

Psychological Underpinnings of Partisan Bias in Tie Formation on Social Media

Mohsen Mosleh^{1, 2, 3}, Cameron Martel³, and David G. Rand^{3, 4, 5}

¹ Oxford Internet Institute, University of Oxford

² Department of Management, University of Exeter Business School

³ Sloan School of Management, Massachusetts Institute of Technology

⁴ Institute for Data, Systems, and Society, Massachusetts Institute of Technology

⁵ Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Individuals preferentially reciprocate connections with copartisans versus counter-partisans online. However, the mechanisms underlying this partisan bias remain unclear. Do individuals simply prefer viewing politically congenial content, or do they additionally prefer socially connecting with copartisans? Is this driven by preference for in-party ties or distaste for out-party ties? In a Twitter (now called X) field experiment, we created bot accounts varying by partisanship and whether they identified as bots or humans. We randomly assigned Twitter users ($N = 3,013$) to be followed by one of these accounts. We found evidence for social motivation—users were much more likely to reciprocate links to copartisan relative to counter-partisan accounts when the accounts identified as humans versus bots. We also found evidence for both in-party preference and out-party dispreference—users were as likely to follow back copartisan accounts as they were unlikely to followback counter-partisan accounts, compared to politically neutral accounts. A follow-up survey experiment ($N = 990$) provides further evidence for distinct roles of issue polarization, out-party animosity, and in-party affinity in moderating follow-back decisions online.

Public Significance Statement

We find that politically active Twitter (now called X) users preferentially follow back copartisan accounts because of social preferences for connecting with shared partisans—above and beyond a motivation to view congenial content. In-party preference and out-party dispreference both contribute to preferential political social tie formation. The psychological underpinnings of preferential copartisan follow-back are multifaceted: Increased issue polarization and (to a lesser extent) out-party dislike predict decreased follow back of counter-partisan and politically neutral accounts (but not copartisan accounts), whereas in-party liking predicts increased follow back of copartisan and politically neutral accounts (but not counter-partisan accounts). These findings illuminate how and why partisans form online social connections.

Keywords: partisanship, social media, polarization, field experiment, intergroup behavior

Supplemental materials: <https://doi.org/10.1037/xge0001662.supp>

This article was published Online First October 14, 2024.

Tara Mandalaywala served as action editor.

Mohsen Mosleh  <https://orcid.org/0000-0001-7313-5035>

Cameron Martel  <https://orcid.org/0000-0003-3181-4309>

David G. Rand  <https://orcid.org/0000-0001-8975-2783>

Mohsen Mosleh and Cameron Martel contributed equally to this article. Full preregistrations, experimental materials, data, and code can be accessed at https://osf.io/n7dym/?view_only=0fe35aeb3bcd4a718c8fd028ede85edc (Mosleh et al., 2023). A preprint of this article has been posted on PsyArXiv (<https://osf.io/preprints/psyarxiv/87bv2>). The ideas and data presented here were disseminated as a presentation at the Conference on Digital Experimentation at Massachusetts Institute of Technology (2023, November) and as a poster at the Polarization Workshop (2023, December).

Mohsen Mosleh and David G. Rand have received research funding from Google and Meta. The authors gratefully acknowledge funding from Google through the Google Research Scholar Award, gifts from

Google and Meta to support other research, National Science Foundation Awards 2047152 and 2217770, the John Templeton Foundation, the Alfred P. Sloan Foundation Grant 2021-16891, and funding via the National Science Foundation Graduate Research Fellowship under Grant 174530.

Mohsen Mosleh played an equal role in conceptualization, formal analysis, investigation, methodology, validation, visualization, writing—original draft, and writing—review and editing. Cameron Martel played an equal role in conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, validation, visualization, writing—original draft, and writing—review and editing. David G. Rand played an equal role in conceptualization, funding acquisition, methodology, writing—original draft, and writing—review and editing.

Correspondence concerning this article should be addressed to Mohsen Mosleh, Oxford Internet Institute, University of Oxford, 1 St Giles', Oxford OX1 3JS, United Kingdom. Email: mohsen.mosleh@oii.ox.ac.uk

There is an abundance of observational evidence showing that Americans are more likely to be connected to copartisans than counter-partisans on social media (Adamic & Glance, 2005; Bakshy et al., 2015; Barberá et al., 2015; Colleoni et al., 2014; Conover et al., 2011). This may be attributable to partisanship being highly correlated with other factors predicting social tie formation, such as similar interests (Aiello et al., 2012; DellaPosta et al., 2015), or could be due to an actual causal preference to associate with copartisans rather than counter-partisans based on partisanship per se (Huber & Malhotra, 2017). In support of the latter, recent field experiments on Twitter (now called X) have demonstrated that there are indeed strong causal effects of shared partisanship on the formation and prevention of online social ties—politically active Twitter users were substantially more likely to reciprocally follow back copartisans compared to counter-partisans (Ajzenman et al., 2023b; Mosleh et al., 2021), and were more likely to block counter-partisans than copartisans (Ajzenman et al., 2023b; Martel et al., 2024).

However, the psychology underlying the preference for reciprocating ties with copartisans but not counter-partisans remains unclear. To what extent is the tendency to preferentially follow back copartisans driven by content versus social preferences? The former content preference theory posits that people engage in selective exposure of political information, such that individuals prefer to see politically concordant content and avoid politically discordant content (Stroud, 2008). Such a motivation may underscore observational evidence that social media users, and more extreme partisans in particular, are more likely to view politically congruent news content (Eady, Nagler, Guess, et al., 2019; Flaxman et al., 2016). Thus, if Twitter users are making follow-back decisions primarily to help cultivate politically congenial news feeds, then selective content exposure preferences may be the driving force behind preferential copartisan social tie reciprocation.

In contrast, the latter social preference theory posits that follow-back decisions are more driven by social motivations and affective preferences to associate with copartisans and not counter-partisans. Indeed, prior experimental work has illustrated political homophily in noninformational contexts such as online dating behavior (Huber & Malhotra, 2017), and a large body of literature on affective polarization indicates that Americans increasingly dislike and distrust counter-partisans (Iyengar et al., 2019). Such social preferences may also augment the value of informational content received online. Information can include both contents (e.g., what is said) and context (e.g., who says it; for another example of information context, see contextual integrity; Nissenbaum, 2004), and these features have separable roles in predicting online behavior such as reposting, commenting, and online reviewing (Rudat & Buder, 2015; Siering et al., 2018; Zhang et al., 2014). Social preferences may inform follow-back decisions through a route of contextual information preference—individuals may selectively reciprocate ties with copartisans because they want information from copartisan peers. Thus, preferential follow-back of copartisan accounts may not be driven solely by a preference for congenial content per se, but by an affective or contextual social preference for connecting with actual fellow partisans.

If social preferences help underscore preferential copartisan social tie formation, a second important psychological question arises: To what extent are these social preferences reflective of in-party preference versus out-party dispreference? That is, to what extent do politically active Twitter users *prefer* forming connections with *copartisans* versus *disprefer* forming connections with *counter-partisans*? Social

psychology research has demonstrated that in-group love and out-group hate can operate independently, and some evidence suggests that in-group love often plays a relatively larger role in prejudicial behavior (Brewer, 1999). Likewise, recent evidence suggests that partisans are more likely to help ingroup members than harm outgroup members under low threat conditions (Amira et al., 2021), and even American politicians are more likely to post positive tweets toward their in-party than negative tweets toward their out-party (Yu et al., 2024). Thus, one route through which preferential copartisan follow-back emerges may be via greater preference for in-party member social ties. Alternatively, increasing out-party hostility in American partisans has been documented over the past several decades and may be a greater motivation for political participation than in-party preference (Iyengar & Krupenkin, 2018). Convergently, the “repulsion hypothesis” proposes that attitudinal similarity does not lead to liking, but rather that dissimilarity leads to disliking (Rosenbaum, 1986). Together, another route through which preferential copartisan follow-back may emerge is instead via greater dispreference for out-party social reciprocation.

Here we shed light on these questions via a field experiment and follow-up online survey experiment. We first conducted a field experiment on Twitter (for a brief review of this approach, see Mosleh et al., 2022) in which we created three similar explicit bot-looking accounts and three similar human-looking accounts, varying only in their expressed partisan identities (Democratic, Republican, or politically neutral). We then randomized a politically balanced (half Democratic, half Republican) set of politically active Twitter users ($N = 3,013$) to be followed by one of our bot accounts and assessed whether each user reciprocated our accounts’ social tie formation by following back our account. Our field experiment addresses our first two questions of interest. First, by comparing copartisan follow-back rates of our bot-looking and human-looking accounts, we can determine whether users appear to follow back accounts based only on their likely informational content or for an additional social contextual motivation. Second, by comparing follow-back rates for the neutral account relative to the copartisan and counter-partisan accounts, we can investigate the relative contributions of in-party preference (preferring the copartisan accounts relative to the neutral accounts) and out-party dispreference (preferring the neutral accounts over the counter-partisan accounts).

