Pre-Analysis Plan: Affective Polarization during Electoral Times: Evidence from Twitter*

Nicolás Ajzenman Bruno Ferman Pedro Sant'Anna

August 26, 2022

Abstract

We study the interplay between political and affective congruence in the formation of social ties on Twitter, in the context of the Brazilian 2022 elections. We create fictitious bot accounts on Twitter that signal two dimensions of identity: first, political preference for one of the two main political candidates (Luiz Inácio Lula da Silva or Jair Messias Bolsonaro); second, preference for a Brazilian soccer club, which we interpret as a dimension of identity that is traditionally affective. The bot accounts randomly follow Twitter accounts with congruent and incongruent identities across these two dimensions. We compute the proportion of follow-backs and blocks received by the bots as measures of the formation of social ties. We run the experiment in the second semester of 2022, both before and after the Brazilian presidential election. Thus, we can study heterogeneous effects of shared identity before, during and after the election. We hypothesize that the salience of the political identity, which is likely to increase as the election approaches, increases affective polarization in the political dimension, by decreasing the relative importance of sharing other affective identities to the formation of ties.

Keywords: Social Identity; Affective Polarization; Brazilian Elections; Social Media.

JEL Codes: C93; D72; D91; Z20.

*We thank Claudio Ferraz and Lorenzo Lagos for their helpful comments and suggestions. This research was approved by the Ethical Compliance Committee on Research Involving Human Beings at Fundação Getulio Vargas with a waiver of informed consent (CEPH/FGV, IRB approval n. 208/2022).

Contents

1 Introduction

We design an experiment to study how political partisanship interacts with other dimensions of identity in the formation of social ties on Twitter – we focus on preference for soccer team, an important dimension of socialization in Brazil. We create fictitious accounts that signal their preferred candidate in the upcoming Brazilian presidential election (either Luiz Inácio Lula da Silva or Jair Messias Bolsonaro, the two candidates that lead pools and that have been the symbols of opposite sides of the political spectrum in Brazil in the last few years) and their preferred soccer team. The accounts then randomly follow Twitter accounts with congruent and non-congruent identities across these two dimensions (political and soccer club preference). We compute the proportion of follow-backs and blocks that each bot receives as measures of the formation of social ties. The bot accounts will be active during the second semester of 2022, both before, during and after the campaign period of the 2022 Brazilian elections.

We interpret preference for soccer club as a dimension of identity that is traditionally relevant to the formation of affective ties – and we call this an "affective identity". The fact that soccer preference has such an affective dimension is particularly true in Brazil, where soccer has a "powerful cultural presence" and is one "central element of social identity" (Murad, 1995). On the other hand, the affective aspect of political identity has only recently become a relevant phenomenon. Therefore, the objective of this research is to study how these two dimensions of identity (political and affective) interact in the decision to establish ties in a social media environment. Our experimental design will allow us to verify if each one of the two dimensions of identity – political and affective – impacts the follow-back and blocking behavior, and if sharing both identities changes this behavior even further.

We are particularly interested in studying how political polarization shapes the formation of social ties. In recent years, the Brazilian political environment has become increasingly polarized, and it is possible that such polarization manifests itself not only on issues or policy-based disagreements but also in terms of identity, with individuals who support one candidate being increasingly less likely to be willing to form ties with individuals from the opposite side. This phenomenon has been called affective polarization in the United States (Iyengar et al., 2019). Our experiment will provide evidence of affective polarization in Brazil, by studying the willingness to form social ties between Twitter users who share or not political identity, even if they share an affective dimension of identity. This way, we will be able to quantify political polarization of social ties relative to another dimension of identity. If affective polarization on the political dimension is strong, we would expect follow-backs to be more responsive to a match of political identity than to a match of soccer club preferences.

Moreover, given that we will perform the experiment during the second semester of 2022, we will also be able to study how political affective polarization – the unwillingness to socialize with politically incongruent individuals, and increased willingness to socialize with politically congruent ones – changes relative to the effect of sharing a more traditionally affective dimension of identity. This will be possible because the 2022 Brazilian presidential election will take place in October, with the campaign period starting officially on August

16^{th}. We hypothesize that, as the election approaches, the political dimension of identity will become more salient and more polarized, possibly making this dimension of identity more relevant to the formation of social ties and decreasing the relative importance of the other identity – soccer club preference.

Finally, we will also create accounts that only signal one of the two dimensions of identity. The objective of doing so is verifying if the two dimensions of identity we study – soccer preference and political identity – are relevant on themselves to the formation of social ties.

The rest of this plan is organized as follows. In section 2, we describe the experimental design, detailing how the accounts will be created, what is our subject pool, and how the treatment will be delivered; section 3 describes our hypotheses; finally, section 4 describes our empirical strategy and section 5 provides further details on our strategy.

