scientific reports

Check for updates

OPEN

Exposure to social bots amplifies perceptual biases and regulation propensity

Harry Yaojun Yan[1,2,3]✉, Kai-Cheng Yang[2,3], James Shanahan[1,3] & Filippo Menczer[2,3]

Automated accounts on social media that impersonate real users, often called "social bots," have received a great deal of attention from academia and the public. Here we present experiments designed to investigate public perceptions and policy preferences about social bots, in particular how they are affected by exposure to bots. We find that before exposure, participants have some biases: they tend to overestimate the prevalence of bots and see others as more vulnerable to bot influence than themselves. These biases are amplified after bot exposure. Furthermore, exposure tends to impair judgment of bot-recognition self-efficacy and increase propensity toward stricter bot-regulation policies among participants. Decreased self-efficacy and increased perceptions of bot influence on others are significantly associated with these policy preference changes. We discuss the relationship between perceptions about social bots and growing dissatisfaction with the polluted social media environment.

Social bots are social media accounts that are controlled at least in part by software on social media. They can be purchased at low cost[1] and operated with various degrees of automation. By posting content and interacting with people, some bots can emulate and deceive real social media users, posing a threat to our social and political life[2]. Social bots have been used to disseminate fake news[3] and inflammatory information[4,5], exploit the private information of users[6], sway public attention on controversial topics[7–9], and create false public support for political and commercial gain[1,10,11], especially during major political events such as elections[12–14]. While the severity of the threats posed by bots is still debated[15], research has demonstrated critical consequences of bot exposure[16,17].

Public awareness of social bots has been on the rise and a majority of Americans believe that bots have negative effects on the public[18,19]. Yet, public perceptions of bots are under-studied[20]. A consensus on the definition of social bots is lacking; people appear to use the term to refer to a variety of entities, from fake profiles and spammers to fully-automated accounts[21]. This ambiguity provides social media users with a scapegoat for their unpleasant online experiences[22]. For example, one may reject accounts with opposing political views by labeling them as bots[17,23]. Such a confirmation bias is only one of many perceptual biases on which users rely when making judgments about online interactions[24]. These biases have a strong evolutionary foundation[25] but make us vulnerable to manipulation[26].

Here we report on two experiments designed to investigate perceptual biases regarding social bots and how they are exacerbated by exposure to bots. In the experiments, participants are instructed to distinguish bot-like social media profiles from authentic users; no feedback is provided. Participants answer questions regarding their perceptions about bots before and after the bot exposure. While these two experiments conceptually replicate each other, they differ slightly in their design. In *Experiment I*, participants were shown profiles containing a mix of non-political and political profiles. Half of the participants were assigned to a condition where the ambiguity between humans and bots was low, while the other half were placed in a condition with high ambiguity. In *Experiment II*, all participants viewed the same set of political profiles, and the overall level of ambiguity between humans and bots was comparable to the high-ambiguity condition in Experiment I. Further methodological details can be found in the "Materials and Methods" section.

Consequently, combining the results of these two experiments allowed for the creation of three analytical conditions: (1) profiles with low human-bot ambiguity and mixed content, (2) profiles with high human-bot ambiguity and mixed content, and (3) profiles with high human-bot ambiguity and political content. By comparing the results from these three conditions, we can also shed light on the impact of varying levels of human-bot ambiguity and explore any differential effects resulting from the presence of political bots as opposed to non-political bots.

[1]The Media School, Indiana University, Bloomington, IN 47405, USA. [2]Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, USA. [3]Observatory on Social Media, Indiana University, Bloomington, IN 47408, USA. ✉email: harryan@iu.edu

We focus on the effects of bot exposure on three perceptions by social media users: estimated prevalence of bots, perceived influence of bots on themselves and others, and assessment of one's own ability to recognize bots. Theories of cognitive biases predict that these perceptions are intertwined and often inaccurate[25,27–30]. More importantly, manipulations of these perceptions drive behavioral changes such as the adoption of different policies[31,32]. Therefore, we also survey preferences for more or less strict countermeasures. To examine the effects of bot exposure, we compare the same set of measures before and immediately after exposure. We find that perceptual biases regarding bots exist and can be amplified by simply raising awareness of the potential threat they pose.

Quantifying perceptual biases about social bots may also help improve the design of machine learning algorithms to detect them[33–35]. These algorithms depend on labeled examples of bot accounts, which are often identified by human annotators[36]. The biases of annotators can therefore propagate through the pipeline and affect downstream tasks. As a few studies have already revealed perceptual biases in human-bot interactions[17,23], more research is needed.

While policy efforts about social bots are still in the nascent stage[37], our findings about bot regulation preferences suggest that policymakers should be cautious when interpreting public opinion data. As demonstrated in this study, public sentiment toward the perceived threats posed by bots can exhibit reactionary and irrational patterns. Recognizing the dynamic and evolving nature of public sentiment can assist policymakers in crafting more effective regulations and strategies concerning social bots.