We next conducted a follow-up survey experiment on Lucid ($N = 990$) to further examine the potential moderators of copartisan tie preference and counter-partisan tie dispreference. Participants completed a series of political measures including issue position questionnaires, out-party disliking, in-party liking, and political knowledge. Participants were then randomly assigned to suppose that either a copartisan, counter-partisan, or politically neutral account on Twitter had followed them and were asked whether they would follow back this user. Our survey experiment then allows us to assess our third question of interest—which political preferences are associated with increased preference for copartisan reciprocation versus decreased preference for counter-partisan reciprocation?

Study 1: Field Experiment

Method

Participants

We first identified a politically balanced set of Twitter users who retweeted recent posts from either mainstream right-leaning

(Fox News) or left-leaning (MSNBC) news outlets. To do this, we used the Twitter API to pull the most recent primary tweets from the Fox News and MSNBC Twitter accounts. We then retrieved the user handles for those who had retweeted these mainstream media posts. This gave just an initial sample of 43,663 users. We excluded users who had more than 15,000 followers (i.e., users who already receive high engagement and are thus less likely to notice our intervention) and users who had fewer than 10 followed accounts (i.e., inactive users who are unlikely to follow back any other accounts). This filtering criteria resulted in 42,272 users eligible for our experiment. From each news outlet's retweeting list, we then randomly sampled 3,000 users. Next, we received up to the 3,200 most recent (re)tweets for each of these users and classified their partisanship based on the links they shared from left- versus right-leaning websites (as per Eady, Nagler, Bonneau, et al., 2019). Finally, we randomly sampled 2,000 users from each party to have a politically balanced total sample of 4,000 users as per our preregistration.

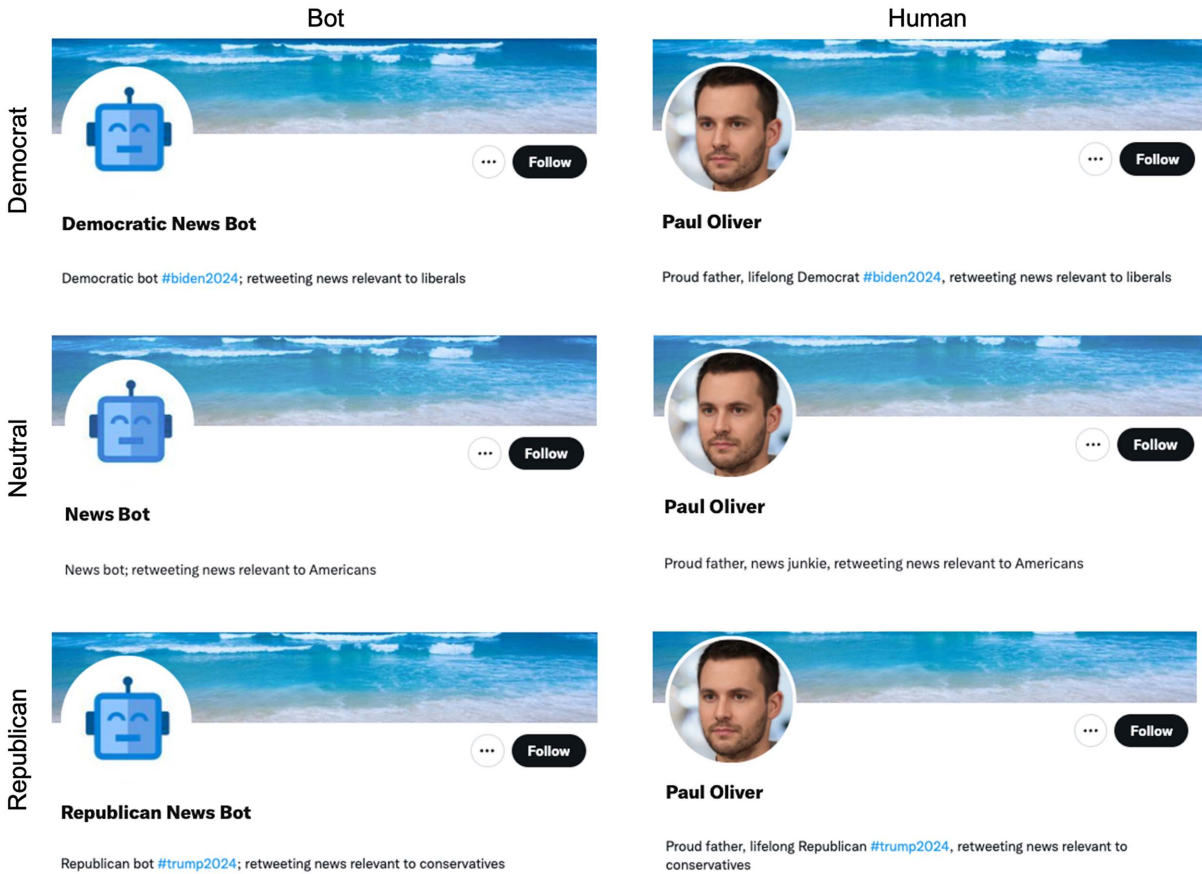
We then used randomization by stratification (i.e., blocking, see Higgins et al., 2016) to assign users with similar characteristics across

experimental conditions. We constructed approximately homogeneous strata of users based on (a) number of days with tweets in the past 2 weeks (as a proxy for recent account activity level), (b) logarithmic-transformed number of followers (as a proxy for account status; we use logarithmic-transformation because follower count is highly skewed), (c) baseline follow-back ratio as measured by the number of accounts with whom the user had reciprocal links, divided by the user's number of followers, (d) estimated user partisanship (Democrat vs. Republican), and (e) partisanship extremity of the user (i.e., absolute value of the continuous partisanship measure from Eady, Nagler, Bonneau, et al., 2019). Additionally, we randomly assigned strata of users to various treatment days, such that we had an equal number of users across experimental conditions for each day of the study.

Procedure

We created three nearly identical explicit bot-looking accounts and three nearly identical human-looking accounts, varying only in their expressed partisan identities (Democratic, Republican, or politically neutral; see Figure 1). We pretested our account profiles

Figure 1
Design of Field Experiment Accounts



Note. We created three explicit bot-looking and three human-looking bot accounts using generative adversarial network-generated pictures. These accounts varied on political partisanship (Democratic, politically neutral, Republican). The bot accounts followed a set of news outlets matching their political partisanship and retweeted randomly from these accounts each day of the experiment (see the Method section for more details). Account profiles were pretested for human versus bot perception, as well as anticipated informativeness/usefulness of tweeted content (see Supplemental Section S1a and S1b). See the online article for the color version of this figure.

to verify that the human-looking accounts were perceived as more human than the bot-looking accounts, and were comparably as human as actual Twitter profiles (see [Supplemental Section S1a and S1b](#) for full pretest analyses). In our initial pretest ($N = 375$; Prolific), we found that our bot-looking accounts were perceived as more likely to be a bot than our human-looking accounts ($p < .001$; [Supplemental Table S1](#)). We did not find evidence that our human-looking accounts were perceived as more or less likely to be a bot compared to several real Twitter profiles or actual conspiratorial Twitter profiles ($ps > .098$).

Importantly, we also pretested the anticipated informativeness of bot versus human accounts. We did so to ascertain that followed users perceived both our explicit bot-looking accounts and human-looking accounts as similarly useful or informative in the quality of content they shared. Conditional on such similar perceived informativeness, we could then more confidently attribute any differences in follow-back rates of bot versus human-looking accounts to social factors beyond content informativeness. Indeed, in our second pretest ($N = 367$; Prolific), we did not find evidence that our human-looking accounts were perceived as any more or less informative/useful than our bot-looking accounts ($p = .709$; [Supplemental Table S2](#)).

Finally, as a verification of our account partisanship manipulation, we tested and found significant evidence that our partisan accounts were perceived as more politically slanted than our politically neutral accounts ($p < .001$; [Supplemental Table S3](#)).

To further increase the credibility of our bot accounts, we initiated each account with approximately 250 politically neutral followers and retweeted a post from a mainstream outlet aligned with their political identification every day. The Democratic accounts retweeted from “MSNBC,” “washingtonpost,” “NBCNews,” and “TheAtlantic”; the neutral accounts retweeted from “Reuters,” “nprnews,” “BBCWorld,” and “AP”; the Republican accounts retweeted from “FoxNews,” “thedispatch,” “NRO,” and “amconmag” Twitter handles. Additionally, the Democratic accounts pinned a tweet from “MSNBC”; the neutral accounts pinned a tweet from “AP”; and the Republican accounts pinned a tweet from “FoxNews.”