2 Experimental Design

In this section, we describe our experimental design. Overall, the experiment will consist on the creation of fictitious accounts on Twitter that signal their preferred candidate in the 2022 Brazilian election and their preferred Brazilian soccer club. The bot accounts randomly follow Twitter users who share or not each identity (political and soccer club preference) with it. After five days active, we compute the number of follow-backs and blocks obtained by each bot. The experiment will be run in waves during the second semester of 2022. We will have 44 waves with 8 bots each.

The next subsection details the creation of the bot accounts. We then describe our subject pool and treatment assignment strategy. Finally, we detail the experimental timeline.

2.1 Profiles

Table 1 describes the elements of our bot accounts. Each bot is characterized by two dimensions of identity: political and soccer-related. We signal political identity by including, in the bot's bio,[1] either the hashtag *#Lula2022* or *#Bolsonaro2022*, and by re-tweeting one post from the candidate supported by the bot.[2] On the other hand, we signal soccer-related identity through the bot's profile picture (which is a picture of a flag with the bot's preferred team logo in a stadium) and by adding the text "Supporter of team X" in the bot's bio.

We also create bot accounts that are neutral in one of these two dimensions. The accounts that are politically neutral only have the soccer team signals, while accounts that are soccer team-neutral have, instead of a specific team's logo as the profile pic, a photo of a soccer stadium outside Brazil (and for which it is not possible to identify the teams) and include the text "Soccer fan" in the bio. Therefore, the accounts that are soccer team-neutral are

[1] "Bio" is the name given to the short description that appears below the account's name on Twitter.

[2] To alleviate concerns that the bots may be amplifying political content, we only re-tweet posts that already have more than 500 re-tweets and that do not include misleading information or hate speech.

still signalling that they are interested in soccer (the only difference is that they do not signal preference for a specific team).

Overall, we have five types of bot accounts: (i) pro-Lula, supporter of a specific soccer team; (ii) pro-Bolsonaro, supporter of a specific soccer team; (iii) pro-Lula, neutral team (but soccer fan); (iv) pro-Bolsonaro, neutral team (but soccer fan); (v) politically neutral, supporter of a specific soccer team. At each wave, we will choose two soccer clubs and activate two accounts of type (i), (ii) and (v) (one for each of the two soccer clubs); and one account of type (iii) and (iv) each, for a total of 8 accounts. We will vary the soccer teams used both across and within waves, so as to alleviate concerns about team's idiosyncrasies affecting the follow-back rates. We will use the six clubs with the largest number of supporters in Brazil: 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.[3] While the bots only support one of these six teams, the subject pool includes individuals who support rivals of these teams – specifically, apart from the six teams listed, we include subjects who support S.C. Internacional (Grêmio's rival), and Botafogo F.R. and Fluminense F.C. (Flamengo and Vasco's rivals).

Table 1: Procedures used to create the bot 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 soccer stadium during a soccer 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 bot accounts contains two information: first, it either says "Supporter of team X" (if the account signals her preferred team) or "Soccer 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 soccer team plays its home matches (and random city landscape for the soccer team-neutral accounts).
Location	The bot accounts' profiles do not include a location.
Website	The bot accounts' profiles do not include an website.
Retweets	The bot account first re-tweets a post from an account related to her preferred soccer team or, in the case of team-neutral accounts, a general tweet about soccer (that isn't specific about any soccer 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 access the bot's profile is the one that signals political identity.
Followers	We asked a group of colleagues to follow the bot accounts before each experimental wave.
Following	One day before following the accounts randomly assigned to it, the bot account will 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 bot accounts. Figure A.2 on the Appendix shows examples of accounts.

[3]The ranking of club supporters comes from a 2022 survey by Sport Track and XP (Sport Track and XP, 2022).

2.2 Sample Selection and Assignment into Treatment

The most important feature of our sample is that we must be able to identify the political identity (either pro-Lula or pro-Bolsonaro) and the preferred soccer team of each subject. First, we use Twitter's API to obtain a sample of users who either tweeted or re-tweet a status containing pro-Lula or pro-Bolsonaro hashtags between May 31^{st} and July 11^{th}, 2022. The list of hashtags we considered is displayed in Appendix table A.1. Then, we inspected if the user's Twitter bio (the short description that the user writes in their profile) signalled the user's preferred Brazilian soccer club. To do this, we first use a simple algorithm that detects terms associated with the 6 most popular Brazilian soccer clubs and their rivals in the bios, and then manually check if the matches are correct. We then remove accounts that were created in 2022 (that are more likely to be inauthentic), accounts that are clearly bots, accounts with less than 10 followers and accounts with a ratio of followers to friends above 20. The objective of doing this is to remove accounts that are very unlikely to follow-back the experimental accounts, and accounts that are not real people. After these procedures, we are left with a sample of 4,652 individual accounts. We note that, due to query restrictions of Twitter's API, this is only a sample of the Brazilian accounts that signal political and soccer club preferences on Twitter.