Results

Overestimation of bot prevalence

The prevalence of bots has been at the center of public discussion since it is closely related to user experience and the financial value of social media platforms. Consensus on a number is unlikely because scholars and researchers cannot agree on the definition of bots[37] and detection is technically challenging[21]. Different stakeholders have provided rather disparate estimates: Twitter's Annual Report in 2021 recognized that 5% of accounts might be spam or false[38]; Elon Musk claimed the number to be as high as 20% (youtube.com/watch?v=CnxzrX9tNoc, 3′ 08″–3′ 30″); these estimates are tainted by potential conflicts of interest[39]. A scholarly estimate of automated accounts on Twitter ranged between 9% and 15% in 2017[40].

Rather than focusing on the *actual* number of bots, we are interested in investigating prevalence as *perceived* by the general public. People commonly have perceptual biases about the prevalence of social phenomena. For example, they tend to believe that a small sample is representative of a larger population[27]. Such a perceptual bias becomes more prominent and is easily manipulated by media messages when the judged matter is deemed undesirable[41–43]. Therefore, we expect participants to overestimate the prevalence of bots and that such bias would be further magnified after experimental exposure to bots (see "Materials and Methods" Section).

Before exposure to bots in the recognition task, participants report that on average 31.9% (SD 18.9%, Median 30.0%) of social media accounts are bots, which is substantially larger than the estimates provided by Twitter, Varol et al.[40], and even Musk. This result suggests that participants tend to overestimate the prevalence of bot accounts. After exposure to bots, the average estimated prevalence goes up to 37.8% (SD 19.2%, Median 35.0%). Figure 1 shows the distributions of the estimates before and after the recognition tasks and the significant increase after exposure (Wilcoxon signed rank test $V = 99{,}448$, $p < 0.001$).

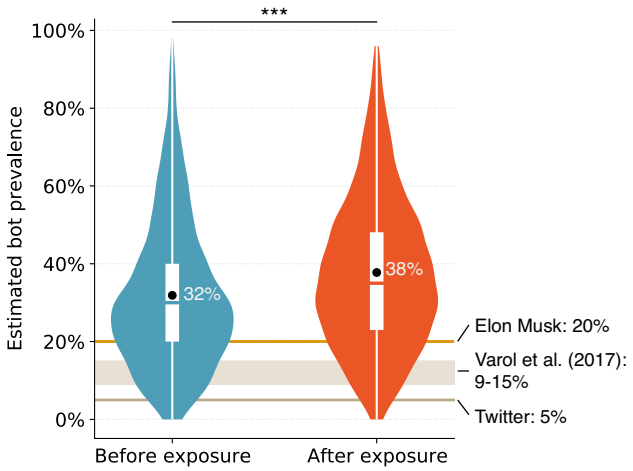

Figure 1. Overestimation of bot prevalence. These violin plots show the distributions of bot prevalence estimates by participants before and after exposure to bots impersonating humans; the difference is significant ($p < 0.001$). The black dots indicate the mean values; the box plots show the 25th, 50th, and 75th percentiles. We also show the estimates of inauthentic and spam accounts by Twitter and Elon Musk together with the estimated prevalence of automated accounts by Varol et al.[40]. See the main text for details.

Misjudged self-efficacy in bot recognition

While people tend to overestimate the prevalence of threats, they may or may not believe they are capable of mitigating such threats. On the one hand, it is beneficial for people to believe in their ability, since such self-efficacy can promote task performance[44–47]. On the other hand, self-assessments of expertise, such as information literacy[48], can be inaccurate and inflated[28,49–51]. To investigate whether self-efficacy in the bot-recognition task is reliable, we asked participants to assess their ability to identify bots before and after exposure and tested the association between their self-assessments and their actual accuracy (see "Materials and Methods" Section).

We find that the majority of participants possess a somewhat high bot-recognition self-efficacy before exposure to bots (Mean 4.8, SD 1.2, Median 5.0 on a 7-point Likert scale). After exposure to bots, participants become less confident (Mean 4.2, SD 1.2, Median 4.3). Bot exposure causes on average a significant decrease in self-efficacy among participants (paired t-test $t = -10.3, p < 0.001$), as shown in Fig. 2a.

We used individual-level regression analysis to assess the relationship between the self-efficacy of participants and their actual accuracy in recognizing bots. The results are reported in Fig. 2b and Table 1. We find that pre-exposure self-efficacy is positively associated with performance ($b = 0.011, S.E. = 0.003, p = 0.002$), although the effect size is rather minimal (adjusted $R^2 = 0.009$). The post-exposure self-efficacy, on the other hand, is not significantly correlated with performance ($b = -0.004, S.E. = 0.003, p = 0.129$), showing that exposure to bots further distorts the self-assessment of the participants. An additional regression shows that participants who report a larger *improvement* in their self-efficacy perform significantly worse in the bot recognition task ($b = -0.009, S.E. = 0.002, p < 0.001, R^2 = 0.011$; see full model results with control variables in Table 1). Overall, these results suggest that self-efficacy is not a reliable predictor for the actual ability to recognize bots, and exposure to bots further exacerbates this unreliability.

