Rooting for the same team: Shared social identities in a polarized context*

Nicolás Ajzenman[†] Bruno Ferman[‡] Pedro C. Sant'Anna[§]

This version: November 18, 2025
First draft: January 16, 2023

Abstract

Can non-political identities bridge political divides? We investigate this through a field experiment on Twitter during the 2022 Brazilian elections. We created fictional accounts signaling political and/or football club identity. These accounts followed real users with varying identity alignments. Sharing political or non-political identity increased follow-backs and reduced blocks. The effect of non-political identity weakens once political identity is revealed, uncovering a previously overlooked phenomenon: political polarization can undermine the integrative effect of cross-cutting identities. Text analysis of tweets during the FIFA World Cup showed that polarization also undermined the cohesive effect of another non-political identity, the national football team.

Keywords: Social Media; Social Identity; Affective Polarization; Brazilian Elections.
JEL Codes: D72; D91; C93; Z20.

*We thank Filipe Campante, Fernanda Estevan, Claudio Ferraz, Thomas Fujiwara, Lorenzo Lagos, Horacio Larreguy, Mario Macis, Mohsen Mosleh, Marcos Nakaguma, Ben Olken, Frank Schilbach, Romain Wacziarg, Ekaterina Zhuravskaya, and seminar participants at Brown, CEPR/Warwick/Princeton/Yale Polecon Symposium 2023, EEA, Exeter, FGV-EESP, Insper, JPAL-SEA, MIT Behavioral Lunch, Monash, NEUDC 2023, NUS, NOVA SBE, NYU, PUC-Chile, Queens, UCSD, UofM, UofT, Universidad de San Andres, Universidad Torcuato Di Tella, and Warwick for their helpful comments and suggestions. We are grateful to Livia Haddad, Luis Lins, and Nícolas de Moura for superb research assistance. This research was approved by the Ethical Compliance Committee on Research Involving Human Beings at Fundação Getulio Vargas (CEPH/FGV, IRB approval n. 208/2022). The experiment was pre-registered at the AEA RCT Registry under ID AEARCTR-0009982.

[†]McGill University. E-mail: nicolas.ajzenman@mcgill.ca
[‡]São Paulo School of Economics - FGV. E-mail: bruno.ferman@fgv.br
[§]MIT Department of Economics. E-mail: p_stanna@mit.edu

1 Introduction

In recent years, political identity has emerged as a particularly divisive cleavage, with negative consequences for interpersonal relations, democratic norms, and social cohesion (Iyengar et al., 2019). Yet individuals possess multiple identities, which potentially cut across divisive dimensions by emphasizing common experiences and a sense of belonging. Social scientists have often pointed to the possibility of such cross-cutting identities to foster cohesion and serve as bridges across political divides (e.g., Depetris-Chauvin et al., 2020). However, when one dimension of identity becomes more and more dominant—as may be the case with growing political polarization—such potential may not be realized. In this paper, we study the interplay between political and non-political identities and show that political polarization undermines these alternative forms of connection, thereby limiting the potential of non-political identities to foster social cohesion.

Our empirical setting examines social media network formation on Twitter (now X) as an indicator of social cohesion—a consequential form of sorting, given how political alignment increasingly structures social connections online. This form of sorting, often manifesting as political homophily, may fuel misinformation, shape content exposure (Levy, 2021), and limit contact with opposing views (Bursztyn et al., 2022). These dynamics can even reinforce polarization and weaken offline cohesion (Enikolopov et al., 2024).

Our focus is Brazil—a highly polarized country—during the 2022 presidential elections. We examine two non-political identities with potential cohesive power: support for a Brazilian football club and for the national team—a key national symbol during the World Cup.[1] We conducted a pre-registered field experiment on Twitter during the second half of 2022, spanning the period before, during, and after Brazil's presidential election campaign. We measure social ties through follow-back and block rates.

We created fictional accounts that signaled support for either Luiz Inácio Lula da Silva or Jair Messias Bolsonaro—representing opposite ends of Brazil's political spectrum—and a preferred football club, which we refer to throughout the paper as the "affective identity," in contrast to the political one. Some accounts remained neutral on one of the two dimensions to isolate the effect of each shared identity independently. The accounts randomly followed real Twitter users with either congruent or non-congruent political and affective identities. Our sample includes 4,620 politically active users. The experiment was conducted in 43 five-day rounds ("waves") during the second half of 2022. At the end of each wave, we recorded whether users followed back or blocked the experimental accounts as a measure of social ties, resulting in a total of 30,194 observations. We document several experimental findings.

First, both identity dimensions shape social ties. Using accounts that signal only one dimension (affective or political), we find that identity congruence increases follow-backs and reduces blocking. Sharing an affective identity raises follow-backs by 13.4 percentage points (58.5%) and reduces blocking by 1.4 points. Political congruence increases follow-

[1] We use "football" rather than "soccer." In Brazil, club rivalries are historically rooted and often divide society along lines largely uncorrelated with political preferences or socioeconomic status (Ronconi, 2022). We treat club preference as a social, non-political identity, given football's central role in Brazilian culture (DaMatta, 1994).

backs by 20 points (119%) and reduces blocking by 12.3 points.

Second, political polarization weakens the cohesive power of other shared identities. Using accounts that signal both identity dimensions, we find that the cohesive effects of shared affective identity become smaller when information on political preferences is revealed. When the account is politically neutral, sharing an affective identity increases follow-backs by 13.4 percentage points; this falls to 8.5 points with political agreement and 4.3 points with disagreement. These results uncover a previously overlooked phenomenon: that political polarization can overshadow other non-political identities, limiting their cohesive effects.

Yet, while disclosure of political preferences weakens the effects of sharing an affective identity, sharing affective identity still fosters (some) ties among counter-partisans and is particularly powerful in preventing blocks. Together, these results suggest that political identities significantly undermine the cohesive potential of shared non-political identities, yet these other identities still play a role in reducing cross-partisan animosity.

We then leverage our six-month experimental window with repeated waves to examine whether changes in salience of political polarization affect cross-partisan behavior. Comparing follow-back rates before and after the official electoral period, we find a small but significant decrease in political homophily, consistent with a decline in the salience of political identity after the election.

Finally, we test whether shared collective national experiences—specifically the Brazilian national team's performance in the 2022 FIFA World Cup—can influence tie formation with counter-partisans on Twitter. The tournament began shortly after the election, shifting public attention from partisan to national identity. This timing created a natural opportunity to test whether moments of collective celebration (e.g., Brazil's victories or tournament progress) or shared disappointment (e.g., match losses or elimination) would strengthen cross-partisan ties, as in Depetris-Chauvin et al. (2020). In doing so, we examine whether a different kind of identity—national identity, which is often seen as more broadly unifying—can succeed where another affective identity (club allegiance) showed limited potential. Although political homophily declined slightly post-election, we find no significant effect of World Cup-related events—positive or negative—on cross-partisan follow-back rates.

Our findings suggest that non-political shared identities have limited power to reduce political segregation, unlike in studies such as Voelkel et al. (2024) or Depetris-Chauvin et al. (2020). This raises the question of why national identity did not foster cohesion in our setting. We hypothesize that national symbols lose their unifying potential when they become politicized. For example, President Trump welcomed the U.S. women's soccer team's 2021 loss to Sweden as a reaction to their "wokeness," a view echoed by many conservatives.[2] We test whether political polarization in Brazil had become so intense that it permeated even national symbols like the football team, eroding their capacity to unify.

To assess this, we collected live Twitter data during the World Cup, taking advantage of the fact that several Brazilian players had publicly expressed political preferences—potentially

[2] See https://www.vox.com/22600500/olympics-conservatives-simone-biles-anti-american and https://www.washingtonpost.com/sports/interactive/2024/american-sports-grievance-culture/. Accessed February 2025.

weakening their appeal among politically opposed fans. We analyzed Brazilians' Twitter reactions to major events involving the national team (e.g., goals, injuries), focusing on how responses varied with the political alignment between fans and the players involved. We show that reactions to both positive and negative events were strongly shaped by partisanship. A striking case was the severe injury of Brazil's star player Neymar: while politically aligned fans expressed concern, those on the opposing side celebrated. In Section 5.2, we show that this is not anecdotal but part of a systematic pattern in Brazilian supporters' behavior during the tournament. Hence, as political polarization permeated players, national identity through the football team was overshadowed by political identities, losing some of its cohesive power. This aligns with our experimental findings that political polarization can undermine the cohesive effect of cross-cutting identities—here, in the context of a different non-political identity: support for the Brazilian national football team.

Our results contribute to several strands of the literature. First, we contribute to the literature on intergroup animosity and the role of shared identities in reducing it. Prior work shows that shared identities can mitigate partisan divisions and intergroup prejudice (Voelkel et al., 2024), with one strand focusing on how sport fosters cohesion between opposing groups (Lowe, 2021; Mousa, 2020). We are particularly related to two papers. Depetris-Chauvin et al. (2020) find that, in Sub-Saharan Africa, individuals are more likely to identify with their nation over their ethnic group after major national football victories. Ronconi (2022) show that, in Latin America, social cohesion improves after matches between rival football clubs—except when players behave violently. Our experimental results complement these studies but suggest that in highly polarized settings, the cohesive potential of shared football identity is substantially weakened. Observational evidence from the World Cup—closer in spirit to Depetris-Chauvin et al. (2020)—further indicates that even identification with the national team may fail to foster cohesion when polarization also permeates the players. More generally, we show that cross-cutting identities, often pointed out as a way of alleviating divisive cleavages (e.g., Dunning and Harrison, 2010), may have limited potential when one divisive dimension dominates—as is the case with increasing political polarization.

Second, our paper relates to research on echo chambers in social media (Zhuravskaya et al., 2020). Digital technology is frequently identified as a driver of polarization (Gentzkow, 2016), and as an environment that amplifies it through the formation of echo chambers (Sunstein, 2001, 2018). While many studies emphasize the role of algorithmic recommendations in creating these chambers (Epstein and Robertson, 2015), our results suggest that echo chambers also emerge from individuals actively sorting into networks with others who share their identities—even when they are otherwise congruent in non-political dimensions. This preference for politically aligned ties may shape the type of content and news individuals consume (Levy, 2021; Halberstam and Knight, 2016) and reduce exposure to dissenting views (Bursztyn et al., 2022), potentially deepening polarization.

Our paper also contributes to the literature on affective polarization—out-group animosity and in-group favoritism based on political preferences (Iyengar et al., 2012, 2019). While most existing work relies on surveys (e.g., Iyengar et al., 2012; Boxell et al., 2022; Wagner, 2021; Reiljan, 2020), we provide a behavioral, revealed-preference measure in a natural setting. Similar approaches have been used to document affective polarization in online dating

(Huber and Malhotra, 2017), holiday gatherings (Chen and Rohla, 2018), and social media (Mosleh et al., 2021). Our design is especially close to Mosleh et al. (2021), who study how shared partisanship shapes tie formation on Twitter in the U.S.

We contribute to this literature in three ways. First, we examine how political identity interacts with non-political identities, showing that political alignment can overshadow other bases of similarity. Second, by analyzing follow-backs and blocks, we separately capture in-group favoritism and out-group avoidance.[3] Third, our longitudinal design allows us to study how social tie formation responds to changes in the salience of identity categories.

The remainder of this paper is organized as follows. In Section 2, we provide relevant background on polarization in Brazil, football, and Twitter, focusing on information relevant to the understanding of the experimental design and its results; then, in Section 3, we detail our experimental design and empirical strategy; then, in Section 4 we present our experimental results; finally, in Section 5, we document how cross-partisan interactions changed over time, focusing on explaining why the shared experience of the World Cup did not significantly reduce political homophily.

2 Background

2.1 Political Polarization in Brazil

Brazil's democracy is facing what some analysts call its most polarizing moment in history (Kingstone and Power, 2017). In the 2022 presidential election, voters chose between right-wing incumbent Jair Bolsonaro and former president Luiz Inácio Lula da Silva of the Workers' Party. Together, they secured over 90% of valid votes in the first round—compared to less than 80% for the top two candidates in previous elections. The run-off was decided by a margin of under two percentage points.[4]

To further illustrate Brazil's current level of affective polarization, we draw on data from the Brazilian Electoral Study (BES), a nationally representative post-election survey. Following Boxell et al. (2022), we measure affective polarization as the gap between respondents' feelings toward their preferred party and all others (Appendix Figure B.1). In 2022, its average level (59.1) surpassed that of the United States in 2020 (56.3) and was notably higher than in France (52.6), Canada (37.7), and Germany (28.5). Brazil also shows a positive polarization trend, smaller than the U.S. (slope = 0.56) but similar to France.

In sum, affective polarization in Brazil has risen sharply and now rivals levels in the U.S. and parts of Latin America, exceeding those in many OECD countries. This polarization

[3]Our finding that political incongruence overrides other identities aligns with Chen and Rohla (2018), who show that Thanksgiving dinners were shorter when attended by individuals from opposing-party precincts. Unlike that study, which may conflate partisanship with other factors, our use of experimentally controlled profiles allows us to isolate the causal effect of political mismatch. We also compare accounts that signal political identity to those that do not, identifying the extent to which polarization overshadows other identities.

[4]Lula won with 50.90% of valid votes, against 49.10% for Bolsonaro.

likely shapes how individuals form social ties across political lines—a central focus of this paper. While survey-based measures offer valuable insights, they can suffer from exaggeration or bias (Bursztyn et al., 2025; Iyengar et al., 2012). In contrast, our experimental design captures behavioral indicators—follow-backs and blocks—within a real-world social media environment.

2.2 Football

Football is by far Brazil's most popular sport: 65% of the population express interest in it (Nielsen Sports, 2022), and 73.1%—including 85.1% of men and 62.5% of women—support a football club (IPEC and O Globo, 2022).

That more Brazilians support a football club than follow the sport underscores its deep social significance. Football is not just entertainment—it's a central part of national and personal identity (Murad, 1995; DaMatta, 1994). This identity can unify, but also divide: in early 2023, at least seven people died in fan violence (G1, 07-11-2023), and unexpected match results have been linked to domestic violence (Arabe, 2022). Similar patterns appear across Latin America (Alabarces, 2003).

Football in Brazil is shaped by long-standing rivalries that foster antagonism between clubs and their supporters (Ronconi, 2022). These rivalries are often regional—for example, Palmeiras vs. Corinthians (São Paulo) and Flamengo vs. Vasco (Rio de Janeiro) are among the most well-known.

Brazilian football rivalries are largely not aligned with social divides. Appendix Figure B.2 shows that supporters of the top clubs are demographically similar, with no club tied to a single trait. All clubs have large, diverse fan bases, preventing identification with any dominant group profile.

This feature is relevant in our context: football club preferences can foster connections among individuals who differ in other identities. In our Twitter sample (detailed in Section 3), supporters of most clubs are not strongly aligned with a single political preference (Appendix Table B.2). This diversity creates meaningful opportunities for cross-cutting ties across political and socioeconomic lines.

2.3 Twitter

Our experiment takes place on Twitter, one of Brazil's most popular social media platforms and a top global market for the company. In 2021, over 17 million Brazilians used Twitter (Statista, 2022b), ranking the country fourth in user count despite being seventh in population.

Most Twitter accounts are public by default, meaning their posts and profile pages are visible to all users. A profile typically includes a photo, background image, short bio, tweet history, and usage metrics such as tweet count, number of followers, and followed accounts ("friends").

Twitter allows one-way connections: users can follow others without reciprocity—unlike platforms like Facebook. Following a public account requires only clicking "follow" on its profile. The followed user then receives a notification showing the follower's profile and can choose to follow back, ignore, or block the account. A follow enables the follower to see the other's tweets, retweets, and likes on their timeline. By the time of our experiment, blocking, by contrast, prevented interaction and visibility, though the blocked account was not notified. However, it can see it was blocked if it visits the blocker's profile.

We interpret follows and blocks as opposite signals of willingness to form social ties. Following suggests openness to connection—such as viewing posts or exchanging messages. Blocking, by contrast, reflects rejection or a desire to create distance. It is a deliberate action that cuts off all interaction between the two accounts within the Twitter environment.

Twitter has become central to political discourse in Brazil and elsewhere, especially during elections (Jungherr, 2016). In the U.S., it has had a causal impact on voting behavior (Fujiwara et al., 2021). While no direct evidence exists for Brazil, 75% of Brazilian Twitter users reported viewing political content in 2018—comparable to Facebook (80%) and above WhatsApp (65%) (LAPOP, 2019). According to DataSenado, 45% said social media influenced their vote (DataSenado, 2019). These patterns make Twitter a compelling setting to study political identity and social connections.

3 Experimental Design and Data

3.1 Experimental Design

Between July and December 2022, we conducted a pre-registered experiment on Twitter. We created fictional accounts that signaled support for a candidate in the 2022 Brazilian presidential election and/or for a Brazilian football club. These accounts randomly followed Twitter users who either shared or did not share the same identity. Five days later, we recorded two outcomes: the number of follow-backs and the number of blocks each account received.

We ran the experiment in five-day waves. In each wave, we activated three types of fictional Twitter accounts: those signaling both political and football identities; those signaling only political identity; and those signaling only football identity. We randomly selected two Brazilian clubs per wave (e.g., clubs A and B) and created eight accounts: pro-Lula supporter of A; pro-Bolsonaro supporter of A; pro-Lula supporter of B; pro-Bolsonaro supporter of B; politically neutral supporter of A; politically neutral supporter of B; pro-Lula with no club preference; and pro-Bolsonaro with no club preference. This design allowed us to assess the role of each identity dimension in tie formation without conditioning on the other.

3.1.1 Fictional Accounts

Appendix Table A.1 summarizes the features of the fictional accounts, each defined by a political identity (pro-Lula, pro-Bolsonaro, or neutral) and a football identity (support for one of Brazil's six most popular clubs or none).[5] Political and football identities were randomly assigned, as detailed in the next subsection.

To signal political identity, we added the hashtag *#Lula2022* or *#Bolsonaro2022* to the account's bio and re-tweeted one post from the corresponding candidate. Neutral accounts included neither a hashtag nor a political retweet. We signaled football identity through the profile picture—a flag with the team's logo in a stadium—and by writing "Supporter of team X" in the bio. For football-neutral accounts, we used a photo of a non-Brazilian stadium (with no identifiable teams) and labeled them "Football fan." These accounts still expressed interest in football, just without signaling team preference. Figure 1 shows examples.[6]

In accounts that signal both identities, the football (affective) identity is more salient than the political one, which appears only in the bio. To assess the robustness of our results, we replicate the experiment using accounts in which political identity is more prominently signaled (see Subsection 4.3).

3.1.2 Sample Selection and Assignment into Treatment

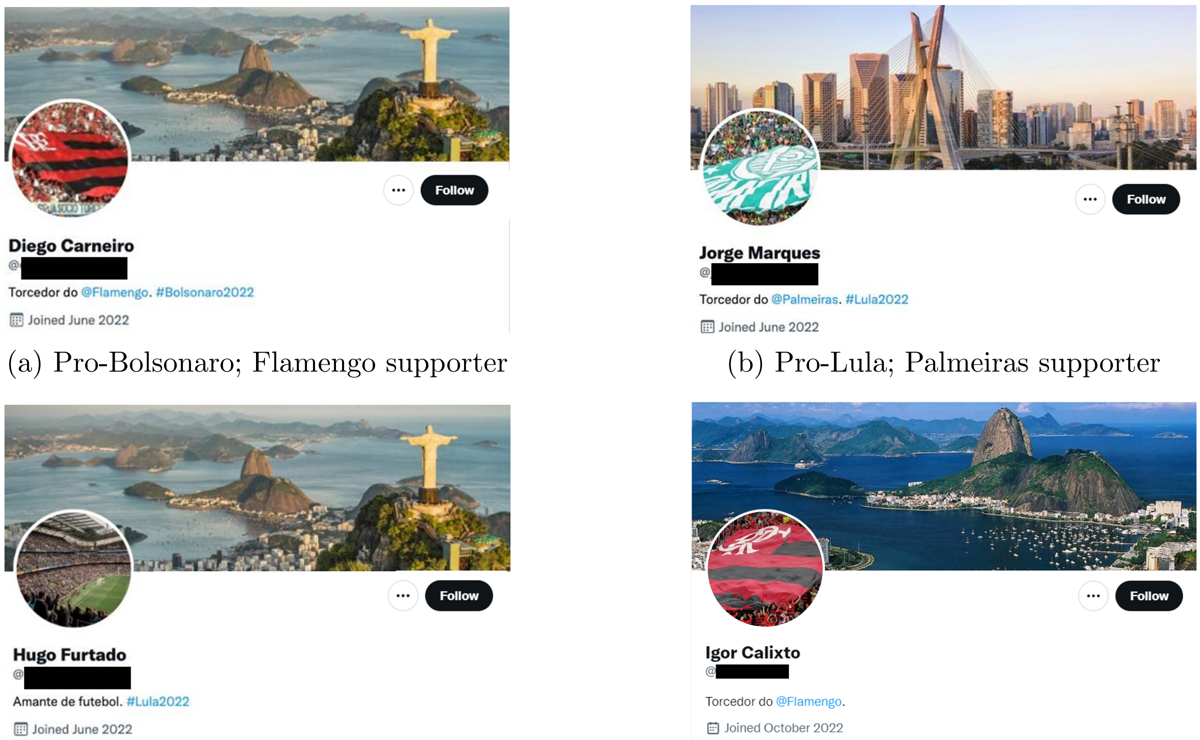
The key requirement for inclusion in our sample is that we must be able to identify both the subject's political identity (pro-Lula or pro-Bolsonaro) and their preferred football team. Appendix Figure A.2 summarizes the sampling procedure.

First, we used Twitter's API to collect users who tweeted or retweeted statuses containing pro-Lula or pro-Bolsonaro hashtags between May 31 and July 11, 2022. The full list of hashtags appears in Appendix Table A.2. As a result, our sample consists of politically engaged users who were discussing politics before the official campaign period, which began on August 16. Next, we inferred football club preference by scanning users' Twitter bios for terms associated with the six most popular Brazilian clubs and their rivals. We used a keyword detection algorithm and manually verified all matches. We then excluded accounts created in 2022 (which are more likely to be inauthentic), accounts with fewer than 10 followers, accounts with a follower-to-friend ratio above 20, and accounts showing signs of automation. Our goal was to remove both inauthentic users and users unlikely to follow back experimental accounts. After these steps, our final sample comprises 4,652 individual accounts. Due to Twitter API constraints, this represents a random subset of Brazilian users

[5] The six clubs are C.R. Flamengo, S.C. Corinthians Paulista, São Paulo F.C., S.E. Palmeiras, Grêmio F.B.P.A., and C.R. Vasco da Gama, based on a 2022 Sport Track/XP survey (Sport Track and XP, 2022). While bots only supported these clubs, the subject pool also included fans of rivals: S.C. Internacional (Grêmio), Botafogo F.R. and Fluminense F.C. (Flamengo/Vasco), and Santos F.C. (Palmeiras, São Paulo, Corinthians).

