

Supplemental Appendix to “Discrimination in the Formation of Academic Networks: A Field Experiment on *#EconTwitter*”

Nicolás Ajzenman Bruno Ferman Pedro C. Sant’Anna

October 30, 2024

A Additional Information on Experimental Design	2
A.1 Procedure for Obtaining the Subject Pool	2
A.2 Experimental Wave Timeline	2
A.3 Details on Experimental Accounts	3
B Additional Figures and Tables	7
B.1 Characteristics of Twitter Profiles of Real PhD Students	7
B.2 Description of Subject-level Variables	8
B.3 Balance and Evidence of No Differential Attrition	10
B.4 Regression Results and Robustness	12
B.5 Evolution of Follow-Back Rates	19
B.6 Organic Follows by Bot Group	21
C Implementation Details	22
D Total Follows and Concerns with Contamination by Twitter’s Algorithm	24

A Additional Information on Experimental Design

A.1 Procedure for Obtaining the Subject Pool

- (i) From January to February 2022, obtain all Twitter users that either tweeted or retweeted a status containing the term “*#EconTwitter*”¹ $\rightarrow$ 14,449 accounts.
- (ii) Remove accounts that no longer exist, accounts that are clearly bots, and protected accounts² $\rightarrow$ 14,055 accounts.
- (iii) Compute follows/friends ratio for the remaining account. Remove accounts with a follows/friends ratio above 15 and accounts with fewer than 10 friends and institutional accounts $\rightarrow$ 10,173 accounts.³ This is our final subject pool.

A.2 Experimental Wave Timeline

Table A.1: Wave timeline

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
...	Create Accounts	Day 0 Follow Friends Rt (x2)	Day 1 Follow subjects	Day 2	Day 3	Day 4
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Day 12	Deactivate Accounts	...	...	...	...	...

Notes: The table shows the timeline of an experimental wave. Accounts are active for 12 days after following the experimental subjects. Each wave starts on a Tuesday, so there will always be accounts from two waves active in the same period (eight of them in the second week and eight in the first week of the wave).

¹The search considered all variations of capital and small letters for the term.

²Note that an account that tweeted a status containing “*#EconTwitter*” at the beginning of January, for instance, may no longer exist at the beginning of March (the account owner may have deleted the account). We identified accounts that were clearly bots by analyzing the accounts’ Twitter bios. In Twitter, “Protected” accounts are the ones that choose not to be public, restraining their information and interaction to the account’s friends.

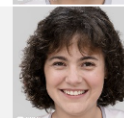
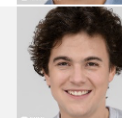
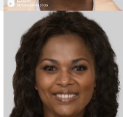
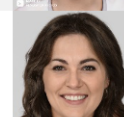
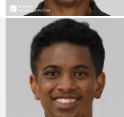
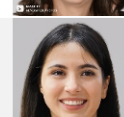
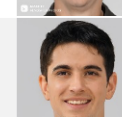
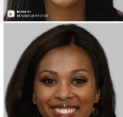
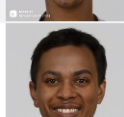
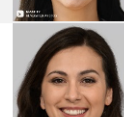
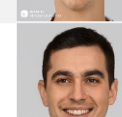
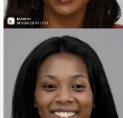
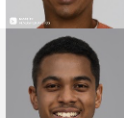
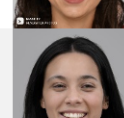
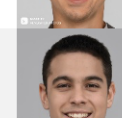
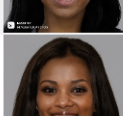
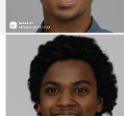
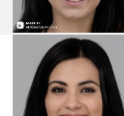
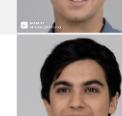
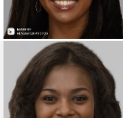
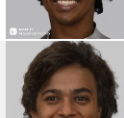
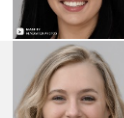
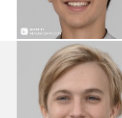
³15 is approximately the follows/friends ratio of the 95th percentile of the subject pool sample after step (ii). We removed accounts with too few friends because those accounts are likely to be inactive or (at least) are extremely unlikely to follow an unknown account.

A.3 Details on Experimental Accounts

Figure A.1: Example of Experimental Accounts



Figure A.2: Profile Pictures Used in the Experiment

1					Wave 6 (lower-ranked) Wave 9 (top-ranked)
2					Wave 2 (top-ranked) Wave 8 (lower-ranked) Wave 10 (lower-ranked)
3					Wave 3 (top-ranked)
4					Wave 1 (lower-ranked) Wave 4 (lower-ranked) Wave 7 (top-ranked)
5					Wave 6 (top-ranked) Wave 9 (lower-ranked)
6					Wave 3 (lower-ranked) Wave 5 (lower-ranked) Wave 10 (top-ranked)
7					Wave 4 (top-ranked) Wave 7 (lower-ranked)
8					Wave 8 (top-ranked)
9					Wave 2 (lower-ranked) Wave 5 (top-ranked)
10					Wave 1 (top-ranked)

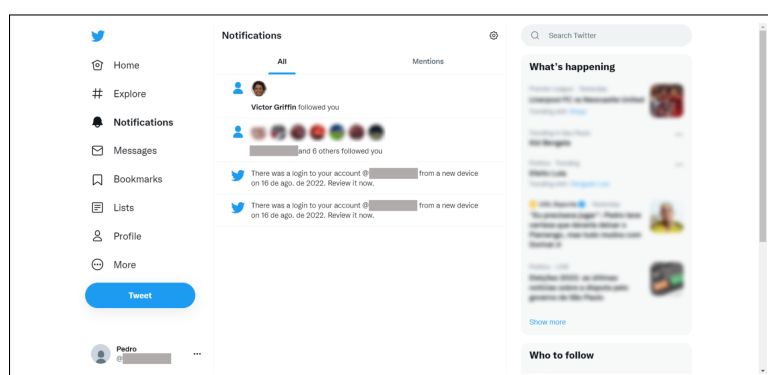
Notes: The picture shows all AI-generated images used as profile pictures in the experiment. The four pictures in each row share the same base image, which is one of the four images (randomly chosen). To construct the other three images from the base, we kept all attributes constant apart from gender or skin tone. The last column in the table indicates the experimental wave(s) and university affiliation for which each set of images was used. In the creation of the Twitter profiles, we cropped the images so the watermark on the left-hand side of the pictures did not appear, even if someone clicked on the profile picture. As discussed on footnote 8 and shown in Appendix Figure A.3, profile pictures appear small in follow notifications.

Table A.2: Procedures used to Create the Bot Accounts

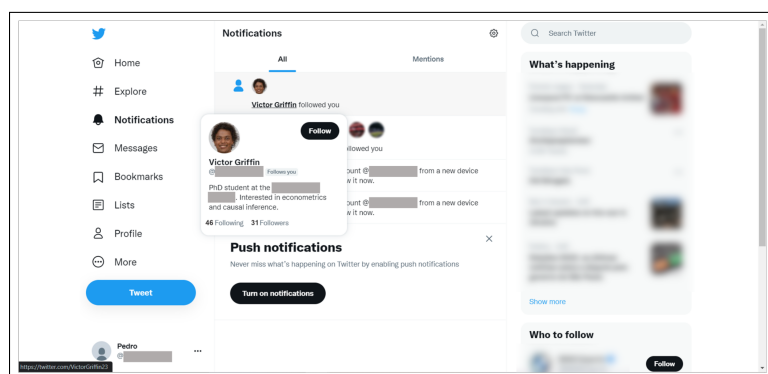
Element of Profile	Procedure
Profile Picture	Use AI generated images from Generated Media Inc. The company’s tool allows us to control several parameters when generating each picture: gender, head pose, age, emotion, skin tone, hair color, hair length, wearing glasses or not, and wearing make-up or not. For each set of four profile pictures, we start from the same “base” face and vary gender (male or female) and skin tone (black or white).
Name	Randomly generated by matching a list of the most common first names and surnames in the US. We excluded from the list all names that are gender-neutral (specifically, we used the NamSor tool to predict the gender of the names in our list, and excluded those with less than 90% confidence in the gender prediction). We also excluded names that are specific to a certain race or ethnicity (to identify race and ethnicity specific names, we used data from Tzioumis (2018), and exclude from the pool of names all names above the 90 percentile of prevalence among Blacks and Whites in the US).
Bio	The Bio from the bot accounts contains two pieces of information: first, the university where they claim to be doing their PhD; second, their research interests. To select the universities, we first considered the ten highest-ranked universities and the universities ranked between positions 79 and 100 in the same ranking that made their list of PhD students publicly available, based on the 2017 USNews rank of graduate universities in economics. We randomly selected 5 universities from each of these two sets. The interest was also randomly assigned from a list designed by the authors. Generally, the bio from a bot account was something like: “PhD student at University X. Interested in labor economics and economics of education.” We decided not to use the university’s Twitter handles (for instance, @UniversityX) because this would likely affect follow recommendations made by Twitter’s algorithm, which could bias the experiment (if Twitter recommended the bot accounts to users from the same university, bots from some universities could get a disproportional volume of followers for reasons unrelated to discrimination). It is harder for the algorithm to target recommendations when the Twitter handle is not used. Importantly, we do not explicitly say in the bio that the student is doing his or her PhD in economics. This is implicit from the interests listed.
Background Image	A landscape from the city where the student claims to be doing the PhD. We have a single landscape for each city.
Location	The bot accounts’ profiles did not include a location.
Website	The bot accounts’ profiles did not include a website.
Retweets	Before following subjects, the bot account retweets two statuses from accounts of academic journals in the field of economics. These two retweets are randomly chosen.
Followers	We asked a group of economic professors and graduate students to follow the bot accounts one day before the bot account followed the accounts randomly assigned to it.
Following	One day before following the accounts randomly assigned to it, the bot account follows all professors and graduate students we had asked to follow the account. It also follows some accounts from academic journals and other institutions related to the field of economics.