We now describe the treatment assignment procedure. In each experimental wave, we will have 8 active bot accounts: four accounts that are part of the main experiment, and that signal both their preferred soccer team and their political preference; and four accounts that are neutral in one of the two dimensions (i.e., two accounts that are "soccer fans", but do not signal a specific team; and two accounts that signal a specific team, but not a political identity). The objective of the four neutral accounts is to study if each dimension of identity (soccer preference and political identity) are relevant to the formation of social ties when they are the only salient identity; however, we are primarily interested in the interaction between the two dimensions, i.e., we focus on the bots that signal both identities. In each wave, we will randomly choose two soccer clubs for the bot accounts.[4] Then, within a wave, three bots will signal preference for each of these two teams. For example, in the pilot wave, we had three bots who supported Flamengo (one pro-Lula, one pro-Bolsonaro and one politically neutral), and three bots who supported Palmeiras (again, one pro-Lula, one pro-Bolsonaro and one politically neutral), plus one pro-Bolsonaro and one pro-Lula bot that did not signal a preferred soccer club.

Each bot will follow approximately 100 subjects during each wave. Following the suggestion of Athey and Imbens (2017), we perform block-randomization to define the treatment assignment. Specifically, the treatment assignment to each bot is done by stratifying the subjects in terms of their political identity, preferred soccer team and whether the subjects' number of followers is above or below the median. First, for the bots who signal their preferred soccer club, we restrict the sample of subjects to the ones who either support the same team as the bot, or who support a rival team. We only consider regional (intra-state) rivalries; the list of rivalries is described in Appendix Table A.2. Given that we are interested in studying the effect of matching bot and subjects' identities on follow-backs and blocks,

[4]Throughout the experiment, we randomly sample teams with a probability equivalent to the proportion of each team's supporters in our sample.

we have four strata in terms of bot-subjects identity pairs (congruence in both dimensions, incongruence in both dimensions, or congruence in a single dimension), and each pair is further divided into two smaller strata (above or below the median number of followers). We sample the same proportion of subjects from each stratum. Each subject may be treated (i.e., followed by a bot) more than once, but never in subsequent waves: after being treated in a wave, a subject only returns to the subject pool after 3 waves.

Apart from following the experimentally assigned accounts, each bot account will also follow one account from someone who knows about the experiment. This person will then inform us whether they received a notification of the follow. The objective of doing so is to guarantee that the follow is being notified to the users.[5] If an account is shadow-banned, we will simply drop it from the analysis.

2.3 Timeline

As described in the previous section, the experiment will be run in waves. In each wave, 8 bots will be active: 4 of them will signal both their political identity and soccer club preference, while the other 4 will be neutral in one of the two dimensions of social identity. Within each wave, we will use the following timeline:

(i) **Day 0:** Creation (or reboot) of accounts according to the procedures described in Table 1. The account re-tweets a post related to its sportive identity (either a post from its preferred club official account – if the bot has a preferred club – or a general post about soccer that does not favor any club), and then a post from its preferred political candidate. The political post must have at least 500 re-tweets and not contain misleading information or hate speech.

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

(iii) **Day 5:** After five days active, we will compute the number of followers and blocks for each account and delete all information in the account, rebooting it to be used in the next wave.

Specifically, the procedure we use to reboot accounts (so that we can use them again in another wave) is as follows: first, we remove all followers and follows from the account. This procedure can easily be done on Twitter and the user who is removed as follower or un-followed is not notified that this happened. We then wait one day[6] and access the bot

[5] On Twitter, a concern we have is with the so-called "shadow-ban". This is a type of punishment Twitter may deploy against users whose behavior on the platform seems suspicious. In practice, what happens is that all activity from a shadow-banned user is "hidden" to other users, including notifications of follows. Therefore, we guarantee that no bot account is shadow-banned before using the results from any experimental wave.

[6] This is done because, in some cases, Twitter takes a few hours to update its system and remove the original follow notification that the subjects received. By waiting one day, we more than guarantee that no one will notice the account changing its information.

account, removing all of its information (re-tweets, profile and background picture, name, etc) and replacing it with the information of the new account. Importantly, we always randomize which account from a previous wave will be "transformed" into each account of the new wave. For instance, an account that was pro-Lula and supported team A in wave 1 may become an account that is pro-Bolsonaro and soccer team-neutral in posterior waves.

We start one wave every Tuesday and every Friday, which means that we always have two overlapping waves at each moment. The specific timeline is displayed in Appendix figure A.1. We plan on running 44 experimental waves between July and December 2022. This period is particularly interesting because the Brazilian presidential election of 2022 will be held during the second semester of the year (specifically, the first round will happen on October 3^{rd}). We will use the differential timing of the experimental waves to study the heterogeneous effect of shared identity on the formation of social ties when political identity is more or less salient.