[6] We used only male names for the accounts. While gender identity is politically relevant, only 20% of our sample are women, leaving us underpowered to test gender interactions. Appendix Table B.12 reports gender-disaggregated results. Further research should explore how political identity interacts with gender in polarized contexts.

Figure 1: Examples of Fictional Accounts



(a) Pro-Bolsonaro; Flamengo supporter

(b) Pro-Lula; Palmeiras supporter

(c) Pro-Lula; Neutral-Team

(d) Politically Neutral, Flamengo supporter

Notes: The figures show examples of fictional accounts used in the experiment. Political identity is signaled by the hashtag *#Lula2022* or *#Bolsonaro2022* on the account's bio. Football club identity is signaled by the profile picture and the text "Supporter of [club's official Twitter account]" on the bio. When football club identity is not signaled (Panel c), the account still signals interest in football through its profile picture (a neutral football stadium) and the text "Football lover" on its bio.

who signal both political and football club identities.[7]

Using Twitter's API, we collected observable characteristics for each subject, including tweet, follower, and friend counts; account location (when available), which we recoded to the regional level; verification status; number of likes ('favorites"), and account creation date. We also predicted gender based on first names, using Brazilian Census data tabulated by Meireles (2021). Appendix Tables B.1 and B.2 report descriptive statistics. The sample is politically balanced, with approximately 55% Lula supporters and 45% Bolsonaro supporters. Each football club includes a substantial share of supporters of both candidates. Although some clubs lean toward one side, none are politically homogeneous: in all cases, at least

[7] As pre-registered, we manually screened for bots before the experiment. Afterward, we used the *Botometer* API to estimate automation likelihood (Sayyadiharikandeh et al., 2020; Yang et al., 2020), excluding accounts with scores above 0.85. We missed only 39 likely bots (less than 1% of the final sample). The median Botometer score is 0.13. In Section 4.3, we show that results are unchanged when restricting to likely human accounts.

27% of users support the minority candidate. This confirms that Brazilian football clubs are not strongly aligned with political preferences and that their supporter bases are politically diverse. In Subsection 4.3.1, we show that our results hold when excluding any single club, including those with skewed political composition.

In each experimental wave, we activated eight fictional accounts: four that signaled both a football club and a political candidate, and four that were neutral in one dimension—two labeled as "football fans" without a team, and two that signaled a team but not a political identity. Each wave featured two randomly selected football clubs.[8] Three accounts per wave signaled support for each of the two selected clubs.

Each fictional account follows about 100 subjects per wave. Following Athey and Imbens (2017), we use block randomization to assign subjects, stratifying by political identity, preferred football club, and follower count (above/below the sample median). For accounts that signal a club, we restrict the sample to users who support either that club or a regional (intra-state) rival (see Appendix Table A.3). We define four account-subject identity pairings: congruent in both dimensions, incongruent in both, or congruent in only one (political or football). Each pairing is further split by follower count, yielding eight strata. We draw equal proportions of subjects from each. Subjects may be treated (i.e., followed) more than once but never in consecutive waves. After being treated, a subject reenters the pool only after three waves, reducing the risk of "learning" about the experiment.

The "treatment" in our experiment is receiving a follow notification from a fictional Twitter account. Variation comes from whether the subject and the account match or differ in political and/or football identity. Appendix Figure A.3 shows an example. Notification visibility depends on whether users access Twitter via mobile or desktop.[9] On both platforms, the profile photo is immediately visible. On mobile, the bio—indicating political affiliation—is also shown. On desktop, it appears only if the user clicks or hovers. However, to follow back or block the account, desktop users must do one of these actions and will thus see the bio.

In addition to following assigned subjects, each fictional account also followed a collaborator aware of the experiment, who then confirmed whether the follow triggered a notification. This procedure ensured that follow notifications were actually being delivered.[10] If an account was shadow-banned, we excluded it from the analysis, as specified in our pre-analysis plan. Across all waves, 12 accounts (5.1% of those created) were shadow-banned. This status was not correlated with political identity: among the 12, 3 were pro-Lula, 4 pro-Bolsonaro, and 5 were politically neutral.

[8] Clubs were sampled with probabilities proportional to their representation in our subject pool.

[9] According to Statista (2022a), 80% of Twitter users access the platform via mobile. In our sample, 72% of subjects used only the mobile app during the experiment, based on live-streamed tweet data.

[10] A "shadow-ban" on Twitter hides an account's activity—including follow notifications—due to suspicious behavior. We verified that each account was not shadow-banned before including it in the analysis.

3.1.3 Timing

As described in the previous section, the experiment was conducted in waves. In each wave, we activated eight fictional accounts: four that signaled both political identity and football club preference, and four that were neutral in one of the two dimensions. Each wave followed the timeline described below:

(i) **Day 0:** Accounts are created per Appendix Table A.1. Each re-tweets a post related to its football identity—either from the club's official account or a neutral football source—and a post from its preferred political candidate. It follows 15 "elite" accounts aligned with its identities (e.g., the candidate and club) and is followed by five colleagues aware of the experiment.

(ii) **Day 1:** Each fictional account follows the subjects assigned to it according to the procedure described in the previous section.

(iii) **Day 5:** After five active days, we record each account's number of followers and blocks, then delete all information and relaunch it for the next wave. The five-day window is based on prior Twitter research showing that over 95% of follow-backs occur within this period (Ajzenman et al., Forthcoming).

We launched one wave every Tuesday and Friday, resulting in two overlapping waves at all times (Appendix Figure A.1). In total, we ran 43 waves between July and December 2022. The Brazilian presidential election took place during the second half of the year, with the first round on October 3 and the second on October 29. We exploit variation in wave timing to examine how the salience of political identity affects the impact of shared identity on tie formation.

In each wave, we track follow-backs daily via Twitter's API, using the fifth-day count as the main measure. Blocks are assessed only at day five because Twitter's API lacks direct block data. To detect blocks, we compare the set of accounts each fictional account follows with the assigned follow list. Differences occur for three exclusive reasons: (i) the fictional account was blocked by the subject; (ii) the subject's account was suspended or deactivated; or (iii) the subject unfollowed the fictional account. We manually verify these cases by viewing profiles from the fictional account's perspective. A block is recorded only if confirmed on day five.[11]

3.2 Empirical Strategy

We are interested in studying the effect of identity congruence in the formation of social ties on Twitter. In most of our analysis, we present results that pool all experimental waves, comparing the follow-back and block rates of subjects who shared or did not share political and/or affective identity with the fictional account.

[11]If a subject blocks then unblocks the fictional account, we treat this as follower removal, not a block. Our block measure includes only persistent blocks through the wave's end.

To formally test significance, we use pre-registered specifications with wave and strata fixed effects. We first focus on experimental accounts signaling a single identity dimension (political or affective) and following subjects who agree or disagree on that dimension. Our outcomes—follow-backs and blocks—capture subjects' responses to being followed by fictional accounts. Therefore, we limit the analysis to the assigned subject–fictional account pairs. Let Y_{ijst} be an indicator equal to one if subject i in strata s interacted with fictional account j during wave t, where "interaction" denotes either a follow-back or a block. We then estimate the following equation:

$$Y_{ijst} = \alpha + \beta \times \text{identity\_congruence}_{ij} + X_{ijt}\delta + \varphi_{st} + \epsilon_{ijst} \quad (1)$$

where identity_congruenceij is an indicator equal to one if the fictional account and subject share the identity dimension under study; φst denotes strata $\times$ wave fixed effects; and ϵ_{ijst} is the error term. X_{ijt} is a vector of controls from the fictional account, subject, and wave, interacted with strata dummies.[12] Controls include the subject's number of followers and tweets, account creation year, gender, and location, plus the Google trend index of the fictional account's football club at wave t (interacted with the identity congruence indicator when applicable). This trend controls for the salience of the football identity across waves.

We also analyze accounts signaling both identity dimensions. This yields four possible subject–account pairs: congruent in both, congruent only in affective or political identity, and incongruent in both. To study these treatment arms, we estimate the following equation:

$$\begin{aligned} Y_{ijst} = \beta_0 + \beta_1 \times \text{political\_congruence}_{ij} + \beta_2 \times \text{affective\_congruence}_{ij} + \\ \beta_3 \times \text{political\_congruence}_{ij} \times \text{affective\_congruence}_{ij} + \\ X_{ijt}\lambda + \phi_{st} + \varepsilon_{ijst} \quad (2) \end{aligned}$$

where political_congruence$_{ij}$ is an indicator equal to one if fictional account j and subject i share political preferences, affective_congruence$_{ij}$ equals one if fictional account j and subject i share preference for football club, and the other variables have the same definition as before. We include the same covariates as before.

Coefficient β_1 represents the effect (in percentage points) of sharing political identity on follow-backs or blocks for subjects who do not share affective identity with the bot. Similarly, β_2 measures the effect of sharing affective identity among subjects who do not share political identity. Finally, β_3 captures the difference in the effect of sharing affective identity between subjects who do and do not share political identity with the bot.

We present standard errors clustered at the fictional account level. To verify the reliability of our inference method given the number of clusters, we conducted the simplest assessment proposed by Ferman (2022). We simulated data under the null hypothesis of no treatment effect, using Bernoulli draws with the pilot's average follow-back rate. Reassuringly, the rejection rate at the 5% significance level closely matched the nominal 5% in all cases.

[12] We include strata $\times$ wave fixed effects following Bruhn and McKenzie (2009). Among strata fixed effects, a "misfit" dummy identifies subjects assigned to treatment in violation of proportional assignment within strata; these misfits are globally reassigned per Carril (2017).

3.3 Balance and Attrition

Appendix Table B.3 presents summary statistics for treated subjects across the eight treatment arms. Pre-treatment characteristics are balanced: for all variables, we cannot reject the null of equality across treatments at standard significance levels. We perform a joint test and report its F-statistic in the table's last column.

The table also reports attrition rates by treatment arm. We define attrition as cases where a subject was assigned treatment but could not be treated. This occurs for three reasons: the subject's account was suspended by Twitter for policy violations, the subject deactivated their account, or the subject set their account to private. In the first two cases, the account is not found on Twitter; in the third, the account is accessible but we did not follow it, per IRB guidelines.

Overall, there was no differential attrition across treatment arms, with rates around 9%. Appendix Table B.4 shows no significant differences in characteristics between attrited subjects across arms. Thus, pooling waves and estimating Equations (1) or (2) cross-sectionally should not raise attrition concerns. Including wave fixed effects ensures comparisons involve statistically similar accounts, since attrition occurs at wave onset.

Attrition might affect our analysis of heterogeneous effects over time, which compares subject behavior across waves. Because attrition can vary, the subject pool differs between waves. Appendix Table B.5 confirms this: subjects who experienced attrition are disproportionately more likely to support Jair Bolsonaro and have higher follower counts and Twitter activity than those who never attrited. Therefore, for analyses of heterogeneous effects over time, we restrict the sample to subjects who remained active throughout the experimental period.

4 Experimental Results

4.1 Effects of Political or Affective Congruence on the Formation of Ties

We begin by examining whether sharing each identity—affective or political—affects social tie formation on Twitter. For this, we restrict the analysis to experimental accounts signaling a single identity dimension. Results for this subset appear in Table 1, with Panel A showing follow-backs and Panel B showing blocks.

We first find that football clubs are, unconditionally, a relevant dimension of socialization in our setting. The first three columns of Table 1 present estimates of β_1 from Equation (1) for experiments using politically neutral accounts. For completeness and in line with our pre-registration, we also show results without controls, wave and strata fixed effects, and the additional covariates described in Section 3.2.

A subject is about 14 percentage points (pp) more likely to follow back a fictional account supporting their own team versus a rival—over a 50% increase relative to the 22.9% follow-

Table 1: Effect of Shared Identity on Follow-Backs and Blocks—Experimental Accounts Signaling a Single Dimension of Identity

Panel A: Follow Backs, Fictional Accounts Signaling a Single Dimension of Identity						
	Dependent Variable: Follow-Backs (1 = Yes)					
	Politically-neutral Fictional accounts			Football-neutral Fictional accounts		
	(1)	(2)	(3)	(4)	(5)	(6)
Football club congruence	0.1337***	0.1413***	0.1406***			
	(0.0133)	(0.0134)	(0.0130)			
Political congruence				0.2000***	0.1994***	0.1979***
				(0.0148)	(0.0147)	(0.0133)
Average (incongruent pairs)	0.229			0.168		
Wave, Strata Fixed Effects	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observations	7,388	7,388	7,388	7,678	7,678	7,678
Panel B: Blocks, Fictional accounts Signaling a Single Dimension of Identity						
	Dependent Variable: Blocks (1 = Yes)					
	Politically-neutral Fictional accounts			Football-neutral Fictional accounts		
	(1)	(2)	(3)	(4)	(5)	(6)
Football club congruence	-0.0126***	-0.0126***	-0.0132***			
	(0.0032)	(0.0032)	(0.0043)			
Political congruence				-0.1225***	-0.1225***	-0.1224***
				(0.0062)	(0.0061)	(0.0060)
Average (incongruent pairs)	0.023			0.130		
Wave, Strata Fixed Effects	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observations	7,199	7,199	7,199	7,492	7,492	7,492

Notes: The table shows regression estimates of sharing identities on follow-backs (Panel A) and blocks (Panel B) for experimental accounts signaling a single identity dimension (political or affective). Estimates of β_1 in Equation (1) are presented without controls, with wave and strata fixed effects, and with additional controls interacted with treatment (bot's football club, Google trend index, and subject's followers and friends). The sample excludes shadow-banned accounts as pre-registered. Block observations are fewer due to missing data in one wave caused by a technical issue. Standard errors clustered at the fictional account level appear in parentheses. Significance codes: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

back rate for rival teams. Block results tell a similar story despite low block rates for politically neutral accounts. Subjects block 2.3% of rival accounts, and sharing a football club reduces this by 1.26 pp (significant at 1%). Thus, football club preferences are a significant factor in social tie formation in our setting. This quantitatively supports sociological and anthropological observations that football club allegiance is a key socialization dimension in Brazil (e.g., Murad, 1995; DaMatta, 1982).

We also find that political identity significantly influences social tie formation. Columns 4–6 of Table 1 report the effect of shared political identity for experimental accounts signaling only political preference. Sharing political identity increases follow-back probability by 20 pp (from 16.8% to 36.8%) and reduces block probability by about 12 pp (from 13% to 0.7%), both significant at 1%. These results confirm that both political and football-related identities independently shape socialization in our setting.

4.2 The Interplay between Political and Affective Congruence

So far, we have discussed results for accounts signaling exclusively political or affective identity. Examining experimental accounts that signal both identity dimensions enables us to study how these identities interact in shaping social tie formation.

Results appear in Figure 2 (follow-backs) and Figure 3 (blocks), using raw data without additional controls. The top panel of each figure shows average follow-back or block rates for all eight treatment arms—four with fictional accounts signaling both identities and four signaling a single dimension—displayed simultaneously. The *x-axis* indicates political identity congruence: the three left columns show disagreement, the three right columns agreement, and the two center bars represent cases where political identity is not signaled (affective identity only), corresponding to column 1 of Table 1. Bar colors indicate the football club relationship: shared preference, rival clubs, or no club signal (gray bars, matching column 4 of Table 1). The bottom panel reports estimated differences and p-values for key two-by-two treatment comparisons, with standard errors clustered by fictional account. Appendix Tables B.6 and B.7 present difference-in-means tests with wave and strata fixed effects plus controls, yielding similar results.

The figures show that sharing either identity dimension significantly increases follow-back likelihood and reduces blocking. Subjects least likely to reciprocate are those sharing neither political nor football club identity, with only a 16% follow-back rate. Similarly, those most likely to block (14.6%) share neither identity. Sharing either identity boosts follow-backs and lowers blocks, but the effect sizes differ when conditioning on the other identity, suggesting political identity largely overshadows affective identity. We explore this further in the following subsections.

4.2.1 Effects of shared identity when the other identity is revealed

We begin by examining whether congruence in one identity dimension significantly affects tie formation probability, conditional on sharing or not sharing the other identity.

Among subject–fictional account pairs supporting the same football club (blue bars), follow-back likelihood is 20.3% when they disagree politically versus 40.8% when they agree—a 20.5 pp difference highly significant (p-value < 0.001), comparable to the effect observed with football-neutral accounts. Similarly, among pairs supporting rival clubs (red bars), follow-backs rise from 16% with political disagreement to 32.4% with agreement—a 16.4 pp increase, also significant at 1%. In all cases, sharing political identity roughly doubles follow-back probability. This pattern is illustrated in Figure 2, comparison (a).

Similarly, political identity strongly affects blocking regardless of the fictional account's preferred club. When club preference is unknown, blocking probability is 12.3 pp lower if the fictional account and subject share political identity versus opposing identities ($p < 0.001$). For rival club supporters, the difference is 13.5 pp ($p < 0.001$), and for same-club supporters, it remains sizable at 7.9 pp ($p < 0.001$). Although same-club pairs show a significant block reduction—discussed later—blocking primarily targets politically opposed accounts. As Figure 3 illustrates, counter-partisans tend to avoid one another via blocks.

Overall, sharing political identity has a strong effect on follow-backs, regardless of whether the fictional account and subject support the same club, rival clubs, or if the bot's football preference is unknown. This suggests that political identity's impact on follow-backs is neither significantly offset nor reinforced by affective identity information. Its effects on blocks is also large regardless of the other identity.

The same cannot be said of the effect of sharing affective identity conditional on information on the bot's political identity. The effect of sharing affective identity depends on whether the bot's political identity is revealed. When fictional accounts do not signal political identity, sharing a football club increases follow-back probability by 13.4 pp. This effect diminishes when political identity is known. Among politically congruent pairs (Figure 2, right bars), follow-back rates are 32.4% for rival clubs versus 40.8% for same-club supporters—a difference of 8.4 pp ($p < 0.001$). This effect is significantly smaller than when political identity is unknown, implying a nearly 40% reduction in affective congruence impact. For politically opposed pairs, sharing a club increases follow-backs by only 4.3 pp (from 16%), a 68% reduction compared to when political identity is not signaled. This pattern appears in Figure 2, comparison (b).

During the analyzed period, political identity overshadowed other identity dimensions, notably football club preference, in social tie formation. Knowledge of a fictional account's political identity diminished the effect of shared affective identity, especially among politically opposed individuals, weakening potential ties that might form without political signaling. This suggests that in highly polarized contexts like ours, political preferences can reduce the bonding power of other shared identities and even disrupt social ties that would otherwise form.

4.2.2 Congruence in affective identity and the formation of social ties among counter-partisans

Despite political divergence reducing the impact of shared affective identity, this dimension can still foster ties among politically opposed individuals. Specifically, sharing a football club preference increases follow-back probability by 4.4 percentage points compared to pairs supporting rival clubs ($p < 0.001$). It also raises follow-back likelihood by 3.5 percentage points relative to cases where fictional accounts do not signal a preferred club, a difference significant at the 1% level. Thus, shared football identity promotes social ties even across political divides.

For blocks, the effect is even stronger. Among politically opposed individuals, sharing affective identity reduces blocking probability by 6.1 percentage points compared to rival clubs ($p < 0.001$), a 42% reduction. It also lowers blocking by 4.5 percentage points relative to cases where football club information is not provided ($p < 0.001$). Thus, sharing affective identity significantly decreases blocking risk among counter-partisans, although the overall block rate remains relatively high.[13]

Although political identity often overshadows affective identity in tie formation, our re-

[13]These effects persist when fictional accounts signal political identity more prominently (Appendix B.6).

