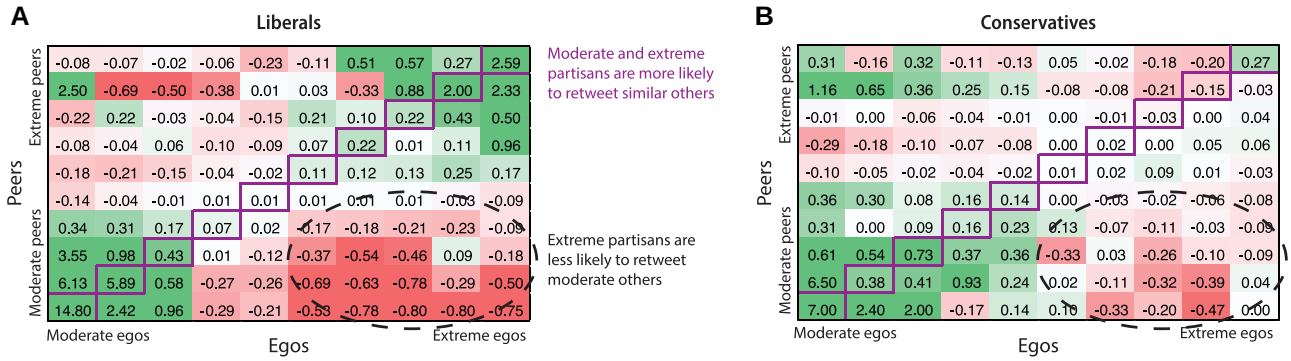

Fig. 3. The political affiliations of egos and their peers were categorized into 10 bins, each with a width of 0.2. To determine the extent to which the political positions of peers—those being retweeted—are overrepresented or underrepresented, we computed the difference between actual data and random behavior. This calculation was normalized by the expected number of users in each bin according to random behavior, facilitating a bin-by-bin comparison. Positive values indicate overrepresentation, while negative values indicate underrepresentation. For both A) liberals and B) conservatives, we see that (i) moderate and extreme partisans are more likely to retweet similar others and (ii) extreme partisans are less likely to retweet moderate others than vice versa.

political distribution of available users. If a certain network has more extreme users, content produced by these users is more likely to be shared. Before conducting an actual statistical evaluation of acrophily, and given the complexity of seeing acrophily in the data, we were interested in visualizing retweet patterns in the network. To do that, we compared the political affiliation of the users that participants actually retweeted to what they would have retweeted if choosing at random. We categorized each user's political affiliation, which ranges from 0 to 2, into 10 bins with a width of 0.2 each. We also categorized the bins of the retweeted users in the same way. For each bin combination, we then calculated the difference between the actual count of retweets and the count of retweets that would have been achieved if users were chosen at random. We normalized these findings by dividing by the random number. The result was a 10×10 grid in which each cell provided an indication whether users in this cell were either overrepresented in retweets or underrepresented.

Figure 3 provides a descriptive visualization of users' retweet patterns. Two results appear. The first is that users (both liberals and conservatives) tended to homophilize: to retweet other users who share their political affiliation at a higher rate than random. This is seen by the positive numbers on the diagonal of the matrix. Results also suggest a process of acrophily: extreme egos are less likely to retweet moderate peers (reflected by the red numbers in the bottom right quadrant), compared to the probability of moderate users retweeting extremes.

Evaluating acrophily

To begin our analysis, we first performed a test for homophily in the data (reported in Supplementary material). Having established the occurrence of homophily, we then turned to testing our hypotheses related to acrophily. To measure acrophily, we had to compare users' actual behavior to a simulated behavior given the data. Looking only at users' retweet behavior could not reveal this tendency, because even if egos retweeted alters that were more extreme than they are, this could be driven by the distribution of political affiliation of the actual network. For example, because the political affiliation distribution of conservative egos is skewed toward the extreme, even if users were randomly retweeting each other, they would appear to exhibit the opposite of acrophily, because there are more moderate alters available to

retweet. Therefore, acrophily can only be detected when comparing participants' actual behavior to a hypothetical simulation.

We conducted the simulations following the methodology of previous analyses of acrophily (18). We performed the simulations separately for each retweet count because the network was based on retweets between users and the number of retweets between an ego and an alter could be considered as an indirect indication of tie strength. We ran 1,000 simulations for each retweet count, ranging from one to five. The simulations were based on three different strategies that determined the substitution procedure of the alters for each ego: complete homophily, acrophily, and complete acrophily.