To avoid spillover across conditions in the field, each bot account blocked all users assigned to other bot conditions. This prevents our accounts in one condition from being exposed to or suggested by the Twitter recommendation algorithm to users in other conditions. All users were assigned to only one condition, and thus followed by one of our six bot accounts.

We then followed users over 14 days (from January 24, 2023 to February 6, 2023). Each time a user followed back one of our accounts, we recorded and removed that connection. We did so to maintain a constant number of followers for our accounts (i.e., to avoid the possibility that follow-back probability from subsequent users could be affected by the number of followers our accounts had at a given time), and to avoid mutual connections with other users (which could also potentially affect probability of follow-back rates through triadic closure; see [Mosleh et al., 2024](#)).

We successfully followed $N = 3,013$ Twitter users ($Mdn = 176$ followers, $Mdn = 403$ followed accounts, 10,985 total [re]tweets, 52% Republican; to approximate demographic information, we used user profile pictures to estimate users’ perceived age and sex; see [An & Weber, 2016](#); [Chakraborty et al., 2017](#); [Kteily et al., 2019](#) and inferred this information for 61% of users—of those, 42% were inferred to be female, 58% were inferred to be male, and median inferred age was 47). We initially planned to follow 4,000

users—however, our ability to follow more people was limited by Twitter after we followed 3,013 users (504 followed by copartisan, human-looking account; 504 followed by co-partisan, bot-looking account; 499 followed by counter-partisan, human-looking account; 497 followed by counter-partisan, bot-looking account; 498 followed by neutral, human-looking account; and 511 followed by neutral, bot-looking account).

Transparency and Openness

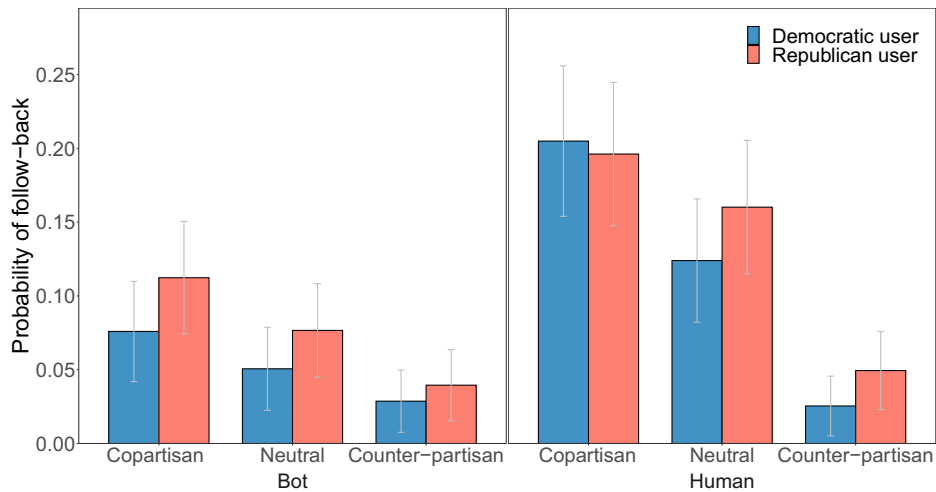
We received ethical approval and obtained a waiver for informed consent from the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects Protocol No. 1907910465. Our field experiment procedures and analyses were preregistered at https://aspredicted.org/blind.php?x=5H6_HBH. Data and analysis code necessary for reproducing our results are available at https://osf.io/n7dym/?view_only=0fe35aeb3bcd4a718c8fd028ede85edc.

Results

Follow-back rates across conditions are shown in [Figure 2](#). We began by investigating whether preferential tie formation with copartisans relative to counter-partisans is driven solely by an interest in receiving politically aligned content from another account (i.e., content motivation), or additionally by a preference for connecting with, or receiving content from, a fellow copartisan peer (i.e., social motivation). To do so, we ran a linear model predicting whether the user followed back our account, using a dummy variable for co-partisanship and a dummy variable for whether the account was politically neutral (i.e., using counter-partisan as the holdout), a dummy variable for explicit bot condition, dummy variable for user Republican partisanship (calculated via [Eady, Nagler, Bonneau, et al., 2019](#)), and all interactions between experimental condition dummies and user partisanship.

Based on prior work, we expected that users in our sample will be significantly more likely to reciprocate followership with copartisan accounts compared to counter-partisan accounts—and this is what we found, $b = 0.179$, 95% CI [0.128, 0.231], $SE = .026$, $t(3001) = 6.81$, $p < .001$; see [Supplemental Table S4](#).

To evaluate the relative importance of social versus content motivations, we asked whether (and to what extent) this preferential follow-back is weaker when our account appeared to be a bot (and thus there are no social motivations) versus a human. Indeed, we did find a highly significant interaction between whether our account looks like a bot, and whether it is a copartisan, $b = -0.132$, 95% CI [-0.205, -0.059], $SE = .037$, $t(3001) = -3.55$, $p < .001$. Specifically, although users preferentially followed-back copartisans over counter-partisans in both cases, this effect was much larger for the human-looking accounts, $b = 0.179$, 95% CI [0.121, 0.238], $SE = .030$, $t(1495) = 6.02$, $p < .001$ relative to the explicit bot accounts, $b = 0.047$, 95% CI [0.003, 0.091], $SE = .022$, $t(1506) = 2.12$, $p = .035$. This interaction is also evident when assessed in relative terms rather than just absolute differences (given that bot-looking accounts were followed back less overall than human-looking accounts). We conducted a logistic regression for robustness, and again found a significant interaction between bot appearance and co-partisanship on follow-back, such that the effect of co-partisanship on follow-back rate is greater for human-looking than bot-looking accounts, $b = -1.267$,

Figure 2*Probability of Follow-Back by Partisanship and Human-Looking Versus Explicit Bot Accounts*

Note. Users are less likely to reciprocate links with copartisan and neutral explicit bots compared to copartisan and neutral human-looking accounts, respectively. The magnitude of the effect of copartisanship and counter-partisanship are almost the same, relative to the neutral condition follow-back probability. Shown here is the probability of Democratic and Republican users following back our accounts in each experimental condition. Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

CI $[-2.547, -0.019]$, $SE = .635$, $t(3001) = -1.995$, $p = .046$; see [Supplemental Table S5](#). Furthermore, there were no significant differences between Democrats and Republicans in any of these effects ($p > .258$ for all interaction terms involving user partisanship).

These results suggest that although preferential follow-back of co-partisans is to some extent driven by a congenial content motive (as indicated by the significant co-partisan effect for the explicit bot accounts), much of this effect is about being motivated to connect with another *person* who shares their partisanship.

Given these findings, we next investigated the nature of the social motivations that appear to be driving preferential follow-back. To what extent is preferential follow-back of copartisans about (positive) preference for reciprocating connections from copartisans, versus (negative) distaste for reciprocating connections from counter-partisans?

To address this question, we compared the follow-back probabilities for copartisan and counter-partisan accounts with our neutral control condition. We used a linear probability model to predict whether a user follows back our account, using a dummy variable for copartisanship and a dummy variable for counter-partisanship (i.e., using the neutral control as the holdout), as well as a dummy variable for the explicit bot condition, a dummy variable for user Republican partisanship, and all interactions.

Comparing the shared partisanship conditions to the neutral condition, we found evidence of both preference for reciprocating ingroup ties and preference for *not* reciprocating outgroup ties. Follow-back probability was significantly higher in the copartisanship condition compared to the neutral condition, $b = 0.081$, 95% CI $[0.030, 0.132]$, $SE = .026$, $t(3001) = 3.09$, $p = .002$; and follow-back probability was significantly lower in the counter-partisanship condition compared to the neutral condition, $b = -0.099$, 95% CI $[-0.150, -0.047]$, $SE = .026$, $t(3001) = -3.73$, $p < .001$.

Furthermore, the absolute magnitude of these two effects was not significantly different, as determined by a Wald test comparing co-partisan and counter-partisan dummy variables: $F(1) = 0.97$, $p = .32$. Thus, we did not find evidence for a difference in the size of in-party reciprocation preference and out-party reciprocation dispreference.

We also did not find evidence of partisan asymmetries between Democratic and Republican users in these effects, interaction between co-partisanship and user partisanship: $b = -0.045$, 95% CI $[-0.117, 0.027]$, $SE = .036$, $t(3001) = -1.232$, $p = .218$; interaction between counter-partisanship and user partisanship: $b = -0.012$, CI $[-0.084, 0.060]$, $SE = .037$, $t(3001) = -0.333$, $p = .739$.

Altogether, then, our field experiment showed that social media users prefer co-partisan social tie formation not solely because of a preference for congenial political content, but also because of a social preference for connecting with copartisan *humans*. This social motivation is equal parts in-party preference and out-party dispreference.