Notes: The table summarizes the procedures used to create the bot accounts.

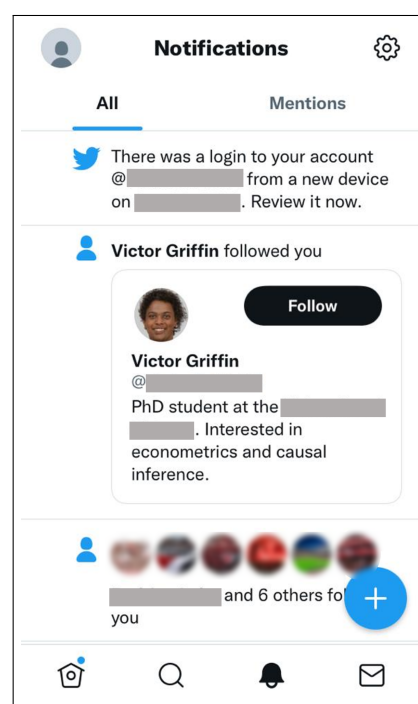
Figure A.3: Example of Treatment Notifications on Desktop and Mobile Twitter Apps



(a) Desktop notification



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



(c) Mobile app notification

B Additional Figures and Tables

B.1 Characteristics of Twitter Profiles of Real PhD Students

Table B.1: Summary Statistics of Real-life Twitter Profiles of First- and Second-year Economics PhD Students

	2021 Cohort			2020 Cohort		
	Median	Mean	Std.Dev	Median	Mean	Std.Dev
Tweets	94.5	357.33	917.26	47	174.58	252.40
Following	331.5	471.29	495.88	279	382.32	327.60
Followers	152.5	372.96	507.11	115	401.00	807.61
Website	0.0	0.29	0.46	0	0.26	0.45
Background Image	1.0	0.71	0.46	0	0.47	0.51
Location	0.0	0.42	0.50	1	0.53	0.51
Profile Pic is a Self-Portrait	1.0	0.83	0.38	1	0.89	0.32

Notes: The table shows descriptive statistics of Twitter profiles from first and second-year PhD students from three universities. For the first-year (2021) cohort, 24 out of 59 students had profiles we could find (40.67%) as of January 6th, 2022. For the second-year (2020) cohort, we could find profiles for 18 out of 56 students (32.14%).

B.2 Description of Subject-level Variables

Table B.2: Description of Variables at the Subject Level

Variable	Description	N (%)
Gender	Whether the account belongs to someone identified as male or female. To obtain this information, we used the users' full name to predict its gender, using the NamSor tool, ¹ which accurately predicts gender based on full names. We only considered predictions done with above 90% confidence, and assigned as missing the gender information for the accounts with confidence below this threshold. We manually checked a randomly selected subsample of 100 accounts, and obtained 98% accuracy.	8,316 (58.4%)
Nfavorites	Number of tweets marked as "favorite" (i.e., "liked") by the user.	14,055 (100%)
Nfollows	Number of accounts the user follows.	14,055 (100%)
Nfriends	Number of friends the user has, i.e., number of accounts that follow the user.	14,055 (100%)
Verified	Indicator variable equal to one if the account is verified, a "badge" provided by Twitter to signal that the account is authentic.	14,055 (100%)
Continent	The continent in which the user lives. We obtained this information via the "location" information from Twitter. This information is provided by the user and can, in principle, be anything (it does not have to be a real location and does not have to be correct). We classified the "real" location given by region: North America, South and Central America, Europe, Asia, Africa, Oceania. If a person indicated more than one place from different continents, we classified the location as missing. At the end of the procedure, we manually checked a random subsample of 100 accounts and obtained 100% accuracy.	8,931 (62.7%)
Profession	The user's profession. We classified professions using the user's account description (or "bio"). The list of professions/areas of work is: professor (which is subdivided into "assistant", "associate" and "other"); PhD student; Post-Doc; Other academic position (for instance, Research Fellow, Research Assistant, etc.); Industry/Tech; Government; Non-profit/Multi-lateral Organization; Journalist. We first searched for keywords related to each profession, and then manually verified the matches. At the end of the procedure, we checked a random subsample of 100 accounts and obtained 99% accuracy.	8,555 (60.1%)
Race/Ethnicity	We manually classify perceived race or ethnicity from subjects' profile pictures.	8,983 (63.9%)

Continued on next page

Table B.2: Description of Variables at the Subject Level (Continued)

Variable	Description	N (%)
University Affiliation	Indicator variable equal to one if a user is affiliated to a highly ranked university. To obtain this information, we also consider the user’s account description (“bio”) and search for keywords associated with the highly ranked universities. We obtain this variable for top-ten and top-twenty US universities according to the USNews Ranking.	14,055 (100%)
Year of Account Creation	The year in which the account was created. This information is provided by Twitter’s API and is therefore perfectly precise.	14,055 (100%)
Concern about the Lack of Diversity in Economics	Indicator variable equal to one if a user follows at least one Twitter account dedicated to this topic. The list of accounts we consider are: @AEACSMGEP (the American Economic Association Committee on the Status of Minority Groups in the Economics Profession); @AEACSWEP (the AEA Committee on the Status of Women in the Economics Profession); @ResearchInColor (a foundation whose objective is to “diversify economics by increasing the number and retention of scholars of color in economic disciplines through mentoring and financial support); @SadieCollective (a collective that is “addressing the pipeline pathway for Black women in economics and related fields”) and @weconpol (“an inclusive community for women interested in econ, policy and development”).	14,055 (100%)
Background Picture	Indicator variable equal to one if the user has a background picture (banner).	14,055 (100%)
Share of Tweets/Retweets via Mobile App	We live-streamed tweets and retweets from the sample of subjects during two weeks in October 2022 (October 06 th -12 th and 18 th -24 th). For each tweet, we collected the source (e.g., mobile app, desktop, etc.). For each user, we compute the share of tweets and retweets in this period that were sent via Twitter’s mobile app.	13,263 (94.4%)

¹ We chose this tool for a few reasons: first, it has already been used in academia, including to predict names using Twitter data (e.g., [Hridoy et al. \(2015\)](#)); second, it has been shown to be at least as accurate as similar tools ([Sebo, 2021](#)); third, its database includes names from a variety of countries, and permits the analysis of full names.

Notes: The table lists and describes the variables obtained for all users on #EconTwitter (the universe of accounts that tweeted or retweeted a message containing the term *#EconTwitter* between January and February 2022). Column N (%) shows the number of accounts and the percentage of the total pool for which we were able to obtain each piece of information. Table 1 shows descriptive statistics for the final experimental sample (which excludes bots, institutional accounts, and very big or very small accounts, as specified in the pre-analysis plan).