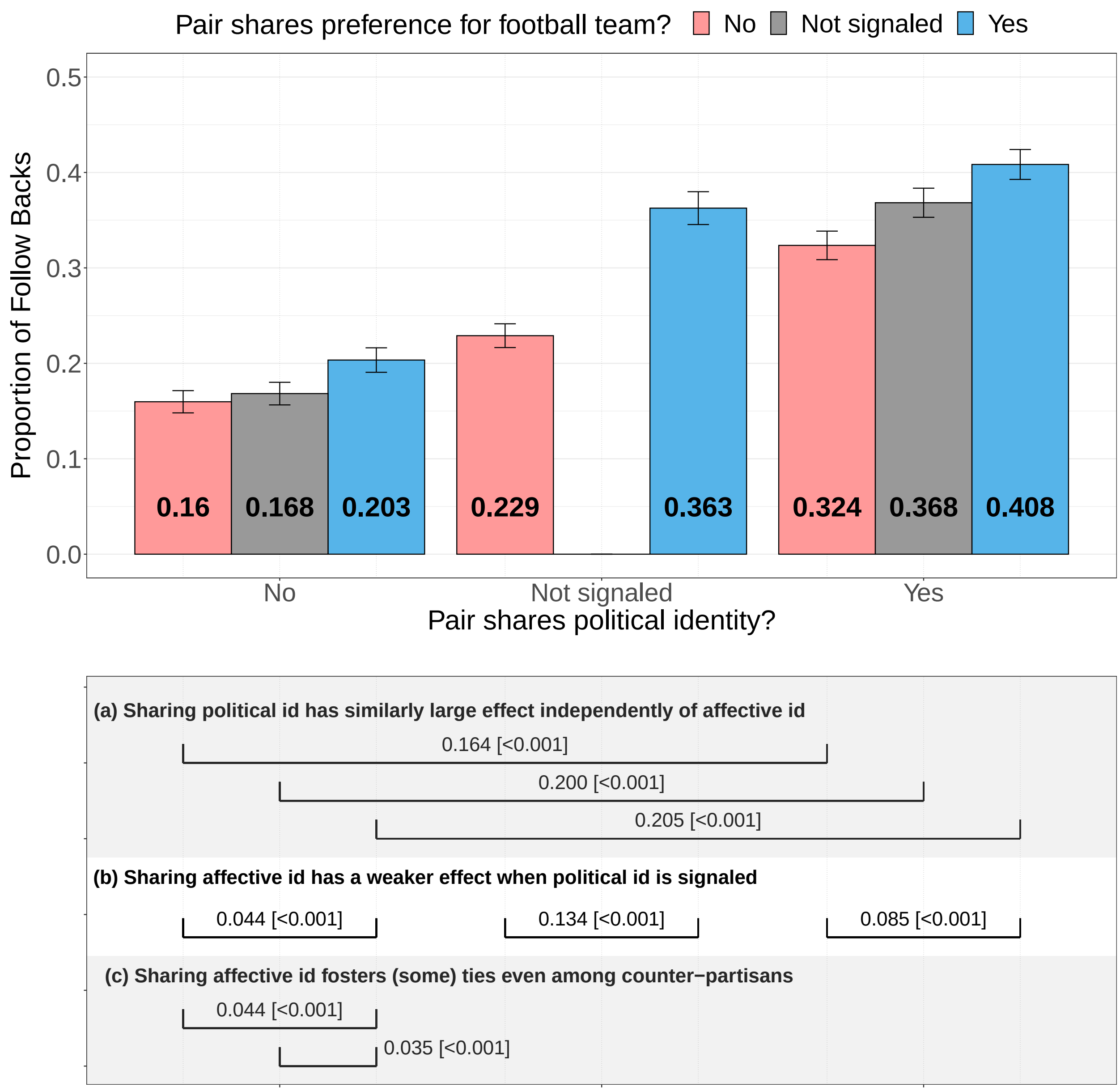
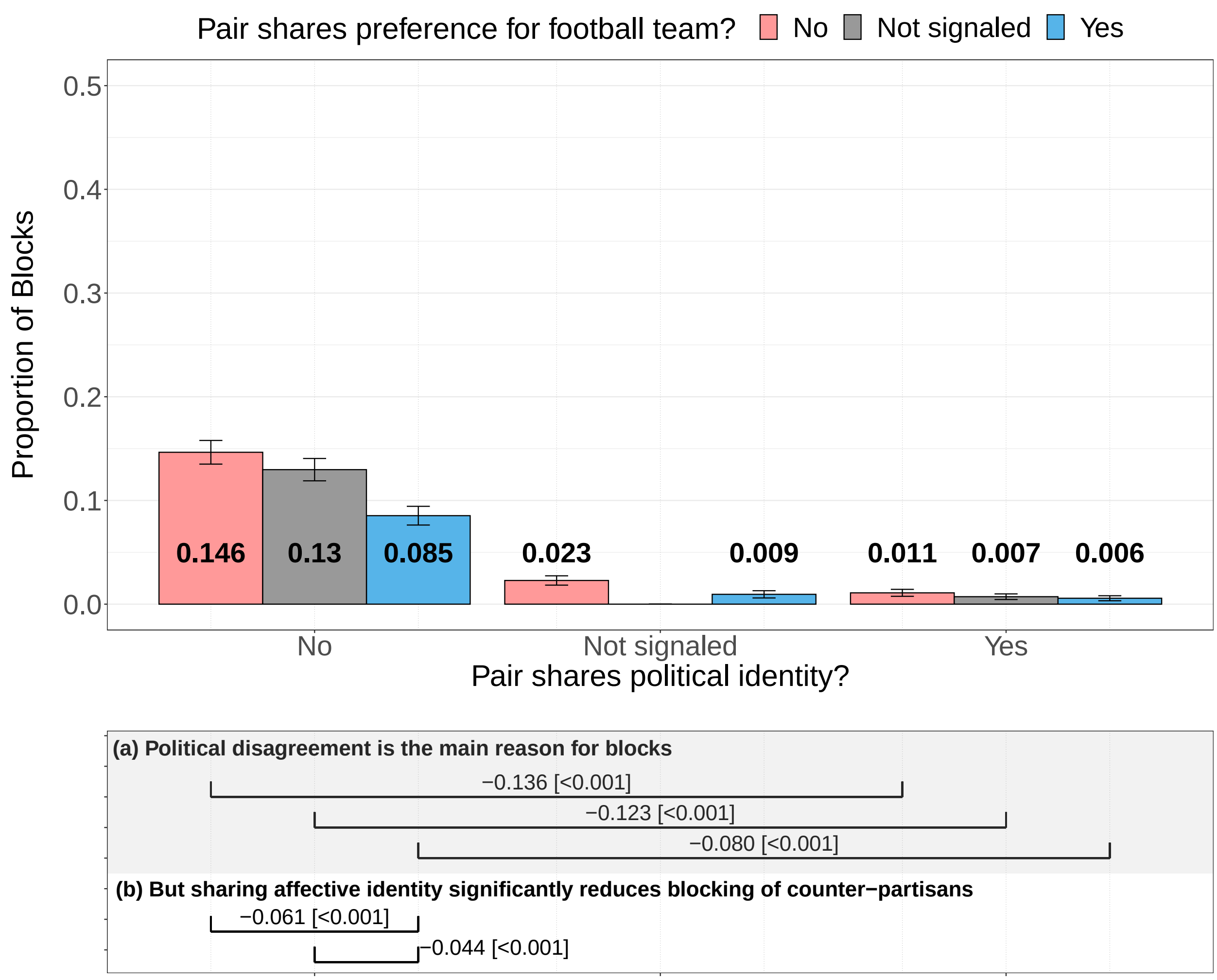


Figure 2: Effect of Shared Political and Affective Identity on Follow-Backs

Notes: The figures display the effect of sharing political and affective (football club) identity on follow-back rates across all eight treatment arms (fictional accounts signaling one or both identities). The sample pools subject–fictional account pairs from all waves, excluding shadow-banned accounts (as pre-registered), totaling 30,194 observations (4,620 unique subjects). The x-axis indicates political identity sharing or absence thereof, while colors denote shared football club preference or its absence. Bars show average follow-back rates; error bars represent 95% confidence intervals. The bottom panel reports estimated average differences and p-values (in brackets) for selected treatment pairs, with standard errors clustered at the fictional account level. Full difference tests with controls appear in Appendix Table B.6.

Figure 3: Effect of Shared Political and Affective Identity on Blocks



Notes: The figures show the effect of sharing political and affective (football club) identity on block rates across all eight treatment arms (fictional accounts signaling one or both identities). The sample pools subject–fictional account pairs from all waves, excluding shadow-banned accounts (as pre-registered), totaling 30,194 observations (4,620 unique subjects). The x-axis indicates political identity sharing or absence thereof, while colors denote shared football club preference or its absence. Bars show average block rates; error bars represent 95% confidence intervals. The bottom panel reports estimated average differences and p-values (in brackets) for selected treatment pairs, with standard errors clustered at the fictional account level. Full difference tests with controls appear in Appendix Table. B.7.

sults indicate that even amid intense polarization, sharing identities like football club preference can foster (some) connections and reduce avoidance among political opponents. This aligns with evidence that football promotes social cohesion (Depetris-Chauvin et al., 2020; Ronconi, 2022). Highlighting common interests, such as football preference, may help bridge political divides. Our findings also resonate with research on interventions to reduce affective polarization, especially cross-partisan conversations. Santoro and Broockman (2022) show these conversations' effectiveness depends on topic; our experiment suggests emphasizing shared identities can enhance impact, albeit modestly in polarized settings.

Overall, both dimensions of identity significantly influence social tie formation. While political identity tends to overshadow affective identity—diminishing the impact of shared football club preference on follow-back decisions—congruence in affective identity can still generate social ties. This is notable given our sample consists of politically engaged individuals (using political hashtags months before the election), suggesting that similarities in less politically correlated dimensions can help bridge political divides despite the overshadowing effect.

4.3 Validity, Interpretation, and Robustness

This section presents robustness checks and additional analyses that further validate the results and interpretations discussed above.

4.3.1 Effect of Specific Clubs

A potential concern is that specific football clubs signal other traits influencing subjects' decisions, rather than pure congruence in club preference—such as association with a particular politician. To address this, we repeat our analysis excluding one fictional account club (and its rivals) at a time. Appendix Table B.10 shows stable point estimates across these sub-samples, indicating results are not driven by any specific club.

Some subjects support football clubs not represented by any fictional account, meaning they can only be assigned to follow fictional accounts supporting rival clubs—always in the out-group affective dimension. To address this, we reanalyzed the data excluding these subjects. Appendix Table B.11 shows results remain qualitatively and numerically consistent with the main analysis.

4.3.2 Automated Accounts

Concerns may arise if a large proportion of our subjects are automated accounts ("bots"). Following our pre-analysis plan, we manually excluded accounts likely to be fictional before the experiment. Post-experiment, we used the *Botometer* API to estimate each subject's automation probability.[14] The median score in our final sample was 0.13, indicating a 13% chance of automation, with fewer than 1% (39 accounts) above 0.85—none of which were

[14]For details on *Botometer*'s algorithm, see Sayyadiharikandeh et al. (2020) and Yang et al. (2020).

manually removed. Appendix Tables B.8 and B.9 report results for subjects below the median *Botometer* score (unlikely bots), which are nearly identical to the main sample. Estimates remain stable across various *Botometer* score thresholds.

A related concern is that some subjects may perceive the experimental accounts as inauthentic. Overall, given the high take-up of the experiment, it is clear many users considered the accounts realistic. Nevertheless, the interpretation of our results could be questioned if users saw accounts that did not share their political identity as more likely to be fictional than those who did. While we cannot directly assess this perception, we provide indirect evidence suggesting this was not the case. Specifically, we use a Bayesian Classifier algorithm to determine whether a subject's most recent tweets before being followed contained political content. Subjects tweeting about non-political topics may be more likely to expect followers from users with differing political identities and thus be less suspicious of experimental accounts with opposing political identity. Therefore, if follow-back and blocking behavior are similar between users who recently tweeted about politics and those who tweeted about other topics, this offers indirect evidence that doubts about account authenticity do not drive our results.

Appendix Figure B.7 presents results for this analysis. Due to data limitations, we report findings starting from wave 11, the first wave for which we collected subjects' most recent tweets before treatment. Appendix Figure B.8 shows results for the full sample within this time frame. We restrict the analysis to subjects who tweeted within seven days prior to treatment. Reassuringly, subjects who had just tweeted about politics behaved similarly to those who tweeted about other topics and generally matched the overall patterns observed in the main analysis. We also conduct a similar heterogeneity analysis by classifying users' bios based on political content, using keyword searches to identify political references. Subjects without political content in their bios should be less suspicious of politically divergent followers. Approximately half of our subjects had political references in their pre-experiment bios. Appendix Figure B.9 shows that subjects with and without political bios behaved similarly. These findings together suggest that suspicions regarding the experimental accounts' authenticity do not drive our results.

These exercises also help rule out another alternative story: that follows by counter-partisans lead subjects to differentially update about some other (undesired) characteristic of this potential friend. For instance, a pro-Lula user might think that someone with a pro-Bolsonaro hashtag who follows them is not smart, even compared to other potential pro-Bolsonaro Twitter friends. This type of story would only be a concern if the updated belief is specific to the act of following a counter-partisan that signals this characteristic. To the extent that less openly political subjects have less reason to make this type of judgement, the two checks described above also suggest this type of concern is unwarranted in our setting.

4.3.3 Demand for follow-backs: Information versus social connections

Why do subjects in our experiment form or reject ties with the experimental accounts? Our preferred explanation is that Twitter ties represent social connections—subjects want to become virtual friends" with the accounts. An alternative view is that subjects seek information and consider these accounts as sources with a preferred slant. To explore this, we ran an auxiliary experiment using accounts explicitly labeled as automated information sharers (see Appendix Figure B.10). These accounts followed randomly selected subjects as in the original experiment. We compare follow-back and block rates between these information" accounts and original-style accounts. Conducted one year later, this auxiliary experiment observed generally lower follow-back and block rates. This approach parallels Mosleh et al. (2023), who find that U.S. users prefer following same-party human-identified accounts over fictional information sharers, indicating a social motivation to follow rather than mere interest in partisan content.

We find that subjects do not solely seek information. Appendix Figure B.11 and Table B.13 show that subjects sharing political and affective identities with fictional accounts are 10 pp (56%) more likely to follow original accounts than informational ones. Subjects disagreeing on affective identity follow both types at similar rates, indicating social motivation strongly influences in-group follow-backs. Blocking results mirror this: subjects disagreeing in both dimensions are 4 pp (65%) more likely to block original than informational accounts, though not significantly. This suggests social motives also affect blocking. Overall, if subjects cared only about information, polarization in social tie formation would be lower, with less pronounced in-group following and out-group blocking.

5 Polarization Trends: Post-Election Salience and the Effect of the World Cup

5.1 Experimental effects over time

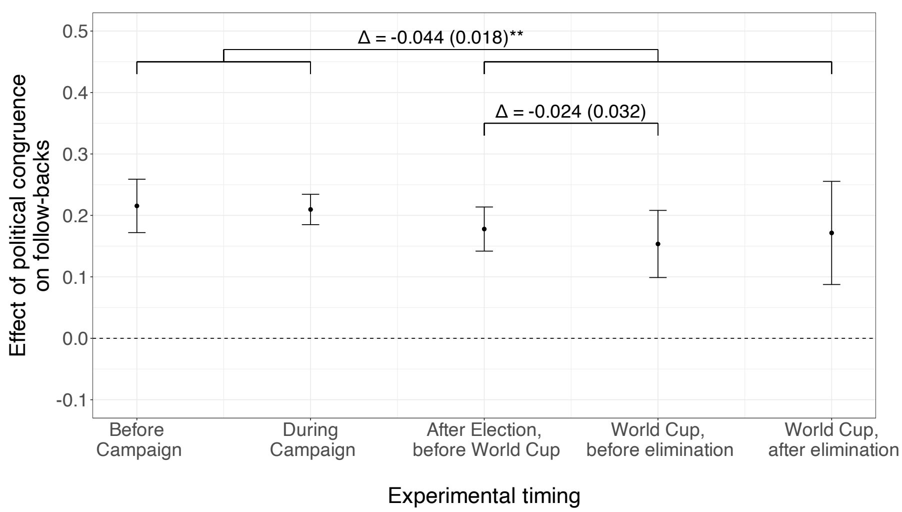
During our experimental period (August to December 2022), two key events shaped the salience of political and non-political identities in Brazil: the official electoral calendar—with the second voting round on October 30th and campaign period ending a day earlier—and the 2022 FIFA World Cup, starting twenty-one days post-election. These events allow two complementary tests. First, we expect political identity salience to peak around elections, potentially increasing online political homophily. Second, the widely popular World Cup, a major national symbol, could have heightened shared national identity, possibly reducing political homophily. According to established literature (Depetris-Chauvin et al., 2020), shared national identity may be a stronger cohesive force than football team allegiance, especially following positive shared experiences like wins.

This section addresses two questions to test our hypothesis. First, did online political homophily decrease after the election, as its salience waned? Second, did the shared national experience of the World Cup promote cross-partisan ties, and did this effect vary with the national team's performance?

Figure 4 shows follow-back results over time, plotting the estimated effect of political congruence for all experimental accounts before, during, and after the official electoral period. As described in Section 3, we restrict the sample to subjects active throughout the experiment—those who tweeted within seven days before each treatment wave—to avoid follow-back declines caused by reduced Twitter activity (27,701 observations). Equivalent block results appear in Appendix Figure C.1.[15]

We observe a small but significant drop in political congruence effects on follow-backs post-election, from 21.2 pp before/during to 16.8 pp after—a 4.4 pp decrease significant at 5%. This suggests reduced political salience lowers homophily.[16] Blocks show a similar decline, supporting this interpretation.

Figure 4: Effect of Congruence in Political Identity on Follow-Backs at Different Times



Notes: The figure displays point estimates and 95% confidence intervals for the effect of political identity congruence on follow-backs across experimental waves grouped by period: before, during, and after the official electoral period. The post-election period is further split into before the World Cup, during positive results for Brazil, and after Brazil's elimination. Timing details appear in Appendix Figure A.1. The sample pools data from all waves within each period, restricted to subjects active throughout the experiment (tweeting within seven days before each treatment), totaling 27,701 observations. Brackets above point estimates show differences between periods with standard errors (in parentheses), clustered at the bot-account level.. Significance codes: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

Nevertheless, political congruence continues to strongly affect follow-backs and blocks post-election. Does this effect lessen as shared national identity gains salience during the World Cup? While the literature emphasizes shared identities as a path to inter-group cohesion (Voelkel et al., 2024; Depetris-Chauvin et al., 2020), we find the World Cup had, at best, only a minor impact on reducing political homophily or politically motivated blocking.

Focusing on the post-election period, the World Cup began 21 days after the election. Comparing political congruence effects on follow-backs before and after the World Cup's

[15] Appendix Table B.8 confirms that these results remain consistent under this and other sample restrictions, accounting for natural user inactivity during the experiment.

[16] Appendix Figure C.2 confirms election-related search trends align with the campaign period.

start reveals stability: political congruence raises follow-backs by 17.8 pp after the election but before the World Cup, and by 16.3 pp after the World Cup begins ($p = 0.51$).

As Depetris-Chauvin et al. (2020) suggest, the cohesive effect of shared national identity may occur only during positive team performance. Brazil's 2022 World Cup included a successful streak before elimination in the quarterfinals. If shared national identity reduces political homophily during such periods, we would expect lower polarization during Brazil's winning streak. However, comparing political congruence effects on follow-backs after the election but before the World Cup versus during Brazil's successful run reveals only a non-significant 2.4 pp decrease ($p = 0.45$). Similar results for blocks indicate that even during national team success, shared national identity had limited effect on fostering cross-partisan ties.

Given the importance of the World Cup to Brazilians,[17] we interpret this as further evidence that, in polarized contexts, political identities can overshadow other identities—such as national identity—limiting their potential to foster cross-partisan ties.

5.2 Why did the World Cup not foster cross-partisan connections?

Why, contrary to expectations, did the shared national identity highlighted by the World Cup fail to significantly foster cross-partisan cohesion in our experiment? We hypothesize that political polarization permeated the national team's players and staff, influencing supporters' reactions and hindering cross-partisan ties. Anecdotal evidence worldwide suggests national symbols lose unifying power when politicized. For instance, former President Trump claimed Americans were "happy" about the U.S. women's soccer team's defeat in July 2021, attributing it to "wokeism,"[18] and Brazil was no exception.

This section presents evidence supporting this explanation using observational Twitter data from the tournament. We show that criticism or support for players and the coach varied with political identity congruence between supporters and team members. We interpret this as further evidence that in polarized settings, political identities can overshadow others—such as national identity—impacting social interactions even in non-political domains.

5.2.1 Background, Data and Methods

Using Twitter's API, we identified all Brazil-based users who tweeted or retweeted statuses containing the hashtags pro-Lula or pro-Bolsonaro during the week before the first round of the 2022 presidential election (September 25–October 1, 2022). The hashtag list matches that used to define our experimental subject pool (see Appendix A). This yielded approximately 200,000 accounts, nearly perfectly evenly split.

[17] Over 85% of Brazilians expressed interest in the 2022 edition (TGMResearch, 2022). The national football team is a core element of Brazilian identity (DaMatta, 1994).

[18] See https://www.vox.com/22600500/olympics-conservatives-simone-biles-anti-american and https://www.washingtonpost.com/sports/interactive/2024/american-sports-grievance-culture/. Accessed February, 2025.

After identifying these accounts, we randomly sampled 10% of pro-Lula and 10% of pro-Bolsonaro users. Using Twitter's API, we collected all tweets and retweets from these users within two hours before and after Brazil's World Cup matches. For Brazil's debut against Serbia, this resulted in 230,953 tweets from 17,701 users. We classify tweet content using two methods: keyword searches for straightforward categories (e.g., specific players) and a Bayesian Classifier algorithm for abstract categories (e.g., politics).[19] For each game, we estimate the difference in the likelihood that pro-Lula versus pro-Bolsonaro users post at least one tweet on a given topic within five-minute intervals during the game. The estimating equation is:

$$\mathbb{1}\{\text{Tweeted about topic}\}_{it} = \lambda_t + \sum_{k=\underline{t}}^{\bar{t}} \beta_k \times \mathbb{1}\{k = t\} \times \mathbb{1}\{\text{Pro-Lula}\}_i + \varepsilon_{it} \qquad (3)$$

where the dependent variable is an indicator equal to one if user i at time interval t tweeted about the topic under study (for instance, tweets about Neymar), λ_t represents interval fixed effects (which account for variation on the frequency of tweets during different moments in the game), and ε_{it} is an idiosyncratic error term. We are interested in the β_t, which represents the difference in the likelihood (in percentage points) that a pro-Lula user tweeted about the topic under study at interval t, relative to a pro-Bolsonaro user. We present results with standard errors clustered at the user level, and also report uniform confidence bands (using the plug-in method from Montiel Olea and Plagborg-Møller, 2019).

We focus on two Brazil 2022 World Cup games: the debut against Serbia and the knockout match against Croatia, illustrating how political identities shaped interactions. In the opener, Richarlison—opposed to Bolsonaro—scored two winning goals, while Bolsonaro supporter Neymar was injured.[20] We examine whether fan reactions differed by political alignment. In the Croatia match, Neymar scored late, but Brazil lost on penalties, with many blaming coach Tite—seen as a Lula supporter. Comparing partisan reactions in these games helps assess political influence. Other matches offered fewer such opportunities.

5.2.2 Results

Overall, we find that players' and staff's political identities influenced fan reactions to World Cup events, potentially undermining the unifying effect national identity might have had during the tournament.

We begin with Brazil's opening game against Serbia. Appendix Figure C.3 shows tweet and retweet counts (on any topic) by pro-Lula and pro-Bolsonaro users in five-minute intervals, highlighting Richarlison's two goals. Before kickoff, both groups tweeted at similar rates—around 800 tweets per five minutes. After the game started, tweeting increased

[19] These algorithms are standard (e.g., Alrababa'h et al., 2021). For each category, we manually labeled 2,000 tweets per match to train a Naïve Bayesian Classifier to predict categories for the full dataset.

[20] Richarlison's political stance was public and widely reported (O Globo, 09-13-2022; UOL, 11-22-2022), while Neymar publicly supported Bolsonaro (G1, 09-29-2022). For instance, Bolsonaro rallied at Neymar's Institute and Neymar dedicated his first World Cup goal to Bolsonaro.

slightly for both, maintaining parallel trends. However, following Richarlison's first goal, pro-Lula accounts exhibited a pronounced spike in tweet volume unmatched by pro-Bolsonaro accounts. This disparity persisted until over an hour after Richarlison's second goal, when pro-Lula tweeting rates converged with pro-Bolsonaro levels.

These trends show reactions to Richarlison's goals differed between pro-Lula and pro-Bolsonaro users. Given his public criticism of Bolsonaro, political identities likely shaped these interactions, reducing celebrations among opponents. Estimating Equation (3) for tweets about Richarlison, Figure 5a shows pro-Lula users were 1–3 pp more likely to tweet about him after his first goal—a 25% increase from a 7.8% baseline—significant and lasting two hours.

This result supports the interpretation that Twitter users sharing Richarlison's political identity were more engaged with his goals. To explore tweet content, we use two strategies. First, Appendix Figures C.4a and C.4b present word clouds of the most frequent words in Richarlison-related tweets by pro-Lula and pro-Bolsonaro users, respectively. Pro-Lula tweets feature terms linked to Richarlison's social activism—such as "science" and "ambassador," reflecting his role as a COVID-19 ambassador for the University of São Paulo. Additionally, words like "Lula" and "voter" appear frequently among pro-Lula users, indicating emphasis on the player's political affiliation.[21]

Second, to analyze content systematically, we use a Bayesian classifier algorithm to identify tweets with political content. Figure 5c shows the difference in tweet rates about Richarlison by pro-Lula and pro-Bolsonaro accounts, distinguishing tweets with and without political content. This reveals that pro-Lula accounts are more likely to tweet about Richarlison overall, especially emphasizing political themes. Thus, pro-Lula users not only tweeted more about Richarlison after his goals but also frequently highlighted his political identity.