Study 2: Survey Experiment

Our field experiment provided evidence that Twitter users engage in preferential co-partisan social tie formation because of social considerations, above and beyond purely content-based motivations. Relative to our politically neutral control account, we also found similar levels of copartisan preference and counter-partisan dispreference. However, it is unclear from our field results what particular social motives may be associated with these (dis)preferences. Specific political individual differences may be particularly predictive of following-back copartisans or declining to reciprocate ties with counter-partisans, respectively. In a follow-up survey experiment, we examined how four political covariates—issue polarization, out-party

disliking, in-party liking, and political knowledge—may be associated with distinct routes of preferential tie reciprocating behavior.

Method

Participants

We recruited 990 participants from Lucid ($M_{\text{age}} = 47.14$; 513 female participants, 475 male participants, two nonbinary participants or respondents who preferred not to answer; for full descriptive statistics and information on reporting of sex and race, see [Supplemental Table S7](#)). To be eligible for the survey, participants were required to report having a Twitter account and must have correctly answered two trivial attention screeners (e.g., captcha).

Procedure

Participants first completed Twitter usage and attention check screeners. Participants next answered basic demographics questions (e.g., age, sex, race). In random order, participants next completed the following question blocks: (a) political orientation (e.g., partisanship) and out-party dislike versus in-party like feelings thermometer (“How would you rate [Republican Party/Democratic Party] voters?” [0 = *very cold*, 100 = *very warm*]; adapted from [Weisberg & Rusk, 1970](#)), (b) issue position polarization items (11 items; health insurance, government job assurance, government services provision; government support for Black Americans; military spending; abortion; assault rifle ban; gay marriage; immigration criminality; Affordable Care Act repeal; investment in environmentally friendly technology; measures adapted from [American National Election Studies, 2021](#); [Berinsky et al., 2021](#); items typically scaled 1 = *liberal response*, 7 = *conservative response*, 8 = “*don’t know, haven’t thought much about this*” [recorded as not applicable, as preregistered]; abortion item rescaled independently), (c) political knowledge (four multiple choice items, e.g., “Whose responsibility is it to decide if a law is constitutional or not?”; adapted from [Tappin et al., 2021](#)). Participants next completed the cognitive reflection test ([Frederick, 2005](#)) and several questions on their usage of Twitter (e.g., “How often are you on Twitter?” “Which of these accounts have you retweeted in the past [if any]?”).

For the main survey task, participants were asked to suppose they were currently on Twitter. Participants were then asked to suppose that they were followed on Twitter by a below account. Participants were randomly assigned to be followed by one of three accounts—a Democratic, pro-Biden account (bio: “Proud father, lifelong Democrat #biden2024, liberal news junkie”; note that at the time of the experiment, Biden was the presumptive Democratic Party presidential nominee), a Republican, pro-Trump account (bio: Proud father, lifelong Republican #trump2024, conservative news junkie), or a politically neutral account (bio: Proud father, lifelong sports fan, news junkie). Each condition had 330 participants. All accounts had the same profile name, handle, and picture (Thomas Maddocks, @TMaddocks, White, male-presenting generative adversarial network-generated picture). Follower notification was presented as a mock Twitter “Notifications” page showing that Thomas Maddocks had followed the participant. All participants were then asked, “How would you respond to this notification that Thomas Maddocks follows you?” (1 = *follow this user back*, 0 = *ignore this notification*). Participants were then asked to write several sentences about why they either decided to follow back or

ignore the user that followed them. Finally, participants were asked to select the most important reason why they decided to either follow the user back or ignore the user (depending on what they initially answered). Full survey materials and stimuli are available at https://osf.io/n7dym/?view_only=0fe35aeb3bcd4a718c8fd028ede85edc.

Transparency and Openness

We obtained an exempt evaluation from the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects Protocol No. E-4591. Our survey experiment procedures and analyses were preregistered at https://aspredicted.org/blind.php?x=72D_YGP. Data and analysis code necessary for reproducing our results are available at https://osf.io/n7dym/?view_only=0fe35aeb3bcd4a718c8fd028ede85edc.

Results

We used our survey experiment to shed more light on the psychology underlying the observed copartisan preference and counter-partisan dispreference. We began by examining the follow-back probabilities for copartisan, neutral, and counter-partisan accounts in our survey experiment. Using a linear regression predicting follow-back using a copartisanship dummy and counter-partisanship dummy (i.e., neutral as the holdout), we found that participants were less likely to follow back a counter-partisan account relative to the neutral control, $b = -0.205$, $SE = .037$, $t(987) = -5.545$, $p < .001$; [Supplemental Table S10](#), and nominally, though not significantly, more likely to follow-back a copartisan account compared to the neutral control, $b = 0.062$, $SE = .037$, $t(987) = 1.673$, $p = .095$. Unlike in our field experiment results, the magnitude of these effects was significantly different, such that out-party distaste was stronger than ingroup preference (Linear hypothesis test: $b = -0.143$, $p = .026$; [Supplemental Table S11](#)).

One possible explanation for this difference from our field experiment is that our field experiment only assessed users who had previously retweeted from Fox News or MSNBC, whereas the only inclusion criterion for the survey experiment was having a (self-reported) Twitter account. Thus, the average user in the field experiment was likely more politically engaged than the average participant in the survey experiment. Indeed, when restricting our analyses for the survey experiment to participants who self-reported retweeting from these accounts in the past, the results are similar to the field experiment: we did not find a significant difference in the magnitude of copartisan follow-back, $b = 0.138$, $SE = .061$, $t(355) = 2.286$, $p = .023$; [Supplemental Table S27](#) and counter-partisan nonfollow-back, $b = -0.201$, $SE = .062$, $t(355) = -3.265$, $p = .001$; linear hypothesis test comparing the two coefficients: $b = -0.063$, $p = .550$; [Supplemental Table S28](#).

To dig deeper we then examined whether the effect of shared partisanship on follow-back probability is moderated by four key political covariates: issue polarization (11-item political issues measure; polarization measured as absolute value of difference from center scale position), out-party disliking, in-party liking (both measured via 0–100 feelings thermometers), and political knowledge (four multiple-choice questions).

For each moderator, we conducted a linear regression predicting follow-back, using a copartisanship dummy, counter-partisanship dummy (i.e., using neutral control as the holdout), the moderator,

and interactions between the treatment dummies and the moderator. We also examined a model including all four moderators in the same model, allowing for interactions between each of our shared partisanship dummies and each of our moderators.

Participants with greater issue polarization were less likely to follow-back accounts in the neutral condition, $b = -0.097$, $SE = .026$, $t(984) = -3.77$, $p < .001$; see [Supplemental Table S16](#). Issue polarization did not moderate the effect of counter-partisanship, counter-partisan and issue polarization interaction: $b = -0.016$, $SE = .036$, $t(984) = -0.446$, $p = .655$. However, we found that relative to the neutral condition, more issue-polarized participants were relatively more likely to follow back copartisan accounts, co-partisan and issue polarization interaction: $b = 0.118$, $SE = .037$, $t(984) = 3.203$, $p = .001$.

We found highly similar results when examining the moderation of shared partisanship follow-back rates by out-party dislike. Participants with greater out-party animosity were less likely to follow back accounts in the neutral condition, $b = -0.126$, $SE = .025$, $t(984) = -5.099$, $p < .001$; see [Supplemental Table S17](#). Out-party animosity did not moderate the effect of counter-partisanship, counter-partisan and out-party animosity interaction: $b = -0.033$, $SE = .035$, $t(984) = -0.942$, $p = .346$. Interestingly however, we found that relative to the neutral condition, participants with greater out-party dislike were marginally more likely to follow back