Gap in perceived bot influence on others vs. self

People commonly assume stronger media effects on others than on themselves when facing adversarial media messages. This so-called "third-person perception" (TPP)[29] is moderate but robust in the context of mass media messaging[52–54]. The perceptual gap associated with TPP can be further magnified when the judged matter is threatening to existing social norms[55]. Surveys have also reported on TPP in new media environments, such as internet pornography[56], Facebook trolls[57], and online fake news[58]. However, the change in an individual user's TPP caused by direct interactions with media content is less studied[59]. Following these previous studies, we hypothesize that TPP could also be observed in the context of interactions with social bots and amplified after exposure.

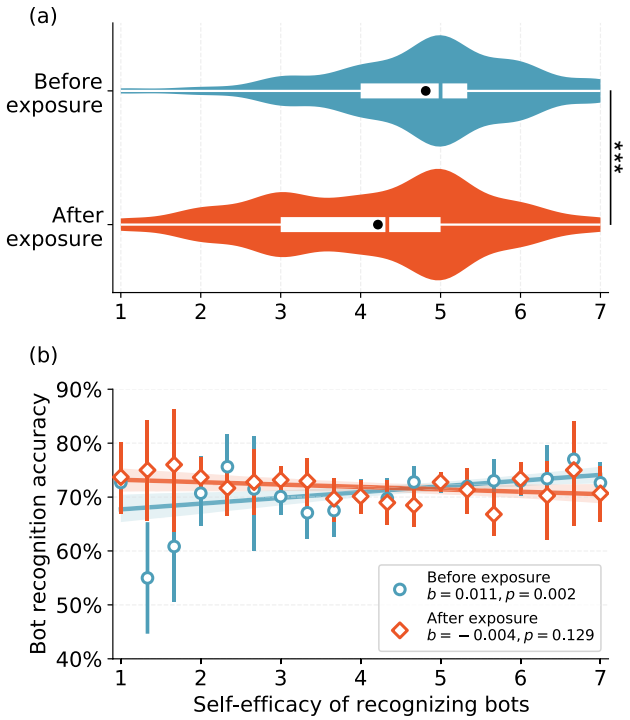

Figure 2. Self-efficacy and bot recognition accuracy. (**a**) The violin plots show the distributions of self-reported bot-recognition efficacy by participants before and after exposure to bots impersonating humans; the difference is significant ($p < 0.001$). Self-efficacy is in the range 1–7 where a larger value indicates higher confidence. The black dots indicate the mean values; the box plots show the 25th, 50th, and 75th percentiles. (**b**) Relation between participant self-efficacy and actual bot-recognition accuracy before and after exposure. The coefficients estimated from the linear regressions and the p-values are annotated.

	Before exposure	After exposure	Changes
(Intercept)	1.29***	1.28***	1.26***
High-ambiguity mixed bots[a, c]	− 0.47***	− 0.47***	− 0.47***
High-ambiguity political bots[b, c]	− 0.36***	− 0.35***	− 0.32***
Age	− 0.08***	− 0.09***	− 0.09***
Education	− 0.01	− 0.01	− 0.01
Independent[d]	− 0.02	− 0.02	− 0.02
Republican[d]	− 0.12*	− 0.12*	− 0.12*
Self-efficacy			
Before exposure	0.05*		
After exposure		− 0.01	
Change scores			− 0.04*
R^2	0.094	0.089	0.093

Table 1. Generalized linear models of self-efficacy predicting performance in the bot-recognition task. The dependent variable is the proportion of accurate answers out of twenty trials. We also carried out the regression at the trial level, and the results are consistent except for the self-efficacy changes (the last column), where the association with the actual accuracy yields $p = .051$ $^{*}p < .05$; $^{***}p < .001$. We use beta distribution as the link function [a]The second condition of *Experiment I* [b]*Experiment II* [c]Low-ambiguity mixed bots (i.e., the first condition of *Experiment I*) as the reference group [d]Democrat as the reference group.

We measure perceived influence on a five-point scale, where a larger value indicates a stronger effect (see "Materials and Methods" Section). As shown in Fig. 3, participants perceive stronger bot influence on others (Mean 3.1, SD 0.9) than on themselves (Mean 2.0, SD 0.9) before exposure to bots. The difference is significant (paired t-test $t = 31.6, p < 0.001$). After exposure, we observe significant increases in the perceived influence on selves ($t = 9.4, p < 0.001$) as well as on others ($t = 11.4, p < 0.001$). However, TPP persists (influence on others: Mean 3.5, SD 1.0; on self: Mean 2.3, SD 1.0; gap: $t = 31.7, p < 0.001$). In fact, the gap between the perceived influence of bots on others versus on selves widens slightly after exposure: $t = 2.3, p = 0.021$. These results suggest that exposure to bots amplifies not only the perceived bot influence but also the TPP bias of the participants.