The first strategy was a *complete homophily* simulation in which egos preferred alters whose political affiliation ratings were closest to them, regardless of whether they were more or less extreme. In the *complete homophily* simulation, we iterated through each ego one at a time, substituting the ego's actual chosen alter with the alter closest to them based on absolute difference in political ratings (Fig. 4). For example, imagine an ego with a tie connection of five retweets with another alter. In the complete homophily simulation, we substituted this alter with another alter who had the closest political affiliation rating to the ego among all alters who had five retweet connections in the network. The second is an *acrophily* simulation in which egos prefer only alters whose political ratings are similar or more extreme than them. Notice, however, that in this simulation, egos first will retweet users who are similar to them and only after that try to find others who are more extreme. In the acrophily substitution, we again selected each ego one at a time and substituted alters without replacement, only we replaced the actual alter with an alter that was similar or more extreme than the ego in terms of their political affiliation, starting from the closest ego and moving toward more extreme (Fig. 4). This meant that in the acrophily substitution a more extreme alter would always be selected so long as more extreme alters remained in the substitute alter pool. For example, if an ego had a political rating of 1.0, and the alter pool consisted of three alters, one with a rating of 0.95, one with a rating of 1.1, and one with a rating of 1.2, the current ego would then be paired with the 1.1 alter, despite the alter with a rating of 0.95 being closer to the ego's 1.0 rating. The alter with a rating of 1.1 would then be removed from the alter pool, meaning the alter pool would then consist of only the alter ratings of 0.95 and 1.2. If the next ego also had a rating of 1.0, we would repeat the process such that the second ego would be paired with the alter with the rating of 1.2. Only in a scenario in which all remaining alters were less extreme than the ego was an ego paired with a less extreme alter. A third and more extreme strategy is a *complete acrophily* simulation in which egos strongest preference is for the more extreme alters in the network. The difference between the *complete acrophily* and the *acrophily* strategies is the starting point. The first selected alter in the *acrophily* strategy is the alter closest to the ego, whereas the first selected alter in the *complete acrophily* strategy is the most extreme user available in the network (Fig. 4).

After running these three simulations, we consolidated the data by taking the average of all alters that an ego retweeted for all of our simulations separately for each retweet number (1–5) and testing whether that average alter score was more or less extreme than the score for ego. We calculated a binary outcome (more extreme or not) rather than a continuous outcome of the difference because the difference variable is highly impacted by the user's political affiliation and therefore provides misleading information on the degree of acrophily. The result of the simulation was therefore a dataset with each line representing a separate ego-alter connection, the retweet number between that connection, and whether the alter was more extreme than the ego in each of the three simulations.

To evaluate acrophily, we compared the probability that egos actually retweeted a more extreme alter to the probability that they would have retweeted a more extreme alter in each of our three hypothetical simulations: a homophily simulation, an acrophily simulation (in which users acrophilize but first retweet closer alters), and a complete acrophily simulation (in which users acrophilize but first retweet the most extreme alters; Fig. 5). In line with the acrophily findings, we expected that users would exhibit a preference for more extreme others higher than that predicted by homophily simulations.

We conducted our comparisons for each political group separately to simplify the results of the model. For each political group, we conducted a mixed linear model comparing the likelihood of retweeting a more extreme alter in the actual data and the three simulated strategies. In line with recent recommendations, we chose a linear model because it produced a better fit and more interpretable results (50). However, a generalized linear model produced very similar outcomes (reported in Supplementary material). To learn more about how these differences change as a function of the retweet connection, we added a covariate with

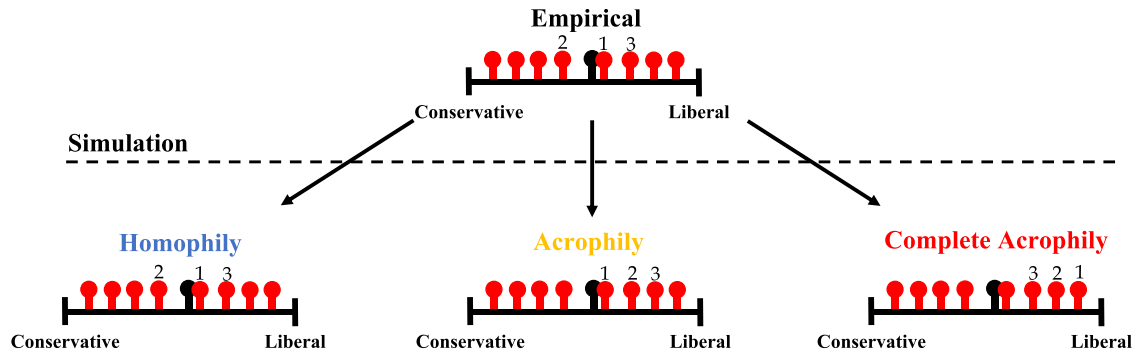

Fig. 4. An illustration of the simulations done for the network data in study 2. The ticker in the center of the scale represents a user's hypothetical political rating. The rest of the tickers represent potential peers that the user could retweet. The numbers above the tickers represent the order of retweets. We present three different simulations, each simulation representing a different tie-selection strategy. In homophily, the user retweets the users who are most similar to them in terms of political views. In acrophily, the user retweets users who are similar or more extreme in terms of political view, but the user's preference is to first retweet users who are closer to them. In complete acrophily, the user again prefers more extreme peers, but this time the preference is to start with the most extreme user in the network.