B.3 Balance and Evidence of No Differential Attrition

Table B.3: Balance Table

Variable	Treatment Arm (Bot's characteristics)								F Stat [p-value]
	Male				Female				
	White		Black		White		Black		
	Top-ranked	Lower-ranked	Top-ranked	Lower-ranked	Top-ranked	Lower-ranked	Top-ranked	Lower-ranked	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Number of followers	1,176.3 (3,841.1)	1,268.6 (4,655.4)	1,190 (3,688.7)	1,356.8 (5,375)	1,297.3 (5,739)	964.7 (3,282.6)	1,323.7 (5,105.4)	1,467.7 (7,180)	0.0079 [1.00]
Number of friends	1,221.3 (1,590.3)	1,254.5 (1,918.4)	1,406.2 (2,193.9)	1,341.4 (3,087)	1,344.6 (4,099.7)	1,227.9 (1,370.4)	1,403.2 (2,780.2)	1,421.4 (2,276.2)	0.0091 [1.00]
Number of statuses ('tweets')	20,004.7 (92,373.2)	17,642.6 (65,217.5)	22,446.5 (69,613.1)	19,747.7 (68,805)	16,310.8 (47,260.7)	18,150.2 (56,881)	20,993.9 (76,448.8)	21,859.3 (105,441.2)	0.007 [1.00]
Number of favorited statuses ('likes')	21,156.6 (64,822.6)	20,104.4 (54,783.8)	23,046.6 (65,082.2)	25,402 (72,248.4)	23,033.7 (60,450.7)	20,377.7 (53,227.9)	23,116.1 (65,656)	21,344 (56,885.5)	0.0073 [1.00]
Number of lists	24.766 (97.3)	27.324 (113.1)	25.844 (101.8)	25.796 (91.1)	27.185 (112)	19.585 (65.9)	32.052 (155.7)	42.824 (254.5)	0.0226 [1.00]
Share above bottom 20% followers	0.787 (0.41)	0.788 (0.409)	0.804 (0.397)	0.793 (0.405)	0.795 (0.404)	0.807 (0.395)	0.823 (0.382)	0.808 (0.394)	0.0081 [1.00]
Share above bottom 20% friends	0.779 (0.415)	0.795 (0.404)	0.812 (0.391)	0.801 (0.399)	0.802 (0.399)	0.809 (0.393)	0.809 (0.393)	0.809 (0.393)	0.0066 [1.00]
Share above bottom 20% statuses	0.778 (0.416)	0.78 (0.414)	0.807 (0.395)	0.796 (0.403)	0.801 (0.399)	0.802 (0.398)	0.806 (0.396)	0.818 (0.386)	0.0102 [1.00]
Account is verified	0.008 (0.089)	0.014 (0.117)	0.008 (0.088)	0.01 (0.101)	0.008 (0.089)	0.006 (0.075)	0.015 (0.121)	0.009 (0.095)	0.009 [1.00]
Year of account creation	2,014.6 (4.176)	2,014.5 (4.164)	2,014.5 (4.178)	2,014.3 (4.118)	2,014.6 (4.112)	2,014.6 (4.059)	2,014.6 (4.179)	2,014.4 (4.158)	0.0063 [1.00]
Has background picture	0.734 (0.442)	0.724 (0.447)	0.771 (0.42)	0.773 (0.419)	0.741 (0.439)	0.755 (0.43)	0.78 (0.415)	0.759 (0.428)	0.019 [1.00]
Follows diversity accounts	0.154 (0.362)	0.156 (0.363)	0.176 (0.381)	0.159 (0.366)	0.176 (0.381)	0.16 (0.367)	0.172 (0.377)	0.146 (0.353)	0.0079 [1.00]
Female	0.174 (0.379)	0.179 (0.383)	0.178 (0.383)	0.178 (0.383)	0.175 (0.38)	0.173 (0.378)	0.175 (0.38)	0.167 (0.374)	8e-04 [1.00]
Affiliated to Top-10 University	0.031 (0.172)	0.052 (0.221)	0.033 (0.178)	0.042 (0.201)	0.042 (0.201)	0.037 (0.19)	0.034 (0.182)	0.044 (0.205)	0.0111 [1.00]
Profession									
Graduate Student	0.216 (0.412)	0.208 (0.406)	0.229 (0.42)	0.216 (0.412)	0.216 (0.412)	0.23 (0.421)	0.211 (0.409)	0.223 (0.416)	0.0032 [1.00]
Professor	0.211 (0.408)	0.214 (0.411)	0.193 (0.395)	0.2 (0.4)	0.21 (0.408)	0.195 (0.397)	0.205 (0.404)	0.2 (0.4)	0.0031 [1.00]
Works in Industry/Tech	0.067 (0.25)	0.104 (0.306)	0.084 (0.277)	0.09 (0.286)	0.091 (0.288)	0.074 (0.262)	0.081 (0.273)	0.081 (0.274)	0.015 [1.00]
Other	0.076 (0.265)	0.049 (0.217)	0.058 (0.233)	0.07 (0.256)	0.055 (0.228)	0.076 (0.265)	0.073 (0.26)	0.067 (0.25)	0.0152 [1.00]
Region									
Europe	0.233 (0.423)	0.243 (0.429)	0.234 (0.424)	0.251 (0.434)	0.257 (0.437)	0.236 (0.425)	0.26 (0.439)	0.25 (0.433)	0.0053 [1.00]
Canada/US	0.243 (0.429)	0.228 (0.42)	0.255 (0.436)	0.254 (0.436)	0.24 (0.427)	0.224 (0.417)	0.255 (0.436)	0.232 (0.422)	0.0076 [1.00]
Other	0.131 (0.337)	0.147 (0.354)	0.159 (0.366)	0.151 (0.358)	0.133 (0.339)	0.159 (0.366)	0.142 (0.349)	0.149 (0.357)	0.0081 [1.00]
Number of treated observations	881	873	765	881	875	881	880	884	
%	0.127	0.126	0.111	0.127	0.126	0.127	0.127	0.128	
Attrition (not treated)	16	25	16	17	22	17	18	19	0.0048 [1.00]
% of assigned to treatment	0.018	0.028	0.02	0.019	0.025	0.019	0.02	0.021	

Notes: This table presents descriptive statistics of treated subjects in the experiment for each of the eight treatment arms. For each treatment arm and each pre-treatment variable, the table presents the mean value among treated subjects, as well as its standard deviation (in parentheses). The last column in the table displays an F-statistic of a joint test of difference in means across the treatment arms, along with the test's p-value. 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 description of the variables is in the Appendix. 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 deactivated their account, were suspended by Twitter, or chose to make their profile private). The F-statistic displayed for this row is obtained from a regression of the attrition indicator on the treatment indicators. The last row shows the percentage of participants assigned to treatment that could not be treated.