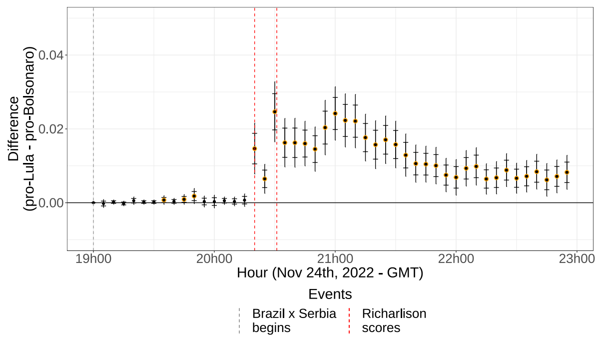
While Richarlison opposed Bolsonaro, Neymar—Brazil's most famous player—supported him during the elections. After failing to score and sustaining an injury that sidelined him for two games, how did pro-Lula and pro-Bolsonaro users react? We analyze Neymar-related tweets in Figures 5b and 5d. Pro-Lula users were slightly more likely to tweet about Neymar from the start and increasingly so after his injury, often with political content.

We find that pro-Lula users were notably likelier to celebrate Neymar's injury. Using the same Bayesian classification method as before, we categorize tweets about Neymar based on whether they express positive sentiments regarding this event, which sidelined a key player for the rest of the tournament.[22]

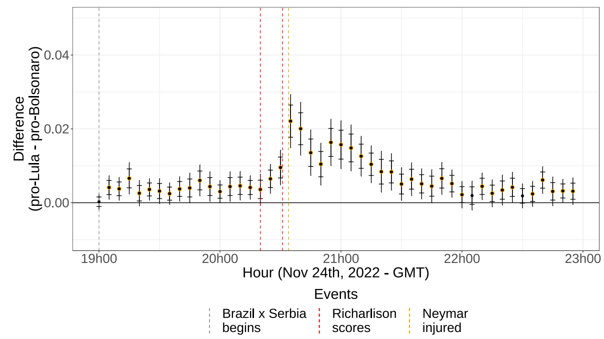
[21]To illustrate, we present three pro-Lula tweets in this context (translated by the authors): (i) "Richarlison: 2 goals, not *bolsonarista* and doesn't owe to the IRS, I am so happy"; (ii) "The only one who isn't *bolsominion*! Wonderful Richarlison" (iii)"And who scored? One of the few decent players of this tiny national team, Richarlison, c'mon! He voted for Lula, knows where he comes from, and honors his country's jersey!"

[22]We reproduce some of the tweets explicitly celebrating Neymar's injury as examples: (i) "Who needs Neymar when they have Richarlison? Apart from being against Bolsonaro, he has no debt with the IRS."; (ii) "Neymar is crying, I'm smiling."; (iii) "Neymar got hurt, cries, and supporters shout: 'So what? I'm not an orthopedist'."; (iv) "Brazil won, Richarlison scored, Neymar left the game crying. I couldn't be happier!"; (vi) "The game got so good without Neymar, the tax evader who supports a coup. I hope he doesn't return until the end of the World Cup."; (vii) "The tax evader is out of the World Cup? I can't believe God can

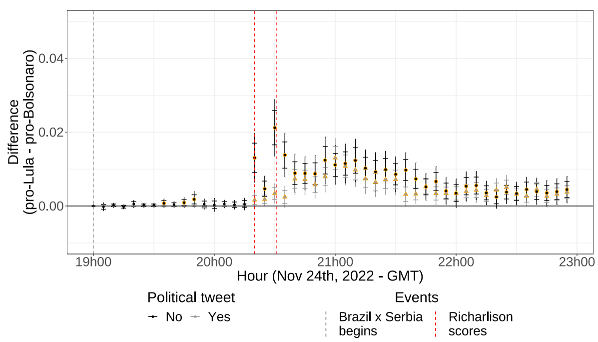
Figure 5: Difference in the number of tweets about Neymar and Richarlison between pro-Lula and pro-Bolsonaro Twitter users during Brazil × Serbia



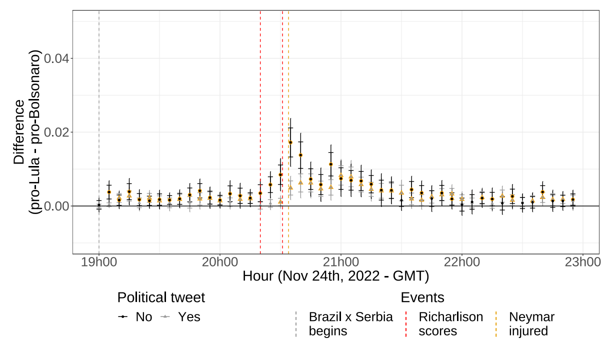
(a) Tweets about Richarlison



(b) Tweets about Neymar



(c) Tweets about Richarlison, Heterogeneity by Political Content of Tweet



(d) Tweets about Neymar, Heterogeneity by Political Content of Tweet

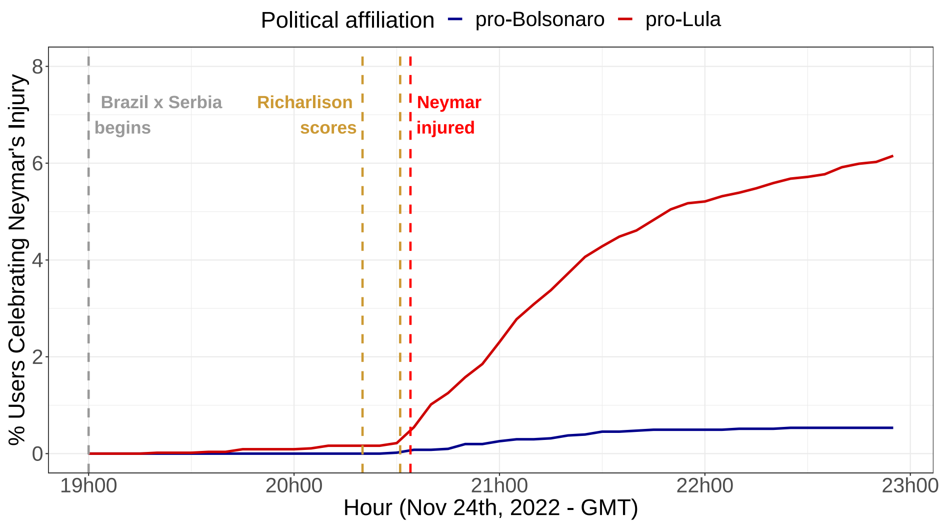
Notes: The top two figures plot the difference in the likelihood that a pro-Lula and pro-Bolsonaro account posts a tweet about a specific topic for every five-minute interval around the 2022 World Cup game between Brazil and Serbia. Figure 5a displays results for tweets about Richarlison, while 5b displays results for tweets about Neymar. We estimate Equation (3) as described in the text. Figures 5c and 5d plot a similar exercise but separate the analysis between tweets with or without political content. To classify tweets according to their content, we use a Bayesian Classifier algorithm. In all cases, data comes from a 10% random sample of all Brazilian Twitter users that tweeted or re-tweeted a status containing a pro-Lula or pro-Bolsonaro hashtag in the week before the first round of the 2022 presidential election. The error bars with ticks represent 95% confidence intervals, while the extended bars represent 95% uniform sup-t confidence bands, estimated using Montiel Olea and Plagborg-Møller (2019)'s plug-in estimator. Standard errors are clustered at the user level. Point estimates marked in orange denote estimates significant at the 5% level (point-wise).

Figure 6 shows the cumulative percentage of users posting tweets celebrating Neymar's injury by political affiliation. By the match's end, 6% of pro-Lula accounts posted such tweets compared to 0.5% of pro-Bolsonaro users. This highlights that polarization influenced fan interactions, with pro-Lula users disproportionately celebrating an event that might negatively affect the team.[23]

be that good."

[23]Word clouds of Neymar-related tweets by political group support these findings (Appendix Figures C.4c

Figure 6: Cumulative percentage of Twitter users that posted tweets celebrating Neymar's injury by user's political affiliation



Notes: The figure plots the cumulative percentage of Twitter users in our sample that posted any tweet classified as celebrating Neymar's injury, by their political affiliation (pro-Lula or pro-Bolsonaro), during Brazil's debut match against Serbia in the 2022 FIFA World Cup. Data comes from a 10% random sample of all Brazilian Twitter users that tweeted or re-tweeted a status containing a pro-Lula or pro-Bolsonaro hashtag in the week before the first round of the 2022 presidential election. To classify tweets according to whether they celebrated Neymar's injury, we use a Bayesian Classifier algorithm as described in the text.

The analyses of Neymar and Richarlison tweets converge on the conclusion that political identities mediated interactions with the national team in a polarized Brazil. Together with our experimental findings, this case study suggests that high affective polarization causes political identities to overshadow others, eroding cross-partisan social ties and altering engagement with national symbols. Political identities limit the unifying potential of collective experiences such as national team support.

This game marked a positive outcome for Brazil, which literature suggests can foster national cohesion. A contrasting case is the knockout match against Croatia, where a draw in regular time was followed by Neymar's late goal in overtime, but Brazil lost on penalties, resulting in elimination. Appendix Figure C.3 shows pro-Lula users tweeted more than pro-Bolsonaro users from the start, but no clear differences emerged following key events such as Brazil's goal and elimination.

Finally, to explore reactions to Brazil's elimination, we focus on criticism of coach Tite, perceived as a Lula supporter. Appendix Figure C.5a shows pro-Bolsonaro users tweeted about Tite 1–2 pp more than pro-Lula users every five minutes for two hours after the penalty loss. Word clouds (Appendix Figure C.6) reveal both groups criticized Tite, mentioning "vestiário" (locker room) and negative terms like arrogant and dumb. However,

and C.4d).

pro-Bolsonaro tweets contained political words such as "comunista" and "Lula," absent in pro-Lula tweets. Meanwhile, "Neymar" appeared frequently in pro-Lula tweets, indicating criticism of the pro-Bolsonaro player. These patterns confirm that political identities strongly shaped reactions.

Overall, these case studies show that political identities shaped interactions with Brazil's 2022 World Cup team amid strong affective polarization. This effect appeared in celebration and defeat. Political identity drove criticism after loss, consistent with research on out-group animosity (Hewstone et al., 2002), and influenced reactions even during victory, despite literature suggesting team success fosters cohesion (Depetris-Chauvin et al., 2020; Alrababa'h et al., 2021). Our findings suggest affective polarization limits this cohesion, as supporters identify only with politically aligned players. This explains why political homophily persisted during the World Cup and complements our experiment by showing political identity obstructs ties based on other shared identities.

6 Conclusion

Political homophily has noticeably increased in recent years. When one identity dimension—such as political affiliation—dominates public life, the potential for diverse social bonds may be diminished. This study examines the interaction between political and non-political identities, revealing that heightened political polarization weakens other bases for connection and thus restricts broader social cohesion. Using the formation of social media networks on Twitter (now X) as a case study, we investigate how political alignment shapes online social ties—a critical aspect of contemporary social sorting.

Our results demonstrate that in contexts of intense affective polarization, political identities can overshadow other identity dimensions in tie formation, with political identity actively undermining connections that might otherwise form. For example, online political homophily and animosity toward counter-partisans remained high despite a shared national event—the World Cup. We find political identities overshadowed national identity, likely because polarization extended to national team players. While sporting events can foster cohesion, their unifying potential is limited when political identities mediate interactions.

This observation carries important implications. First, by showing that affective polarization can overshadow other identity dimensions, we highlight a mechanism through which it shapes social interactions. People not only sort by political preference but also devalue other identities in forming ties, reducing opportunities for engagement with dissenting views and increasing segregation and polarization. Additionally, our findings inform debates on social media and polarization. While many argue that algorithms drive echo chambers (Sunstein, 2018), our experiment reveals that individuals actively self-select politically similar connections online. This behavior further limits exposure to opposing views and exacerbates polarization.

Nonetheless, it is important to note that sharing other identities, such as football club preference, can still foster social ties—even among political opponents. Although political identity tends to overshadow affective congruence, a modest positive effect of shared non-political identities persists. This suggests that highlighting such shared identities could be a potential pathway to reducing animosity and fostering cross-partisan ties. Future research should explore how best to leverage these alternative identities to mitigate polarization.

This paper has limitations suggesting future research directions. It is unclear if our findings apply beyond election periods when political identities are most salient. Further studies should explore whether and when political overshadowing fades. Also, our politically engaged Brazilian sample limits generalizability. Still, confirming that political identities hinder ties based on other identities in polarized contexts is crucial to understanding the effects of political polarization.

References

Ajzenman, Nicolás, Bruno Ferman, and Pedro C Sant'Anna, "Discrimination in the Formation of Academic Networks: A Field Experiment on #EconTwitter," *American Economic Review: Insights*, Forthcoming.

Alabarces, Pablo, *Futbologías: fútbol, identidad y violencia en América Latina*, CLACSO, Consejo Latinoamericano de Ciencias Sociales, 2003.

Alrababa'h, Ala, William Marble, Salma Mousa, and Alexandra A Siegel, "Can exposure to celebrities reduce prejudice? The effect of Mohamed Salah on islamophobic behaviors and attitudes," *American Political Science Review*, 2021, *115* (4), 1111–1128.

Arabe, Isadora Bousquat, "Own goal: impact of soccer matches on domestic violence in Brazil," Master's thesis, Universidade de São Paulo 2022.

Athey, Susan and Guido W Imbens, "The econometrics of randomized experiments," in "Handbook of Economic Field Experiments," Vol. 1, Elsevier, 2017, pp. 73–140.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro, "Cross-country trends in affective polarization," *The Review of Economics and Statistics*, 2022, pp. 1–60.

Bruhn, Miriam and David McKenzie, "In pursuit of balance: Randomization in practice in development field experiments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 200–232.

Bursztyn, Leonardo, Georgy Egorov, Ingar K Haaland, Aakaash Rao, and Christopher Roth, "Justifying dissent," Technical Report, National Bureau of Economic Research 2022.

_ , Ingar Haaland, Nicolas Röver, and Christopher Roth, "The Social Desirability Atlas," 2025.

Carril, Alvaro, "Dealing with misfits in random treatment assignment," *The Stata Journal*, 2017, *17* (3), 652–667.

Chen, M Keith and Ryne Rohla, "The effect of partisanship and political advertising on close family ties," *Science*, 2018, *360* (6392), 1020–1024.

DaMatta, Roberto, "Esporte na sociedade: um ensaio sobre o futebol brasileiro," *Universo do futebol: esporte e sociedade brasileira. Rio de Janeiro: Pinakotheke*, 1982, pp. 19–42.

_ , "Antropologia do óbvio: Notas em torno do significado social do futebol brasileiro," *Revista USP*, 1994, (22), 10–17.

DataSenado, "Redes Sociais, Notícias Falsas e Privacidade na Internet," 2019.

Depetris-Chauvin, Emilio, Ruben Durante, and Filipe Campante, "Building nations through shared experiences: Evidence from African football," *American Economic Review*, 2020, *110* (5), 1572–1602.

Dunning, Thad and Lauren Harrison, "Cross-cutting cleavages and ethnic voting: An experimental study of cousinage in Mali," *American Political Science Review*, 2010, *104* (1), 21–39.

Enikolopov, Ruben, Maria Petrova, Gianluca Russo, and David Yanagizawa-Drott, "Socializing Alone: How Online Homophily Has Undermined Social Cohesion in the US," *Available at SSRN*, 2024.

Epstein, Robert and Ronald E Robertson, "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections," *Proceedings of the National Academy of Sciences*, 2015, *112* (33), E4512–E4521.

Ferman, Bruno, "Assessing inference methods," *arXiv preprint arXiv:1912.08772*, 2022.

Fujiwara, Thomas, Karsten Müller, and Carlo Schwarz, "The effect of social media on elections: Evidence from the United States," Technical Report, National Bureau of Economic Research 2021.

G1, "Levantamento mostra que sete torcedores morreram durante brigas de torcidas em 2023," https://g1.globo.com/sp/sao-paulo/noticia/2023/07/11/levantamento-mostra-que-sete-torcedores-morreram-durante-brigas-de-torcidas-em-2023.ghtml 07-11-2023. Acessed: 07-26-2023.

_ , "Neymar declara apoio a Bolsonaro," https://g1.globo.com/sp/santos-regiao/eleicoes/2022/noticia/2022/09/29/neymar-declara-apoio-a-jair-bolsonaro.ghtml 09-29-2022. Acessed: 02-01-2023.

Gentzkow, Matthew, "Polarization in 2016," *Toulouse Network for Information Technology Whitepaper*, 2016, pp. 1–23.

Halberstam, Yosh and Brian Knight, "Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter," *Journal of Public Economics*, 2016, *143*, 73–88.

Hewstone, Miles, Mark Rubin, and Hazel Willis, "Intergroup bias," *Annual Review of Psychology*, 2002, *53* (1), 575–604.

Huber, Gregory A and Neil Malhotra, "Political homophily in social relationships: Evidence from online dating behavior," *The Journal of Politics*, 2017, *79* (1), 269–283.

IPEC and O Globo, "Pesquisa de Opinião Pública sobre Torcidas de Futebol," 2022.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes, "Affect, not ideologya: A social identity perspective on polarization," *Public Opinion Quarterly*, 2012, *76* (3), 405–431.

_ , Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J Westwood, "The origins and consequences of affective polarization in the United States," *Annual Review of Political Science*, 2019, *22* (1), 129–146.

Jungherr, Andreas, "Twitter use in election campaigns: A systematic literature review," *Journal of Information Technology & Politics*, 2016, *13* (1), 72–91.

Kingstone, Peter and Timothy J Power, *Democratic Brazil divided*, University of Pittsburgh Press, 2017.

LAPOP, "AmericasBarometer," 2019.

Levy, Ro'ee, "Social media, news consumption, and polarization: Evidence from a field experiment," *American Economic Review*, 2021, *111* (3), 831–70.

Lowe, Matt, "Types of contact: A field experiment on collaborative and adversarial caste integration," *American Economic Review*, 2021, *111* (6), 1807–44.

Meireles, Fernando, *genderBR: Predict Gender from Brazilian First Names* 2021. R package version 1.1.2.

Mosleh, Mohsen, Cameron Martel, and David Rand, "Psychological underpinnings of partisan bias in tie formation on social media," 2023.

_ , _ , Dean Eckles, and David G Rand, "Shared partisanship dramatically increases social tie formation in a Twitter field experiment," *Proceedings of the National Academy of Sciences*, 2021, *118* (7).

Mousa, Salma, "Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq," *Science*, 2020, *369* (6505), 866–870.

Murad, Maurício, "O lugar teórico da sociologia do futebol," *Revista Pesquisa de Campo-Núcleo de Sociologia do Futebol-UERJ*, 1995, (2).

Nielsen Sports, "World Football Report," 2022.

O Globo, "Richarlison critica uso político de camisa da Seleção: 'Faz perder a identidade'," https://oglobo.globo.com/esportes/noticia/2022/09/richarlison-evita-divisoes-politicas-de-camisa-da-selecao-faz-a-gente-perder-a-identidade.ghtml 09-13-2022. Acessed: 01-30-2023.

Olea, José Luis Montiel and Mikkel Plagborg-Møller, "Simultaneous confidence bands: Theory, implementation, and an application to SVARs," *Journal of Applied Econometrics*, 2019, *34* (1), 1–17.

Reiljan, Andres, "'Fear and loathing across party lines' (also) in Europe: Affective polarisation in European party systems," *European Journal of Political Research*, 2020, *59* (2), 376–396.

Ronconi, Juan Pedro, "Divided for Good: Football Rivalries and Social Cohesion in Latin America," 2022.

Santoro, Erik and David E Broockman, "The promise and pitfalls of cross-partisan conversations for reducing affective polarization: Evidence from randomized experiments," *Science Advances*, 2022, *8* (25).

Sayyadiharikandeh, Mohsen, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer, "Detection of novel social bots by ensembles of specialized classifiers," in "Proceedings of the 29th ACM international conference on information & knowledge management" 2020, pp. 2725–2732.

Sport Track and XP, "Convocados/XP Football Report," 2022.

Statista, "Twitter: Statistics & Facts," 2022.

— , "Twitter users in Brazil from 2017 to 2025," https://www.statista.com/forecasts/1146589/twitter-users-in-brazil#statisticContainer 2022. Accessed: 10-24-2022.

Sunstein, Cass R, *Echo chambers: Bush v. Gore, impeachment, and beyond*, Princeton University Press Princeton, NJ, 2001.

— , *# Republic: Divided democracy in the age of social media*, Princeton University Press, 2018.

TGMResearch, "TGM Global World Cup Survey 2022," https://tgmresearch.com/football-world-cup-2022-in-brazil.html 2022. Acessed: 01-30-2023.

UOL, "Destaque da Seleção, atacante Richarlison vira voz política entre jogadores," https://congressoemfoco.uol.com.br/temas/esporte/destaque-da-selecao-atacante-richarlison-vira-voz-politica-entre-jogadores/ 11-22-2022. Acessed: 01-30-2023.

Voelkel, Jan G, Michael N Stagnaro, James Y Chu, Sophia L Pink, Joseph S Mernyk, Chrystal Redekopp, Isaias Ghezae, Matthew Cashman, Dhaval Adjodah, Levi G Allen et al., "Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity," *Science*, 2024, *386* (6719), eadh4764.

Wagner, Markus, "Affective polarization in multiparty systems," *Electoral Studies*, 2021, *69*, 102199.

Yang, Kai-Cheng, Onur Varol, Pik-Mai Hui, and Filippo Menczer, "Scalable and generalizable social bot detection through data selection," in "Proceedings of the AAAI conference on artificial intelligence," Vol. 34 2020, pp. 1096–1103.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, "Political effects of the internet and social media," *Annual Review of Economics*, 2020, *12* (1), 415–438.