On each wave, we compute follow-backs once a day using Twitter's API. In our main analysis, we will use the final follow-back measure, computed on the fifth day since the bot followed the subjects. On the other hand, we only compute blocks at the end of each wave (i.e., on the fifth day). This happens because Twitter's API does not allow us to directly compute blocks. The procedure we use to compute blocks is as follows: first, we use Twitter's API to obtain, for each bot account, the set of accounts followed by it. We then compare this set with the set of accounts assigned to be followed by the bot. The difference between the two sets can be due to three mutually exclusive reasons: (i) the bot was indeed blocked by a subject; (ii) the subject was suspended or deactivated their account; (iii) the subject removed the bot from its followers. To assess which one of the three happened for each subject in this difference set, we manually enter these subjects' profiles from the bot's Twitter account. From the profiles, we can easily see which of the three cases happened. We only classify the subject as having blocked the bot if we see, on the fifth day, a block using this procedure.[7]

3 Hypotheses

This section describes the hypotheses that we will test in the experiment. Our main analysis focuses only on the bots that signal both dimensions of identity (political identity and preference for soccer club). In this analysis, we will be interested in estimating the effect that a shared identity (either in one or both dimensions) has on the likelihood of a follow-back or block by the subject to the bot. Then, in a secondary analysis, we will explore the data on the bots that are neutral in one of the two dimensions.

Moreover, we highlight that we are interested in studying how congruence (or lack thereoff) in political and soccer-club identity affects the formation of social ties on Twitter. Therefore, we will not study specifically how sharing a specific candidate or soccer club

[7]A fourth possibility is that a subject blocked a bot, but then unblocked it. We do not treat this as a block but as a follower removal. Thus, in our measure of blocks, there are only subjects that blocked a subject and kept it this way until the end of the wave.

preference affects behavior; rather, we will aggregate all candidates and all clubs and just consider whether bots and subjects share identities in either dimension.

3.1 Main analysis: bots that signal both dimensions of identity

Our main analysis considers only the bots that signal both their political identity and their preferred soccer team. In this case, we are interested in whether or not having a shared identity with the bot has an effect on the subject's likelihood of following-back or blocking the bot. We interpret follow-backs as evidence of an increased willingness to form a social tie with the bot; on the other hand, blocks are interpreted as a desire not to socialize with the bot. Given that the bots have two salient dimensions of identity, we will be able to test if sharing only one identity (either political or sportive) – and therefore having an unmatched identity in the other dimension – or sharing both identities have a differential effect on the probability of follow-back. Specifically, we have the three following hypotheses:

A1 ***(Follow-backs when subjects share political identity, but not soccer-related identity).*** *The probability that a subject follows-back a bot account with whom they share political identity, but not soccer-related identity, is greater than the probability of follow-back by a subject who does not share either identity with the bot;*

A2 ***(Follow-backs when subjects share soccer-related identity, but not political identity).*** *The probability that a subject follows-back a bot account with whom they share a preferred soccer team, but not political identity, is greater than the probability of follow-back by a subject who does not share either identity with the bot;*

A3 ***(Follow-backs when subjects share both dimensions of identity).*** *The probability that a subject follows-back a bot account with whom they share both political identity and preferred soccer team is greater than the probability of follow-back by a subject who does not share either identity with the bot. This probability is also greater than the probability of follow-back when the bot and subject only share one dimension of identity.*

We have a similar set of hypotheses for our second outcome, blocks:

B1 ***(Blocks when subjects share political identity, but not soccer-related identity).*** *The probability that a subject blocks a bot account with whom they share political identity, but not soccer-related identity, is lower than the probability of block by a subject who does not share either identity with the bot;*

B2 ***(Blocks when subjects share soccer-related identity, but not political identity).*** *The probability that a subject blocks a bot account with whom they share a preferred soccer team, but not political identity, is lower than the probability of block by a subject who does not share either identity with the bot;*

B3 ***(Blocks when subjects share both dimensions of identity).*** *The probability that a subject blocks a bot account with whom they share both political identity and preferred soccer team is lower than the probability of block by a subject who does not share either identity with the bot. This probability is also lower than the probability of block when the bot and subject only share one dimension of identity.*

Recall that all subjects either support Lula or Bolsonaro, and that the bot accounts will only follow subjects who either support the same soccer team as they do, or support a rival team. Therefore, when there is an identity match between bot and subject in only one dimension, this directly implies that bot and subject have opposite identities in the other dimension. Thus, when we consider only our main treatments (bots that signal both identities), we can only study the effect of each dimension of identity (political and sportive) conditional on sharing the other identity, or having an opposite identity in this other dimension. For this reason, we also create bots that are neutral in one dimension of identity. We describe how we analyse the data from this type of bot in the next section.

3.2 Secondary analysis: bots that signal only one dimension of identity

Apart from the bots that signal both their preferred political candidate and their preferred soccer club, we will create bots that are neutral in one of the two dimensions. Specifically, in each wave there will be two politically neutral bots (one supporting each soccer club used in the wave) and two soccer club-neutral bot (each supporting one of the two political candidates). The soccer club-neutral bots will still signal interest for soccer, as described on section 2.