Propensity for more stringent bot regulation

We asked the participants about their views regarding regulations and other countermeasures toward social bots (see "Materials and Methods" Section). The results are presented in Fig. 4a. We find that the preference for stricter regulations among participants is significantly amplified by exposure to bots ($t = 5.1, p < 0.001$). The group favoring the strictest measures grows from 27.1% to 38.6%.

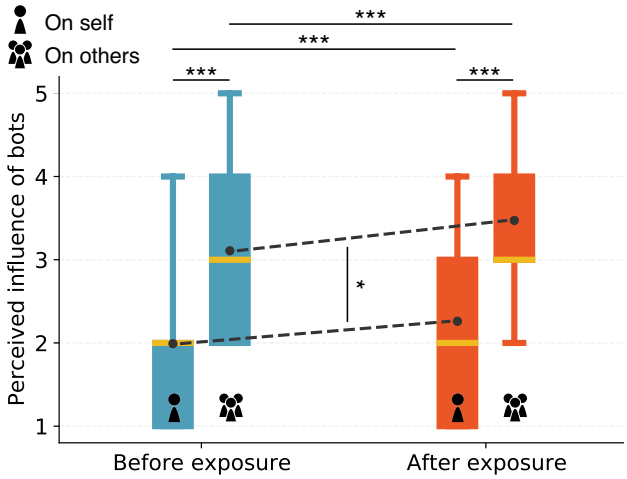

Figure 3. Stronger perceived bot influence on others than on selves. The box plots show perceived influence of bots on others versus on participants themselves before and after exposure to bots impersonating humans. The magnitude of the perceived influence is in the 1–5 range. The perceived influence of bots on others is significantly stronger than on participants themselves, indicating third-person perceptions, both before and after exposure ($p < 0.001$). The perceived influence of bots on self and others both significantly increase after exposure ($p < 0.001$). The gap between perceived influence on others and on participants themselves is slightly but significantly larger after exposure ($p = 0.021$).

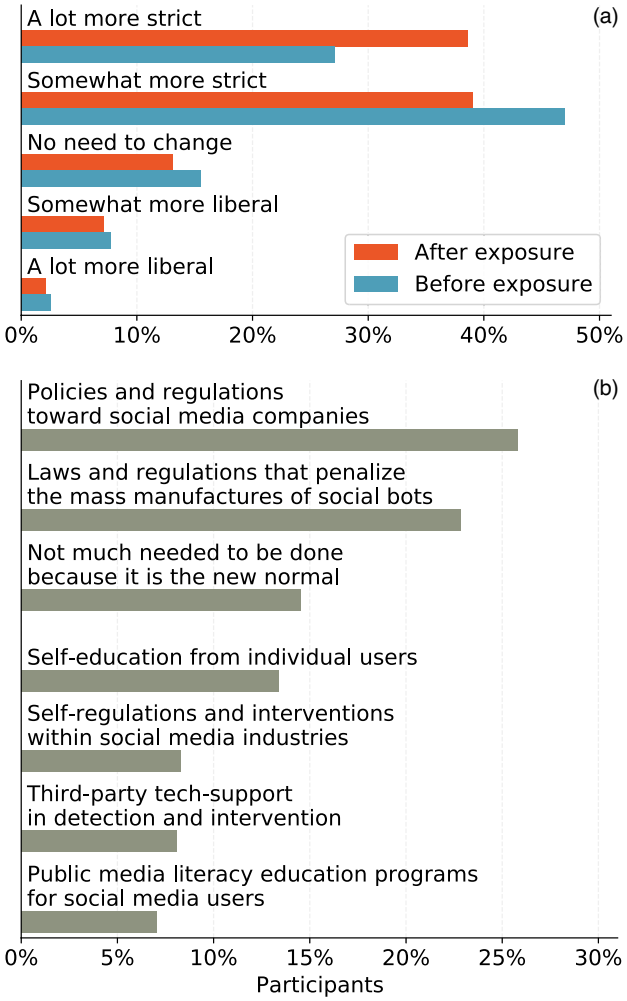

Figure 4. Preferences for social bot countermeasures. (**a**) Percentages of participants preferring more strict or more liberal restrictions before and after exposure to bots. (**b**) Percentages of participants who rank different countermeasures as their top choices.

We also investigate whether changes in regulation preference are predicted by the changes in the three measures explored in previous sections while controlling for demographic factors. The overall model contributions to the change in preference are small ($R^2 < 0.05$, see Table 2). However, we find a significant association between the decrease in bot-recognition self-efficacy and the dependent variable ($b = -0.11, p = 0.005$), whereas the change in perceived prevalence is not a significant predictor. While TPP (the perceptual gap between bot influence on others and on self) is not a significant predictor, the increase in perceived influence on others caused by exposure is significantly associated with the preference toward more stringent regulation ($b = 0.10, p < 0.015$).