Online Appendix to
“Rooting for the same team: Shared social identities in a polarized context”

Nicolás Ajzenman   Bruno Ferman   Pedro C. Sant’Anna

November 18, 2025

A Additional Information on Experimental Design **2**

A.1 Procedures to Create Experimental Accounts 2

A.2 Pro-Lula and Pro-Bolsonaro Hashtags 3

A.3 Football Club Rivalries 3

A.4 Experimental Timeline 4

A.5 Procedure to Obtain the Subject Pool 4

A.6 Follow Notification 5

B Additional Figures and Tables: Twitter Experiment **6**

B.1 Affective Polarization in Brazil: Comparative Electoral Studies Survey Measure 6

B.2 Characteristics of Brazilian Football Club Supporters 7

B.3 Descriptive Statistics of the Subject Pool 8

B.4 Balance, Attrition, and Take-up 10

B.5 Main Results: Comparison of Results across Treatment Arms and Robustness 13

B.6 Experiment with Fictional accounts with more Salient Political Identity . . . 16

B.7 Other Robustness Exercises 19

B.8 Demand for Information *versus* Social Connections 25

C Additional Results for the Analysis of Formation of Ties over Time **28**

A Additional Information on Experimental Design

A.1 Procedures to Create Experimental Accounts

Table A.1: Procedures Used to Create the Experimental Accounts

Element of Profile	Procedure
Profile Picture	For the accounts that signal their preferred team, the profile picture is a photo of the team's logo in a flag inside a stadium; for the team-neutral accounts, the profile picture is a photo of the interior of a foreign football stadium during a football game (we chose photos in which the teams that were playing could not be identified). In all cases, we have a set of possible images, which are randomly chosen to construct each bot.
Name	Randomly generated by matching a list of the most common male first names and surnames in Brazil.
Bio	The Bio from the fictional accounts contains two pieces of information: first, it either says "Supporter of team X" (if the account signals her preferred team) or "football fan" (if the account is team-neutral); second, it includes either the hashtag "#Lula2022" or "#Bolsonaro2022" (depending on the bot's political identity). For the politically-neutral accounts, we merely remove this second part.
Background Image	A landscape from the city where the account's preferred football team plays its home matches (and random city landscape for the football team-neutral accounts).
Location	The fictional accounts' profiles do not include a location. 25.5% of subjects' profiles do not include a location.
Website	The fictional accounts' profiles do not include a website. 82.7% of subjects' profiles do not include a website url.
Retweets	The fictional accountfirst re-tweets a post from an account related to her preferred football team or, in the case of team-neutral accounts, a general tweet about football (that isn't specific about any football team). Then, the account re-tweets a post from its preferred political candidate. The post must necessarily have more than 500 re-tweets and not include any misleading information or hate speech. This way, the first post that is seen when someone accesses the bot's profile is the one that signals political identity.
Followers	We asked a group of colleagues to follow the fictional accounts before each experimental wave so that the fictional accounts have some followers when subjects receive the notifications.
Following	One day before following the accounts randomly assigned to it, the fictional accountwill follow a set of "elite" accounts related to its political identity and preferred team (for instance, it will follow the team's official profile, the profile of its preferred candidate and of some of its allies).

Notes: The table summarizes the procedures used to create the fictional accounts. Figure 1 shows examples of accounts.

A.2 Pro-Lula and Pro-Bolsonaro Hashtags

Table A.2: List of pro-Lula and pro-Bolsonaro hashtags used to build the subject pool

Pro-Lula	Pro-Bolsonaro
#Lula2022	#Bolsonaro2022
#Lula22	#Bolsonaro22
#Lula13	#FechadoComBolsonaro
#LulaPresidente	#BolsonaroReeleito
#LulaNoPrimeiroTurno	#BolsonaroNoPrimeiroTurno
#VamosJuntosPeloBrasil	#BolsonaroOrgulhoDoBrasil
#JuntosComLula	#JuntosComBolsonaro
#BrasilComLula	#BrasilComBolsonaro

A.3 Football Club Rivalries

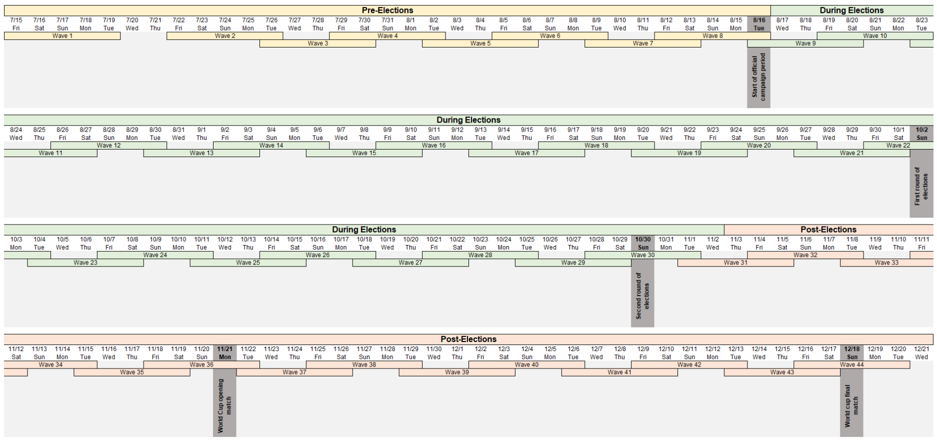
Table A.3: Football club rivalries

	Botafogo	Flamengo	Fluminense	Vasco	Corinthians	Palmeiras	Santos	São Paulo	Grêmio	Internacional
Flamengo	X	✓	X	X						
Vasco	X	X	X	✓						
Corinthians					✓	X	X	X		
Palmeiras					X	✓	X	X		
São Paulo					X	X	X	✓		
Grêmio									✓	X

Notes: The table displays the football club rivalries we considered when constructing the sample of subjects. The X mark indicates a rivalry. A fictional account that signals support for team A will only follow subjects whose preferred football club is either team A or team A's rival. We restricted ourselves to regional (inter-state rivalries). The clubs in the rows are the ones that a fictional account may support, while the clubs in the columns are the ones that subjects may support.

A.4 Experimental Timeline

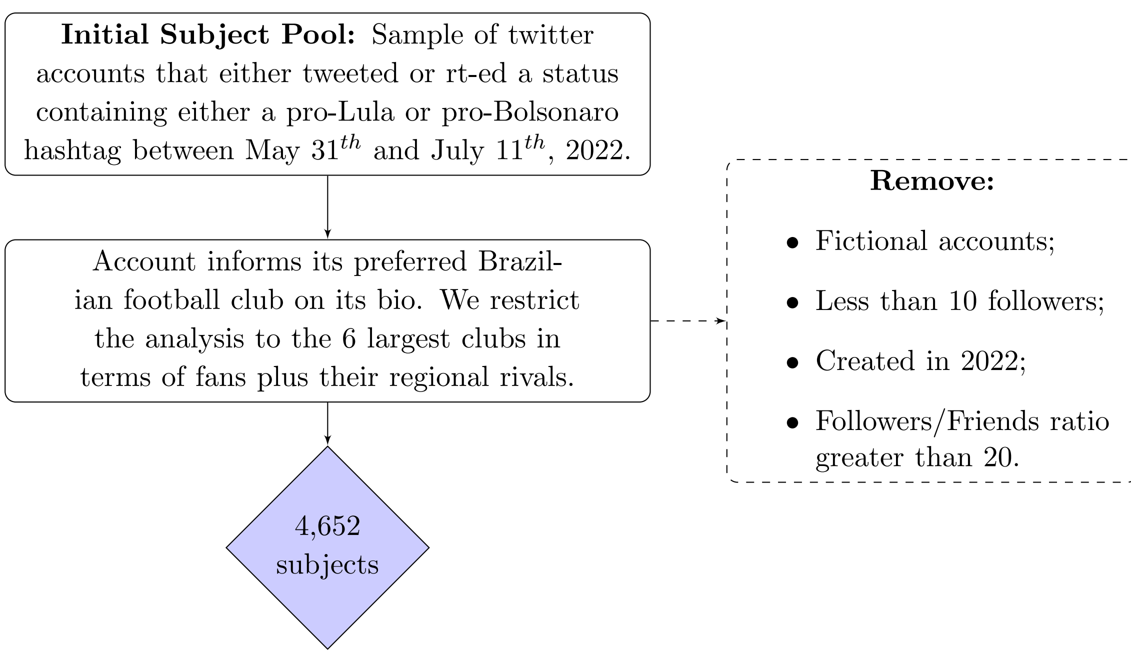
Figure A.1: Experimental Timeline



Notes: The table shows the experimental timeline. We consider that each wave starts at the moment in which the fictional accounts follow the subjects. The table also shows the periods we define as before, during, and after the election period, along with relevant dates.

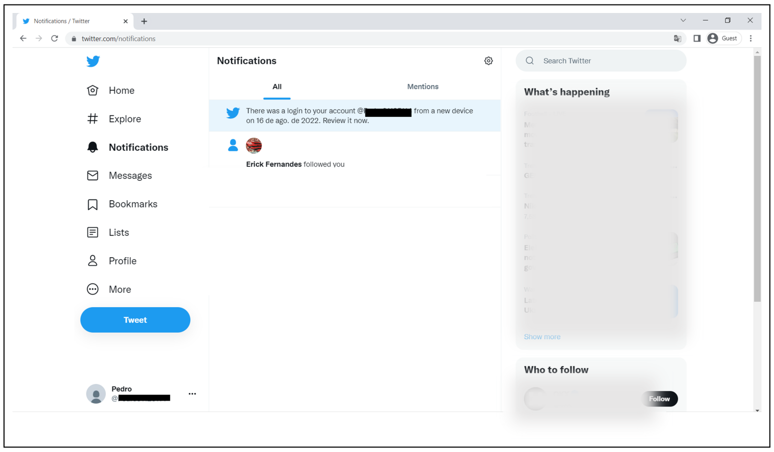
A.5 Procedure to Obtain the Subject Pool

Figure A.2: Procedure to obtain the subject pool

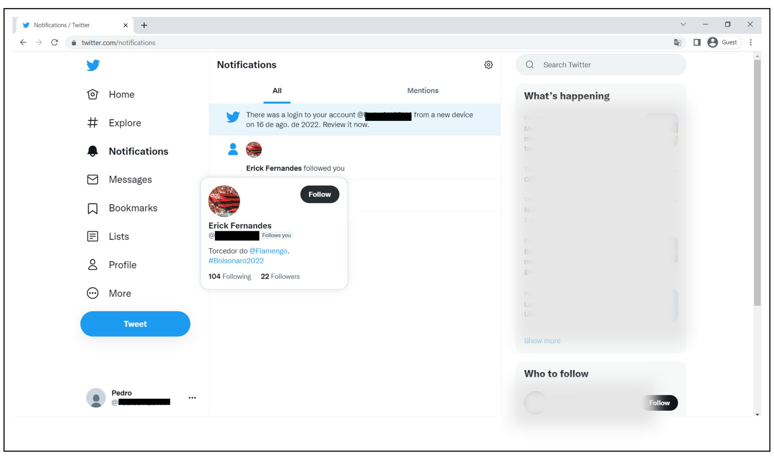


A.6 Follow Notification

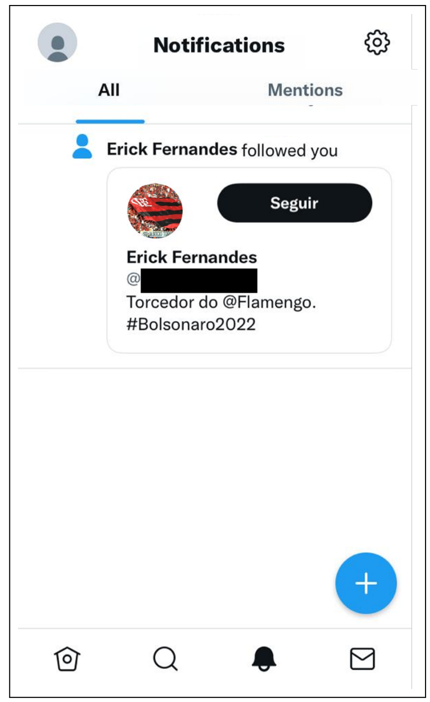
Figure A.3: Example of treatment notifications on desktop and mobile Twitter apps



(a) Desktop Notification



(b) Desktop Notification (after hovering the mouse's cursor over the bot's profile)

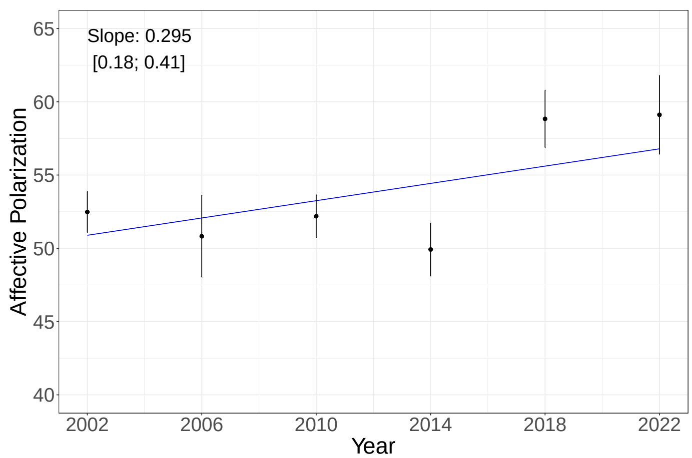


(c) Mobile app notification

B Additional Figures and Tables: Twitter Experiment

B.1 Affective Polarization in Brazil: Comparative Electoral Studies Survey Measure

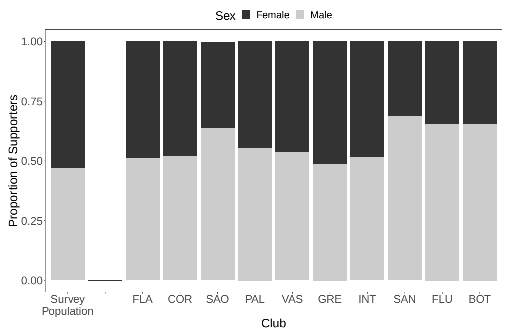
Figure B.1: Trends in Affective Polarization, Brazil (Boxell et al. (2022)'s method)



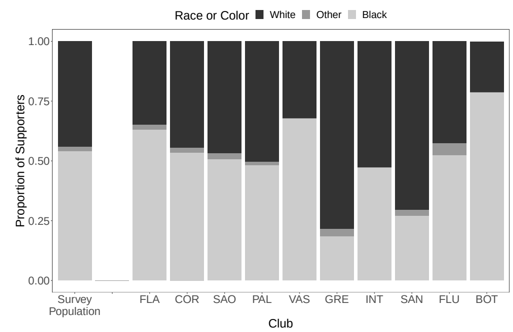
Notes: The figure presents trends in affective polarization in Brazil, using data from the Brazilian Electoral Study (BES), a national post-electoral survey undertaken since 2002. Following Boxell et al. (2022), we estimate affective polarization as the mean difference between in-party and out-party feeling among respondents who claim to identify with a given party. The question from which we construct the measures of in- and out-party feeling is "I'd like to know what you think about each of our political parties. After I read the name of a political party, please rate it on a scale from 0 to 10, where 0 means you strongly dislike that party and 10 means that you strongly like that party. If I come to a party you haven't heard of or you feel you do not know enough about, just say so. The first party is PARTY A." Error bars display 95% confidence intervals for the affective polarization index in each election year, and the blue line displays a fitted bivariate linear regression line with affective polarization as the dependent variable and the survey year as the independent one. The plot reports the slope (change per year) and estimated 95% confidence interval computed using heteroskedasticity-robust standard errors in the top-left.

B.2 Characteristics of Brazilian Football Club Supporters

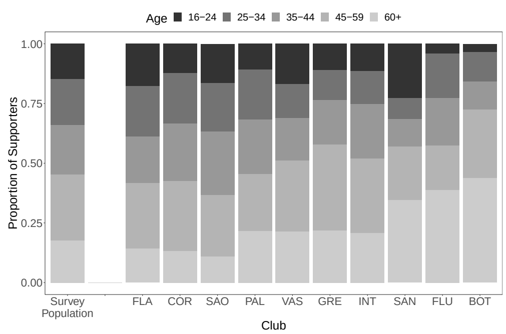
Figure B.2: Characteristics of Brazilian Football Club Supporters



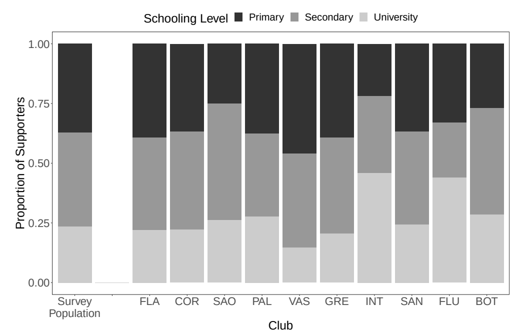
(a) Sex



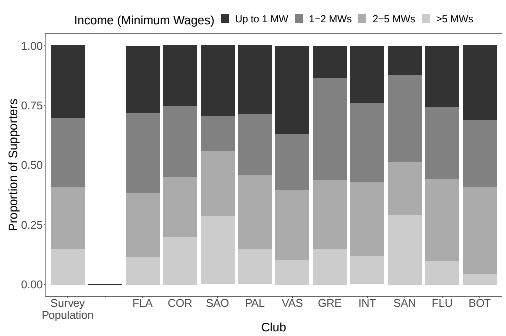
(b) Race



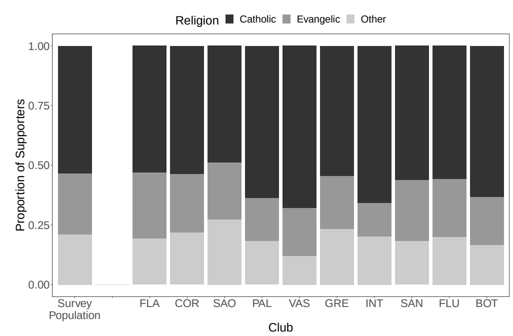
(c) Age



(d) Education Level



(e) Income



(f) Religion

Notes: The figures show the proportion of supporters of each of the six most popular Brazilian clubs and its rivals across socio-economic characteristics. Data comes from IPEC and O Globo (2022). The left-most bar in each plot shows the proportion with each characteristic in the survey population. Clubs are ordered by number of supporters.

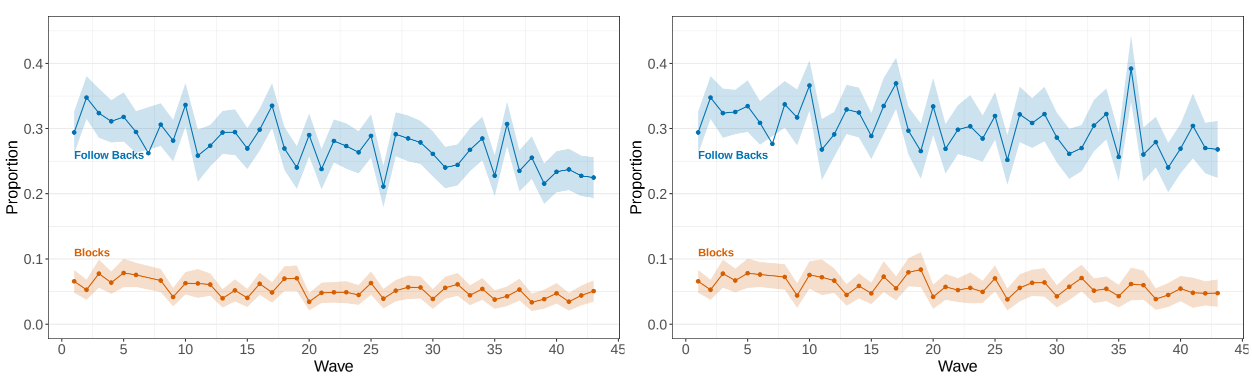
B.3 Descriptive Statistics of the Subject Pool

Table B.1: Descriptive Statistics of the Subject Pool - Numerical Variables

Variables	Mean	Median	Std. Deviation	Min	Max	Obs.
Number of followers	2047.66	662	5685.92	11	141490	4652
Number of friends	2289.28	1057	5078.55	8	137451	4652
Number of statuses (tweets + rts)	25439.26	8050	58732.38	4	1665213	4652
Number of favorites (likes)	42152.19	17398	72984.66	0	1618281	4652
Year of account creation	2015.28	2016	4.66	2006	2022	4652
Botometer score	0.2	0.13	0.2	0	0.98	3878

Notes: The table shows summary statistics for the subject pool in the experiment. '*Botometer* score' is a number between 0 and 1 generated by the *Botometer* API, which determines the probability that each subject is classified as an automated account. A higher score means that the account is more likely to be automated.

Figure B.3: Evolution of Treatment Take-up



(a) All Subjects (b) Active Subjects

Notes: The figures display the evolution of experimental take up across experimental waves. The first figures considers all subjects, while the second is conditional on subjects who were active (i.e., tweeted or re-tweeted) at least 24 hours before treatment. The shaded areas correspond to 95% confidence intervals.

Table B.2: Descriptive Statistics of Subject Pool

Variables	% Classified	N	%
Political Identity	100		
Bolsonaro		2069	44.48
Lula		2583	55.52
Affective Identity			
Corinthians	100	566	12.17
pro-Bolsonaro		156	27.56
pro-Lula		410	72.44
Palmeiras	100	485	10.43
pro-Bolsonaro		293	60.41
pro-Lula		192	39.59
São Paulo	100	403	8.66
pro-Bolsonaro		219	54.34
pro-Lula		184	45.66
Santos	100	165	3.55
pro-Bolsonaro		74	44.85
pro-Lula		91	55.15
Flamengo	100	1342	28.85
pro-Bolsonaro		641	47.76
pro-Lula		701	52.24
Vasco	100	447	9.61
pro-Bolsonaro		179	40.04
pro-Lula		268	59.96
Botafogo	100	245	5.27
pro-Bolsonaro		102	41.63
pro-Lula		143	58.37
Fluminense	100	172	3.70
pro-Bolsonaro		69	40.12
pro-Lula		103	59.88
Grêmio	100	258	5.55
pro-Bolsonaro		118	45.74
pro-Lula		140	54.26
Internacional	100	210	4.51
pro-Bolsonaro		80	38.10
pro-Lula		130	61.90
Region	64.23		
Center-West		216	7.23
pro-Bolsonaro		117	54.17
pro-Lula		99	45.83
Northeast		379	12.68
pro-Bolsonaro		122	32.19
pro-Lula		257	67.81
North		123	4.12
pro-Bolsonaro		58	47.15
pro-Lula		65	52.85
Southeast		1746	58.43
pro-Bolsonaro		760	43.53
pro-Lula		986	56.47
South		418	13.99
pro-Bolsonaro		199	47.61
pro-Lula		219	52.39
Foreign		106	3.55
pro-Bolsonaro		67	63.21
pro-Lula		39	36.79
Gender	81.17		
Female		844	22.35
pro-Bolsonaro		268	31.75
pro-Lula		576	68.25
Male		2932	77.65
pro-Bolsonaro		1462	49.86
pro-Lula		1470	50.14
Has background pic.	100	3930	84.48
pro-Bolsonaro		1689	42.98
pro-Lula		2241	57.02
Has website	100	804	17.28
pro-Bolsonaro		253	31.47
pro-Lula		551	68.53

Notes: The table displays summary statistics for the subject pool. Figure A.2 describes the procedure used to obtain the subjects. The variable political identity is obtained accordingly to the hashtag used by the subject, while affective identity is obtained from information in the subject's bios. Region is created using self-declared information in the "location" field of the profile, which we recode to the regional level. % Classified is the percentage of all subjects for which we were able to obtain the variable. For each variable, we indicate the number of subjects (N) and the proportion of subjects in each category (the proportion is relative to the number of classified subjects). Finally, for each category, we show the proportion of subjects who are pro-Lula or pro-Bolsonaro. The variable Gender is obtained by using Brazilian Census data (organized by Meireles (2021)) to compute the proportion of men and women with each given name in the sample. A gender is assigned to a subject if at least 90% of his or her name's occurrences in the 2010 census were of an specific gender.