Using these two sets of bots, we will test the hypothesis that each dimension of identity is relevant to the formation of social ties. This time, since the bot is neutral in the other dimension, we do not have to worry that the analysis is conditional on subject and bot either sharing or having opposite identities on the other dimension. Specifically, we test the following hypotheses:

N1 ***(Follow-backs when subjects share political identity).*** *The probability that a subject follows-back a bot accounts with whom they share political identity is greater than the probability of follow-back by a subject who has an opposite political identity to the bot;*

N2 ***(Follow-backs when subjects share soccer-related identity).*** *The probability that a subject follows-back a bot accounts with whom they share a preferred soccer team is greater than the probability of follow-back by a subject who has an opposite soccer-related identity to the bot (i.e., who supports a rival soccer club).*

N3 ***(Blocks when subjects share political identity).*** *The probability that a subject blocks a bot account with whom they share political identity is lower than the probability of block by a subject who has an opposite political identity to the bot;*

N4 ***(Blocks when subjects share soccer-related identity).*** *The probability that a subject blocks a bot account with whom they share a preferred soccer team is lower than the probability of block by a subject who has an opposite soccer-related identity to the bot (i.e., who supports a rival soccer club).*

To test the first and third hypotheses, we use only the data from soccer club-neutral bots; to test the other two hypotheses, we use only data from the politically-neutral bots. By testing these hypotheses, we will be able to know if each dimension of identity we consider is relevant in itself for socialization.

4 Research Plan

4.1 Identification and Estimation

4.1.1 Main Analysis

As discussed in section 3.1, our main analysis will restrict attention to the bots that signal both dimensions of identity (political and soccer preference), and our primary outcome of interested is follow-backs, i.e., follows that happen after the bot account randomly followed the subjects. Thus, we restrict our analysis to the experimentally assigned pairs users-bot accounts.

Our outcome of interest in this section is

$$Y_{ijst} := \mathbb{1}\{\text{user } i \text{ interacted with bot account } j \text{ at wave } t\},$$

where s denotes the strata user i belongs to. Y_{ijst} is an indicator equal to 1 if the user "interacted" with the bot account, in response to the bot account following the account. As discussed in the previous section, we will analyze two types of interactions: follow-backs and blocks. Therefore, Y_{ijst} either indicates a follow-back or a block, depending on the specification. Given that the follows from bot accounts were randomly assigned, the causal effects of congruence in identity between bot and subject are identified and can be estimated by comparing the probability of follow-back (or blocks) between each set of pairs bot-subject. Specifically, we estimate by OLS the following equation:

$$\begin{aligned} Y_{ijst} = \alpha + \beta_1 \times \textit{political\_congruence}_{ij} + \beta_2 \times \textit{soccer\_congruence}_{ij} + \\ \beta_3 \times \textit{political\_congruence}_{ij} \times \textit{soccer\_congruence}_{ij} + \\ X_{ijt}\lambda + \delta_t + \theta_s + \phi_{st} + \varepsilon_{ijst} \quad (1) \end{aligned}$$

where $\textit{political\_congruence}_{ij}$ is an indicator equal to one if bot j and subject i share political preferences, $\textit{soccer\_congruence}_{ij}$ equals one if bot j and subject i share preference for soccer club, δ_t, θ_s and ϕ_{st} represent, respectively, wave, strata and strata $\times$ wave fixed effects[8] and

[8] We include strata fixed effects following the suggestion from Bruhn and McKenzie (2009). We also include strata $\times$ wave fixed effects to account for possible differences in the behavior of subjects from different strata at different moments in time. Moreover, note that, among the strata fixed effects, there will be a misfits dummy.

ε_{ijst} is the error term. X_{ijt} is a vector of control variables from the bot, subjects and waves (interacted with the treatment dummies). Specifically, we include in this vector the number of followers and tweets from the subject; the year he or she created the account; the subject's gender and location; and the bot's soccer team.

Since our outcome is an indicator variable, equation 1 represents a linear probability model, and coefficients β_1, β_2 and β_3 can be interpreted as the difference (in percentage points) in follow-backs (or blocks) between a subject who does not share either identity with the bot and a subject who shares political identity, soccer-related identity, or both (respectively). Hence, to test hypotheses **A1** through **A3**, we will test if the coefficients β_1 through β_3 are positive, respectively, when our outcome is follow-backs; similarly, to test hypotheses **B1** through **B3**, we will test if the coefficients β_1 through β_3 are positive, respectively, when our outcome is blocks.[9] Moreover, to test the second part of hypothesis **A3** – which says that the effect on follow-backs of sharing both identities is greater than the effect of sharing a single identity – we will test if β_3 is significantly greater than β_1 and β_2. We will use a similar procedure to test the second part of hypothesis **B3**.