Table B.4: Differential Attrition

Variable	Treatment Arm (Bot's characteristics)								F Stat [p-value]
	Male				Female				
	White		Black		White		Black		
	Top-ranked	Lower-ranked	Top-ranked	Lower-ranked	Top-ranked	Lower-ranked	Top-ranked	Lower-ranked	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Number of followers	845.4 (852.1)	760.8 (1,384.3)	1,686.8 (3,703.5)	1,204.2 (3,590.2)	947.6 (1,769.5)	1,444.6 (4,352)	621.5 (766.1)	1,094.5 (2,048.1)	0.113 [0.997]
Number of friends	1,857.2 (1,741.9)	1,479.1 (1,802.1)	2,020.1 (3,518.2)	1,454.7 (3,314.5)	1,700.1 (1,820.6)	1,879.8 (4,549.1)	1,568 (1,589.7)	1,232.2 (1,388.5)	0.0586 [1.00]
Number of statuses ('tweets')	72,216.1 (170,092.4)	74,999.8 (190,548.1)	33,770.6 (56,954)	19,775.5 (41,656.9)	64,219 (160,830.6)	31,836.8 (74,665.9)	30,754.6 (62,575.3)	39,077.7 (127,206.9)	0.1703 [0.99]
Number of favorited statuses ('likes')	76,527.5 (192,158.3)	57,006.5 (122,428.2)	26,767.8 (34,879.2)	21,817.2 (31,661)	43,991.5 (81,500.1)	27,468.9 (84,383.6)	67,903.7 (152,098.8)	27,099.6 (46,041.1)	0.2199 [0.979]
Number of lists	42.625 (110.6)	15.68 (49.8)	15.75 (34.6)	5.647 (9)	8.136 (23.2)	7.765 (12.3)	8.722 (16.4)	32.579 (86.8)	0.3441 [0.929]
Account is verified	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 [-]
Year of account creation	2,015.5 (5.007)	2,015.6 (4.803)	2,016.1 (2.778)	2,018.1 (3.436)	2,016.5 (4.351)	2,014.8 (4.72)	2,014.6 (4.461)	2,016.3 (4.042)	0.363 [0.919]
Has background picture	0.625 (0.5)	0.72 (0.458)	0.875 (0.342)	0.706 (0.47)	0.773 (0.429)	0.882 (0.332)	0.944 (0.236)	0.789 (0.419)	0.3792 [0.91]
Follows diversity accounts	0.062 (0.25)	0.12 (0.332)	0.062 (0.25)	0.059 (0.243)	0.091 (0.294)	0.118 (0.332)	0 (0)	0.105 (0.315)	0.1283 [0.996]
Female	0.125 (0.342)	0.16 (0.374)	0 (0)	0.176 (0.393)	0.182 (0.395)	0.059 (0.243)	0.111 (0.323)	0 (0)	0.3491 [0.926]
Affiliated to Top-10 University	0 (0)	0.04 (0.2)	0 (0)	0.059 (0.243)	0.091 (0.294)	0 (0)	0 (0)	0 (0)	0.3074 [0.947]
Profession									
Graduate Student	0.062 (0.25)	0.16 (0.374)	0.188 (0.403)	0.176 (0.393)	0.227 (0.429)	0.059 (0.243)	0.056 (0.236)	0.158 (0.375)	0.2176 [0.979]
Professor	0.062 (0.25)	0.08 (0.277)	0 (0)	0 (0)	0 (0)	0.176 (0.393)	0 (0)	0.053 (0.229)	0.5069 [0.825]
Works in Industry/Tech	0.188 (0.403)	0.04 (0.2)	0.062 (0.25)	0 (0)	0.136 (0.351)	0.118 (0.332)	0.111 (0.323)	0.316 (0.478)	0.571 [0.776]
Other	0 (0)	0.04 (0.2)	0 (0)	0 (0)	0.045 (0.213)	0.059 (0.243)	0 (0)	0 (0)	0.1939 [0.985]
Region									
Europe	0.188 (0.403)	0.32 (0.476)	0.188 (0.403)	0.059 (0.243)	0.091 (0.294)	0.118 (0.332)	0.056 (0.236)	0.158 (0.375)	0.4176 [0.886]
Canada/US	0.312 (0.479)	0.2 (0.408)	0.188 (0.403)	0.235 (0.437)	0.273 (0.456)	0.176 (0.393)	0.167 (0.383)	0.158 (0.375)	0.1004 [0.998]
Other	0.188 (0.403)	0.16 (0.374)	0.25 (0.447)	0.294 (0.47)	0.136 (0.351)	0.294 (0.47)	0.111 (0.323)	0.263 (0.452)	0.188 [0.987]
Number of untreated observations	16	25	16	17	22	17	18	19	0.0048 [1.00]

Notes: This table presents descriptive statistics of subjects that attrited, i.e., that were assigned to treatment but that were not treated in the experiment, for each of the 8 treatment arms. This could happen for one of three reasons: the users deactivated their account, were suspended by Twitter, or chose to make their profile private. For each treatment arm and each pre-treatment variable, the table presents the variable mean among attrited subjects, as well as the standard deviation (in parentheses). The last column in the table displays an F-statistic of a joint test of difference in means across the treatment arms, along with the test's p-value. 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, i.e., there is no differential attrition. The description of the variables is in the Appendix. The F-statistic displayed in the last row is obtained from a regression of the attrition indicator on the treatment indicators.

B.4 Regression Results and Robustness

Table B.5: Effect of Bot Characteristics on Follow-Backs

<i>Dependent Variable:</i>	<i>Follow Backs (1=Yes)</i>			
	Full Sample		Excluding Same University	
	(1)	(2)	(3)	(4)
Bot's Gender (1=Female)	0.044 (0.008)	0.043 (0.008)	0.044 (0.008)	0.044 (0.008)
Bot's Race (1=Black)	-0.021 (0.008)	-0.022 (0.008)	-0.023 (0.008)	-0.023 (0.008)
Bots' University Affiliation (1=Top-Ranked)	0.036 (0.008)	0.035 (0.008)	0.033 (0.008)	0.032 (0.008)
Controls	No	Yes	No	Yes
Wave, Strata Fixed Effects	Yes	Yes	Yes	Yes
Observations	6920	6920	6735	6735
Dep. Variable Mean	0.187	0.187	0.188	0.188

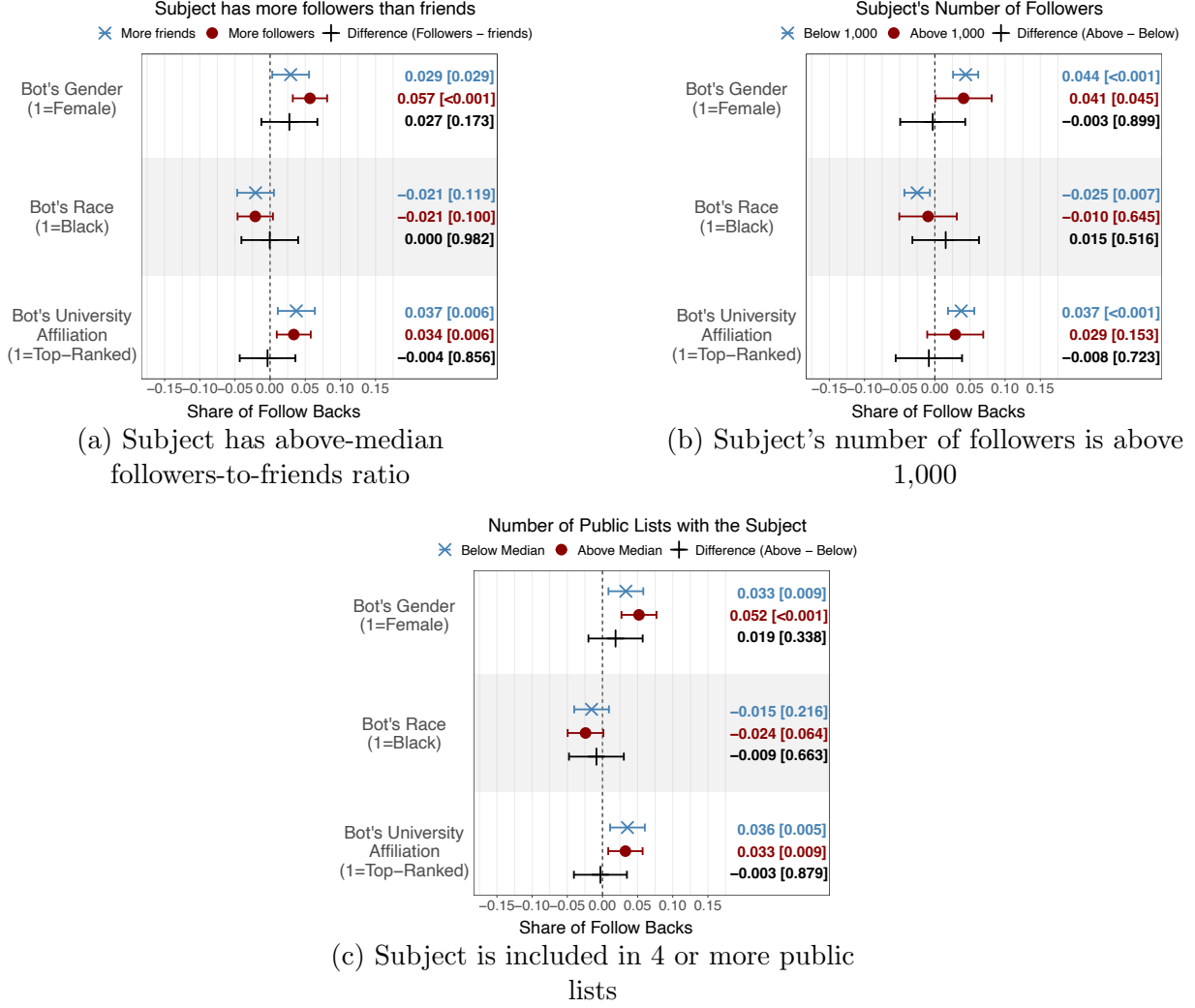
Notes: The table displays regression results for the main experiment, using data from the ten experimental waves, excluding shadow-banned accounts. The first two columns display results using the full sample of subjects, while the last two exclude subjects that claim (in their Twitter bio) to be affiliated with the universities used in each wave. The dependent variable is an indicator equal to one if a Twitter user (subject) that was followed by a bot subsequently followed the bot back. The controls used in specifications (2) and (4) are (at the subject level): continent, profession, gender, affiliation to Top-10 university, year of account creation, has background picture, follows accounts addressing the lack of diversity in academia, has a verified account, number of Twitter followers and number of Twitter friends. Clustered standard errors at the bot-account level are in parentheses.