B.4 Balance, Attrition, and Take-up

Table B.3: Balance Table

	Treatment Arm								
	Both Dimensions				Affectively Neutral Accounts		Politically Neutral Accounts		
Variable	In-politics; In-affective	In-politics; Out-affective	Out-politics; In-affective	Out-politics; Out-affective	In-politics; Neutral-affective	Out-politics; Neutral-affective	Neutral-politics; In-affective	Neutral-politics; Out-affective	F Stat [p-value]
Number of followers	1,858.1	1,939.5	1,826.5	2,032.1	2,077.4	1,962	1,839	2,032.2	0.0137 [1.00]
	(4,899.3)	(4,816.2)	(4,584.2)	(5,387.3)	(6,220)	(5,001.6)	(5,067.6)	(5,987.4)	
Number of friends	2,190.1	2,191.7	2,074.4	2,302.9	2,313.7	2,221.4	2,132.8	2,312.9	0.0146 [1.00]
	(4,548.3)	(3,898.6)	(3,958.3)	(4,779.3)	(5,556)	(4,368.3)	(4,587.9)	(5,494.9)	
Number of statuses ('tweets + rts')	24,448	24,873.2	25,061.3	24,909.8	24,775.2	25,720.6	24,168.4	26,130.1	0.0055 [1.00]
	(55,867.4)	(51,507.3)	(56,480.8)	(53,622.5)	(51,148.7)	(50,935.8)	(60,306.6)	(64,734.1)	
Number of favorited statuses ('likes')	43,139.1	43,136.6	44,731.2	40,517.3	44,915.1	41,968.2	42,112	42,084.4	0.0154 [1.00]
	(87,867.7)	(73,385.3)	(83,492.9)	(63,641.3)	(82,595.3)	(69,698.8)	(81,036.1)	(71,119.9)	
Number of lists	4.024	4.164	4.133	4.33	4.056	3.619	3.157	4.184	0.0119 [1.00]
	(24.8)	(20.1)	(28.8)	(25.2)	(20.5)	(13.4)	(10)	(19.1)	
Account is verified	0.001	0.001	0.002	0.001	0.002	0.002	0	0.002	0.0129 [1.00]
	(0.033)	(0.028)	(0.043)	(0.023)	(0.039)	(0.046)	(0.018)	(0.04)	
Year of account creation	2,015.1	2,015.1	2,015.2	2,015.1	2,015	2,015.2	2,015.1	2,015	0.0104 [1.00]
	(4.599)	(4.689)	(4.582)	(4.653)	(4.585)	(4.59)	(4.599)	(4.655)	
Has background picture	0.839	0.843	0.841	0.838	0.839	0.83	0.833	0.838	0.0054 [1.00]
	(0.368)	(0.363)	(0.366)	(0.368)	(0.368)	(0.376)	(0.373)	(0.368)	
Gender (1=Female)	0.173	0.172	0.175	0.184	0.169	0.186	0.188	0.179	0.0138 [1.00]
	(0.378)	(0.377)	(0.38)	(0.387)	(0.375)	(0.389)	(0.391)	(0.383)	
Region									
Center-West	0.043	0.036	0.041	0.031	0.042	0.041	0.045	0.041	0.0214 [1.00]
	(0.202)	(0.186)	(0.198)	(0.172)	(0.2)	(0.199)	(0.207)	(0.198)	
Northeast	0.065	0.064	0.075	0.064	0.067	0.07	0.082	0.06	0.0311 [1.00]
	(0.246)	(0.244)	(0.264)	(0.245)	(0.25)	(0.255)	(0.275)	(0.238)	
North	0.021	0.017	0.024	0.023	0.023	0.021	0.032	0.016	0.0443 [1.00]
	(0.144)	(0.128)	(0.154)	(0.15)	(0.15)	(0.143)	(0.177)	(0.126)	
Southeast	0.311	0.335	0.303	0.329	0.311	0.295	0.329	0.335	0.0458 [1.00]
	(0.463)	(0.472)	(0.459)	(0.47)	(0.463)	(0.456)	(0.47)	(0.472)	
South	0.082	0.072	0.082	0.07	0.072	0.071	0.09	0.071	0.0283 [1.00]
	(0.274)	(0.258)	(0.274)	(0.255)	(0.259)	(0.258)	(0.286)	(0.257)	
Foreign	0.02	0.02	0.02	0.02	0.021	0.018	0.015	0.016	0.0113 [1.00]
	(0.139)	(0.141)	(0.141)	(0.14)	(0.144)	(0.133)	(0.12)	(0.126)	
Number of treated observations	3783	3761	3790	3794	3845	3833	3003	4385	
%	0.125	0.125	0.126	0.126	0.127	0.127	0.099	0.145	
Attrition (not treated)	379	415	396	384	346	363	356	378	0.0367 [1.00]
% of assigned to treatment	0.091	0.099	0.095	0.092	0.083	0.087	0.106	0.079	
Always active (tweeted every week)	2863	2830	2900	2923	2901	2908	2300	3353	0.0091 [1.00]
% of treated	0.757	0.752	0.765	0.77	0.754	0.759	0.766	0.765	
Active 1 day before treatment	2965	2948	2947	2994	3030	2969	2253	3435	0.0319 [1.00]
% of treated	0.784	0.784	0.778	0.789	0.788	0.775	0.75	0.783	

Notes: The table displays average and standard deviations for subject-level variables across the eight treatment arms in the experiment. The F-statistic is computed from a regression of the pre-treatment variable on the treatment indicators. For all pre-treatment variables, we cannot reject the null hypothesis of equality of means across all eight treatments. The row "Number of treated obs." shows the number of treated observations (i.e., accounts followed by a bot) for each treatment arm, while "%" shows the percentage treated among all treated participants. The row "Attrition" shows the number of participants assigned to each treatment that could not be treated (either because they de-activated their account, were suspended by Twitter, or chose to make their profile private). The row "Always active" show the number and proportion of subjects that tweeted at least once in the seven days before every experimental wave (not only those in which they were specifically treated), while "Active 1 day before treatment" show the number of subjects who had Twitter activity (tweets or re-tweets) one day before treatment.

Table B.4: Balance Table - Attrited subjects

	Treatment Arm								
	Both Dimensions				Affectively Neutral Accounts		Politically Neutral Accounts		
Variable	In-politics; In-affective	In-politics; Out-affective	Out-politics; In-affective	Out-politics; Out-affective	In-politics; Neutral-affective	Out-politics; Neutral-affective	Neutral-politics; In-affective	Neutral-politics; Out-affective	F Stat [p-value]
Number of followers	2,632.7	2,230.8	2,639.5	2,037.3	3,466.9	1,919.2	2,968.2	3,806.9	0.3768 [0.916]
	(6,739.1)	(5,147.4)	(6,707.1)	(4,794.8)	(8,311.2)	(3,758.9)	(8,108.2)	(9,896.4)	
Number of friends	2,913.8	2,474.6	2,944.9	2,357.5	3,666.6	2,281.9	2,975.4	3,832.6	0.3385 [0.936]
	(5,836.6)	(4,405.6)	(6,389.1)	(4,600.4)	(7,612.2)	(3,614.7)	(7,165.1)	(8,813.5)	
Number of statuses ('tweets + rts')	32,189.6	27,035.1	29,585.8	25,845.2	34,110.6	20,760.1	23,967.5	25,613.6	0.1369 [0.995]
	(102,192.6)	(54,890.2)	(97,306.3)	(58,804.4)	(105,093.3)	(35,136.7)	(55,960.2)	(48,499)	
Number of favorited statuses ('likes')	44,705.6	45,693	39,562.4	39,206.7	38,030.3	40,212.8	34,530.7	44,454.2	0.1215 [0.997]
	(68,391)	(83,868.8)	(65,426.9)	(66,475.1)	(62,098.3)	(66,975.1)	(66,181.8)	(72,858.7)	
Number of lists	3.011	2.949	4.869	2.227	4.107	1.683	2.408	2.705	0.1077 [0.998]
	(12.4)	(11.7)	(48.4)	(8.2)	(16)	(5.7)	(7.5)	(8.5)	
Account is verified	0	0	0	0	0	0	0	0.003	0.1114 [0.998]
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0.052)	
Year of account creation	2,017.4	2,017.5	2,017.6	2,017.4	2,017.1	2,017.3	2,018	2,017.7	0.1194 [0.997]
	(4.514)	(4.545)	(4.687)	(4.653)	(4.742)	(4.704)	(4.527)	(4.455)	
Has background picture	0.868	0.913	0.866	0.854	0.874	0.889	0.832	0.862	0.2075 [0.984]
	(0.339)	(0.282)	(0.341)	(0.353)	(0.333)	(0.315)	(0.374)	(0.345)	
Gender (1=Female)	0.108	0.137	0.096	0.112	0.126	0.13	0.168	0.16	0.2175 [0.981]
	(0.311)	(0.345)	(0.295)	(0.316)	(0.333)	(0.336)	(0.374)	(0.367)	
Region									
Center-West	0.018	0.036	0.025	0.023	0.031	0.026	0.064	0.039	0.2406 [0.975]
	(0.135)	(0.187)	(0.157)	(0.151)	(0.173)	(0.161)	(0.244)	(0.193)	
Northeast	0.063	0.036	0.056	0.049	0.045	0.058	0.046	0.061	0.0716 [0.999]
	(0.244)	(0.187)	(0.229)	(0.217)	(0.207)	(0.234)	(0.21)	(0.239)	
North	0.026	0.024	0.033	0.026	0.014	0.026	0.029	0.033	0.0537 [1.00]
	(0.16)	(0.154)	(0.178)	(0.159)	(0.118)	(0.161)	(0.168)	(0.179)	
Southeast	0.311	0.328	0.263	0.339	0.287	0.304	0.26	0.298	0.1537 [0.993]
	(0.464)	(0.47)	(0.441)	(0.474)	(0.453)	(0.461)	(0.439)	(0.458)	
South	0.071	0.063	0.076	0.102	0.104	0.087	0.061	0.085	0.1438 [0.995]
	(0.258)	(0.243)	(0.265)	(0.302)	(0.306)	(0.283)	(0.239)	(0.28)	
Foreign	0.021	0.034	0.018	0.036	0.02	0.032	0.026	0.028	0.0751 [0.999]
	(0.144)	(0.181)	(0.132)	(0.188)	(0.139)	(0.176)	(0.159)	(0.164)	
Attrition (not treated)	379	415	396	384	356	378	346	363	
% of assigned to treatment	0.091	0.099	0.095	0.092	0.106	0.079	0.083	0.087	

Notes: This table shows the average and standard deviations (in parentheses) of pre-treatment variables for subjects that suffered attrition at some point of the experiment. The last column in the table reports a F-test of joint equality of means across all treatment arms.

Table B.5: Differences between accounts that ever suffered attrition or did not

Variable	Never Attrited	Ever Attrited	T Stat [p-value]
Political identity (1=pro-Bolsonaro)	0.415	0.582	8.9246 [0.00]***
	(0.493)	(0.494)	
Number of followers	1,888.5	2,745.1	3.3879 [0.001]***
	(5,357.4)	(6,923.8)	
Number of friends	2,159	2,886.7	3.2766 [0.001]***
	(4,833.2)	(6,059.9)	
Number of statuses ('tweets + rts')	24,035.8	31,123.1	2.3951 [0.017]**
	(50,720.5)	(82,900.1)	
Number of favorited statuses ('likes')	41,183.4	46,212.4	1.7953 [0.073]*
	(72,139.2)	(74,208.5)	
Number of lists	4.143	4.235	0.0751 [0.94]
	(19.9)	(34.4)	
Account is verified	0.002	0.001	-0.3066 [0.759]
	(0.04)	(0.034)	
Year of account creation	2,014.9	2,016.8	10.3188 [0.00]***
	(4.568)	(4.748)	
Has background picture	0.841	0.861	1.5468 [0.122]
	(0.366)	(0.346)	
Gender (1=Female)	0.23	0.19	-2.3476 [0.019]**
	(0.421)	(0.393)	
Region			
Center-West	0.073	0.065	-0.662 [0.508]
	(0.261)	(0.247)	
Northeast	0.134	0.094	-2.7021 [0.007]***
	(0.34)	(0.292)	
North	0.038	0.057	1.68 [0.093]*
	(0.192)	(0.232)	
Southeast	0.578	0.615	1.5306 [0.126]
	(0.494)	(0.487)	
South	0.143	0.126	-1.0113 [0.312]
	(0.35)	(0.332)	
Foreign	0.033	0.043	0.9493 [0.343]
	(0.18)	(0.203)	
Number of observations	3782	851	
%	0.816	0.184	

Notes: The table compares average characteristics of subjects that never suffered attrition throughout all experimental waves ("never attrited") and those that suffered attrition at some point ('ever attrited'). Standard deviations are in parentheses. A subject is considered to have suffered attrition if we cannot find its account or cannot follow it on Twitter, which can happen if the user is suspended, deactivated its accounts, or made it private. The last column of the table displays the t-statistic and p-value of a test of difference in means for the respective variable between the two groups. Significance codes: $^{***}: p < 0.01$, $^{**}: p < 0.05$, $^{*}: p < 0.1$.

B.5 Main Results: Comparison of Results across Treatment Arms and Robustness

Table B.6: Differences in Average Follow-Back Rate Across Treatment Arms

i/j	Out; Out		Out; No Signal		Out; In		No Signal; Out		No Signal; In		In; Out		In; No Signal		In; In	
Out-politics; Out-affective	$\Delta_{raw}(j-i)$ (Std. Error)	$\Delta_{FE,Controls}(j-i)$ (Std. Error)	0.009 (0.011)	0.012 (0.01)	0.044*** (0.009)	0.044*** (0.009)	0.069*** (0.011)	0.045*** (0.01)	0.203*** (0.014)	0.188*** (0.013)	0.164*** (0.011)	0.164*** (0.011)	0.209*** (0.013)	0.21*** (0.011)	0.249*** (0.012)	0.244*** (0.012)
Out-politics; No signal affective					0.035*** (0.012)	0.035*** (0.012)	0.061*** (0.012)	0.037*** (0.011)	0.194*** (0.014)	0.18*** (0.013)	0.155*** (0.012)	0.151*** (0.011)	0.2*** (0.015)	0.199*** (0.015)	0.24*** (0.012)	0.228*** (0.012)
Out-politics; In-affective							0.026** (0.012)	-0.002 (0.012)	0.159*** (0.014)	0.145*** (0.013)	0.12*** (0.013)	0.121*** (0.012)	0.165*** (0.014)	0.162*** (0.013)	0.205*** (0.012)	0.204*** (0.012)
No signal politics; Out-affective									0.134*** (0.013)	0.141*** (0.013)	0.095*** (0.012)	0.117*** (0.011)	0.139*** (0.014)	0.163*** (0.012)	0.179*** (0.012)	0.193*** (0.013)
No signal politics; In-affective											-0.039*** (0.015)	-0.033*** (0.013)	0.006 (0.016)	0.021 (0.014)	0.046*** (0.015)	0.042*** (0.013)
In-politics; Out-affective													0.045*** (0.014)	0.046*** (0.012)	0.085*** (0.011)	0.079*** (0.011)
In-politics; No signal affective															0.04*** (0.014)	0.031** (0.013)

Notes: The table displays differences in average follow-back rate between treatment arms. Each column or row represents one of the eight treatment arms in the experiment (the same ones displayed in Figure 2). The treatment arms are defined by whether fictional account and subject have congruent or incongruent identities in the political and affective (football club preference) dimensions. For each dimension (political or affective) we denote congruence using the term "in", and incongruence with the term "out" (as in "in-group" and "out-group" ties). A third option is that the fictional account does not signal the dimension. For each treatment arm, we first inform the relationship between fictional account and subject's political identity, and then affective (for example, "in; out" means that fictional account and subject share political identity and support rival clubs). Each table cell shows estimates and standard deviations for the difference in the average follow-back rate between the column and the row-treatment arm. In each cell, we report the raw difference between the groups, and the estimate including wave and strata fixed effects. Standard errors clustered at the bot-account level are in parentheses. Significance codes: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

Table B.7: Differences in Average Blocking Rate Across Treatment Arms

i/j	Out; Out		Out; No Signal		Out; In		No Signal; Out		No Signal; In		In; Out		In; No Signal		In; In	
Out-politics; Out-affective	$\Delta_{raw}(j-i)$ (Std. Error)	$\Delta_{FE,Controls}(j-i)$ (Std. Error)	-0.017* (0.009)	-0.018** (0.009)	-0.061*** (0.008)	-0.06*** (0.008)	-0.124*** (0.008)	-0.123*** (0.008)	-0.137*** (0.007)	-0.127*** (0.007)	-0.136*** (0.007)	-0.135*** (0.007)	-0.139*** (0.007)	-0.141*** (0.008)	-0.141*** (0.007)	-0.14*** (0.007)
Out-politics; No signal affective					-0.044*** (0.008)	-0.044*** (0.007)	-0.107*** (0.006)	-0.104*** (0.006)	-0.12*** (0.006)	-0.121*** (0.006)	-0.119*** (0.006)	-0.119*** (0.006)	-0.123*** (0.006)	-0.123*** (0.006)	-0.124*** (0.006)	-0.123*** (0.006)
Out-politics; In-affective							-0.062*** (0.006)	-0.063*** (0.006)	-0.076*** (0.005)	-0.072*** (0.006)	-0.074*** (0.005)	-0.074*** (0.005)	-0.078*** (0.005)	-0.077*** (0.005)	-0.08*** (0.005)	-0.08*** (0.005)
No signal politics; Out-affective									-0.013*** (0.003)	-0.013*** (0.003)	-0.012*** (0.003)	-0.013*** (0.003)	-0.016*** (0.003)	-0.015*** (0.002)	-0.017*** (0.003)	-0.016*** (0.003)
No signal politics; In-affective											0.001 (0.003)	0 (0.003)	-0.002 (0.002)	-0.002 (0.002)	-0.004 (0.002)	-0.003 (0.002)
In-politics; Out-affective													-0.004* (0.002)	-0.003 (0.002)	-0.005** (0.002)	-0.006*** (0.002)
In-politics; No signal affective															-0.002 (0.002)	-0.002 (0.002)

Notes: The table displays differences in average blocking rate between treatment arms. Each column or row represents one of the eight treatment arms in the experiment (the same ones displayed in Figure 3). The treatment arms are defined by whether fictional account and subject have congruent or incongruent identities in the political and affective (football club preference) dimensions. For each dimension (political or affective) we denote congruence using the term "in", and incongruence with the term "out" (as in "in-group" and "out-group" ties). A third option is that the fictional account does not signal the dimension. For each treatment arm, we first inform the relationship between fictional account and subject's political identity, and then affective (for example, "in; out" means that fictional account and subject share political identity and support rival clubs). Each table cell shows estimates and standard deviations for the difference in the average blocking rate between the column and the row-treatment arm. In each cell, we report the raw difference between the groups (column − row), and the estimate including wave and strata fixed effects. Standard errors clustered at the bot-account level are in parentheses. Significance codes: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

Table B.8: Main Results for Different Sub-samples: Experimental accounts that signal both dimensions of identity

Panel A: Follow Backs							
	Dependent Variable: Follow Backs (1 = Yes)						
	Full Sample			Never attrited	Tweeted every week	Active (1 day)	Unlikely to be automated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Political congruence	0.1639***	0.1643***	0.1476***	0.1439***	0.1622***	0.1606***	0.1398***
	(0.0108)	(0.0106)	(0.0139)	(0.0139)	(0.0165)	(0.0166)	(0.0172)
Affective congruence	0.0437***	0.0424***	0.0512***	0.0473***	0.0597***	0.0551***	0.0532***
	(0.0087)	(0.0087)	(0.0114)	(0.0129)	(0.0145)	(0.0136)	(0.0154)
Political congruence × Affective congruence	0.0411***	0.0387***	0.0503***	0.0531***	0.0364*	0.0521**	0.0461*
	(0.0129)	(0.0127)	(0.0170)	(0.0184)	(0.0211)	(0.0200)	(0.0267)
Wave, Strata Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes	Yes
Observations	15,128	15,128	15,128	13,257	9,953	11,854	6,814
R^2	0.04856	0.08886	0.09909	0.09795	0.10527	0.10199	0.10824

Panel B: Blocks							
	Dependent Variable: Blocks (1 = Yes)						
	Full Sample			Never attrited	Tweeted every week	Active (1 day)	Unlikely to be automated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Political congruence	-0.1355***	-0.1354***	-0.1267***	-0.1062***	-0.1193***	-0.1469***	-0.1329***
	(0.0072)	(0.0073)	(0.0081)	(0.0080)	(0.0092)	(0.0092)	(0.0115)
Affective congruence	-0.0611***	-0.0609***	-0.0652***	-0.0518***	-0.0623***	-0.0797***	-0.0859***
	(0.0076)	(0.0076)	(0.0093)	(0.0093)	(0.0115)	(0.0115)	(0.0139)
Political congruence × Affective congruence	0.0559***	0.0553***	0.0578***	0.0457***	0.0534***	0.0707***	0.0722***
	(0.0078)	(0.0078)	(0.0097)	(0.0096)	(0.0120)	(0.0121)	(0.0150)
Wave, Strata Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes	Yes
Observations	14,737	14,737	14,737	12,945	9,718	11,501	6,645
R^2	0.05768	0.06426	0.06790	0.05552	0.06102	0.07730	0.07583

Notes: The table presents regression estimates for the effect of sharing identities on follow-backs (Panel A) and blocks (Panel B), for different sub-samples of subjects, considering only the accounts that signaled both dimensions of identity. The sample excludes shadow-banned accounts, as pre-registered and discussed in the text. The first three columns show estimates using the full sample, estimating Equation (2) with and without wave and strata fixed effects and additional controls. The controls used are bot's football club, clubs' Google Trends index, subjects' region, gender, number of followers and number of tweets. The remaining columns perform similar estimates using sub-samples of subjects. A subject suffers attrition if we cannot follow it during a wave (because its account was de-activated, suspended, or made private). The sample of "never attrited" subjects is composed exclusively of subjects that did not suffer this type of attrition at any wave. Subjects that tweeted at least once in the seven days before every treatment wave are considered always active. Active subjects are those who tweeted or re-tweeted a status one day before treatment. Finally, the last column considers the sub-sample composed of subjects with below median score from the *Botometer* API (specifically, subjects with less than 13% chance of being automated accounts), which estimates the probability that a Twitter account is automated. Standard errors clustered at the fictional account level are in parentheses. Significance codes: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

Table B.9: Main Results for Different Sub-samples: Experimental accounts that signal a single dimension of identity