We will also report estimates without control variables.

4.1.2 Secondary Analysis

To test hypotheses **N1** through **N4**, we will restrict our sample to the subjects that were followed by the bots who are neutral in one of the two dimensions of identity, and estimate a model of the form:

$$Y_{ijst} = \alpha + \beta_1 \times identity\_congruence_{ij} + X_{ijt}\lambda + \delta_t + \theta_s + \phi_{st} + \varepsilon_{ijst} \tag{2}$$

were $identity\_congruence_{ij}$ is an indicator equal to one if bot and subject share identity (in the dimension we are studying) and the other variables have the same definitions as before.

To test hypotheses **N1** and **N3**, we restrict the sample to subjects that were followed by soccer club-neutral bots. On the other hand, to test hypotheses **N2** and **N4**, we restrict our sample to subjects that were followed by politically-neutral bots. As in the main analysis, we also report estimates without controls.

4.2 Inference and Power

We will present standard errors clustered at the bot-account level.

Using the results from the pilot we conducted, we compute the minimum detectable effect of our experiment. Considering our main analysis, for a power fixed at 80% and a 5% significance level, we can detect an effect of sharing political identity, sharing soccer-related identity, and sharing both identities of at least 3, 3 and 4.5 percentage points, respectively, for follow-backs, and of at least 2, 2 and 3.2 percentage points for blocks.

[9]Note that our hypotheses are that sharing identities *increases* follow-backs and *decreases* blocks. Thus, we will be able to perform one-sided tests (with null hypothesis in each case being that $\beta_k \leq 0, k = 1, 2, 3$ for follow-backs and $\beta_k \geq 0, k = 1, 2, 3$ for blocks).

We also perform the simplest assessment proposed by Ferman (2022) to verify if our inference method is reliable. We simulate our data under the null hypothesis of no treatment effects, using Bernoulli draws with parameter equal to the average follow-back rate in the pilot. Reassuringly, we obtained a rate of rejection of the null under a nominal significance level of 5% that was very close to 5%.

4.3 Heterogeneity: Formation of social ties before, during and after the election

Since our experiment will be run during the second semester of 2022 – when the Brazilian presidential elections will also be held – we can explore the timing of follows to study heterogeneous effects of shared identity over time. We have two strategies to study this heterogeneity: first, we will use the official campaign period to divide the sample into three periods (before, after, and during the election); second, we will use Google Trends data as a measure of how salient the political campaign is during each experimental wave. We discuss each method in what follows.

4.3.1 Official campaign dates

For our first analysis, we divide our experimental timeline into three periods: before elections, which goes until the official start of the campaigning period (on August 16^{th}); during elections, which goes from the beginning of the campaign until the last round of voting for president (which can either be on October 3^{rd}, if the election is decided in the first round or October 30^{th} if there is a second round);[10] and after elections, which goes from the day after the last voting happens until the end of our experiment.

To study how the timing affects follow-back and block decisions, we estimate equation (1) including interaction terms between timing indicators and the identity congruence dummies. Specifically, let $B_t := \mathbb{1}\{t \in \{1, 2, \ldots, 8\}\}$ be an indicator equal to one if the wave happened before the election, and $A_t := \mathbb{1}\{t \in \{31, 32, \ldots, 44\}\}$ be an indicator equal to one if the

[10] In Brazil, presidential elections are decided by plurality rule with the possibility of a run-off election if no candidate obtains over 50% of valid votes in the first round. The official election calendar can be accessed at `https://www.tse.jus.br/eleicoes/calendario-eleitoral/calendario-eleitoral`. Access on July 10^{th}, 2022.

wave happened after elections. We then estimate the model:

$$\begin{aligned} Y_{ijst} = \alpha + \beta_1 \times political\_congruence_{ij} + \beta_2 \times soccer\_congruence_{ij} + \\ \beta_3 \times political\_congruence_{ij} \times soccer\_congruence_{ij} + \\ \beta_{b,1} \times \big(B_t \times political\_congruence_{ij}\big) + \beta_{b,2} \times \big(B_t \times soccer\_congruence_{ij}\big) + \\ \beta_{b,3} \times \big(B_t \times political\_congruence_{ij} \times soccer\_congruence_{ij}\big) + \\ \beta_{a,1} \times \big(A_t \times political\_congruence_{ij}\big) + \beta_{a,2} \times \big(A_t \times soccer\_congruence_{ij}\big) + \\ \beta_{a,3} \times \big(A_t \times political\_congruence_{ij} \times soccer\_congruence_{ij}\big) + \\ X_{ijt}\lambda + \delta_t + \theta_s + \phi_{st} + \varepsilon_{ijst} \quad (3) \end{aligned}$$

Note that we omit the during election indicator. Thus, the coefficients $\beta_{b,1}$ through $\beta_{b,3}$ represent the differential effect of shared identity on follow-backs (or blocks) before and during the election, while the coefficients $\beta_{a,1}$ through $\beta_{a,3}$ represent the differential effect of shared identity on follow-backs after and during the election. Our prior is that the official campaign period will increase the salience of the political identity, making it more relevant to the formation of ties relative to the affective identity (soccer-club preference). Thus, we expect that $\beta_{b,1}$ and $\beta_{a,1}$ to be negative, while $\beta_{b,2}$ and $\beta_{a,2}$ to be positive.

