

Supplementary - “I Have No Idea What a Social Bot Is”: On Users’ Perceptions of Social Bots and Ability to Detect Them

Daniel Kats
daniel.kats@nortonlifelock.com
NortonLifeLock Research Group
Culver City, CA, USA

Mahmood Sharif
mahmoods@cs.tau.ac.il
Tel Aviv University
Tel Aviv, Israel

1 STUDY PROTOCOL

1.1 Introduction

[Consent form, study procedure, etc.]

Q1. On which of the following online social networks do you currently have an account? [Options: Facebook; Instagram; LinkedIn; Reddit; TikTok; Twitter; YouTube.]

Study description. The goal of this study is to explore users’ usage habits of different online social networks, assess their familiarity of different aspects of online social networks, and learn about their experiences.

1.2 OSN Usage Habits

Please answer the following questions about your usage patterns of various online social networks.

Q2. Roughly speaking, how often do you **visit** each of the following online social networks? [For each OSN the participant indicated having an account on, provide the following options: Less than once a month; Less than once a week but at least once a month; Less than once a day but at least once a week; About once a day; Several times a day.]

Q3. Roughly speaking, how much time do you **spend** on each of the following online social networks? [For each OSN the participant indicated having an account on, provide the following options: Less than 1 hour a week; Less than an hour most days but at least 1 hour a week; At least 1 hour most days.]

Q4. Roughly speaking, how often do you **interact** with content (e.g., like, comment, or share) on each of the following online social networks? [Same options as Q2.]

Q5. Roughly speaking, how often do you **create** content (e.g., Tweet, post videos or images, post) on each of the following online social networks? [Same options as Q2.]

Q6. What **content categories** do you usually create or consume on the following online social networks? [For each OSN the participant indicated having an account on, provide the following options (mostly taken from Twitter’s list of categories, and extended to accommodate other OSNs): Arts & Culture; Business & Finance; Careers (e.g. Education, College Life, Professional Topics); Entertainment; Fashion & Beauty; Food; Friends & Family; Gaming; Hobbies & Interests; Humor & Memes; Movies & TV; Music; News; Outdoors; Science; Sports; Technology; Travel.]

1.3 Familiarity With Social Bots

Please answer the following questions as honestly as you can. There are no correct or wrong answers, and your responses will not impact your compensation in any way.

Q7. Are you familiar with the notion of social bots on online social networks? [Options: Yes; No.]

Q8. [If familiar] In two or three sentences, please describe what you believe a social bot is on an online social network: [Free text response.]

Q9. [If unfamiliar] In two or three sentences, please describe what you believe a social bot is likely to be on an online social network: [Free text response.]

Q10. What do you believe are the goals of social bots on online social networks? Please respond in one or two sentences. [Free text response.]

Q11. Do you believe that social bots influence **your** behavior on online social networks? [Options: Definitely not; Probably not; Maybe; Probably yes; Definitely yes.]

Q12. Why do you think that social bots do (not) influence **your** behavior? [Free text response.]

Q13. Do you believe that social bots influence **other users’** behavior on online social networks? [Options: Definitely not; Probably not; Maybe; Probably yes; Definitely yes.]

Q14. Why do you think that social bots do (not) influence **other users’** behavior? [Free text response.]

1.4 Social Bot-Prevalence Estimate

Q15. Based on your experience, what percentage of posts on each of the following social networks is made by social bots? [For each OSN the participant indicated having an account on, provide the following options: <10% of posts; 10%–25% of posts; 25%–50% of posts; >50% of posts.]

1.5 Account-Classification - Stage #1

For each of the following ten Twitter accounts, please indicate whether you believe they are social bots or not.

[For participants who indicated not being familiar with bots, show:] While you have indicated not owning a Twitter account, please do not be discouraged. Try to categorize the accounts based on your best guess, according to the accounts’ characteristics. Again, we remind you that your responses will not impact your compensation at the end of the study.

Q16–Q25. Do you believe that the following account is a social bot? [Options: Yes; No. Randomized between the first and second set of ten accounts. The ten accounts within the group are also presented in a randomized order.]

Q26. What are the characteristics that led you to believe that certain accounts are social bots? [Free text response.]