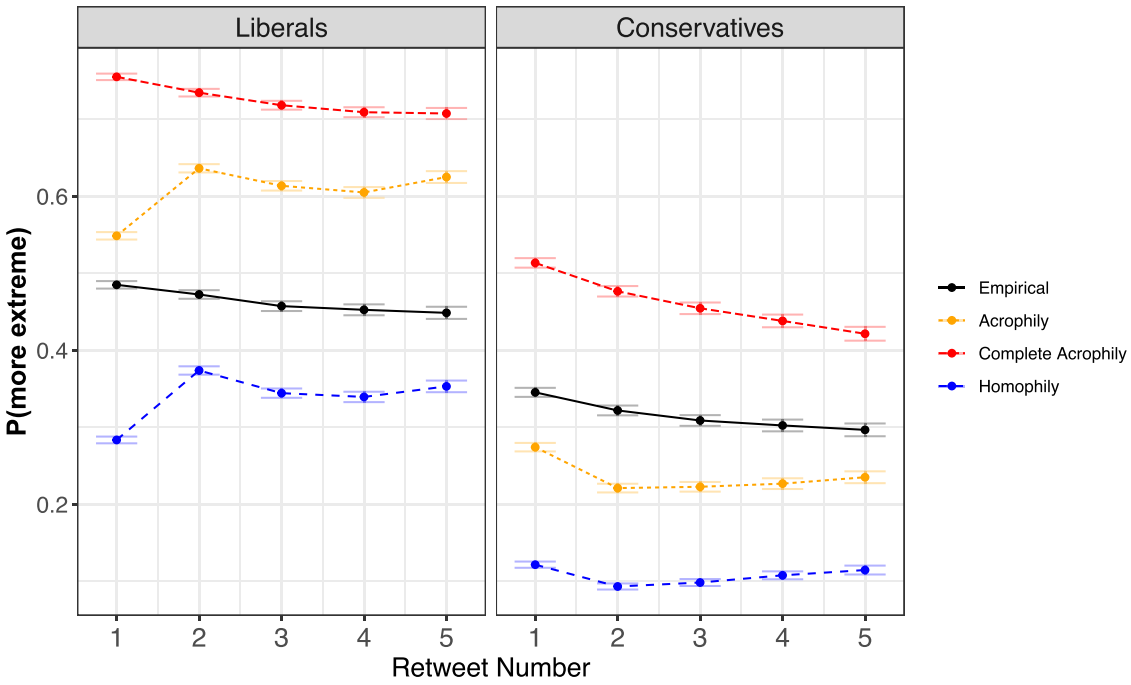

Fig. 5. Comparison of the probability of retweeting a more extreme user in our empirical data compared to the three simulations. The x axis represents the number of retweets between two users. The y axis is the probability of retweeting a more extreme alter. Error bars represent a 95% CI. Results of the liberal sample suggest that liberals' probability of retweeting a more extreme alter was more extreme than homophily but less extreme than acrophily. For the conservative participants, the probability of retweeting a more extreme alter was higher than the acrophily model.

the number of retweets. Finally, we added a within-user intercept nested within a retweet number slope. Starting with the Liberal sample, the baseline probability of retweeting a more extreme users was 46.90%, but this again could be mainly driven by the shape of the political affiliation distribution. Liberal users have 33.37% chance of retweeting extreme users when using our homophily simulation, which, as hypothesized, is significantly less than the actual retweet behavior [$b = -0.13$, $t = -93.96$, $P < 0.001$, $R^2 = 0.08$, 95% CIs = (−0.1381, −0.1325)]. Liberals were 59.98% likely to retweet extremes compared to our acrophily model, which is significantly more than the actual retweet behavior [$b = 0.13$, $t = 90.80$, $P < 0.001$, $R^2 = 0.08$, 95% CIs = (0.1279, 0.1335)] and 73.22% likely to retweet extremes in our complete acrophily simulation [$b = 0.26$, $t = 182.73$, $P < 0.001$, $R^2 = 0.08$, 95% CIs = (0.2603, 0.2659)]. These results suggest that egos were more likely to retweet extreme alters as seen by the comparison to the homophily simulation, but not more likely to retweet extreme alters compared to the acrophily simulations. Results also suggested that the tendency to retweet more extreme users increased with retweet number; however, this effect was quite weak [$b = 0.003$, $t = 5.48$, $P < 0.001$, $R^2 = 0.08$, 95% CIs = (0.001, 0.004)].

We then turned to our conservative sample. The baseline probability of retweeting a more extreme user was 33.28%, but this again could be mainly driven by the shape of the political affiliation distribution, which is extreme leaning. As hypothesized, conservative users were 12.05% likely to retweet extreme users in our homophily simulation, which is significantly lower than the actual retweet behavior [$b = -0.21$, $t = -151.12$, $P < 0.001$, $R^2 = 0.09$, 95% CIs = (−0.2149, −0.2194)]. Conservatives were also 25.24% likely to retweet extremes compared to our acrophily model, which is also significantly lower than the actual retweet behavior [$b = -0.08$, $t = -57.24$, $P < 0.001$, $R^2 = 0.09$, 95% CIs = (−0.0831, −0.0776)]. However, conservatives were 48.26% likely to retweet extremes in our complete acrophily, which was higher than their actual retweeting tendency [$b = 0.14$, $t = 106.72$, $P < 0.001$, $R^2 = 0.09$, 95% CIs = (0.1471, 0.1526)]. These results suggest that the conservative egos were more likely to retweet more extreme alters than in the acrophily simulation, but not more than in the complete acrophily simulation. Unlike in our liberal sample, results suggested that the tendency to retweet more extreme users decreases with retweet number [$b = -0.008$, $t = -15.06$, $P < 0.001$, $R^2 = 0.09$, 95% CIs = (−0.009, −0.007)].