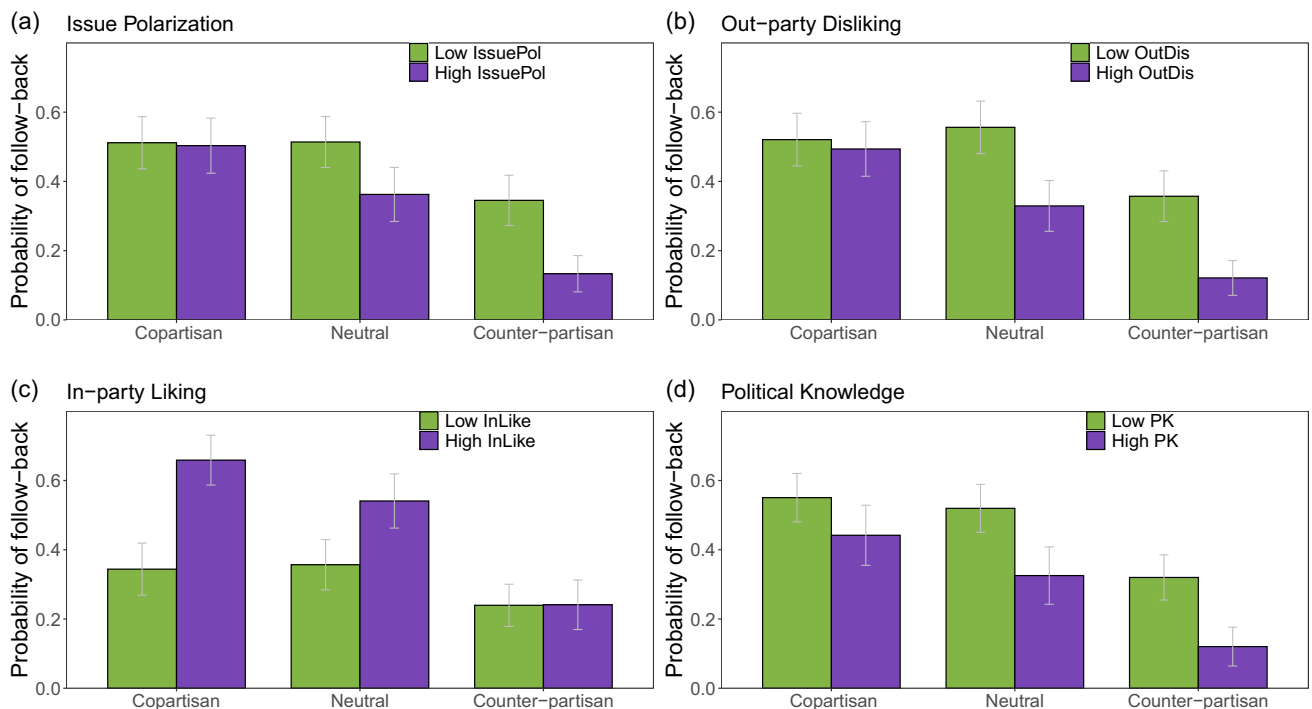
copartisan accounts, co-partisan and out-party animosity interaction: $b = 0.068$, $SE = .036$, $t(984) = 1.90$, $p = .058$. Combined, these findings were such that participants with greater issue polarization and (to a lesser extent) out-party animosity followed back co-partisans at about the same rate, but were relatively less likely to reciprocate followership of neutral and counter-partisan accounts (see [Figure 3a](#) and [3b](#)).

In contrast, participants with greater in-party liking were more likely to follow back accounts in the neutral condition, $b = 0.110$, $SE = .026$, $t(984) = 4.19$, $p < .001$; see [Supplemental Table S18](#)—perhaps due to being overall more agreeable or sociable relative to people lower in in-party liking. This was true to a similar extent in the copartisan condition, co-partisan and in-party liking interaction: $b = 0.053$, $SE = .036$, $t(984) = 1.472$, $p = .141$. Examining the counter-partisan condition, however, we found that relative to the neutral condition, participants higher in in-party liking were less likely to follow back counter-partisan accounts, counter-partisan and in-party liking interaction: $b = -0.099$, $SE = .037$, $t(984) = -2.696$, $p = .007$. This suggests that individuals with greater in-party liking were more likely to follow back co-partisan and neutral accounts, but no more or less likely to follow-back counter-partisan accounts (see [Figure 3c](#)).

We also conducted comparable analyses examining the potential moderation role of political knowledge. Participants higher in

Figure 3

Probability of Social Tie Reciprocation by Political Covariate Moderator



Note. Shown here are the probabilities of participants following back our accounts in each experimental condition. (a) Median split by issue polarization. Participants with greater issue polarization are less likely to follow back neutral and counter-partisan accounts, relative to copartisan accounts. (b) Median split by out-party disliking. Participants with greater out-party animosity are marginally less likely to follow back neutral and counter-partisan accounts, relative to copartisan accounts. (c) Median split by in-party liking. Participants with greater in-party affinity are more likely to follow back neutral and copartisan accounts, relative to counter-partisan accounts. (d) Median split by political knowledge. We do not find evidence that political knowledge significantly moderates follow-back rate by shared partisanship. Error bars indicate 95% confidence intervals. See the online article for the color version of this figure.

political knowledge were less likely to follow-back the neutral condition account, $b = -0.081$, $SE = .026$, $t(984) = -3.145$, $p = .002$; see [Supplemental Table S15](#)—however, we did not find evidence that political knowledge significantly moderates the follow-back rate or copartisan or counter-partisan accounts relative to neutral accounts ($ps > .352$; [Figure 3d](#)).

In our model including all four political variables, we again found significant evidence that, relative to the neutral condition, participants higher in in-party liking were less likely to follow-back counter-partisan accounts (interaction: $b = -0.092$, $SE = .035$, $t(975) = -2.636$, $p = .009$; [Supplemental Table S14](#)). We also again found significant evidence that relative to the neutral condition, more issue-polarized participants were more likely to follow back copartisan accounts, $b = 0.098$, $SE = .041$, $t(975) = 2.382$, $p = .017$. We did not find significant evidence of moderation by out-party disliking or political knowledge.

As a final preregistered assessment, we conducted a principal component analysis on our four political covariates. This yielded two components of interest—a first component we summarize as “overall polarization” (high loadings for issue polarization, out-party dislike, and political knowledge) and a second component of “in-party liking” (high loading for in-party like; see [Supplemental Table S8](#) for variable loadings). In line with our models examining individual political attributes, we found that relative to the neutral condition, more “overall polarized” participants were relatively more likely to follow back copartisan accounts, co-partisan and overall polarization interaction: $b = 0.077$, $SE = .026$, $t(981) = 2.931$, $p = .003$; see [Supplemental Table S12](#), following the pattern observed for issue polarization and (marginally) out-party animosity as moderators. Likewise, relative to the neutral condition, participants higher in the principal component “in-party liking” were less likely to follow back counter-partisan accounts, counter-partisan and in-party liking interaction: $b = -0.085$, $SE = .035$, $t(981) = -2.418$, $p = .016$; see [Supplemental Figure S1](#).

Free Response Text Analysis

As exploratory analyses, we also examined what free response motivations participants provided as to why they either followed back or ignored accounts across our experimental conditions. First, we used generative pre-trained transformer 4 to filter for free response answers that were longer than only a few words and considered “coherent”—this yielded a total of 515 coherent responses about why participants followed back or ignored accounts in our experiment (see [Supplemental Section S2h](#)). Then, we assessed whether the motivations participants mentioned in their responses matched those expected from our experimental manipulation—namely, if participants discussed copartisanship as a reason for following back concordant accounts and counter-partisanship as a reason for ignoring discordant accounts. To do so, we prompted generative pre-trained transformer 4 to indicate whether each participant’s response mentioned the account belonging to either a different party or their same party (for prompt, see [Supplemental Section S2h](#)). As expected, we found that 50% of participants in the copartisan condition who followed back the copartisan account mentioned shared partisanship as a motivation for following back the account ([Figure 4a](#)). And likewise, we found that 58% of participants in the counter-partisan condition who ignored the counter-partisan

account mentioned different partisanship as a motivation for ignoring the account ([Figure 4b](#)).

Next, we examined other possible motivations for following back or ignoring accounts that participants may have mentioned in their free response explanations. Based on previous work examining motivations for following and blocking others online ([Martel et al., 2024](#)), we provided generative pre-trained transformer 4 a list of potential motivations with which to classify free text responses (see [Supplemental Section S2h](#)). Participants who followed back accounts were most likely to mention motivations of wanting to see information the accounts posted (76% in copartisan condition, 62% in neutral condition, 58% in counter-partisan condition) and making friends (54% in copartisan condition, 55% in neutral condition, 35% in counter-partisan condition) in their free response explanations. Interestingly, participants were also likely to mention curiosity as a motivation for following back—particularly among those who followed back counter-partisan accounts (66%). This provided some suggestive evidence of curiosity as a motivation for reciprocating cross-party ties. Participants who ignored accounts were most likely to mention not wanting to see information the accounts posted, particularly in the counter-partisan condition (75%). This, coupled with the high proportion of those ignoring counter-partisans mentioning the motivation of different partisanship, is in line with our field experimental findings suggesting that both informational content motives and contextual social motives contribute to tie reciprocation decisions online. Finally, we also saw that considering the account as a stranger was a frequently selected motivation for ignoring accounts in our experiment—particularly in the neutral condition (55%).