We also asked participants to rank their preferences among countermeasures targeting different stakeholders after exposure to bots (see "Materials and Methods"). The results are shown in Fig. 4b. Overall, a majority of participants rank top-down options, such as legislative regulations targeting social media platforms (25.8%) and penalizing bot operators (22.8%), as their preferred policies. The countermeasures that have received the most attention from platforms and researchers, including company self-regulation (8.3%), third-party support (8.1%), and media literacy campaigns (7.1%), have less support. A substantial portion of the participants (14.5%) are in favor of accepting the existence of bot manipulation as the "new normal".

Discussion

Our experimental design has some limitations. First, the participants were asked the same questions twice. Since we did not have a separate control group, we cannot exclude the potential confounding effect of sequential testing. Second, the profiles selected in the experiments may not be representative of real-life bot exposure. Third, two of our findings regarding third-person perception changes and regulation preferences are statistically significant but have small effect sizes.

With these caveats, this study demonstrates how exposure to social bots significantly distorts perceptions of bot prevalence and influence. First, we find that potentially inflated estimates about bot prevalence on social media are further amplified after bot exposure. According to the "law of small numbers" bias[27], this overestimation can be attributed in part to participants extrapolating from the few examples in the experiment to the

Predictors	Model I[1]		Model II[2]	
	B	*p*	*B*	*p*
(Intercept)	0.05	0.443	0.04	0.478
Age	− 0.04	0.33	− 0.04	0.338
Education	− 0.06	0.112	− 0.06	0.108
Independent[a]	− 0.06	0.535	− 0.03	0.56
Republican[a]	− 0.12	0.259	− 0.11	0.302
Change scores[b]				
Perceived prevalence	0.06	0.153	0.04	0.308
Self-efficacy	**− 0.11**	**0.005**	**− 0.1**	**0.008**
Third-person perceptions	0.07	0.073		
Perceived influence on self			− 0.00	0.988
Perceived influence on others			**0.10**	**0.015**
R^2/ R^2 adjusted	0.03/0.02		0.034/0.022	

Table 2. Ordinary least-square models predicting changes of policy restrictions. Significant values with *p* below 0.05 are in [bold]. [1]Model I includes the discrepancy between perceived influences on others and on selves as a single independent variable to measure the effect of third-person perceptions. [2]Model II includes the perceived influence of bots on others and on selves as separate independent variables. [a]Democrat is the reference group. [b]Change scores are the differences between pre- and post-exposure measures.

entire social media environment. The susceptibility of prevalence estimates to experimental manipulations also underscores the effects of media interactions on perceptions of social reality. For example, heavy consumers of TV entertainment, which frequently features violent stories and scenes, tend to exaggerate the prevalence of violence in the real world[43,60]. Just as people who perceive the world as more dangerous because of TV viewing develop a strong "mean-world" sentiment[61], the overestimation of social bots may exemplify dissatisfaction with a polluted social media environment[22].

Second, our results suggest that the participants generally have an optimistic but unreliable assessment of their own ability to recognize social bots. Prior to bot exposure, participants tended to express confidence in their bot-recognition skills. However, exposure to bots with no feedback created self-doubt. On one hand, participants may have encountered greater difficulty than anticipated in the bot recognition tasks, since our stimuli included some highly ambiguous accounts (see "Materials and Methods" Section: Profile Selection). On the other hand, this result underscores the malleability of one's assessments regarding their own ability to detect bots and the potential recency bias of such judgments. While self-efficacy is positively correlated with actual bot recognition performance before bot exposure[44], the effect size is small, and the correlation is weaker after exposure. Furthermore, those who report larger improvements after the bot recognition task tend to perform worse—exposure actually raises the susceptibility to bot deception among over-confident users. These findings are consistent with the Dunning-Kruger effect[28,51] about the inability to objectively assess one's own expertise.

Third, participants believe that other social media users are more vulnerable to bot influence than themselves, consistent with the third-person perceptions observed in other contexts involving negative media messages[52]. Such a self-other perceptual gap can in part be explained by the typical egotistic bias[62,63]. Bot exposure widens the perceptual gap by weakening one's notion of self-immunity and to a greater extent by elevating the perceived vulnerability of others.

Finally, priming social media users to the threats posed by bots could unintentionally exacerbate a culture of distrust. After bot exposure, the majority of participants express preferences for regulations that target bot operators and social media companies over other options. This is consistent with a growing demand for governmental oversight of social bots[37] and may reflect an increasing public distrust towards social media companies[64]. Our analysis suggests that the public support for more top-down policies may be due to uncertainty about one's vulnerability to bot manipulation and fear of bot influence on others. The support for more stringent policy is susceptible to experimental manipulation and can be seen to stem from common cognitive biases, indicating that such policy preferences are not entirely rational. Regulating bots also raises First Amendment issues in the U.S.[65]. We believe that regulations may play a positive role in countering social media manipulation, but only in combination with other interventions[66], such as information literacy and continued development of bot detection systems.