Panel A: Follow Backs, Affective Identity Only

	Dependent Variable: Follow Backs (1 = Yes)						
	Full Sample			Never attrited	Tweeted every week	Active (1 day)	Unlikely to be automated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Affective congruence	0.1337***	0.1413***	0.1454***	0.1548***	0.1747***	0.1604***	0.1483***
	(0.0133)	(0.0134)	(0.0187)	(0.0196)	(0.0212)	(0.0213)	(0.0211)
Wave, Strata Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes	Yes
Observations	7,388	7,388	7,388	6,583	4,983	5,688	3,500
R^2	0.02123	0.06732	0.08339	0.09017	0.09770	0.08595	0.08507

Panel B: Blocks, Affective Identity Only

	Dependent Variable: Blocks (1 = Yes)						
	Full Sample			Never attrited	Tweeted every week	Active (1 day)	Unlikely to be automated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Affective congruence	-0.0134***	-0.0126***	-0.0132***	-0.0113**	-0.0129**	-0.0135***	-0.0183**
	(0.0031)	(0.0032)	(0.0043)	(0.0046)	(0.0053)	(0.0048)	(0.0070)
Wave, Strata Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes	Yes
Observations	7,199	7,199	7,199	6,424	4,859	5,516	3,423
R^2	0.00253	0.01003	0.01773	0.01529	0.01757	0.02072	0.02001

Panel C: Follow Backs, Political Identity Only

	Dependent Variable: Follow Backs (1 = Yes)						
	Full Sample			Never attrited	Tweeted every week	Active (1 day)	Unlikely to be automated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Political congruence	0.2000***	0.1994***	0.1979***	0.1880***	0.1982***	0.2076***	0.1797***
	(0.0148)	(0.0147)	(0.0133)	(0.0135)	(0.0164)	(0.0162)	(0.0185)
Wave, Strata Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes	Yes
Observations	7,678	7,678	7,678	6,823	5,079	5,999	3,418
R^2	0.05092	0.08798	0.10159	0.09892	0.10616	0.10787	0.10123

Panel D: Blocks, Political Identity Only

	Dependent Variable: Blocks (1 = Yes)						
	Full Sample			Never attrited	Tweeted every week	Active (1 day)	Unlikely to be automated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Political congruence	-0.1225***	-0.1225***	-0.1224***	-0.1075***	-0.1188***	-0.1386***	-0.1236***
	(0.0062)	(0.0061)	(0.0060)	(0.0058)	(0.0074)	(0.0068)	(0.0085)
Wave, Strata Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes	Yes
Observations	7,492	7,492	7,492	6,668	4,961	5,830	3,324
R^2	0.05894	0.06775	0.07063	0.06323	0.07162	0.08179	0.08295

Notes: The table presents regression estimates for the effect of sharing identities on follow-backs (Panel A and C) and blocks (Panel B and D), for different sub-samples of subjects, considering only the accounts that signaled either affective (top two panels) or political (bottom two panels) identity. The sample excludes shadow-banned accounts, as pre-registered and discussed in the text. The first three columns show estimates using the full sample, estimating Equation (2) with and without wave and strata fixed effects and additional controls. The controls used are bot's football club, clubs' Google Trends index, subjects' region, gender, number of followers and number of tweets. Controls involving bot's football club are not included for the treatment arms with fictional accounts that only signal political identity. The remaining columns perform similar estimates using sub-samples of subjects. A subject suffers attrition if we cannot follow it during a wave (because its account was de-activated, suspended, or made private). The sample of "never attrited" subjects is composed exclusively of subjects that did not suffer this type of attrition at any wave. Subjects that tweeted at least once in the seven days before every treatment wave are considered always active. Active subjects are those who tweeted or re-tweeted a status one day before treatment. Finally, the last column considers the sub-sample composed of subjects with below median score from the *Botometer* API (specifically, subjects with less than 13% chance of being automated accounts), which estimates the probability that a Twitter account is automated. Standard errors clustered at the fictional account level are in parentheses. Significance codes: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

B.6 Experiment with Fictional accounts with more Salient Political Identity

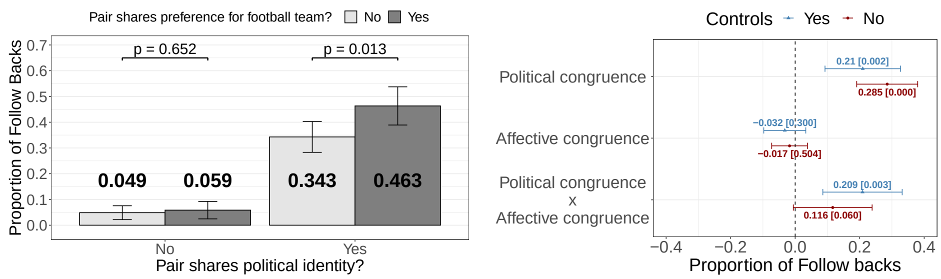
Figure B.4: Examples of Fictional Accounts - More salient political identity

(a) Pro-Bolsonaro; São Paulo supporter

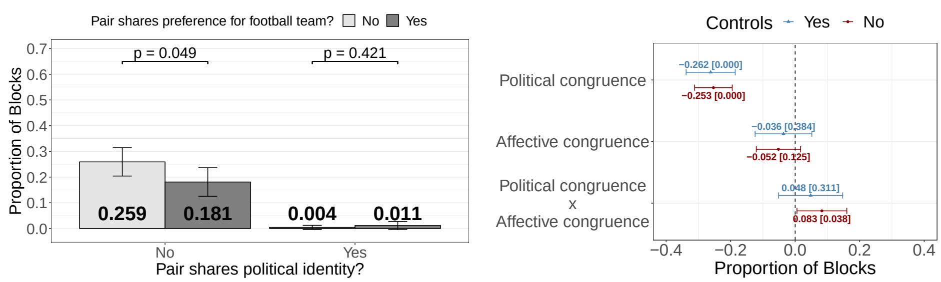
(b) Pro-Lula; Palmeiras supporter

Notes: The figures show examples of fictional accounts used in the extra experiment, in which the political identity signal was more salient.

Figure B.5: Effect of shared political and affective identity on the formation of social ties: Fictional accounts with more salient political identity



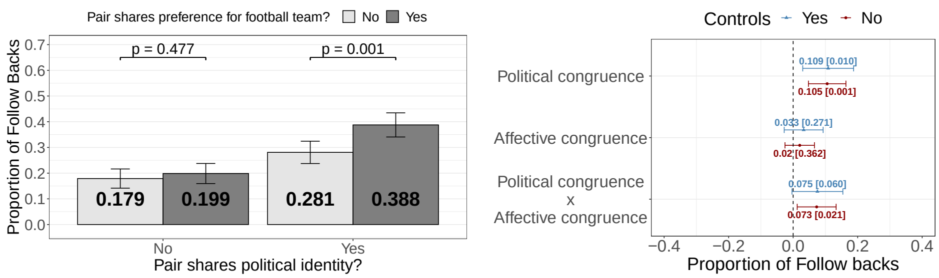
(a) Follow Backs



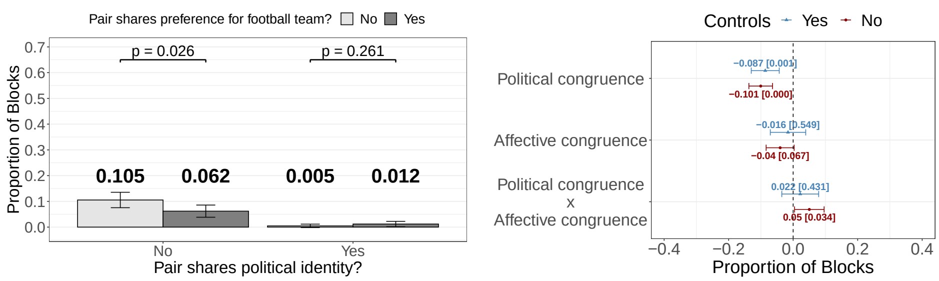
(b) Blocks

Notes: The figures show the effect of sharing political and affective (football club) identity on the rate of follow-backs and blocks for the experiment with fictional accounts with a more salient political identity. The figure on the left shows the average rate of follow-backs or fictional accounts for the entire experiment, excluding shadow-banned accounts. The p-value on these plots is the p-value of a simple t-test of difference in means between the two groups indicated by the bracket. The left-hand side plot shows the coefficients estimated from equation (2), which includes wave and strata fixed effects. The controls used are the bot's football club, the google trend index of the clubs, subject's number of followers and statuses, interacted with the treatment indicator. The plots show 95% confidence intervals (error bar), coefficient estimates and p-values (in brackets). Confidence intervals and p-values are computed using standard errors clustered at the fictional accountlevel.

Figure B.6: Results of the main experiment for the same waves as experiment with more salient political identity



(a) Follow Backs



(b) Blocks

Notes: The figures show the effect of sharing political and affective (football club) identity on the rate of follow-backs and blocks for the fictional accounts of the original experiment, restricting the analysis for the waves in which we conducted the extra experiment with fictional accounts with more salient political identity. The figure on the left shows the average rate of follow-backs or fictional accounts for the entire experiment, excluding shadow-banned accounts. The p-value on these plots is the p-value of a simple t-test of difference in means between the two groups indicated by the bracket. The left-hand side plot shows the coefficients estimated from equation (2), which includes wave and strata fixed effects. The controls used are the bot's football club, the google trend index of the clubs, subject's number of followers and statuses, interacted with the treatment indicator. The plots show 95% confidence intervals (error bar), coefficient estimates and p-values (in brackets). Confidence intervals and p-values are computed using standard errors clustered at the fictional account level.

B.7 Other Robustness Exercises

Table B.10: Main Results Excluding Fictional accounts' Football Clubs

Panel A: Follow Backs, Affective Identity Only

	Dependent Variable: Follow Backs (1 = Yes)						
Excluded Club:	-	Flamengo	Corinthians	São Paulo	Palmeiras	Vasco	Grêmio
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Affective congruence	0.1454***	0.1402***	0.1596***	0.1519***	0.1057***	0.1617***	0.1419***
	(0.0187)	(0.0236)	(0.0197)	(0.0202)	(0.0217)	(0.0193)	(0.0200)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,388	5,167	6,567	5,756	6,148	6,649	6,653
R^2	0.08339	0.08352	0.08533	0.09057	0.08148	0.08547	0.08590

Panel B: Blocks, Affective Identity Only

	Dependent Variable: Blocks (1 = Yes)						
Excluded Club:	-	Flamengo	Corinthians	São Paulo	Palmeiras	Vasco	Grêmio
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Affective congruence	-0.0132***	-0.0159***	-0.0119**	-0.0142***	-0.0145***	-0.0134***	-0.0116**
	(0.0043)	(0.0048)	(0.0049)	(0.0046)	(0.0053)	(0.0047)	(0.0046)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,199	4,978	6,473	5,661	5,959	6,460	6,464
R^2	0.01773	0.01653	0.01857	0.02276	0.01953	0.01848	0.01998

Panel C: Follow Backs, Both Dimensions of Identity

	Dependent Variable: Follow Backs (1 = Yes)						
Excluded Club:	-	Flamengo	Corinthians	São Paulo	Palmeiras	Vasco	Grêmio
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Political congruence	0.1476***	0.1429***	0.1434***	0.1497***	0.1520***	0.1500***	0.1493***
	(0.0139)	(0.0169)	(0.0153)	(0.0155)	(0.0149)	(0.0155)	(0.0151)
Affective congruence	0.0512***	0.0266*	0.0520***	0.0542***	0.0605***	0.0619***	0.0469***
	(0.0114)	(0.0150)	(0.0119)	(0.0134)	(0.0131)	(0.0118)	(0.0118)
Political congruence × Affective congruence	0.0503***	0.0693***	0.0583***	0.0466**	0.0331*	0.0387**	0.0522***
	(0.0170)	(0.0203)	(0.0183)	(0.0206)	(0.0187)	(0.0181)	(0.0176)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,128	10,464	13,205	11,928	12,755	13,647	13,641
R^2	0.09909	0.09980	0.10082	0.10415	0.09794	0.09836	0.10029

Panel D: Blocks, Both Dimensions of Identity

	Dependent Variable: Blocks (1 = Yes)						
Excluded Club:	-	Flamengo	Corinthians	São Paulo	Palmeiras	Vasco	Grêmio
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Political congruence	-0.1267***	-0.1159***	-0.1227***	-0.1283***	-0.1293***	-0.1284***	-0.1268***
	(0.0081)	(0.0101)	(0.0086)	(0.0096)	(0.0087)	(0.0090)	(0.0088)
Affective congruence	-0.0652***	-0.0530***	-0.0592***	-0.0682***	-0.0728***	-0.0639***	-0.0674***
	(0.0093)	(0.0107)	(0.0100)	(0.0109)	(0.0101)	(0.0103)	(0.0101)
Political congruence × Affective congruence	0.0578***	0.0437***	0.0527***	0.0599***	0.0636***	0.0590***	0.0607***
	(0.0097)	(0.0112)	(0.0105)	(0.0113)	(0.0106)	(0.0106)	(0.0103)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,737	10,073	13,011	11,731	12,364	13,256	13,250
R^2	0.06790	0.06653	0.06637	0.07301	0.07053	0.06778	0.06943

Notes: The table presents regression estimates for the effect of sharing affective identity on follow-backs (Panel A and C) and blocks (Panel B and D), considering only the accounts that signaled only affective identity (top two panels), or accounts that signaled both dimensions (bottom two panels). Specifically, it shows OLS estimates of specification 2, excluding one of the bot's clubs at a time. The sample excludes shadow-banned accounts, as pre-registered and discussed in the text. Standard errors clustered at the fictional account level are in parentheses. Significance codes: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

Table B.11: Experiment Results Excluding Clubs Not Signaled by Fictional accounts

Panel A: Fictional accounts signaling affective Identity Only				
Dependent Variables:	Follow Backs (1 = Yes)		Blocks (1 = Yes)	
Sample:	Full	Excluding non-signaled Clubs	Full	Excluding non-signaled Clubs
	(1)	(2)	(3)	(4)
Affective congruence	0.1454***	0.1636***	-0.0132***	-0.0123**
	(0.0187)	(0.0204)	(0.0043)	(0.0048)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	7,388	5,949	7,199	5,784
R^2	0.08339	0.08361	0.01773	0.01779

Panel B: Fictional accounts Signaling both Dimensions of Identity				
Dependent Variables:	Follow Backs (1 = Yes)		Blocks (1 = Yes)	
Sample:	Full	Excluding non-signaled Clubs	Full	Excluding non-signaled Clubs
	(1)	(2)	(3)	(4)
Political congruence	0.1476***	0.1454***	-0.1267***	-0.1209***
	(0.0139)	(0.0171)	(0.0081)	(0.0096)
Affective congruence	0.0512***	0.0455***	-0.0652***	-0.0577***
	(0.0114)	(0.0131)	(0.0093)	(0.0103)
Political congruence × Affective congruence	0.0503***	0.0529***	0.0578***	0.0515***
	(0.0170)	(0.0195)	(0.0097)	(0.0106)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	15,128	12,326	14,737	11,964
R^2	0.09909	0.09614	0.06790	0.06223

Notes: The table presents regression estimates for the effect of sharing identity on follow-backs and blocks, considering treatment arms with fictional accounts that signaled affective identity only (Panel A) or both dimensions of identity (Panel B). Columns (2) and (4) present results for a subsample of subjects that excludes those who support a club that was not among the six clubs signaled by fictional accounts during the experiment. Standard errors clustered at the fictional account level are in parentheses. Significance codes: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

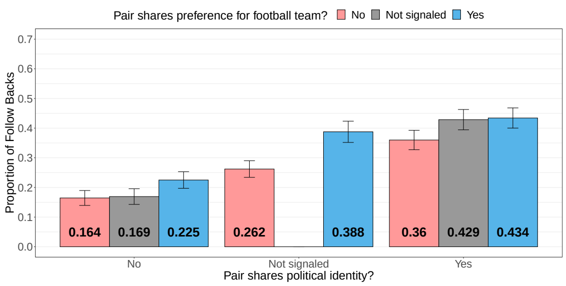
Table B.12: Experiment Results by Subjects' Gender

Panel A: Fictional accounts signaling affective Identity Only				
Dependent Variables:	Follow Backs (1 = Yes)		Blocks (1 = Yes)	
Subjects' Gender:	Female	Male	Female	Male
	(1)	(2)	(3)	(4)
Affective congruence	0.1263***	0.1439***	-0.0298***	-0.0078*
	(0.0410)	(0.0239)	(0.0096)	(0.0043)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	1,351	4,649	1,322	4,519
R^2	0.11726	0.09154	0.08234	0.01817

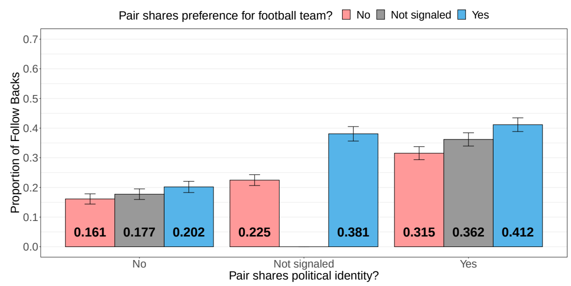
Panel B: Fictional accounts Signaling both Dimensions of Identity				
Dependent Variables:	Follow Backs (1 = Yes)		Blocks (1 = Yes)	
Subjects' Gender:	Female	Male	Female	Male
	(1)	(2)	(3)	(4)
Political congruence	0.2409***	0.1448***	-0.1718***	-0.1183***
	(0.0324)	(0.0169)	(0.0183)	(0.0089)
Affective congruence	0.0050	0.0554***	-0.0657***	-0.0732***
	(0.0251)	(0.0149)	(0.0207)	(0.0118)
Political congruence × Affective congruence	0.0107	0.0597**	0.0496**	0.0678***
	(0.0463)	(0.0240)	(0.0225)	(0.0120)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	2,662	9,658	2,595	9,407
R^2	0.12504	0.10883	0.10914	0.06581

Notes: The table presents regression estimates for the effect of sharing identity on follow-backs and blocks, considering treatment arms with fictional accounts that signaled affective identity only (Panel A) or both dimensions of identity (Panel B), separately by subjects' gender. We predict subjects' gender from their first names using data from the 2010 Brazilian Census organized by Meireles (2021). A gender is assigned to a subject if at least 90% of the occurrences of their name in the 2010 census correspond to that gender. Standard errors clustered at the fictional account level are in parentheses. Significance codes: $^{***}: p < 0.01$, $^{**}: p < 0.05$, $^{*}: p < 0.1$.

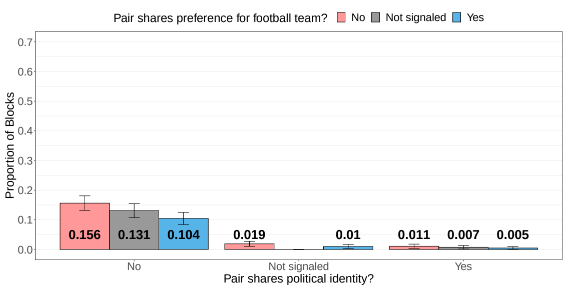
Figure B.7: Heterogeneity on type of content posted before treatment



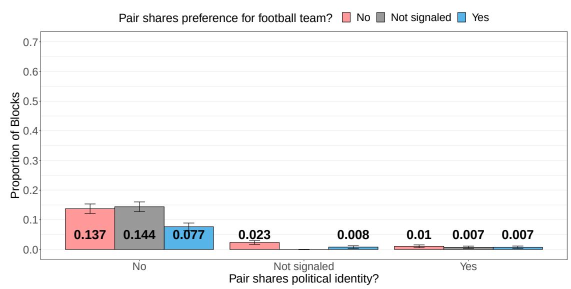
(a) Follow Backs, subjects whose last tweet before treatment had political content



(b) Follow Backs, subjects whose last tweet before treatment did not have political content



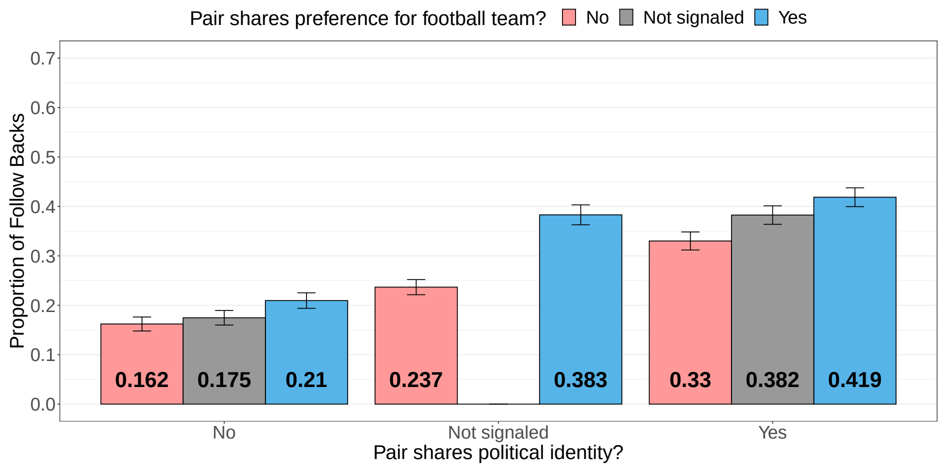
(c) Blocks, subjects whose last tweet before treatment had political content



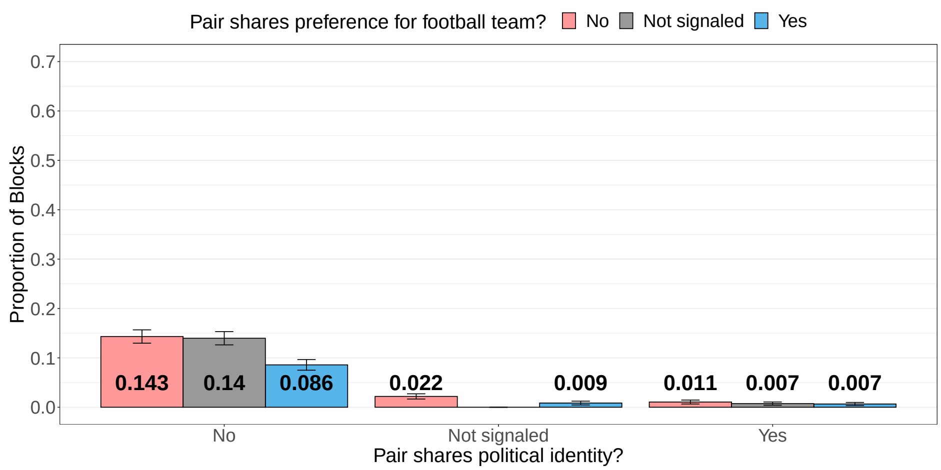
(d) Blocks, subjects whose last tweet before treatment did not have political content

Notes: The figures show the effect of sharing political and affective (football club) identity on the rate of follow-backs and blocks, for all eight treatment arms in the main experiment (fictional accounts that signal both or a single dimension of identity). The x-axis shows whether fictional account and subject share political identity (or show that this dimension is not signaled by the fictional accounts), while the colors show whether fictional account and subject share preference for football club (or show that this dimension is not signaled by the bot). Each bar shows the average follow-back rate (panels a and b) and block-rate (panels c and d) for each of these treatment arms. The figures report results for two sub samples of subjects: the ones whose last tweet before treatment had political content and the ones whose last tweet before treatment had other type of content. To classify tweets' content, we use a Naive Bayesian Classifier Algorithm. This analysis is restricted to waves 11 to 43 due to data constraints. We also restrict the analysis to subjects who tweeted at most one week before treatment. The error bars represent 95% confidence intervals.