Acrophily and out-group animosity

To learn more about individual-user tendencies for acrophily, we created a by-user acrophily coefficient. Following previous analyses of acrophily (18), we assumed that a larger tendency for acrophily would mean that the average political affiliation of the actual retweeted alters by a certain ego would be greater than the average political affiliation of the simulated retweeted user in the homophily simulation. We therefore took the average political affiliation of all alters a user retweeted and computed the difference from the hypothetical political affiliation in the homophily simulation. Larger positive numbers in this coefficient indicated stronger acrophily. To learn more about how such acrophily tendency is associated with behavior on social media, we randomly selected a sample of 37,326 users from our simulation and utilized the Twitter API to retrieve 200 of their most recent English tweets as well as their user account information. These tweets were selected based on the users' most recent activity at the calendar time of data collection in 2023. Only tweets and user information that was still publicly available were included in the sample. The final sample consisted of 35,668 users and a total of 7,114,377 tweets.

Using this sample, we recorded two types of variables: tweet-level and user-level variables (see Supplementary material for full list). Motivated by findings from study 1, which showed an

association between acrophily and stronger negative feelings toward the out-group, we investigated whether the acrophily coefficient correlates with the interaction of two aspects of social media communication that could be capturing this phenomenon: the expression of negative sentiments and the usage of third-person plural pronouns (e.g. "they," "them"). We selected the expression of negative sentiments as an indicator of animosity and third-person plural pronouns as a linguistic marker for referring to out-groups. By analyzing the interaction between these variables, we aimed to determine whether acrophily is related to the expression of negative feelings in general, or if it is also specifically associated with how these negative emotions are conveyed in discussions about out-groups.

To capture the sentiments expressed by users at an individual level, we considered the tweet information of the 200 most recent tweets and retweets of each user. We then extracted the sentiments from all tweets using VADER (42), which is a tool specifically developed for sentiment analysis in social media and is especially suited for short texts such as those posted on Twitter (51). For each tweet, VADER produces a continuous sentiment score ranging from – 1 (extremely negative) to +1 (extremely positive). We further used LIWC2015 (43) to count the frequency of third-person plural pronouns in users' tweets. To address the skewness of this distribution and mitigate the influence of outliers, we applied a logarithmic transformation to this variable.

To test the association between acrophily and out-group animosity at the individual level, we performed a linear regression model including the acrophily coefficient as the DV and the tweets' sentiments, the log-transformed frequency of using "they" terms, and their interaction as predictors. We also incorporated users' political affiliation strength and the total collected tweet count per user as control variables. As expected and consistent with the findings from study 1, our results indicate that participants displaying higher levels of acrophily produced both more negative content and also more content referring to others [sentiment main effect: $b = -0.08$, $t(35{,}662) = -4.52$, $P < 0.001$, 95% CIs = (−0.11, −0.04); third-person main effect: $b = 0.01$, $t(35{,}662) = 2.33$, $P = 0.02$, 95% CIs = (0.00, 0.02), interaction: $b = -0.14$, $t(35{,}662) = -3.43$, $P < 0.001$, $R^2 = 0.81$, 95% CIs = (−0.23, −0.06)]. Interestingly, the significant interaction suggests that individuals with higher levels of acrophily are more likely to express negative sentiments, particularly those who refer more to out-groups. Essentially, the negativity in content associated with acrophily becomes more pronounced when these individuals use language that references others, highlighting a targeted form of animosity toward out-groups.

General discussion

The goal of the current project was to examine the occurrence of acrophily—attraction to extremes—on social media. In study 1, we conducted an online preregistered experiment where participants evaluated either politically moderate or extreme user profiles and were asked whether they were likely to follow them. We found that participants preferred extreme copartisans over moderate ones. In study 2, we looked at retweet behavior of a large user sample on Twitter. We found that liberals tended to retweet users that were on average more extreme than pure homophily but less extreme than just acrophily. Conservatives' strategy, however, was even more extreme than our acrophily strategy.

Our findings of acrophily are congruent with and can be explained by previous literature that examined how people evaluate political extremes. Previous research suggests that extremes tend to be perceived as more representative of the political normative view (18), more committed to their party (24), more authentic (25, 26), and more coherent in their views, which seems to be attractive (23). We further tested some of these perceptions in study 1, reported in the Supplementary material, and provide some empirical support for these ideas. For example, we found that participants evaluated extreme profiles as more aligned with the mainstream view of their party supporters (strongest predictor of acrophily), more confident, interesting, and intelligent, and more likely to provide solid arguments in discussions. All of these seem to be contributing to acrophily.

The current project not only provides evidence for acrophily but is also the first to examine the association between acrophily and affective polarization. In study 1, we found that acrophily was associated with out-group animosity, measured by both reported feelings toward the opposite political party and one's own party. In study 2, we analyzed the content of users' tweets and found that users' tendency for acrophily was associated with both the expression of negative emotions as well as the use of third-person language. Furthermore, it was particularly associated with the co-occurrence at an individual level of expressing negative emotions and using third-person language, providing further evidence for out-group animosity. These findings provide further evidence of the connection between group identity and preference and affective polarization, a connection that only recently has received empirical support (38).