While our current study has provided insights into the immediate effects of exposure to social bots, it is possible that these effects could be reactive and temporary. However, it is essential to recognize that social bots have long been and will continue to be an integral component of social media platforms. As the long-term effects of social media gradually unfold[67,68], there is a growing interest in understanding the enduring impact of bots. To address this important aspect, we are actively considering the development of multi-wave experiments and public opinion tracking polls. These future studies will allow us to explore how the evolving landscape of social bots, possibly powered by state-of-the-art artificial intelligence technologies[69], affects user behavior, policy preferences, and other relevant outcomes over time.

Materials and methods

Bot recognition task We conducted two experiments with the same pre-test questionnaire, followed by slightly different bot recognition tasks and post-test questionnaires. After finishing the pre-test questionnaires, participants were instructed to view 20 Twitter user profiles, half of which were bot-like and the other half were real users (the selection process is explained below). Participants were directed to the actual profiles through links to twitter.com. After viewing each profile, participants were asked to label it as human or bot.

Profile selection Following Yan et al.[17], we relied on Botometer scores[36] and expert coding for profile selection. Botometer is a widely used machine-learning tool for bot detection (botometer.org). It generates a score for each profile ranging from 0 to 5, with higher scores indicating more likely automated accounts. Scores close to 2.5 suggest high ambiguity. We started with an initial profile pool that consisted of randomly sampled 28,558 followers of U.S. congresspeople from both Republican and Democratic parties. We then sampled a total of 1,561 profiles with low (below 0.1) or high (above 4.9) bot scores, and 785 ambiguous profiles (bot scores around 2.5). During the final selection, expert coding by two authors was used to label the political nature and bot-likeness of the profiles independently. The coders placed particular emphasis on profiles with high ambiguity, as they presented challenges for Botometer. The expert coders underwent training to evaluate multiple profile heuristics systematically. Specifically, they considered factors such as the consistency of screen names and handles, the authenticity of profile and background pictures, the profile descriptions, the numbers of followers and friends, total tweet counts, tweet frequency, the percentage of original tweets and retweets, as well as the diversity and content authenticity of tweets in the timelines. Notably, two expert coders yielded fully consistent results. The final profile pool consisted of a total of 40 profiles that included even portions of political and non-political profiles, bot-like and authentic users, and low-ambiguity and high-ambiguity profiles. In addition to bot scores and expert coding, the human/bot classification of the 40 profiles was additionally corroborated by a crowdsourcing strategy, which used the majority of answers from a partisanship-balanced subsample of participants. All of the accounts were still active immediately after the experiment.

Experiment I used a mixed between/within-subject design. Participants ($N = 308$) were assigned to either a low-ambiguity or a high-ambiguity condition. In the low-ambiguity condition, profiles were selected with bot scores below 0.1 or above 4.9. In this condition, the human/bot labels generated by the three approaches mentioned above were fully consistent. In the high-ambiguity condition, half of the selected profiles had ambiguous scores (close to 2.5); the profiles in this condition were labeled by expert coding and crowdsourcing. In both conditions, only half of the profiles exhibited clear political partisanship.

Experiment II adopted a pure within-subject design: all participants ($N = 656$) viewed the same 20 profiles, all of which exhibited clear partisanship (e.g., including identity markers such as "#Republican" or "#VoteBlue"), and half of which had ambiguous scores. This experiment included additional policy-related questions in the post-test questionnaire. Analyses of perceptions about bot prevalence, self-efficacy, and bot influence are based on merged data from the two experiments, while the policy analysis is based on Experiment II only.

Sampling This study enlisted participants from Amazon Mechanical Turk (MTurk). Previous research has indicated that samples recruited from MTurk can effectively capture participants with a wide range of backgrounds[70]. However, these samples tend to over-represent digitally active individuals and exhibit higher levels of digital literacy[71]. While this may constrain the generalizability of MTurk samples to other research domains, we consider active internet and social media users as the target population in the context of the current study.

Prevalence of bots We measured the perceived prevalence of bots before and after the bot recognition task with the same question: "According to your estimation, what percentage of accounts on social media do you think are social bots?" Participants answered the questions with a slider ranging from 0 to 100 percent.

Self-efficacy We measured the self-efficacy in recognizing bot profiles by asking participants to what extent they agree with the following three statements: (1) "I will recognize social bots if I encounter them"; (2) "I believe I can succeed at telling social bots apart from real users"; and (3) "When facing social bots that highly resemble regular users, I can still find clues to weed them out." The answers included seven options, ranging from "Strongly disagree" to "Strongly agree." The same questions were asked before and after the bot recognition task (Cronbach $\alpha = 0.88$ and 0.93, respectively).