We perform a similar analysis with the sample of subjects followed by bots who are neutral in one dimension of identity, adding the interaction terms to equation (2).

4.3.2 Google Trends

While the analysis using the official campaign periods is interesting, our experimental design allows us to study the effect of salience on an even more disaggregated time frame. This might be interesting because, even within the campaign period, there might be moments where politics is particularly salient (such as debates). Using data from Google Trends, which measures the search volume of specific terms on Google, we will have a daily measure of how salient the elections are for Brazilian voters.

Specifically, we will obtain daily Google Trends data for two terms, "Bolsonaro" and "Lula", restricting the search location to Brazil. Google Trends provides an index from 0 to 100 of the volume of searches for one or more terms. We will first obtain the daily average of the search index for the two terms. Then, we will compute the two-day rolling average of this average search volume. This will be our measure of political salience.

For each wave t, we will consider the google trends index for the first day of the wave (which is a two-day rolling average of that day and the next). The two first days of each wave are the days in which follow-backs and blocks are more likely, so that the index will more accurately measure the salience of political preference at the moment in which the subject is treated. We will denote this measure by $trend_t$. Then, we estimate the following equation:

$$
\begin{aligned}
Y_{ijst} = \alpha + \beta_1 \times political\_congruence_{ij} + \beta_2 \times soccer\_congruence_{ij} + \\
\beta_3 \times political\_congruence_{ij} \times soccer\_congruence_{ij} + \\
\beta_{t,1} \times \left(trend_t \times political\_congruence_{ij}\right) + \beta_{t,2} \times \left(trend_t \times soccer\_congruence_{ij}\right) + \\
\beta_{t,3} \times \left(trend_t \times political\_congruence_{ij} \times soccer\_congruence_{ij}\right) \\
X_{ijt}\lambda + \delta_t + \theta_s + \phi_{st} + \varepsilon_{ijst} \quad (4)
\end{aligned}
$$

Note that the coefficients $\beta_{t,1}$ through $\beta_{t,3}$ represent the marginal effect of sharing a political or sportive identity (or both) for a given salience, as measured by $trend_t$. Our prior is that, when political salience is greater, sharing political identity will be more relevant relative to the affective identity, so that $\beta_{t,1}$ will be positive, while $\beta_{t,2}$ and $\beta_{t,3}$ will be negative.

One potential concern with the heterogeneity analysis (both when considering the official campaign period and the google trends index) is that the salience of the soccer clubs may also vary during the experiment timeline. For instance, if a club wins an important match, the behavior of subjects who support that club or its rivals may change.

We will test this using the politically neutral bot accounts. Specifically, we will first test if the average rate of follow-backs and blocks is constant across waves for all politically-neutral bots; then, we will test if these rates are constant for each club across waves. If we do not reject the null hypothesis in each of these cases, we will have evidence that the behavior regarding the affective identity does not change over time, so the concern described in the previous paragraph will be minimized. On the other hand, if there is evidence of a differential pattern of behavior for the politically neutral bots over time, we will deal with this by controlling for the bot's soccer club google trend index in all regressions. We will do this using the same strategy we use for the google trends index of political salience: for each club, we obtain the average google trend index for the first two days of each wave, and interact this index with the treatment dummies in every regression.

5 Assessing the experiment

5.1 Attrition

Considering our initial sample of subjects, attrition can happen for two reasons: first, users may deactivate their accounts; second, Twitter may suspend users who do not follow the platform's rules and policies. This attrition is not a concern within each wave, since we know right before each wave which subject's accounts still exist and which were suspended or deactivated.

However, it may be a concern for the heterogeneity analysis, since our aim is to compare results across waves. Thus, if the subject pool in different waves is not comparable, we will not be able to argue that the differential effects obtained in different waves is due to the proximity of the election or other events that make some dimensions of identity more salient; the difference could possibly be due to differences in the sample.

To alleviate this concern, we will report two sets of evidence. First, we will perform balance tests comparing the attrited subject pool and the subject pool of 'survivors', i.e., accounts that did not suffer attrition at any point during the experiment. If the two sets of accounts are similar, the concern discussed previously will not be relevant. Second, we will report our estimates for all waves using only the sample of 'surviving' accounts. For instance, if an account was treated in the first wave, but was suspended by Twitter after the 15^{th} wave, we would exclude this account from the analysis in this robustness exercise.