1.6 Social Bot Definition

In what follows we define what a social bot is on online social networks. **Please read the definition carefully.**

A social bot on social media is a computer algorithm that automatically or semi-automatically produces content and interacts with humans as well as other bots on social media, possibly trying to emulate or alter human's behavior. Bots on social media can be roughly split into two categories:

1- **Benign social bots:** These are typically overt bots (i.e., they identify as bots) that provide useful functionality or used for entertainment.

2- **Malicious social bots:** These are typically covert bots (i.e., they attempt to masquerade as humans) that engage in harmful tasks, such as spreading disinformation, scams, or spam. In certain cases, malicious bots may also attempt to impersonate other individuals (e.g., public figures).

1.7 Account-Classification - Stage #2

Based on the previously provided definition, please indicate whether you believe the following ten accounts are social bots or not.

Q27–Q36. Do you believe that the following account is a social bot? [Options: Yes; No. Pick the group of ten accounts not presented in the first stage. The ten accounts within the group are also presented in a randomized order.]

Q37. What are the characteristics that led you to believe that certain accounts are social bots? [Free text response.]

1.8 OSNs and Social Bots

As a reminder, we provide the previously provided definition for a social bot on online social networks in what follows. If you remember the definition, please feel free to proceed to the next question. [Show the definition from above.]

Q38. Are you satisfied with how various online social networks are dealing with social bots on their platforms? [For each OSN the participant indicated having an account on, provide the following options: Very unsatisfied; Somewhat unsatisfied; Neutral; Somewhat satisfied; Very satisfied.]

Q39. Please explain why you are satisfied/unsatisfied with how certain online social networks deal with social bots: [Free form text.]

Q40. Assume that online social networks had a tool to help them accurately identify social bots. How do you believe that they should use the tool? [Options: Ban all bot accounts; Allow benign bot accounts only, and show bot indicators on posts; Allow benign bot accounts only, but without showing bot indicators on posts; Allow all bot accounts, and show bot indicators on posts; Allow all bot accounts, but without showing bot indicators on posts.]

1.9 Demographics

[Ask about age, gender, education level, occupation, tech-savviness self-reporting; country of residence; type of area that the participant resides in (college town; rural, suburban; urban; prefer not to answer).]

[Get the participant's Prolific ID, thank participant, and end study.]

2 REGRESSION VARIABLES AND INITIAL MODEL

We initially started with the following model that contains all explanatory variables and several interactions that we believed impact the correctness of participants' classification (descriptions are in-lined before the explanatory variables and marked with a # sign).

```
CorrectClass ~ (1|ID)
#Demographics
+ Male + AgeScaled + NotTechSavvy + IsRural + HighDegree
#OSN usage habits
+ NotHasTwitter + HeavyFacebook + HeavyInstagram
+ HeavyReddit + HeavyLinkedIn + HeavyTikTok
+ HeavyTwitter + HeavyYouTube
#Reported familiarity with bots
+ NotFamiliar
#Time taken to finish study(∈ [0,1])
+ TimeTakenScaled
#Perceptions
+ DefAutomated + DefFake + DefInteractive + DefAbstract
+ DefImpersonate + DefCyclic + DefConfuse + DefCreate
+ DefDontKnow + DefSockpuppet + DefUnintelligible
+ DefAnonymous + GoalBoost + GoalInfluence + GoalBenign
+ GoalConfuse + GoalScam + GoalAdvertise + GoalEarn
+ GoalSupport + GoalDontKnow + GoalUnintelligible
#Account type
+ BotBenign + BotDateScam + BotFake + BotPolitical
+ BotTechScam + NonBotParody + NonBotPopular
+ NonBotUnpopular + NonBotVerified
#Participant saw definition?
+ SawDef
#Interactions
+ SawDef * NotHasTwitter + SawDef * NotFamiliar
+ SawDef * TimeTakenScaled + SawDef * AgeScaled
+ SawDef * IsRural + SawDef * BotBenign
+ SawDef * BotDateScamSawDef * BotFake
+ SawDef * BotPolitical + SawDef * BotTechScam
+ SawDef * NonBotParody + SawDef * NonBotPopular
+ SawDef * NonBotUnpopular + SawDef * NonBotVerified
```