Table B.6: Follow-Back Gaps for Subsets of Experimental Accounts

	Bot's Gender (1=Female)	Bot's Race (1=Black)	Bot's Affiliation (1=Top-Ranked)
Panel A: Excluding male AI-photos with long (non-short) hair			
	0.027 (0.009) [0.004]	-0.021 (0.007) [0.006]	0.044 (0.007) [0.000]
Number of Observations	4992		
Number of Clusters (Bots)	51		
Panel B: Excluding bots of Black students whose names have below-median prevalence among Blacks			
	0.036 (0.009) [0.000]	-0.020 (0.010) [0.045]	0.035 (0.010) [0.001]
Number of Observations	5289		
Number of Clusters (Bots)	54		
Panel C: Excluding bots whose names have below-median prevalence among bot's race			
	0.047 (0.012) [0.000]	-0.030 (0.011) [0.012]	0.035 (0.011) [0.003]
Number of Observations	3343		
Number of Clusters (Bots)	34		

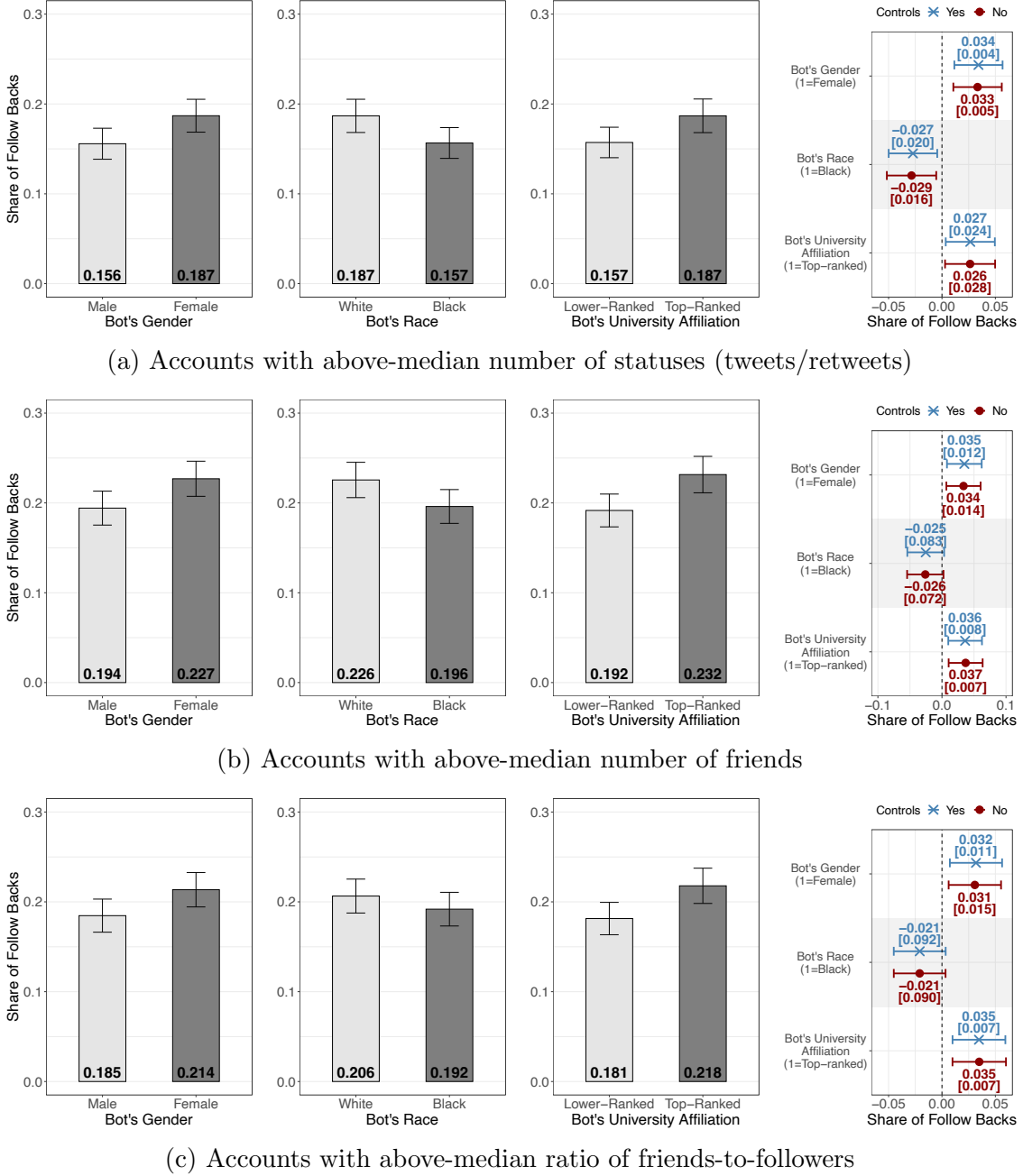
Notes: The table presents regression results of the experiment for different subsets of experimental accounts. We report results for the specification in Equation (1), with each column showing point estimates and confidence intervals (in parentheses) for each of the three bot characteristics. Panel A shows results excluding male photos with long or not-short hair (photo sets 1, 4, 5, 9, and 10). All sets of photos are available in Appendix Figure A.2. We performed this exercise because a referee raised concerns about photos of Black men with non-short hair being easier to identify as AI-generated than their White male counterparts. The last two panels exclude experimental accounts based on the racial prevalence of their names, using data from Tzioumis (2018). Panel B excludes experimental accounts of Black students whose names have a below-median prevalence among Black Americans (we compute the median within the sample of names that were part of the experiment, which already excluded the most racialized White and Black names). Finally, Panel C excludes experimental accounts whose names have a below-median prevalence among their own race. We performed these last two robustness checks to account for the possibility that bot accounts of Black students with names unlikely to belong to Black Americans were easier to identify as inauthentic. Clustered standard errors at the bot-account level are in parentheses.

Figure B.1: Heterogeneity in Follow-Back Behavior by Subject's Account Reach, Alternative Definitions of Reach



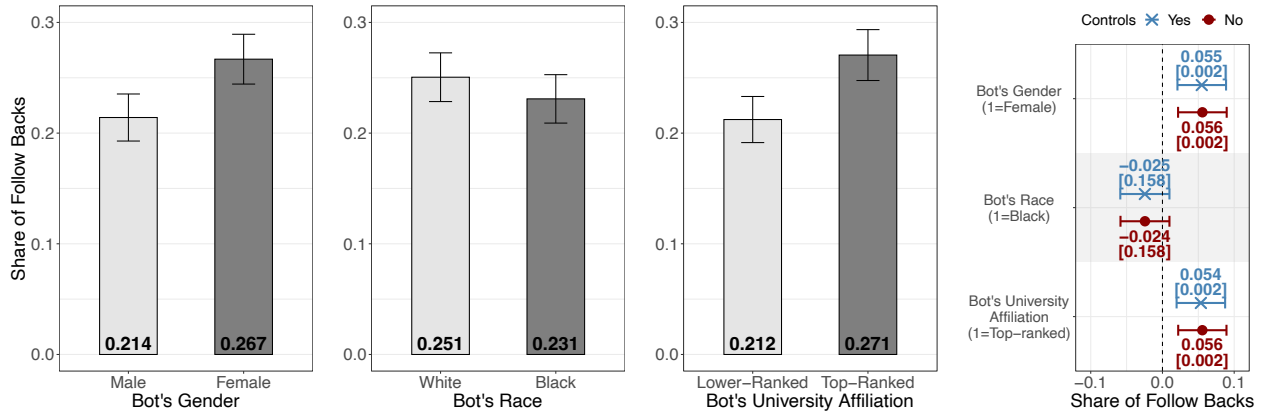
Notes: The figures show estimates of the follow-back behavior of strong and weak accounts analogous to those of Figure 4b, using different definitions of strength. The first panel defines stronger accounts as those with an above-median ratio of followers to friends (the median followers-to-friends ratio in our sample is 0.62). A higher followers-to-friends ratio means that the account has relatively more followers than friends (i.e., accounts it follows). The second panel considers accounts with more than 1,000 followers as strong. The third panel defines accounts listed in at least four public lists as strong. A Twitter list is a selection of Twitter accounts discussing the same topic. Lists can be created by Twitter users themselves, and then be left public—so that other users can subscribe to “follow” this list. Thus, if an account is included in a public list, other users may find this account worth following. Four lists was chosen as a cutoff because this is the median value of the variable in this sample.