Figure B.8: Effect of shared political and affective identity on the formation of social ties, Waves 11-43



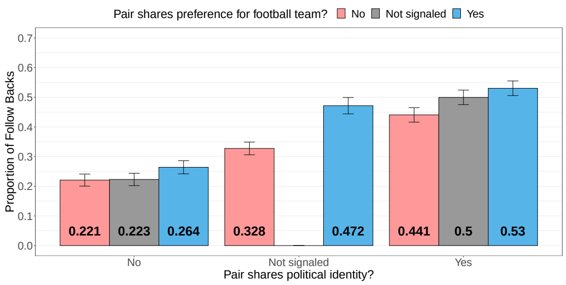
(a) Follow Backs



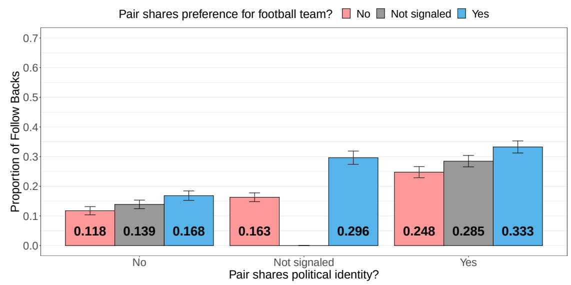
(b) Blocks

Notes: The figures show the effect of sharing political and affective (football club) identity on the rate of follow-backs and blocks, for all eight treatment arms in the main experiment (fictional accounts that signal both or a single dimension of identity). The x-axis shows whether fictional account and subject share political identity (or show that this dimension is not signaled by the fictional accounts), while the colors show whether fictional account and subject share preference for football club (or show that this dimension is not signaled by the bot). Each bar shows the average follow-back rate (panel a) and block-rate (panel b) for each of these treatment arms. This analysis is restricted to waves 11 to 43, and to subjects who tweeted at most one week before treatment, in order to allow comparisons with the heterogeneity analysis of Appendix Figure B.7. The error bars represent 95% confidence intervals.

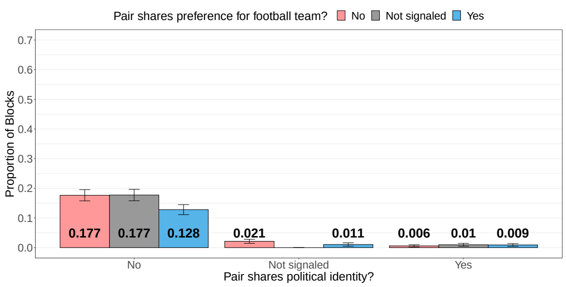
Figure B.9: Heterogeneity on type of content in user's pre-treatment bios



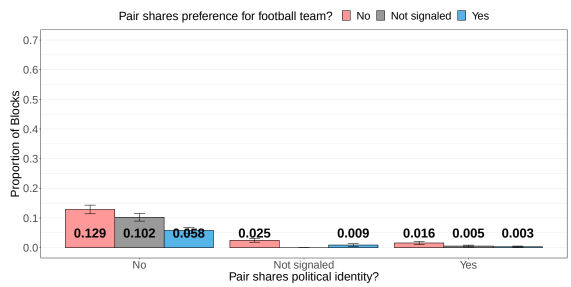
(a) Follow Backs, subjects whose bio had political content



(b) Follow Backs, subjects whose bio did not have political content



(c) Blocks, subjects whose bio had political content



(d) Blocks, subjects whose bio did not have political content

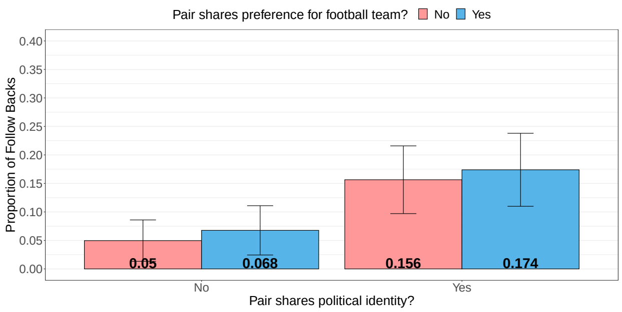
Notes: The figures show the effect of sharing political and affective (football club) identity on the rate of follow-backs and blocks, for all eight treatment arms in the main experiment (fictional accounts that signal both or a single dimension of identity). The x-axis shows whether fictional account and subject share political identity (or show that this dimension is not signaled by the fictional accounts), while the colors show whether fictional account and subject share preference for football club (or show that this dimension is not signaled by the bot). Each bar shows the average follow-back rate (panels a and b) and block-rate (panels c and d) for each of these treatment arms. The figures report results for two sub samples of subjects: the ones whose bio (before treatment) had political content and the ones whose bio (before treatment) had other type of content. To classify bios' content, we use a simple keyword search in a dictionary of words related to the Brazilian elections. The error bars represent 95% confidence intervals.

B.8 Demand for Information *versus* Social Connections

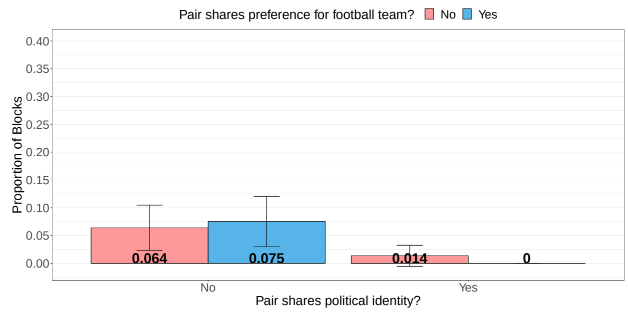
Figure B.10: Examples of Informational Fictional Accounts

Notes: The figure shows an example of informational fictional accountused in the experiment. The informational fictional accounts explicitly signal they are automated in their profile name (with the word "Bot" in parentheses) and that they share information on their political and affective identity in their bio. The bio reads *"fictional account that re-tweets news pieces about Corinthians [its football team] and #Bolsonaro22.*"

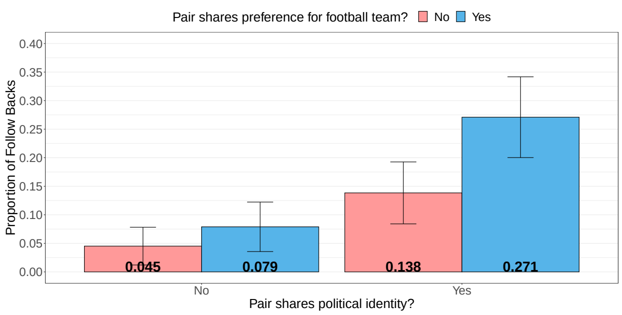
Figure B.11: Effect of shared political and affective identity on Follow Backs, information versus original accounts



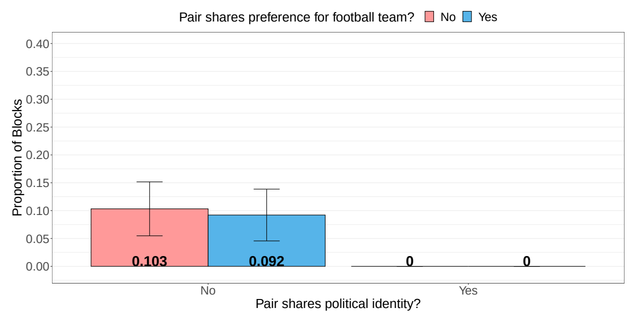
(a) Informational Accounts



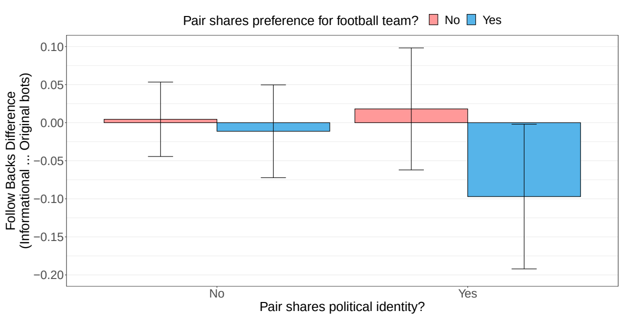
(b) Informational Accounts



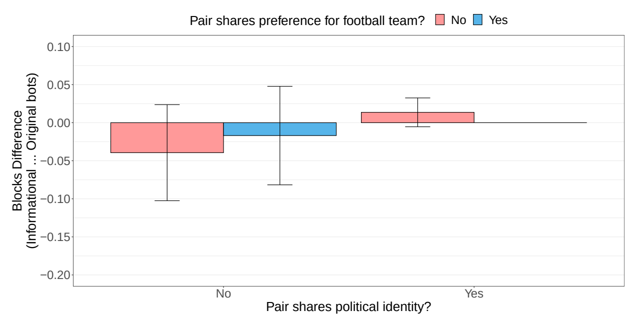
(c) Original Accounts



(d) Original Accounts



(e) Difference (Informational − Original)



(f) Difference (Informational − Original)

Notes: The figures show the effect of sharing political and affective (football club) identity on the rate of follow-backs (left) and blocks (right), separately for the fictional accounts that explicitly say they will share information (top) and the original fictional accounts (middle), as well as the differences (bottom panel). Data comes from four experimental waves conducted between December 13th, 2023, and February 14th, 2024. The plots show 95% confidence intervals (error bars). The bottom plot displays differences between informational and original accounts.

Table B.13: Motivation to establish ties: information versus social ties

Panel A: Follow Backs

	Dependent Variable: Follow Backs (1 = Yes)			
	(1)	(2)	(3)	(4)
Informational Bot	-0.0258**	-0.0175	-0.0145	-0.0165
	(0.0097)	(0.0135)	(0.0115)	(0.0155)
Political congruence		0.1282***		0.1098***
		(0.0177)		(0.0218)
Informational Bot × Political congruence		-0.0174		0.0029
		(0.0268)		(0.0276)
Affective congruence			0.0630***	0.0260
			(0.0179)	(0.0180)
Informational Bot × Affective congruence			-0.0447**	-0.0042
			(0.0217)	(0.0268)
Political × Affective congruence				0.0731
				(0.0441)
Informational Bot × Political × Affective congruence				-0.0801
				(0.0667)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,433	2,433	2,433	2,433
R^2	0.02311	0.05665	0.02639	0.06115

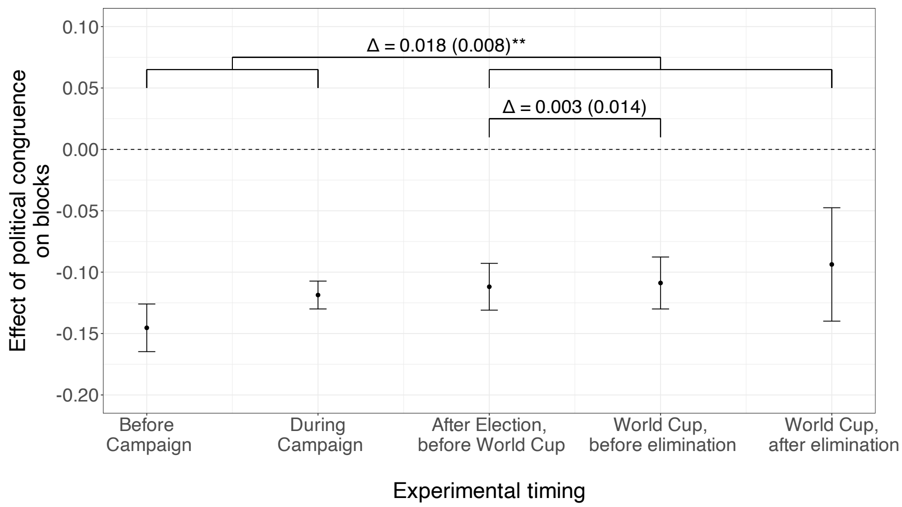
Panel B: Blocks

	Dependent Variable: Blocks (1 = Yes)			
	(1)	(2)	(3)	(4)
Informational Bot	0.0011	0.0032	0.0044	0.0099
	(0.0091)	(0.0201)	(0.0102)	(0.0223)
Political congruence		-0.0902***		-0.0890***
		(0.0164)		(0.0188)
Informational Bot × Political congruence = 1		-0.0033		-0.0099
		(0.0227)		(0.0248)
Affective congruence			-0.0008	0.0014
			(0.0168)	(0.0354)
Informational Bot × Affective congruence = 1			-0.0147	-0.0292
			(0.0229)	(0.0478)
Political × Affective congruence				-0.0045
				(0.0368)
Informational Bot × Political × Affective congruence = 1				0.0284
				(0.0498)
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,433	2,433	2,433	2,433
R^2	0.00408	0.04974	0.00450	0.05044

Notes: The table presents regression estimates for the effect of explicitly informational-sharing accounts on follow-backs (Panel A) and blocks (Panel B). The sample excludes shadow-banned accounts. "Informational Bot" is an indicator equal to one for fictional accounts that explicitly state that they are automated and will share information about their preferred politician and football club. "Political identity" and "Affective identity" are indicators equal to one if fictional account and subject share political or affective identity (respectively). Data for this table comes from four experimental waves conducted between December 2023 and February 2024 (i.e., one year after the original experiment). Standard errors clustered at the fictional accountlevel are in parentheses. Significance codes: *** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$.

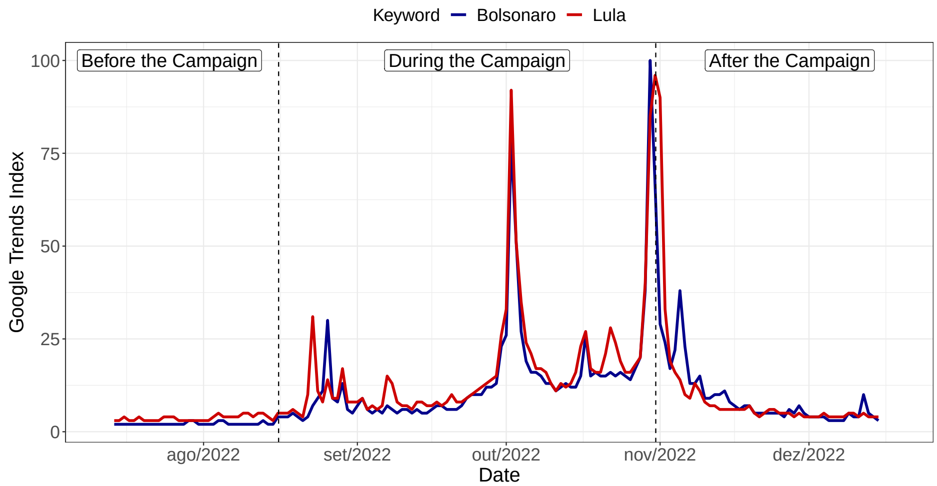
C Additional Results for the Analysis of Formation of Ties over Time

Figure C.1: Effect of Congruence in Political Identity on Blocks at Different Times



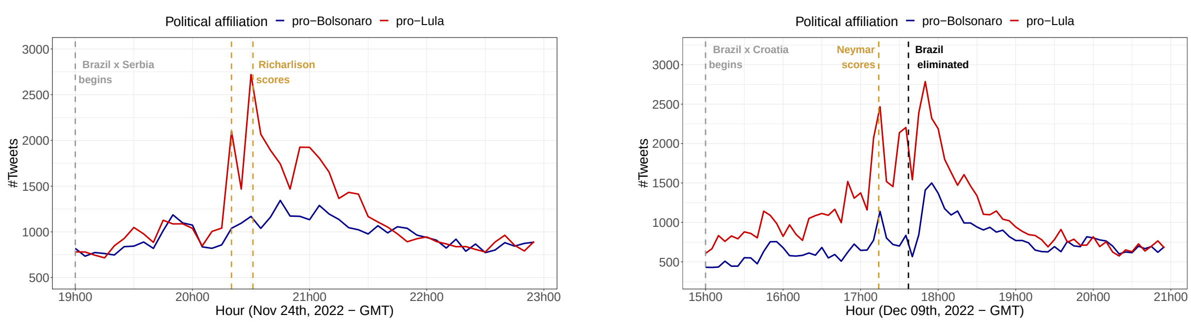
Notes: The figure displays point estimates and confidence intervals for the effect of congruence in political identity on blocks for different sets of experimental waves, ordered by period: before the official electoral period; during the electoral period; and after the electoral period. The after-election period is further divided into before the beginning of the World Cup, during the World Cup and when results for Brazil were positive, and after Brazil's elimination from the Tournament. Timing details are in Appendix Figure A.1. The sample pools data from all experimental waves within each period, restricting the analysis to subjects who were always active during the experimental period (i.e., who tweeted in the seven days before being treated every time they were treated). This gives us a total of 27,701 observations. The brackets above the point estimates display estimates and standard errors (in parentheses) for the difference in the effect of political congruence between the signaled periods. Standard errors are clustered at the bot-account level. Significance codes: $^{***}: p < 0.01$, $^{**}: p < 0.05$, $^{*}: p < 0.1$.

Figure C.2: Google Trend Index during the Experimental Period for the Two Main Presidential Candidates in the Brazilian 2022 Presidential Elections



Notes: The figure displays the Google Trends Index for searches of the terms "Lula" and "Bolsonaro" in Brazil during the experimental period. The periods denoted as "before", "during", and "after" the campaign correspond to official campaign periods as determined by Brazil's Superior Electoral Court.

Figure C.3: Number of tweets during selected Brazil games, 2022 FIFA World Cup



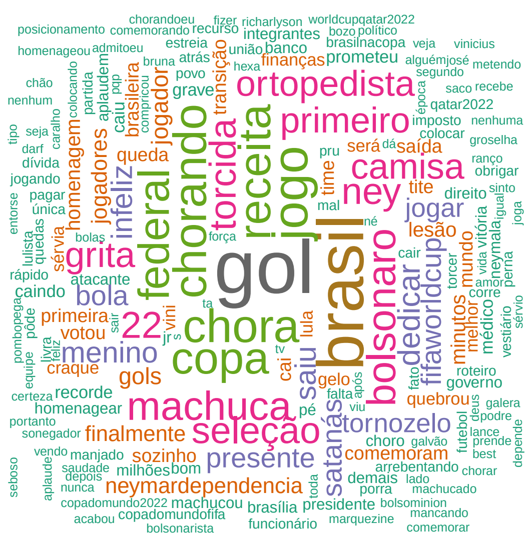
(a) Brazil *vs.* Serbia (b) Brazil *vs.* Croatia

Notes: The figure displays the number of tweets sent by pro-Lula and pro-Bolsonaro users on the day of the match between Brazil and Serbia and Brazil and Croatia in the 2022 FIFA World Cup. Data comes from a 10% random sample of all Brazilian Twitter users that tweeted or re-tweeted a status containing a pro-Lula or pro-Bolsonaro hashtag in the week before the first round of the 2022 presidential election. Tweets are aggregated into intervals of five minutes.

Figure C.4: Word Clouds of Tweets by pro-Lula and pro-Bolsonaro users during Brazil x Serbia

(a) Tweets about Richarlison, pro-Lula users

(b) Tweets about Richarlison, pro-Bolsonaro users

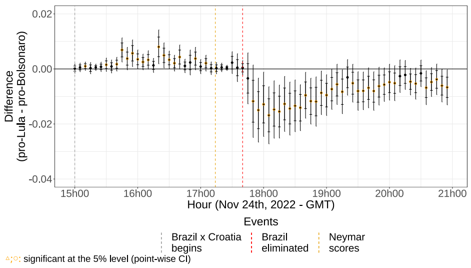


(c) Tweets about Neymar, pro-Lula users

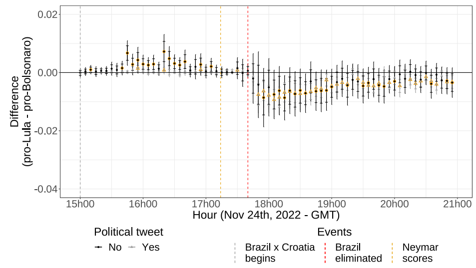
(d) Tweets about Neymar, pro-Bolsonaro users

Notes: The figures show word clouds for tweets and re-tweets posted during Brazil's debut World Cup match against Serbia, from our random sample of users.

Figure C.5: Difference in the number of tweets about Tite between pro-Lula and pro-Bolsonaro Twitter users during Brazil × Croatia



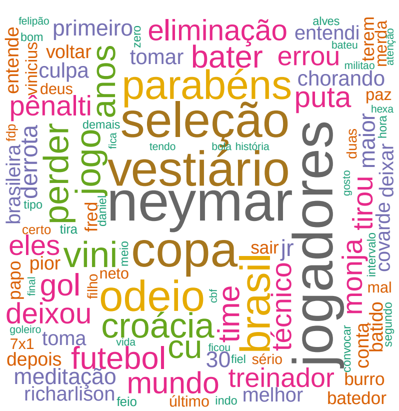
(a) Tweets about Tite



(b) Tweets about Tite, Heterogeneity by Political Content of Tweet

Notes: The left figure plots the difference in the likelihood that a pro-Lula and pro-Bolsonaro account posts a tweet about Brazil's coach Tite for every five minute interval around the 2022 World Cup game between Brazil and Croatia. The right figure plots a similar exercise, but separating the analysis between tweets with political content or not. In both cases, we estimate Equation (**??**) as described in the main text. To classify tweets according to their content, we use a Bayesian Classifier algorithm. In all cases, data comes from a 10% random sample of all Brazilian Twitter users that tweeted or re-tweeted a status containing a pro-Lula or pro-Bolsonaro hashtag in the week before the first round of the 2022 presidential election. The error bars with ticks represent 95% confidence intervals, while the extended bars represent 95% uniform sup-t confidence bands, estimated using Montiel Olea and Plagborg-Møller (2019)'s plug-in estimator. Standard errors are clustered at the user level. Point estimates marked in orange denote estimates significant at the 5% level (point-wise).

Figure C.6: Word Clouds of Tweets by pro-Lula and pro-Bolsonaro users after Brazil x Croatia



(a) Tweets about Tite, pro-Lula users

(b) Tweets about Tite, pro-Bolsonaro users

Notes: The figures show word clouds for tweets and re-tweets posted in the two hours after Brazil's match against Croatia (when Brazil was eliminated), from our random sample of users.