Our findings have far-reaching implications regarding the way segregation and polarization occurs on social media. Assuming that users' tie-selection decisions are driven by acrophily, such decisions lead to a much faster and more extreme segregation compared to segregation driven by homophily (18). Furthermore, if people are disproportionally likely to see extreme users, they are more likely to be exposed to extreme views and eventually to conform to them. In line with this idea, our results suggest that the tendency for acrophily is associated with out-group animosity, suggesting that together with increasing segregation, acrophily is either contributing to or may be a result of increased intergroup hostility. Future research should examine the causal relation between these two variables.

Limitations and future directions

Acrophily seems to be driving interactions on social media. However, the two studies leave many open questions regarding the nature of acrophily. First, opportunities to interact with users on social media are dictated by a mix of user preferences and algorithmic decisions. It is possible that social media algorithms render increased exposure to extreme peers, which in turn contributes to acrophily (31). Given that algorithmic decisions are opaque, it is impossible to know how much of the acrophily findings seen in the analysis was driven by algorithmic preferences. If indeed such algorithms show users more extreme views, it is likely that these views were preferred by them in previous engagements. Additionally, the role that automated accounts, i.e. bots, have in this phenomenon should be studied. Although the relative presence of bots in social media is low, it has been shown that they engage in echo-chamber-like behavior, and their contribution to acrophily remains unexplored (47). Furthermore, we conducted study 1 to specifically address such limitations. Participants in study 1 positively evaluated more politically extreme profiles. Despite the results of study 1, future projects should continue to explore these tendencies in a mix of social media and lab designs.

A second limitation of the current research is the scope of the acrophily findings. Although our analysis reveals a substantial tendency for acrophily in US Twitter data from 2020, it is unclear whether this tendency would be attenuated or increased in other contexts, social media platforms, and outside of social media in general. The US context in 2020 was quite unique, as two significant events coincided: the COVID-19 pandemic and the presidential elections. This unique context could have exacerbated political polarization and, therefore, acrophily. It is also possible that politically extreme users have unique value on Twitter, which is a platform focused on providing news updates, but such tendencies would be reduced in platforms where content sharing is more focused on personal experiences. It is also worth mentioning that users in our Twitter data specifically wrote tweets on political issues, which makes them more engaged in the topic than regular users. Future work should examine these tendencies in other contexts, on different platforms, and outside of digital media. Additionally, since acrophily has been found to co-occur with homophily, future research should explore the contexts or conditions under which one of the two is more likely to be dominant.

Finally, a third limitation of the current project is that it provided correlational evidence for the association between acrophily and out-group animosity, but it did not provide any causal evidence that one of these components is driving the other. Are participants attracted to politically extreme peers because these users satisfy their need for out-group animosity, or whether their attraction to extremes is driving out-group animosity. Future work should manipulate these variables to examine causal connections between them.

Despite these limitations, we believe that the current results point to an extremely important driver of segregation and polarization on social media. Further work is needed to understand acrophily, and to find ways to reduce acrophily, and therefore segregation. Since acrophily seems to occur partly due to a misperception of the norms related to one's political in-group, informing people about the true norms may help reduce acrophily. This can be achieved in several ways. Social media platforms could prioritize information that more accurately represents these norms rather than extreme content. However, this may be challenging as extreme views, thoughts, and feelings tend to generate more engagement, which impacts the platforms' bottom line (31, 52, 53). Other organizations interested in reducing such bias could assist users by helping them filter content to match their preferences and what is most beneficial for them (54, 55). A second way to reduce acrophily may be achieved by reducing affective polarization, as the two seem to be associated. Reducing affective polarization is an objective of various efforts (56), starting from improving spaces for political interactions (57) to using automatized ways such as large language models to facilitate constructive conversations (58).

Materials and methods

Study 1: Evidence for acrophily in reported intentions on social media

Participants

The study was approved by the Harvard University-Area Committee on the Use of Human Subjects before data collection (protocol number IRB23-1645). Based on the results of a pilot study, we recruited N = 400 US citizens who reported frequent use of Twitter through the Connect platform by Cloud Research. Our sample size was calculated to provide us with 90% power to detect our main hypothesis. The sample was balanced in terms of party affiliation and gender. Participants were presented with political in-groups, with assignments made based on the information provided to Connect, which was later confirmed in our study. Due to incorrect assignment by the platform, where Democrats were exposed to Republican profiles or vice versa, or participants who did not identify with either of the two main parties in our experiment, 12 participants were excluded. Therefore, the final sample consisted of 388 participants (age: mean 39.8 years, SD 12.2 years; 194 Democrats, 194 Republicans; 193 female, 190 male, 5 other). All participants successfully completed two attention checks throughout the study.