Figure B.2: Main Results for Subsamples of More Active Accounts



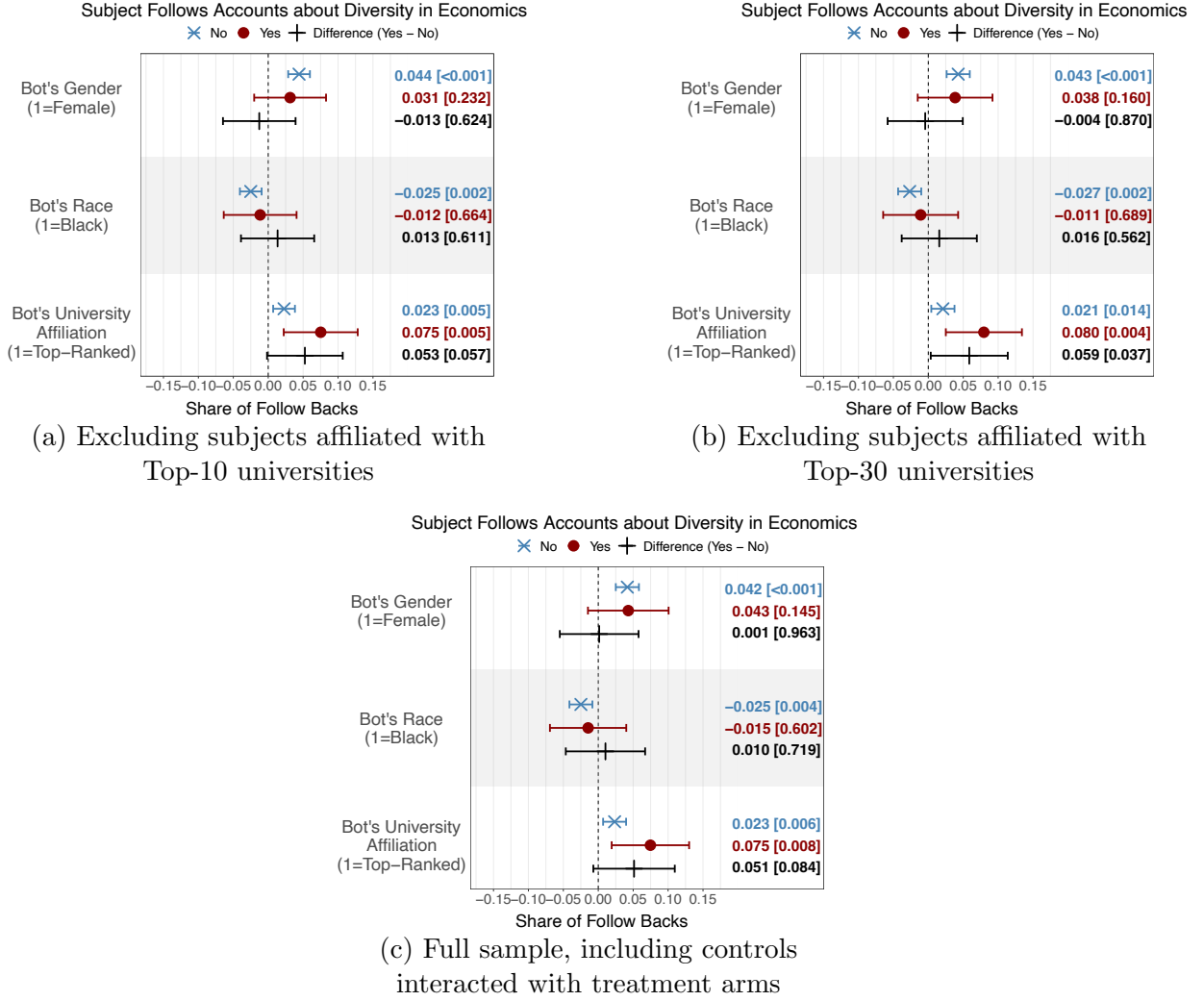
Notes: The figure shows the main results of the experiment for three sub samples of accounts with more intense activity levels on Twitter. We use three definitions for activity: above-median number of statuses, above-median number of friends, and above-median ratio of friends-to-followers. The first three figures show marginal follow-back rates for bot accounts from each group. Data comes from the ten experimental waves, excluding shadow-banned accounts as discussed in the text. The error bars show 95% confidence intervals for the mean follow-back. The last plot shows point estimates and 95% confidence intervals for the coefficients related to the bot's gender, race, and university affiliation, obtained by estimating the pre-registered equation (1) with and without additional subject-level controls (in both cases, including strata and wave fixed-effects). Confidence intervals are computed using standard errors clustered at the bot account level. P-values are in brackets.

Figure B.3: Main Results for the Subsample of Accounts Identified as Scholars



Notes: The figure shows the main results of the experiment for the subsample of accounts classified as scholars (professors, PhD students, post-docs, and other researchers). We classified the profession of 57% of our sample using information available in users' bios; of these, 74% are scholars. The first three figures show marginal follow-back rates for bot accounts from each group. Data comes from the ten experimental waves, excluding shadow-banned accounts as discussed in the text. The error bars show 95% confidence intervals for the mean follow-back. The last plot shows point estimates and 95% confidence intervals for the coefficients related to the bot's gender, race, and university affiliation, obtained by estimating the pre-registered equation (1) with and without additional subject-level controls (in both cases, including strata and wave fixed-effects). Confidence intervals are computed using standard errors clustered at the bot account level. P-values are in brackets.

Figure B.4: Heterogeneity in Follow-Back Behavior by Subject's Concern about the Lack of Diversity in Economics, Subsamples and Adding Controls



Notes: The figures show estimates of the heterogeneity in the follow-back behavior of subjects who publicly display or not concern about the lack of diversity in economics. Estimates are analogous to Figure 4c, but excluding subjects affiliated with top-10 (top-left panel) and top-30 (top-right panel) US universities. In the bottom panel, we show results using the full sample, but adding demeaned controls (including an indicator for affiliation with a top-10 university) interacted with indicators for bots' characteristics. The controls included are: continent (dummies for Europe, US/Canada, and other); profession (dummies for professor, grad student, industry/tech, and other); gender; year of account creation; indicators for affiliation with a top-10 university, having a background picture, and having a verified account; and the number of Twitter followers and friends. The specification on the bottom panel was pre-registered, while the two others, no. Details on the construction of variables are available in Appendix Table B.2.

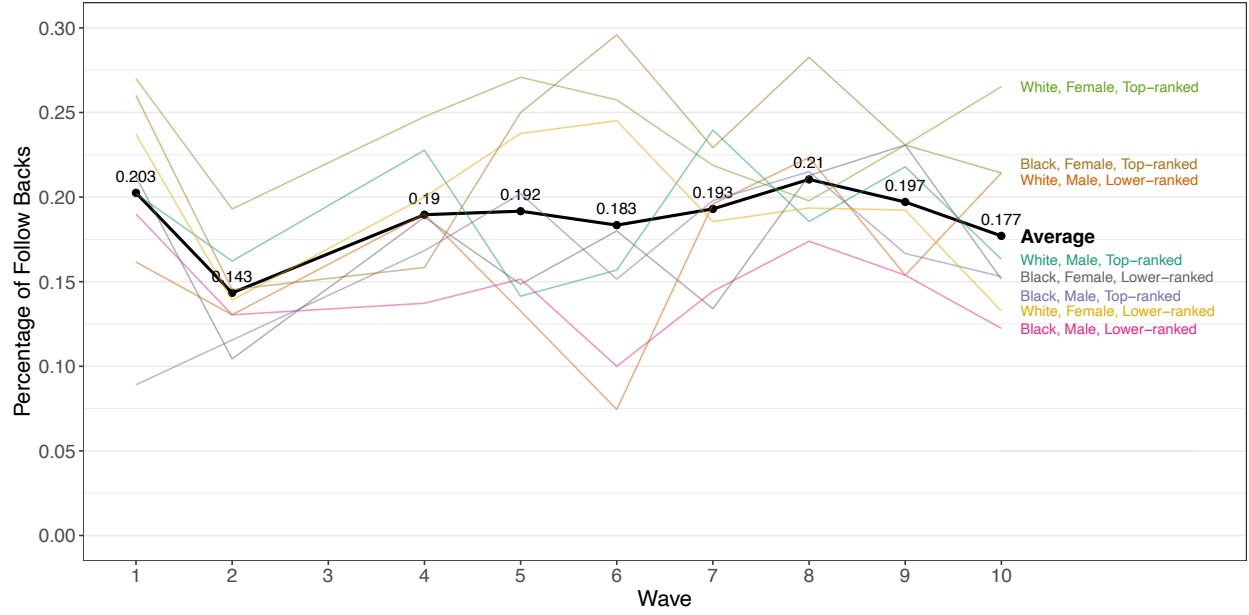
Table B.7: Heterogeneity in Follow-Back Behavior