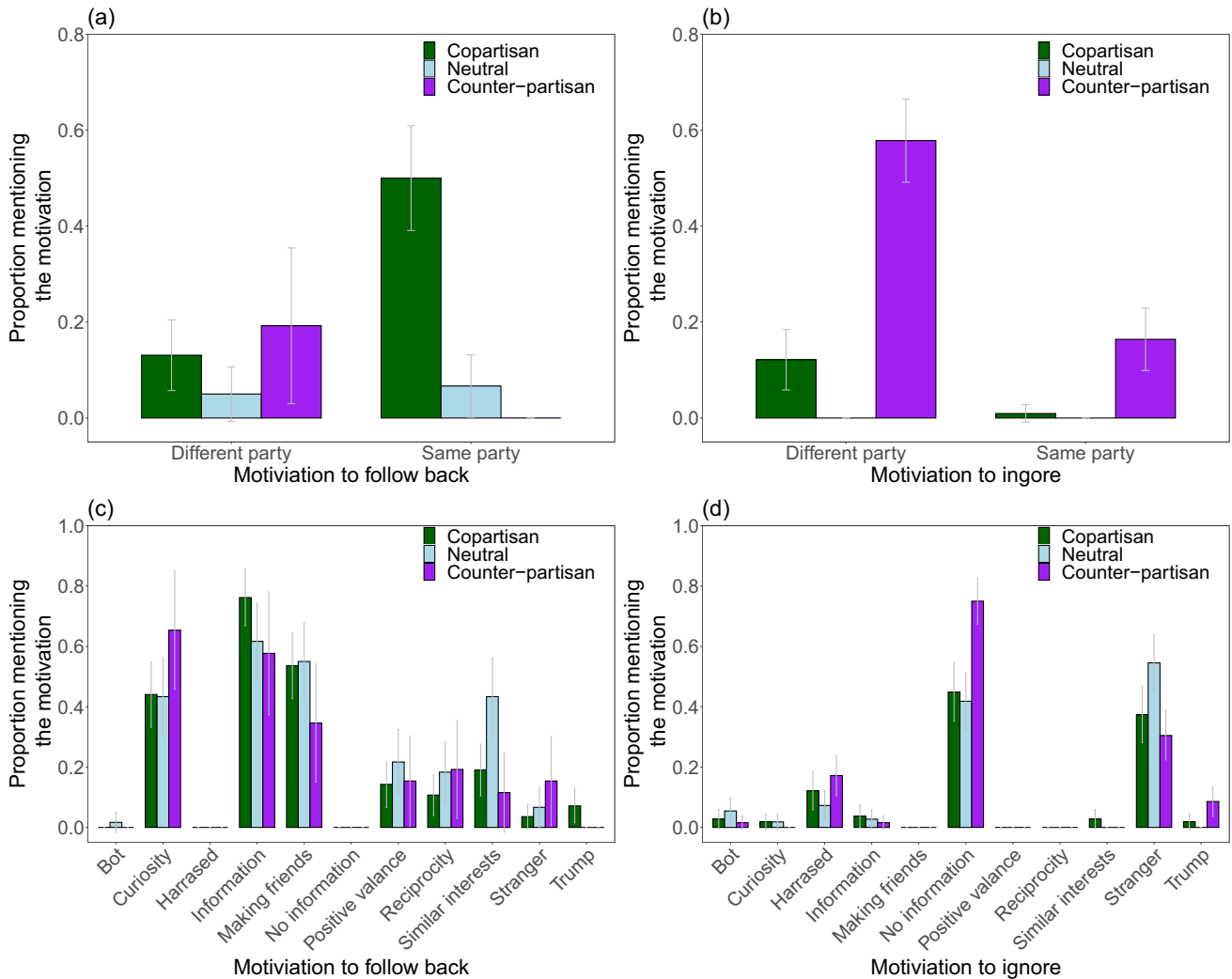
In sum, our survey experiment results suggested that there are distinct psychological mechanisms underscoring out-party dispreference versus in-party preference in follow-back behavior. Individuals greater in issue polarization (and, to a lesser extent, out-party animosity) were less likely to reciprocate social ties with counter-partisan and neutral accounts (but not copartisans); whereas individuals with stronger in-party liking were more likely to follow back copartisan and neutral accounts (but not counter-partisans). We did not observe a partisan asymmetry for these findings (see [Supplemental Table S13](#)), and our results were largely robust to attention, filtering for more politically active users, and filtering for more active Twitter users (see [Supplemental Tables S19–S42](#)). Our exploratory free response text analyses also shed light on a number of additional motivations for following-back and ignoring online accounts. As expected, participants attributed these decisions to shared and differing partisanship—as well as wanting to see (or not see) information from other accounts. Other motivations may also underscore these decisions—such as curiosity for reciprocating cross-party ties and concern about strangers motivating ignoring accounts.

Discussion

Our current work elucidates the psychological foundations of partisan bias in social tie formation on social media. Our Twitter field experiment demonstrates that politically active Twitter users reciprocate social ties with accounts of shared partisanship due to social, rather than purely content based, preferences. Users showed much bigger differences in follow-back rates between copartisans and counter-partisans when the accounts were human-looking compared to explicit bot accounts. This demonstrates that perceiving

Figure 4

Proportion of Participants by Condition and Follow-Back Decision Mentioning Motivations to Follow-Back or Ignore in Free Response Text Data



Note. Shown here are the proportion of participants by experimental condition and follow-back versus ignore decision who mentioned specific motivations to either follow back or ignore accounts, as classified by generative pre-trained transformer 4. (a) Fifty percent of participants in the copartisanship condition who followed back the account mentioned same partisanship as a motivation for following back. (b) Fifty-eight percent of participants in the counter-partisanship condition who ignored the account mentioned different partisanship as a motivation for ignoring. (c) Information and making friends were frequently mentioned motivations for following back accounts; curiosity was also an often-mentioned motivation for follow back, particularly for counter-partisan accounts. (d) Not wanting information (particularly from counter-partisans) and considering the account a stranger were frequently mentioned motivations for ignoring accounts. Error bars reflect 95% confidence intervals. See the online article for the color version of this figure.

accounts as human exacerbates users' preferences to follow back copartisan accounts and not follow back counter-partisan accounts relative to control, showing an additional social preference above and beyond that toward purely congenial content.

We also examined the effect sizes of copartisan preference and counter-partisan dispreference in follow-back behavior relative to a politically neutral control. These analyses revealed that in-party preference and out-party dispreference both contribute to preferential social tie formation in similar magnitudes. To further investigate the social preferences underscoring in-party affiliation and out-party dissociation, we examined political preference moderators of these

phenomena in a follow-up survey experiment. Here, we found evidence that increased issue polarization (and, to a lesser extent, out-party disliking) are associated with decreased social tie reciprocation of counter-partisan and politically neutral accounts, but not of copartisan accounts. Conversely, we found that increased in-party liking predicts increased social tie reciprocation of copartisan and politically neutral accounts, but not of counter-partisan accounts. These patterns suggest that issue polarization and out-party dislike are political attributes associated with decreased reciprocation of ties with neutral and counter-partisan accounts, whereas in-party liking is a political attribute associated with increased reciprocation of ties with

neutral and copartisan accounts. Our exploratory text analyses also provide additional insight on why people follow back and ignore political accounts online. Consistent with our field experiment, we found that shared partisanship and a desire to see information are frequently cited motivations for following back concordant accounts, whereas different partisanship and not wanting to see information are often mentioned motivations for ignoring discordant accounts. Other motivations—such as curiosity and making friends for reciprocating ties, and concern about strangers regarding ignoring accounts—were also frequently mentioned by participants as explanations for their tie-making decisions. Altogether, these findings shed light on the mechanisms motivating online partisan homophily.

Our results have several important theoretical contributions. First, we show the replicability of prior experimental work demonstrating preferential follow-back of copartisans on social media (Ajzenman et al., 2023b; Mosleh et al., 2021). We also provide much more compelling evidence than prior work (due to a much larger sample size) for a lack of difference between politically active Democratic and Republican Twitter users in their level of partisan bias in tie reciprocation on platform. This is perhaps surprising in the context of some observational data which find partisan asymmetries in political homophily—for instance, more conservative individuals tend to show more partisan bias in online social ties (Boutyline & Willer, 2017).

Second, our experiments provide evidence for *why* partisans preferentially follow back copartisans and not counter-partisans. We show that although preferential partisan follow-back is present for both explicit bot and human-looking accounts, the effect is much larger for human-looking accounts. That is, our findings not only replicate prior work demonstrating that individuals prefer connecting and interacting with humans over bots in social interaction settings (e.g., Ishowo-Oloko et al., 2019), but crucially show that human follow-back settings exacerbate the preference that individuals have for reciprocating copartisan ties and declining to reciprocate counter-partisan ties. This result suggests that follow-back behavior is not only driven by a desire for selective exposure to politically congenial content. Rather, heightened selective preferences for human-looking accounts illustrates an additional social motivation to affiliate with, or receive information from, shared partisans (and disaffiliate from counter-partisans) above and beyond a preference for concordant content. This is also further supported by our exploratory text analyses in our survey experiment—participants report shared partisanship, wanting to see information, and wanting to make friends as reasons for following back copartisan accounts, and conversely report different partisanship and not wanting to see information as reasons for ignoring counter-partisan accounts.