Procedure

The experimental design, hypotheses, and planned analyses were pre-registered at: https://aspredicted.org/8WL_7KT. Participants provided consent and then filled out a short online survey measuring their opinions about three putative Twitter profiles as well as their political identification and demographic information. Participants were paid $2.00 for participating in the 5-min study. Participants were presented with three hypothetical Twitter profiles. Each profile included a name, a picture, and a biography. Profiles' pictures were human faces created through the website of Generated Media Inc. (https://generated.photos/). The differences between neutral, moderate, and extreme users were modified in two ways. First, in the users' bios, which were modified to express either a moderate or strong political view, or no political view. Second, extreme partisans had a party's badge over their profile picture (a Democrat donkey or a Republican elephant; see Fig. 1A).

In each trial, participants were exposed to one neutral profile, one moderate in-group, and one extreme in-group profile in a random order (see Supplementary material for details). In a pilot study ($N = 213$; 106 Democrats, 98 Republicans, 9 others), we tested that participants' impressions of the profiles' party affiliations (strong Democrat, strong Republican, moderate Democrat, moderate Republican, independent/other/none) were consistent with our classifications. To achieve this, we coded each response as a numerical value ranging from −2 (indicating a strong Democrat) to 2 (indicating a strong Republican) and found a significant correlation between participants' perceptions and our classifications [Spearman's correlation: $r = 0.78$, $P < 0.001$, 95% CIs = (0.76, 0.81)]. Furthermore, participants were able to distinguish between extreme and moderate partisans (accuracy = 0.70, permutation test: $P < 0.001$). See Supplementary material for details. This validation was extended in study 1, where we further confirmed participants' perceptions of the profiles' party affiliations [$r = 0.80$, $P < 0.001$, 95% CIs = (0.78, 0.82); accuracy = 0.76, $P < 0.001$].

Measures

Follow back

Participants were asked whether they would follow back this person if this person were to follow them on Twitter, using two different methods. Initially, they responded in a binary yes/no format, and subsequently, they indicated how likely they would be to follow the person back on a six-point Likert scale ranging from "definitely not" to "definitely."

Interpersonal attraction

Participants completed an interpersonal attraction scale by indicating whether they agreed or disagreed with the following

statements, using a five-point Likert scale ranging from "I strongly disagree" to "I strongly agree": *I think he could be a friend of mine; he would perfectly fit into my circle of friends; if I wanted to get things done, I could probably depend on him*.

Impressions of profiles
Participants evaluated each profile by indicating whether they agreed or disagreed with the following statements, using a five-point Likert scale ranging from "I strongly disagree" to "I strongly agree": *This person is likely to be nice; this person is likely to be interesting; this person is likely to be intelligent; this person represents the mainstream view of supporters to his party; this person seems self-confident; this person would provide accurate and useful information; this person would provide solid arguments in a discussion; this person seems to be an entertaining user on social media*.

Out-group animosity and in-group positivity
Participants reported their feelings toward the Democratic party, the Republican party, liberal citizens, and conservative citizens on a five-point Likert scale, ranging from "very cold" to "very warm".

Profiles' party and political affiliation
Participants were asked to describe the profiles' party affiliation (multiple choices: strong Democrat, strong Republican, moderate Democrat, moderate Republican, independent, other, and none). Additionally, they were requested to position these profiles on a seven-point scale ranging from "strongly liberal" to "strongly conservative".

Demographics
Participants were asked about their age, gender, nationality, ethnicity, and which social networks they usually use.

Political identification
As they did with the profiles, participants were asked to describe their own party affiliation (strong Democrat, strong Republican, moderate Democrat, moderate Republican, independent, other, and none) and to position themselves on a seven-point scale ranging from "strongly liberal" to "strongly conservative".

Study 2: Observational evidence for acrophily in social media

Evaluating users' political affiliation

To evaluate users' political affiliation, we took a list of 90 news media outlets from the website www.allsides.com, which was at the time of the data collection (2019 May to 2020 December) the full list of all media outlets at the site. AllSides is a media company that attempts to assess the bias of media outlets. Using volunteer raters, supervised by the company staff members, AllSides provides a bias rating from one to five for each media outlet (left, leans left, center, leans right, and right). Of the list of 90 outlets, 22 were considered as "left", 18 as "leans left", 18 as "center", 8 as "leans right", and 23 as "right". Note that the number of outlets is not perfectly balanced across political groups. Nevertheless, we decided to still use the full list for two reasons. The first reason was that our main comparison was between actual and simulated data within each political group, rather than between political groups, and so we wanted to keep as many outlets in each group as possible to utilize the full dataset. The second reason was that it was not clear that a numerically balanced list would better reflect the actual media landscape at the time and might actually lead to loss of information. We estimated users' political affiliation by analyzing all the content that was retweeted by our users and assigned a political affiliation value to each retweeted message based on the assigned bias rating produced by AllSides. We then took an average of all the bias ratings of the retweeted media outlets for each user to determine the user's political affiliation. For example, if a user retweeted one media outlet of −2 (left) and one media outlet of −1 (leans left), then the assumed political affiliation of the user was −1.5. After getting all the political affiliation ratings for our participants, we rescaled the variable to range from −2 to 2, rather than 1 to 5. We also removed participants who were located at the middle of the scale on average (see "Data reduction") and converted liberal ratings from −2–0 to 0–2 so that both political groups would be on a scale of 0-moderate to 2-extreme.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This work was supported by the Templeton World Charity Foundation (Grant number TWCF-2022-31322).