5.2 Active Accounts

Even if the accounts are not formally deactivated, the frequency of their usage may change. Indeed, a user may stop using Twitter for some reason, or at least start checking it less often, which would reduce the likelihood of this user seeing the treatment notification. We can assess this possibility because, before each experimental wave, we know when each user posted their most recent tweet. As a way of verifying that our results are not due to a fall in subjects' engagement on Twitter, we will provide results for the subset of users whose last tweet in each wave in which they were treated happened at most 7 days before treatment.

References

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

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.

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

Iyengar, Shanto, 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.

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

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

Appendix

A Additional Information

A.1 Procedure to obtain the Subject Pool

(i) Obtain a sample of Twitter users who used (tweeted or rt-ed) either a pro-Bolsonaro or pro-Lula hashtag between May 31^{st} and July 11^{th}, 2022.

(ii) Within this sample, obtain all users that signal their preferred Brazilian soccer club in their Bio. We restrict attention to the 6 clubs with the largest number of supporters and to historical rivals of these clubs.[11]

(iii) Manually check that subjects are correctly categorized in terms of political identity and preferred soccer club.

(iv) Exclude accounts that are clearly bots, that were created in 2022, that have less than 10 followers and that have a followers over friends ration above 20.

A.2 Pro-Lula and Pro-Bolsonaro Hashtags

Table A.1: 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

[11]The list of 6 largest clubs (in terms of number of supporters) comes from a 2022 survey made by the companies XP and SportTrack (Sport Track and XP, 2022). The clubs (in order of supporters) are: Flamengo, Corinthians, São Paulo, Palmeiras, Grêmio and Vasco. Apart from these 6 clubs, we also added to our sample supportes of Botafogo F.R. and Fluminense F.C. (who are Flamengo and Vasco's historical rivals), Santos F.C (who is a rival of Palmeiras, São Paulo and Corinthians), and S.C. Internacional (who is a rival of Grêmio).

A.3 Soccer Club Rivalries

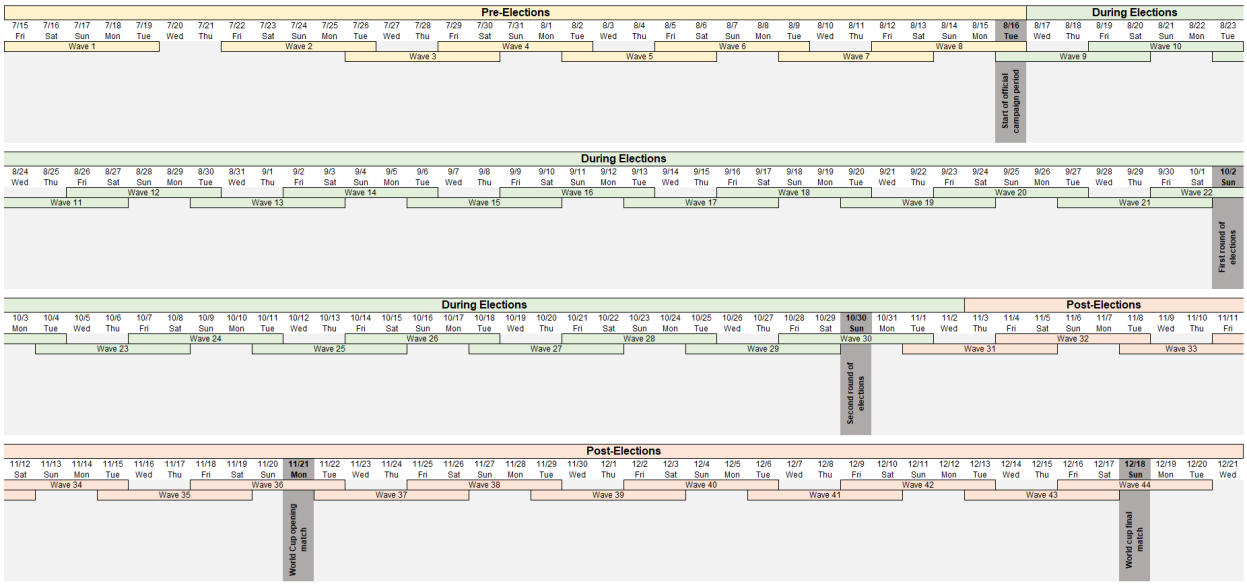
Table A.2: Soccer club rivalries

	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 soccer club rivalries we considered when constructing the sample of subjects. The X mark indicates a rivalry. A bot that signals support for team A will only follow subjects whose preferred soccer 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 bot 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 bot accounts follow the subjects. The table also shows the periods we define as before, during, and after the election period, along with relevant dates.

B Additional Figures and Tables

Figure A.2: Examples of Bot Accounts

(a) Pro-Bolsonaro; Flamengo supporter

(b) Pro-Lula; Palmeiras supporter

(c) Pro-Lula; Neutral-Team