Third, by comparing these effects to a politically neutral control and examining political preferences in our survey experiment, we highlight two distinct routes underpinning these preferences—issue polarization (and, to a lesser extent, out-party dislike) predicts decreased follow-back rates except for with copartisans, while in-party preference predicts increased follow-back rates except for with counter-partisans. Theoretically, our findings are consistent with research on parochial altruism or “the combination of in-group altruism and out-group hostility” (Abbink et al., 2012), which has found that individuals jointly, yet separably, express both favoritism toward in-group members and animosity toward out-group members in social or cooperative settings (e.g., Bernhard et al., 2006). Indeed, our findings imply that issue polarization and out-party disliking are

associated with reduced tie reciprocation for counter-partisans (but predict no additional preference for copartisans), whereas in-party liking is associated with increased tie reciprocation for copartisans (but no additional dispreference for counter-partisans). Our findings also go beyond typical accounts of parochial altruism and intergroup relations by further examining the associations between group preferences and neutral ties. Here, we interestingly find that preferences negatively associated with counter-partisans—issue polarization and out-party disliking—also spillover to predict increased dispreference for neutral accounts. More issue-polarized and out-party-disliking individuals follow back both counter-partisan *and* neutral accounts at lower rates. Likewise, preferences positively associated with copartisans—in-party liking—spillover to predict increased preference for neutral accounts. Individuals with greater in-party liking follow back both copartisan *and* neutral accounts at higher rates. Intergroup preferences predict exacerbated preferential follow-back rates not only on in-group and out-group members, but also upon neutral accounts.

As evident in Figure 3, the confluence of these findings also shows how both outgroup dispreferences and ingroup preferences work in tandem to exacerbate selective tie formation. Participants low in issue polarization, out-party disliking, and in-party liking have similar follow-back rates of copartisan and neutral accounts. Increases in these motivations heighten the predicted difference between copartisan and neutral follow-back rates (albeit via issue polarization and out-party disliking predicting greater reductions in neutral than copartisan follow-back rates, and in-party liking predicting greater increases in copartisan than neutral follow-back rates). Together, this results in the aggregate pattern of preferential tie formation whereby individuals prefer to reciprocate ties with copartisans relative to neutral accounts, and also prefer to reciprocate ties with neutral accounts relative to counter-partisans. As further evidence, this pattern of results, as also seen in our field experimental data, is particularly pronounced among participants high in both out-party disliking and in-party liking (see Supplemental Figure S2).

Fourth, our experiments highlight the importance of incorporating social parameters in future data-driven modeling of polarization in social networks. For instance, recent agent-based models of polarization on social media platforms parameterize agents according to cognitive, primarily content-based attributes such as open-mindedness, thus accounting for confirmation bias and selective exposure (e.g., Wang et al., 2020). These models allow for the consideration of social tie formation based on shared ideological viewpoints—but do not consider the additional social psychological underpinnings we examine in our current work. As per our field experiment, this would suggest that more socially driven or less purely content-driven copartisan or politically neutral accounts may have higher tie reciprocation affinities relative to more purely congenial content-seeking accounts. Furthermore, our survey experiment results suggest that parameters such as issue polarization, out-party dislike, and in-party affinity may not only have distinct consequences on tie formation patterns for partisans but also separately affect tie formation with more neutral, nonpartisan accounts. Indeed, we find that individuals with greater issue polarization and out-party dislike are less likely to follow back neutral accounts, whereas individuals higher in in-party affinity are more likely to follow back neutral accounts. As shown in our text analyses, additional motivations such as curiosity, desire to make friends, and concern about strangers could also be added into more sophisticated models. Indeed, such sociopolitical dynamics may be important to

incorporate in future agent-based models of online polarization to paint a more complete picture of these complex dynamics.

Practically, the current work has important applied implications for the consideration and consequences of online political networks. Prior work shows that political homophily can distort how partisans understand the world and make decisions via preferential information exchange with copartisans (Stewart et al., 2019). Our findings suggest that such preferences may be driven not only by preferences for politically concordant content but also by preferences for associating with politically concordant peers. Critically, we also provide evidence that preferential reciprocation of copartisan versus counter-partisan social ties is driven both by ingroup preference and outgroup dispreference. This finding has key predictions for potential interventions upon partisan social tie preference and attempts to reduce polarization (e.g., Voelkel et al., 2023). Reducing animosity toward counter-partisans may partially increase cross-party connections online, but even in the absence of counter-partisan animosity, our results suggest that in-party preference would still promote substantial partisan homophily.

An important limitation of the current work is that our experiments do not examine the *consequences* of preferential social tie reciprocation among copartisans, nor do we intervene on preferences by experimentally exacerbating in-party liking or out-party disliking. Further research may explore the causal effects of manipulating ingroup love and outgroup hate on preferential tie formation, as well as carefully examine the potential positive and negative consequences of political homophily. Indeed, while such homophily has been linked to negative outcomes such as groupthink (Janis, 2008) and radicalization (Sunstein, 1999), homophilous networks may better facilitate political action (e.g., González-Bailón et al., 2011) and insulate partisan networks from low-quality and toxic content in asymmetric political landscapes (e.g., Mosleh & Rand, 2022).

Constraints on Generality

Our current work also contains several important constraints on generality to note. First, our human-looking accounts in our field and survey experiments were presented only as White, male profiles. Previous field experimental research has observed notable differences in follow-back rates of Black versus White accounts on Twitter (Ajzenman et al., 2023a), as well as differences in treatment effects of interventions delivered by White-presenting versus Black-presenting accounts (Munger, 2017). Asymmetries in follow-back probabilities by perceived account gender have also been observed in the field (Ajzenman et al., 2023a). In our current work, we hold perceived account gender and race constant across all human-looking conditions—thus differences between conditions should be attributed to partisanship, conditional on these attributes. However, it is important for future research on online social tie formation to examine how differences in perceived demographics may potentially moderate copartisan preferences.

Second, Twitter users in our field experiment consisted only of individuals who shared at least one article from Fox News or MSNBC and are not representative of Twitter users nor Americans as a whole. Indeed, we found some evidence of differences in copartisan preference versus counter-partisan dispreference relative to sizes between our field and survey experiments. Likewise, our survey experiment participants were recruited from Lucid and filtered for those self-reporting having a Twitter account. Future

work should assess how our results generalize to more representative populations, as well as on social media platforms other than Twitter.

Third, our experiments were designed and conducted in the years 2022–2023, and the magnitudes of our results may be specific to this time period and particularly reflective of the psychology and preferences of Twitter users in the United States during these years. Indeed, much research suggests that political sectarianism and polarization were, and are, particularly heightened during this time (Card et al., 2023; Finkel et al., 2020; Waller & Anderson, 2021). Thus, the extent of political tie preference among partisans may be especially acute given the timing of this sample. That said, there is evidence to suggest that our general tie reciprocation findings and mechanistic accounts are likely to generalize beyond this time window. First, multiple similar follow-back field experiments in multiple cross-national contexts (e.g., Brazil), have been conducted since 2020 and have found similar patterns of preferences for copartisan ties and dispreferences for counter-partisan ties (Ajzenman et al., 2023a, 2023b; Martel et al., 2024; Mosleh et al., 2021). This, as well as our replication of our field experiment in our survey setting, provide some evidence for cross-time replicability at least within the last several years. Second, the overall degree of polarization or sectarianism should not theoretically change the key predictions of our current work. Given a context with ingroup and outgroup members, our findings suggest that (a) preferential social tie formation can be exacerbated by making accounts human-like and social rather than purely informational; and (b) political motivations, such as issue polarization and out-group animosity are predicted to decrease follow-back propensity for outgroup members and neutral others, whereas ingroup affinity is predicted to increase tie formation with ingroup members and neutral others. The aggregate degree of polarization may change the magnitude of these effects but should not affect their predicted direction. Future work may examine this prediction empirically by assessing follow-back rates and correlates in other time periods or political contexts.

Conclusion

In sum, our results demonstrate the multifaceted underpinnings of partisan assortment on social media. The tendency for partisans to be connected with like-minded others is not just a result of recommendation algorithms but arises from the social preferences of issue polarization, disliking counter-partisans, and liking copartisans. As such, partisan social assortment is likely an enduring feature of social networks—a fact technologists and policy makers must incorporate into their decision making.