Author Contributions

F.Z. and D.D.B.: formal analysis and writing—original draft; G.M. and J.S.: formal analysis; E.F.: data curation and methodology; R.W., E.H., and J.J.G.: writing—review and editing; J.N.: resources, funding acquisition, and investigation; A.G.: conceptualization, formal analysis, funding acquisition, writing—original draft, and writing—review and editing.

Preprints

This manuscript was posted on a preprint: https://osf.io/preprints/osf/cmx4p.

Data Availability

The data and code used for this study are available on OSF at: https://osf.io/jrzb3/.

References

1 Nicholson SP, Coe CM, Emory J, Song AV. 2016. The politics of beauty: the effects of partisan bias on physical attractiveness. *Polit Behav*. 38:883–898.
2 Boutyline A, Willer R. 2017. The social structure of political echo chambers: variation in ideological homophily in online networks. *Polit Psychol*. 38:551–569.
3 Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R. 2015. Tweeting from left to right: is online political communication more than an echo chamber? *Psychol Sci*. 26:1531–1542.
4 Brown JR, Enos RD. 2021. The measurement of partisan sorting for 180 million voters. *Nat Hum Behav*. 5:998–1008.
5 Motyl M, Prims JP, Iyer R. 2020. How ambient cues facilitate political segregation. *Pers Soc Psychol Bull*. 46:723–737.
6 Bishop B. 2009. *The big sort: why the clustering of like-minded America is tearing us apart*. Mariner Books.
7 Stein J, Keuschnigg M, van de Rijt A. 2023. Network segregation and the propagation of misinformation. *Sci Rep*. 13:917.