Third-person perception We measured the TPP of bots by asking two questions about the extent to which participants thought that social bots might have influenced them and average social media users. Participants were given five-point options, ranging from "Not at all" to "A lot," to answer each question. We analyzed the two answers and their discrepancy.

Countermeasure preferences We asked participants before and after the bot recognition task: "If we were to restrict the production and use of bots on social media, what kind of changes would you like to see?" with five-point options ranging from "A lot more liberal" to "A lot more strict." We also asked participants to rank seven specific countermeasures from the most preferred to the least (see wordings in Fig. 4b).

Control variables Our analysis included controls for participants' partisanship, age, and education level (see summary of demographic information in Table 3). We initially considered self-reported variables such as the frequency of Twitter usage ($M = 3.16, SD = 1.28$) and past encounters with social bots ($M = 3.04, SD = 1.04$) as potential control variables. Participants were asked two questions: "How often do you check Twitter?" and "How many times do you think you have encountered social bots on social media before?" Their responses were measured on a five-point scale ranging from "never" to "very frequently." However, the preliminary analysis suggested that self-reported Twitter usage frequency ($r = -0.02, p = 0.508$) and prior encounters with bots ($r = 0.02, p = 0.479$) did not significantly predict performance in bot recognition tasks. Consequently, these variables were not included in the final analysis.

Experimental protocols The research methods and materials were approved by the Institutional Review Board at Indiana University-Bloomington under Protocol #1811295947 prior to the experiments. Informed consent

	Overall	Experiment I	Experiment II
	(N = 964)	(n = 308)	(n = 656)
Partisanship			
Democrat	41.26%	40.39%	41.67%
Republican	23.28%	23.78%	23.03%
Independents	35.44%	38.83%	35.26%
Age (Mean)	34.92	34.38	35.17
Age (SD)	11.45	11.06	11.62
Education			
< College	21.16%	35.71%	14.39%
College or equivalent	50.51%	49.03%	51.21%
> College	28.31%	15.25%	34.45%

Table 3. Demographic information.

was obtained from all participants before they proceeded to participate in the experiments. All methods were performed in accordance with the relevant guidelines and regulations.

Data availability

The dataset used in the current study is available in the GitHub repository at https://github.com/osome-iu/bot_perception_bias.

Received: 7 March 2023; Accepted: 3 November 2023
Published online: 24 November 2023

References

1. Confessore, N., *et al.* The follower factory. The New York Times. (2018). https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html.
2. Ferrara, E. *et al.* The rise of social bots. *Commun. ACM* **59**(7), 96–104 (2016).
3. Shao, C. *et al.* The spread of low-credibility content by social bots. *Nat. Commun.* **9**(1), 4787 (2018).
4. Stella, M., Ferrara, E. & De Domenico, M. Bots increase exposure to negative and inflammatory content in online social systems. *Proc. Natl. Acad. Sci.* **115**(49), 12435–12440 (2018).
5. Uyheng, J. & Carley, K. M. Bots and online hate during the COVID- 19 pandemic: Case studies in the United States and the Philippines. *J. Comput. Soc. Sci.* **3**(2), 445–468 (2020) (**ISSN: 2432-2725.**).
6. Boshmaf, Y., *et al.* The socialbot network: When bots socialize for fame and money. in *Proceedings of the 27th Annual Computer Security Applications Conference. ACM*, pp. 93–102 (2011).
7. Duan, Z. *et al.* Algorithmic agents in the hybrid media system: Social bots, selective amplification, and partisan news about COVID-19. *Hum. Commun. Res.* **48**(3), 516–542 (2022).
8. Marlow, T., Miller, S. & Timmons Roberts, J. Bots and online climate discourses: Twitter discourse on President Trump's announcement of U.S. withdrawal from the Paris Agreement. *Clim. Policy* **21**(6), 765–777 (2021) (**ISSN: 1469-3062.**).
9. Keller, F. B. *et al.* Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Polit. Commun.* **37**(2), 256–280 (2020).
10. Fan, R., Talavera, O. & Tran, V. Social media bots and stock markets. *Eur. Financ. Manag.* **26**(3), 753–777 (2020) (**ISSN: 1468-036X.**).
11. Nizzoli, L. *et al.* Charting the landscape of online cryptocurrency manipulation. *IEEE Access* **8**, 113230–113245 (2020) (**ISSN: 2169-3536**).
12. Ferrara, E. *et al.* Characterizing social media manipulation in the 2020 US presidential election. *First Monday***25**(11) (2020).
13. Ferrara, E. Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*https://doi.org/10.5210/fm.v22i8.8005 *(2017)*.
14. Bastos, M. & Mercea, D. The public accountability of social platforms: Lessons from a study on bots and trolls in the Brexit campaign. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **376**(2128), 20180003 (2018).
15. González-Bailón, S. & De Domenico, M. Bots are less central than verified accounts during contentious political events. *Proc. NatL. Acad. Sci.* **118**(11), e2013443118 (2021).
16. Bail, C. A. *et al.* Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci.* **115**(37), 9216–9221 (2018).
17. Yan, H. Y. *et al.* Asymmetrical perceptions of partisan political bots. *New Med. Soc.* **23**(10), 3016–3037 (2021).
18. Stoking, G., & Sumida, N.. Social Media Bots Draw Public's Attention and Concern. Pew Research Center. (2018). https://www.journalism.org/2018/10/15/social-media-bots-draw-publics-attentionand-concern/.
19. Starbird, K. Disinformation's spread: Bots, trolls and all of us. *Nature* **571**(7766), 449 (2019).
20. Yan, H. Y., & Yang, K. C., *The landscape of social bot research: A critical appraisal* (Handbook of Critical Studies of Artificial Intelligence, OSF Preprints, 2022).
21. Yang, K. C., & Menczer, F., How many bots are on Twitter? The question is difficult to answer and misses the point. The Conversation. (2022). https://theconversation.com/how-many-bots-are-on-twitter-thequestion-is-difficult-to-answer-and-misses-the-point-183425.
22. Halperin, Y. When bots and users meet: Automated manipulation and the new culture of online suspicion. *Glob. Perspect.* **2**(1), 24955 (2021).
23. Wischnewski, M., *et al.* Disagree? You Must be a Bot! How Beliefs Shape Twitter Profile Perceptions. in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–11 (2021).
24. Hills, T. T. The dark side of information proliferation. *Perspect. Psychol. Sci.* **14**(3), 323–330 (2019).