<i>Dependent Variable:</i>	<i>Follow Backs (1=Yes)</i>			<i>Dependent Variable:</i>	<i>Follow Backs (1=Yes)</i>		
Panel A: Subject's gender				Panel C: Subject is concerned about the lack of diversity in economics			
Bot's Gender (1=Female)	0.05 (0.015)	0.049 (0.015)	0.044 (0.015)	Bot's Gender (1=Female)	0.043 (0.008)	0.042 (0.008)	0.042 (0.008)
Bot's Race (1=Black)	-0.027 (0.016)	-0.027 (0.015)	-0.032 (0.015)	Bot's Race (1=Black)	-0.023 (0.008)	-0.024 (0.008)	-0.024 (0.008)
Bot's University Affiliation (1=Top-Ranked)	0.041 (0.016)	0.039 (0.015)	0.035 (0.015)	Bot's University Affiliation (1=Top-Ranked)	0.025 (0.008)	0.025 (0.008)	0.026 (0.008)
Bot's Gender × Female Subject	-0.025 (0.023)	-0.024 (0.023)	-0.025 (0.023)	Bot's Gender × Concerned	0.005 (0.027)	0.002 (0.026)	0.001 (0.028)
Bot's Race × Female Subject	-0.008 (0.023)	-0.008 (0.023)	-0.003 (0.023)	Bot's Race × Concerned	0.013 (0.027)	0.015 (0.026)	0.011 (0.028)
Bot's University × Female Subject	0.011 (0.023)	0.011 (0.023)	0.01 (0.023)	Bot's University × Concerned	0.066 (0.027)	0.062 (0.027)	0.052 (0.029)
Observations	4399	4399	4399	Observations	6920	6920	6920
Dep. Variable Mean	0.189	0.189	0.189	Dep. Variable Mean	0.187	0.187	0.187
Panel B: Subject's account quality				Panel D: Subject's race/ethnicity			
Bot's Gender (1=Female)	0.037 (0.011)	0.037 (0.011)	0.03 (0.012)	Bot's Gender (1=Female)	0.096 (0.042)	0.096 (0.042)	0.098 (0.042)
Bot's Race (1=Black)	-0.024 (0.011)	-0.024 (0.011)	-0.019 (0.011)	Bot's Race (1=Black)	-0.002 (0.042)	-0.002 (0.042)	-0.002 (0.042)
Bot's University Affiliation (1=Top-Ranked)	0.04 (0.011)	0.04 (0.011)	0.044 (0.012)	Bot's University Affiliation (1=Top-Ranked)	-0.023 (0.043)	-0.026 (0.042)	-0.021 (0.042)
Bot's Gender × Strong profile	0.013 (0.018)	0.012 (0.018)	0.025 (0.019)	Bot's Gender × White Subject	-0.061 (0.046)	-0.061 (0.045)	-0.066 (0.046)
Bot's Race × Strong profile	0.006 (0.018)	0.006 (0.019)	-0.005 (0.02)	Bot's Race × White Subject	-0.019 (0.045)	-0.019 (0.044)	-0.022 (0.045)
Bot's University × Strong profile	-0.007 (0.018)	-0.009 (0.018)	-0.017 (0.02)	Bot's University × White Subject	0.067 (0.045)	0.069 (0.045)	0.062 (0.045)
Observations	6920	6920	6920	Observations	4234	4234	4234
Dep. Variable Mean	0.187	0.187	0.187	Dep. Variable Mean	0.187	0.187	0.187
Controls	No	Yes	Yes	Controls	No	Yes	Yes
Interacted Controls	No	No	Yes	Interacted Controls	No	No	Yes
Wave, Strata Fixed Effects	Yes	Yes	Yes	Wave, Strata Fixed Effects	Yes	Yes	Yes

Notes: The table displays estimated coefficients for interaction terms between bot and subject characteristics, using data from the ten experimental waves, excluding shadow-banned accounts. The specification used is the one in Equation (2), both with and without additional controls. The second column in each panel presents regressions with controls, while the third column includes both controls and controls interacted with bots' characteristics. The controls included in the regressions are: continent (dummies for Europe, US/Canada, and other); profession (dummies for professor, grad student, industry/tech, and other); gender; year of account creation; indicators for affiliation with a top-10 university, having a background picture, following accounts addressing the lack of diversity in academia, and having a verified account; and the number of Twitter followers and friends. In all cases, controls are demeaned. The dependent variable is an indicator equal to one if a Twitter user (subject) that was followed by a bot subsequently followed the bot back. Clustered standard errors at the bot-account level are in parentheses.

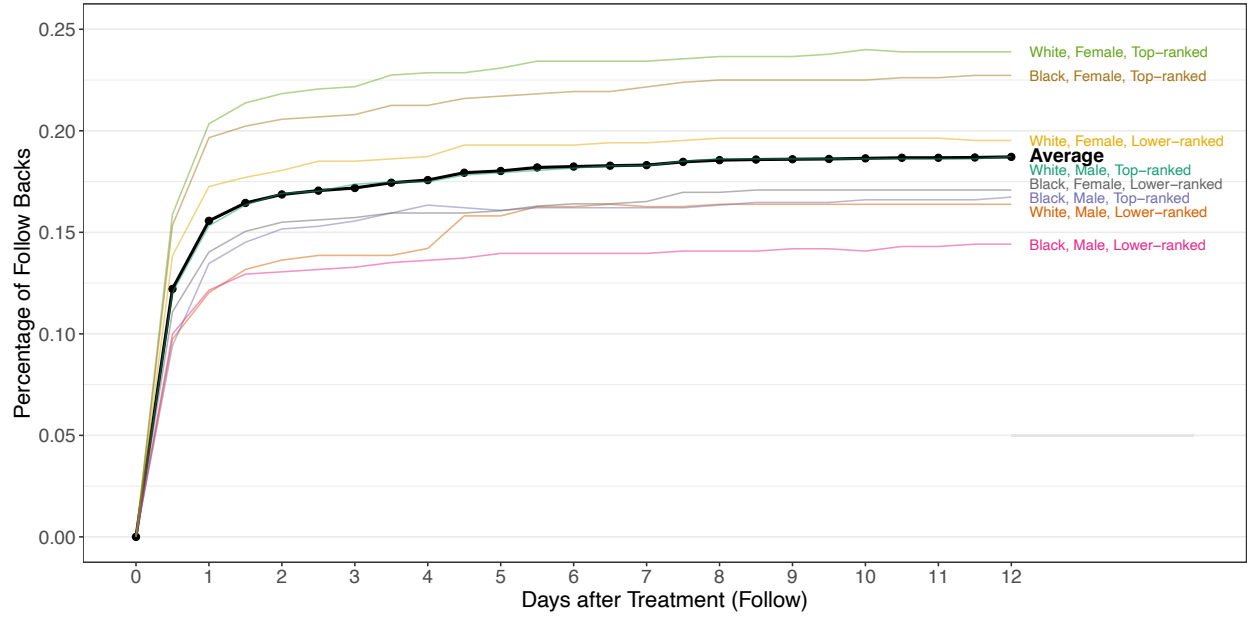
B.5 Evolution of Follow-Back Rates

Figure B.5: Evolution of the Follow-Back Rate across Experimental Waves



Notes: The figure plots the evolution of the follow-back rate in the experiment across experimental waves. The thick black line is the average follow-back rate for all accounts in the wave, while the colored lines represent the follow-back rate in each wave for each bot type. The first wave started on May 26th, 2022, and the last wave started on July 28th, 2022.