References

- Abbink, K., Brandts, J., Herrmann, B., & Orzen, H. (2012). Parochial altruism in inter-group conflicts. *Economics Letters*, 117(1), 45–48. <https://doi.org/10.1016/j.econlet.2012.04.083>
- Adamic, L. A., & Glance, N. (2005). *The political blogosphere and the 2004 U.S. election: Divided they blog* [Conference session]. Proceedings of the 3rd International Workshop on Link Discovery, New York, NY, United States.
- Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., & Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web*, 6(2), 1–33. <https://doi.org/10.1145/2180861.2180866>

- Ajzenman, N., Ferman, B., & Sant'Anna, C. P. (2023a). *Discrimination in the formation of academic networks: A field experiment on #EconTwitter*. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4323614
- Ajzenman, N., Ferman, B., & Sant'Anna, C. P. (2023b). *Rooting for the Same Team: Shared social identities in a polarized context*. Preprint. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4326148
- American National Election Studies. (2021). *ANES 2020 time series study full release* [Dataset and documentation]. Retrieved July 19, 2021, from <https://www.electionstudies.org>
- Amira, K., Wright, J. C., & Goya-Tocchetto, D. (2021). In-group love versus out-group hate: Which is more important to partisans and when? *Political Behavior*, 43(2), 473–494. <https://doi.org/10.1007/s11109-019-09557-6>
- An, J., & Weber, I. (2016). #greysanatomy vs. #yankees: Demographics and hashtag use on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1), 523–526. <https://doi.org/10.1609/icwsm.v10i1.14767>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>
- 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>
- Berinsky, A. J., Margolis, M. F., Sances, M. W., & Warshaw, C. (2021). Using screeners to measure respondent attention on self-administered surveys: Which items and how many? *Political Science Research and Methods*, 9(2), 430–437. <https://doi.org/10.1017/psrm.2019.53>
- Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, 442(7105), 912–915. <https://doi.org/10.1038/nature04981>
- Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology*, 38(3), 551–569. <https://doi.org/10.1111/pops.12337>
- 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>
- Card, D., DellaVigna, S., Funk, P., & Iriberry, N. (2023). Gender gaps at the academies. *Proceedings of the National Academy of Sciences of the United States of America*, 120(4), Article e2212421120. <https://doi.org/10.1073/pnas.2212421120>
- Chakraborty, A., Messias, J., Benevenuto, F., Ghosh, S., Ganguly, N., & Gummadi, K. (2017). Who makes trends? Understanding demographic biases in crowdsourced recommendations. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 22–31. <https://doi.org/10.1609/icwsm.v11i1.14894>
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317–332. <https://doi.org/10.1111/jcom.12084>
- Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 89–96. <https://doi.org/10.1609/icwsm.v5i1.14126>
- DellaPosta, D., Shi, Y., & Macy, M. (2015). Why do liberals drink lattes? *American Journal of Sociology*, 120(5), 1473–1511. <https://doi.org/10.1086/681254>
- Eady, G., Nagler, J., Bonneau, R., & Tucker, J. (2019). *Political information sharing and ideological polarization*. Midwest Political Science Association.
- Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *SAGE Open*, 9(1). <https://doi.org/10.1177/2158244019832705>
- Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, 370(6516), 533–536. <https://doi.org/10.1126/science.abe1715>
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298–320. <https://doi.org/10.1093/poq/nfw006>
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Scientific Reports*, 1(1), Article 197. <https://doi.org/10.1038/srep00197>
- Higgins, M. J., Sävje, F., & Sekhon, J. S. (2016). Improving massive experiments with threshold blocking. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), 7369–7376. <https://doi.org/10.1073/pnas.1510504113>
- 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>
- Ishowo-Oloko, F., Bonnefon, J. F., Soroye, Z., Crandall, J., Rahwan, I., & Rahwan, T. (2019). Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. *Nature Machine Intelligence*, 1(11), 517–521. <https://doi.org/10.1038/s42256-019-0113-5>
- 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>
- Janis, I. L. (2008). Groupthink. *IEEE Engineering Management Review*, 36(1), Article 36. <https://doi.org/10.1109/EMR.2008.4490137>
- Kteily, N. S., Rocklage, M. D., McClanahan, K., & Ho, A. K. (2019). Political ideology shapes the amplification of the accomplishments of disadvantaged vs. advantaged group members. *Proceedings of the National Academy of Sciences of the United States of America*, 116(5), 1559–1568. <https://doi.org/10.1073/pnas.1818545116>
- Martel, C., Mosleh, M., Yang, Q., Zaman, T., & Rand, D. G. (2024). Blocking of counter-partisan accounts drives political assortment on Twitter. *PNAS Nexus*, 3(5), Article pgae161. <https://doi.org/10.1093/pna-snexus/pgae161>
- Mosleh, M., Eccles, D., & Rand, D. G. (2024). *Tendencies toward triadic closure: Field-experimental evidence*. SocArXiv. <https://osf.io/preprints/socarxiv/ys8zw>
- Mosleh, M., Martel, C., Eckles, D., & Rand, D. G. (2021). Shared partisanship dramatically increases social tie formation in a Twitter field experiment. *Proceedings of the National Academy of Sciences of the United States of America*, 118(7), Article e2022761118. <https://doi.org/10.1073/pnas.2022761118>
- Mosleh, M., Martel, C., & Rand, D. G. (2023, July 10). *Ingroup love vs. Outgroup hate: What drives partisan bias in social tie formation on social media?* Open Science Framework. https://osf.io/n7dym/?view_only=0fe35aeb3bcd4a718c8fd028ede85edc
- Mosleh, M., Pennycook, G., & Rand, D. G. (2022). Field experiments on social media. *Current Directions in Psychological Science*, 31(1), 69–75. <https://doi.org/10.1177/09637214211054761>
- Mosleh, M., & Rand, D. G. (2022). Measuring exposure to misinformation from political elites on Twitter. *Nature Communications*, 13(1), Article 7144. <https://doi.org/10.1038/s41467-022-34769-6>
- Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, 39(3), 629–649. <https://doi.org/10.1007/s11109-016-9373-5>

- Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), Article 119. <https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10/>
- Rosenbaum, M. E. (1986). The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, 51(6), 1156–1166. <https://doi.org/10.1037/0022-3514.51.6.1156>
- Rudat, A., & Buder, J. (2015). Making retweeting social: The influence of content and context information on sharing news in Twitter. *Computers in Human Behavior*, 46, 75–84. <https://doi.org/10.1016/j.chb.2015.01.005>
- Siering, M., Muntermann, J., & Rajagopalan, B. (2018). Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems*, 108, 1–12. <https://doi.org/10.1016/j.dss.2018.01.004>
- Stewart, A. J., Mosleh, M., Diakonova, M., Arechar, A. A., Rand, D. G., & Plotkin, J. B. (2019). Information gerrymandering and undemocratic decisions. *Nature*, 573(7772), 117–121. <https://doi.org/10.1038/s41586-019-1507-6>
- Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, 30(3), 341–366. <https://doi.org/10.1007/s11109-007-9050-9>
- Sunstein, C. R. (1999). *The law of group polarization* (John M. Olin Law & Economics Working Paper No. 91). University of Chicago Law School.
- Tappin, B. M., Pennycook, G., & Rand, D. G. (2021). Rethinking the link between cognitive sophistication and politically motivated reasoning. *Journal of Experimental Psychology: General*, 150(6), 1095–1114. <https://doi.org/10.1037/xge0000974>
- Voelkel, J. G., Stagnaro, M., Chu, J., Pink, S., Mernyk, J., Redekopp, C., Ghezae, I., Cashman, M., Adjodah, D., Allen, L., Allis, V., Baleria, G., Ballantyne, N., Van Bavel, J. J., Blunden, H., Braley, A., Bryan, C., Celniker, J., Cikara, M., ... Willer, R. (2023). *Megastudy identifying effective interventions to strengthen Americans' democratic attitudes*. Open Science Framework. <https://osf.io/preprints/osf/y79u5>
- Waller, I., & Anderson, A. (2021). Quantifying social organization and political polarization in online platforms. *Nature*, 600(7888), 264–268. <https://doi.org/10.1038/s41586-021-04167-x>
- Wang, X., Sirianni, A. D., Tang, S., Zheng, Z., & Fu, F. (2020). Public discourse and social network echo chambers driven by socio-cognitive biases. *Physical Review X*, 10(4), Article 041042. <https://doi.org/10.1103/PhysRevX.10.041042>
- Weisberg, H. F., & Rusk, J. G. (1970). Dimensions of candidate evaluation. *The American Political Science Review*, 64(4), 1167–1185. <https://doi.org/10.2307/1958364>
- Yu, X., Wojcieszak, M., & Casas, A. (2024). Partisanship on social media: In-party love among American politicians, greater engagement with out-party hate among ordinary users. *Political Behavior*, 46(2), 799–824. <https://doi.org/10.1007/s11109-022-09850-x>
- Zhang, L., Peng, T. Q., Zhang, Y. P., Wang, X. H., & Zhu, J. J. (2014). Content or context: Which matters more in information processing on microblogging sites. *Computers in Human Behavior*, 31, 242–249. <https://doi.org/10.1016/j.chb.2013.10.031>

Received September 2, 2023

Revision received July 14, 2024

Accepted July 26, 2024 ■