8 Iyengar S, Lelkes Y, Levendusky M, Malhotra N, Westwood SJ. 2018. The origins and consequences of affective polarization in the United States. *Annu Rev Polit Sci*. 22:1–35.
9 Dehghani M, *et al*. 2016. Purity homophily in social networks. *J Exp Psychol Gen*. 145:366–375.
10 Halberstam Y, Knight B. 2016. Homophily, group size, and the diffusion of political information in social networks: evidence from Twitter. *J Public Econ*. 143:73–88.
11 Kossinets G, Watts DJ. 2009. Origins of homophily in an evolving social network. *Am J Sociol*. 115:405–450.
12 McPherson M, Smith-Lovin L, Cook JM. 2001. Birds of a feather: homophily in social networks. *Annu Rev Sociol*. 27:415–444.
13 Diprete TA, Gelman A, Mccormick T, Teitler J, Zheng T. 2011. Segregation in social networks based on acquaintanceship and trust. *Am J Sociol*. 116:1234–1283.
14 Iyengar S, Sood G, Lelkes Y. 2012. Affect, not ideology: a social identity perspective on polarization. *Public Opin Q*. 76:405–431.
15 Bail CA, *et al*. 2018. Exposure to opposing views can increase political polarization. *Proc Natl Acad Sci U S A*. 115:9216–9221.
16 Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ. 2017. Emotion shapes the diffusion of moralized content in social networks. *Proc Natl Acad Sci U S A*. 114:7313–7318.
17 González-Bailón S, *et al*. 2023. Asymmetric ideological segregation in exposure to political news on Facebook. *Science*. 381: 392–398.
18 Goldenberg A, Abruzzo J, Willer R, Halperin E, Gross J. 2023. Homophily and acrophily as drivers of political segregation. *Nat Hum Behav*. 7:219–223.
19 Morrison KR, Miller DT. 2008. Distinguishing between silent and vocal minorities: not all deviants feel marginal. *J Pers Soc Psychol*. 94:871–882.
20 Abrams D, Bown N, Marques JM, Henson M. 2000. Pro-norm and anti-norm deviance within and between groups. *J Pers Soc Psychol*. 78:906–912.
21 Abrams D, Marques J, Bown N, Dougill M. 2002. Anti-norm and pro-norm deviance in the bank and on the campus: two experiments on subjective group dynamics. *Group Process Intergroup Relat*. 5:163–182.
22 Kulibert D, Moss A, Appleby J, O'brien L. 2021. Perceptions of political deviants: a lay theory of subjective group dynamics. PsyArXiv. https://doi.org/10.31234/osf.io/aq652, preprint: not peer reviewed
23 Zimmerman F, Garbulsky G, Ariely D, Sigman M, Navajas J. 2022. Political coherence and certainty as drivers of interpersonal liking over and above similarity. *Sci Adv*. 8(6):eabk1909.
24 Baldassarri D, Gelman A. 2008. Partisans without constraint: political polarization and trends in American public opinion. *Am J Sociol*. 114:408–446.
25 Rigoli F. 2023. Political extremism and a generalized propensity to discriminate among values. *Polit Psychol*. 44:301–318.
26 Klein N, Stavrova O. 2023. Respondents with more extreme views show moderation of opinions in multi-year surveys in the USA and the Netherlands. *Commun Psychol*. 1:37.
27 Goffman E. 1986. *Stigma: notes on the management of spoiled identity. Touchstone edition*. Simon & Schuster.
28 Leonardelli GJ, Pickett CL, Brewer MB. 2010. Optimal distinctiveness theory: a framework for social identity, social cognition, and intergroup relations. *Adv Exp Soc Psychol*. 43:63–113.
29 Goldberg A, Srivastava SB, Manian VG, Monroe W, Potts C. 2016. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. *Am Sociol Rev*. 81:1190–1222.
30 Zuckerman EW. 1999. The categorical imperative: securities analysts and the illegitimacy discount. *Am J Sociol*. 104:1398–1438.
31 Robertson C, Del Rosario K, Van Bavel JJ. 2024. Inside the funhouse mirror factory: how social media distorts perceptions of norms. PsyArXiv. https://doi.org/10.31234/osf.io/kgcrq, preprint: not peer reviewed.
32 Downs A. 1957. *An economic theory of democracy, Nachdr*. Addison Wesley.
33 Bischof D, Wagner M. 2019. Do voters polarize when radical parties enter parliament? *Am J Polit Sci*. 63:888–904.
34 Rabinowitz G, Macdonald SE. 1989. A directional theory of issue voting. *Am Polit Sci Rev*. 83:93–121.
35 Tomz M, Van Houweling RP. 2008. Candidate positioning and voter choice. *Am Polit Sci Rev*. 102:303–318.
36 Gallati L, Giger N. 2020. Proximity and directional voting: testing for the region of acceptability. *Elect Stud*. 64:102024.
37 Lacy D, Paolino P. 2010. Testing proximity versus directional voting using experiments. *Elect Stud*. 29:460–471.
38 Dias N, Lelkes Y. 2022. The nature of affective polarization: disentangling policy disagreement from partisan identity. *Am J Pol Sci*. 66:775–790.
39 Finkel EJ, *et al*. 2020. Political sectarianism in America. *Science*. 370:533–536.
40 Osmundsen M, Bor A, Vahlstrup PB, Bechmann A, Petersen MB. 2021. Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *Am Polit Sci Rev*. 115:999–1015.
41 Chen E, Deb A, Ferrara E. 2022. #Election2020: the first public Twitter dataset on the 2020 US Presidential election. *J Comput Soc Sci*. 5:1–18.
42 Hutto CJ, Gilbert E. 2014. VADER: a parsimonious rule-based model for sentiment analysis of social media text. *Proc Int AAAI Conf Web Soc Media*. 8:216–225.
43 Robinson RL, Navea R, Ickes W. 2013. Predicting final course performance from students' written self-introductions: a LIWC analysis. *J Lang Soc Psychol*. 32:469–479.
44 Yang J, Counts S. 2010. Predicting the speed, scale, and range of information diffusion in twitter. *Proc Int AAAI Conf Web Soc Media*. 4:355–358.
45 Verweij P. 2012. Twitter links between politicians and journalists. *J Pract*. 6:680–691.
46 Schroeder DT, Langguth J, Burchard L, Pogorelov K, Lind PG. 2022. The connectivity network underlying the German's Twittersphere: a testbed for investigating information spreading phenomena. *Sci Rep*. 12:4085.
47 Ferrara E, Chang H, Chen E, Muric G, Patel J. 2020. Characterizing social media manipulation in the 2020 U.S. presidential election. *First Monday*. 25(11). https://doi.org/10.5210/fm.v25i11.11431.
48 Bovet A, Makse HA. 2019. Influence of fake news in Twitter during the 2016 US Presidential election. *Nat Commun*. 10:7.
49 Badawy A, Lerman K, Ferrara E. 2019. Who falls for online political manipulation? Paper presented at: *Companion Proceedings of The 2019 World Wide Web Conference*; May 13–17, 2019, San Francisco, CA. ACM. p. 162–168.
50 Gomila R. 2021. Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *J Exp Psychol Gen*. 150:700–709.
51 Ribeiro FN, Araújo M, Gonçalves P, André Gonçalves M, Benevenuto F. 2016. SentiBench: a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Sci*. 5:23.

52 Schöne JP, Parkinson B, Goldenberg A. 2021. Negativity spreads more than positivity on Twitter after both positive and negative political situations. *Affec Sci*. 2:379–390.
53 Rathje S, Van Bavel JJ, van der Linden S. 2021. Out-group animosity drives engagement on social media. *Proc Natl Acad Sci U S A*. 118:e2024292118.
54 Jia C, Lam MS, Mai MC, Hancock JT, Bernstein MS. 2024. Embedding democratic values into social media AIs via societal objective functions. *Proc ACM Hum Comput Interact*. 8:1–36.
55 Kelly CA, Sharot T. 2023. Knowledge-seeking reflects and shapes well-being. PsyArXiv. https://doi.org/10.31234/osf.io/yd6j5, preprint: not peer reviewed.
56 Hartman R, *et al*. 2022. Interventions to reduce partisan animosity. *Nat Hum Behav*. 6:1194–1205.
57 Navajas J, *et al*. 2019. Reaching consensus in polarized moral debates. *Curr Biol*. 29:4124–4129.e6.
58 Argyle LP, *et al*. 2023. Leveraging AI for democratic discourse: chat interventions can improve online political conversations at scale. *Proc Natl Acad Sci U S A*. 120:e2311627120.