Figure B.6: Evolution of Follow-Backs within an Experimental Wave



Notes: The figure plots the evolution on the follow-back rate of the experimental accounts within each wave. The x axis represents the number of days after the bot follows the subjects (which happens on day 0). The thick black line represents the average follow-back rate across all types of bots, while the colored lines are the follow-back rate for each bot type. Data is pooled across the ten experimental waves, excluding shadow-banned accounts as discussed in the text. In all waves, we collected follow-back data up to twelve days after the treatment.

B.6 Organic Follows by Bot Group

Table B.8: Average Number of Organic Follows by Bot Group

Group	Average Number of Organic Follows	P-value
Bot's Gender		
Female	1.472	0.959
Male	1.457	
Bot's Race		
Black	1.343	0.413
White	1.583	
Bot's Affiliation		
Lower-ranked	1.333	0.364
Top-ranked	1.600	

Notes: The table displays the average number of organic follows (i.e., follows by users who were not followed by the bots) for each group of experimental accounts, across all ten waves (excluding shadow-banned accounts). The p-value in the last column refers to the p-value of a simple test of difference in means between the two types of each group.

C Implementation Details

This Appendix describes how we implemented the experiment in detail, focusing on the activation of accounts in each round and how we followed subjects assigned to each treatment. Most procedures related to account activation and following subjects were done manually to avoid detection by Twitter.

In each wave, we activated eight accounts (one of each group). First (on day “zero”), each account followed a set of “elite” accounts from *#EconTwitter* (accounts from journals and associations) and the accounts of the approximately 30 colleagues who knew about the experiment and who we asked to follow the accounts back. This was done so that the experimental accounts looked more realistic (since they would already have a set of followers and friends). Note that this set of followers and friends was the same for all accounts in a given wave. At this initial moment, all accounts also retweeted two statuses from accounts of economic journals.⁴ The statuses were chosen randomly from all tweets published by these accounts on the week the bot accounts were activated that had garnered more than three retweets. We also randomized the order of account activation in order to avoid introducing any bias related to timing.

After these procedures, the accounts were ready to follow the subjects assigned to them. This happened on the next day (Day “one”). At this time, each account followed the subjects randomly assigned to it. Treatment assignment was performed for each wave and block-randomized using gender, profession, and number of followers. We sampled randomly from within each stratum, assigning the same proportion of users in each stratum to each bot account. In the cases where the number of observations was not divisible by the number of treatments, we reassigned the remaining observations (called misfits) globally by creating a “misfits stratum” and sampling from there (Carril, 2017). The bot accounts always followed subjects on a Thursday (see the timeline in Appendix Table A.1).

We randomized the order of follows both within and across accounts (i.e., the order of subjects to be followed by each account was random, as was the order of accounts following the subjects). This procedure always happened on Thursdays, and the distance between the first and last subject followed in each wave was at most 1h30 (this ensured that the timing at which the treatment notification was received was comparable across waves and subjects within waves). Apart from following all of its subjects, every bot account followed an account of someone who was aware of the experiment. This person would then inform us if he or she received a notification of this follow. If not, we considered the respective bot to have been “shadow-banned” and excluded it from the analysis (as pre-registered). This happened with a single account of the 80 we created during the experiment. We also

⁴The accounts we chose to retweet from are: [@AEA_Journals](#) (journals from the American Economic Association), [@ecmaEditors](#) (Econometrica), [@JPolEcon](#) (Journal of Political Economy), [@QJEHarvard](#) (Quarterly Journal of Economics), [@RevEconStudies](#) (Review of Economic Studies), [@restatjournal](#) (Review of Economics and Statistics), [@JEEA_News](#) (Journal of the European Economic Association), [@EJ_RES](#) (Economic Journal), [@JPubEcon](#) (Journal of Public Economics), [@nberpubs](#) (National Bureau of Economics Working Papers), [@qe_editors](#) (Quantitative Economics), [@EconTheory](#) (Theoretical Economics), [@J_HumanResource](#) (Journal of Human Resources), [@RevOffFinStudies](#) (Review of Financial Studies), [@JofFinance](#) (Journal of Finance), and [@J_Fin_Economics](#) (Journal of Financial Economics).

had eight accounts (one from each group) suspended by Twitter. These accounts were also excluded from the experiment because the suspension happened before we could complete the treatment (this hypothesis was also pre-registered). Thus, the final experimental data includes 71 (non-shadow banned and non-suspended) accounts across ten waves.

During the first five experimental waves, we created new Twitter accounts for the bots at the beginning of every single wave and deleted the accounts from the previous round. After these first five waves, we realized this procedure was unnecessary since we could simply delete all account information and start over with a new identity. Specifically, at the end of the activation period, we removed all followers and friends from the account (except those who knew about the experiment and whom we asked to follow all accounts) and deleted all of its posts. We then changed the account’s name, description, Twitter handle, profile picture, and background picture so that the account started over with a new identity. This new procedure saved time and reduced the likelihood of suspensions by Twitter, so we adopted it starting at wave 6. After the change in procedure, no account was suspended or shadow-banned. To guarantee that no spillovers from previous identities were correlated with the bot type, at each new wave, we randomized which type of bot account would be created using each account (thus, if a specific account represented a White male from a top-ranked university in wave 6, the same account could represent a Black female from a lower-ranked university in wave 8). Every account had an email and a phone number associated with it (for the purpose of Twitter verification).

During each wave, we collected information on follows using Twitter’s API. We collected this information twice a day for twelve days, counting from the beginning of the treatment.

D Total Follows and Concerns with Contamination by Twitter’s Algorithm

An important concern with the experimental design has to do with Twitter’s algorithm, which can suggest accounts for users to follow. If the algorithm’s suggestions were correlated with some of the bots’ characteristics—for instance, if the algorithm was more likely to suggest accounts of top-ranked students—, analyzing total follows would lead to biased estimates of the effects we study. Therefore, following our pre-analysis plan, we restricted our analysis to follow-backs, enabling us to isolate our outcome of interest from the effects of the algorithm.

Still, we have considerable evidence that allows us to rule out any contamination by organic follows motivated by algorithm suggestions. First, the number of follows from accounts our bots had not followed was extremely low (1.46 organic follows per account). In addition, the distribution of organic follows across groups of subjects mirrors the distribution of follow-backs we discussed previously, as seen in Appendix Table B.8.

A related concern is that when people receive a notification from one of the experimental accounts, some of the other accounts could be suggested to them by the algorithm, making the subject suspicious about the veracity of the experimental account. We do not have evidence that such suggestions happened frequently enough to affect our results.⁵ First, if such suggestions were happening, they would have become more frequent over time, as Twitter’s algorithm learned about the similarity of the experimental accounts. In this case, we should expect follow-back rates to decrease over time. More generally, we should also expect declining follow-back rates if the *#EconTwitter* community was becoming more suspicious about the bots for any other reason. In contrast, Appendix Figure B.5 shows that the follow-back rates of the experiment were fairly constant across waves, indicating that there were likely no actions taken by the platform against the accounts, and, more generally, that the *#EconTwitter* community did not learn to recognize the accounts’ lack of authenticity in any other way. We see a similar consistency for each type of bot. Second, no subject from any wave followed an experimental account other than the one that followed him or her. We would expect this to happen if the algorithm was indeed suggesting other bot accounts to treated users. If this occurred, subjects might add other bot accounts, either without noticing the similarity to the account that had followed them or in an effort to actively spoil our results. Moreover, algorithmic suggestions are only highlighted in the Twitter Desktop App, and our results show that most of the Twitter users in our subject pool use the mobile or iPad app more frequently (the median user of our sample tweets from the mobile or iPad app 81% of

⁵Though, this did happen at least once, leading one user to tweet about our experiment, thereby forcing us to finish the experiment earlier than planned. He or she had not been “treated” in the wave in which the post was made. The most likely reason for his or her suspicion is that the algorithm suggested one of the experimental accounts as a potentially interesting user to follow, along with other similar accounts. This further reduces any concern about contamination, since the only suspicion about the experiment came from someone who was not treated. Moreover, the fact that this only happened after we had run ten complete experimental waves (plus three pilot waves) is reassuring: if more people had been suspicious about the experimental accounts, it is likely that we would have seen more tweets on the topic or would perhaps have received direct messages through the bots’ accounts. This did not happen even once.

the time, with 38% of users exclusively using the mobile app). Overall, all of these pieces of evidence suggest that contamination due to Twitter’s algorithm (or other channels) is not a major concern in our setting.