25. Haselton, M. G., Nettle, D., & Andrews, P. W. The Evolution of Cognitive Bias. in *The Handbook of Evolutionary Psychology*. Chap. 25, pp. 724–746. (2015) https://doi.org/10.1002/9780470939376.ch25. https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470939376.ch25
26. Menczer, F. & Hills, T. The attention economy. *Sci. Am.* **323**(6), 54–61. https://doi.org/10.1038/scientificamerican1220-54 (2020).
27. Tversky, A. & Kahneman, D. Belief in the law of small numbers. *Psychol. Bull.* **76**(2), 105 (1971).
28. Dunning, D. The Dunning–Kruger effect: On being ignorant of one's own ignorance. *Adv. Exp. Soc. Psychol.* **44**, 247–296 (2011).
29. Davison, W. P. The third-person effect in communication. *Pub. Opin. Q.* **47**(1), 1–15 (1983).
30. Gunther, A. C. Overrating the X-rating: The third-person perception and support for censorship of pornography. *J. Commun.* **45**(1), 27–38 (1995).
31. Witte, K. & Allen, M. A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Educ. Behav.* **27**(5), 591–615 (2000).
32. Sun, Y., Shen, L. & Pan, Z. On the behavioral component of the third-person effect. *Commun. Res.* **35**(2), 257–278 (2008).
33. Cresci, S. A decade of social bot detection. *Commun. ACM* **63**(10), 72–83 (2020).
34. Yang, K. C., *et al.* Scalable and Generalizable social bot detection through data selection. in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34(01), pp. 1096–1103, (2020) ISSN: 2374-3468.
35. Sayyadiharikandeh, M., *et al.* Detection of novel social bots by ensembles of specialized classifiers. in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 2725–2732 (2020).
36. Yang, K.-C. *et al.* Arming the public with artificial intelligence to counter social bots. *Hum. Behav. Emerg. Technol.* **1**(1), 48–61 (2019).
37. Gorwa, R. & Guilbeault, D. Unpacking the social media bot: A typology to guide research and policy. *Policy Internet* **12**(2), 225–248 (2020).
38. Twitter, Inc. Fiscal Year 2021 Annual Report. Retrieved from https://investor.twitterinc.com/financial-information/annual-reports/default.aspx.2021. https://s22.q4cdn.com/826641620/files/docfinancials/2021/ar/FiscalYR2021TwitterAnnual-Report.pdf.
39. Hals, T. Elon Musk files countersuit under seal vERSUs Twitter over 44 billion deal. Reuters. (2022). https://www.reuters.com/legal/transactional/judge-orders-oct-17-21-trial-over-twitters-lawsuitagainst-musk-2022-07-29.
40. Varol, O. *et al.* Online human-bot interactions: Detection, estimation, and characterization. in *Proceedings of the International AAAI Conference on Web and Social Media* (2017).
41. Kuang, J. *et al.* Bias in the perceived prevalence of open defecation: Evidence from Bihar India. *PLoS ONE* **15**(9), e0238627 (2020).
42. Chia, S. C. How peers mediate media influence on adolescents' sexual attitudes and sexual behavior. *J. Commun.* **56**(3), 585–606 (2006).
43. Morgan, M. & Shanahan, J. The state of cultivation. *J. Broadcast. Electro. Med.* **54**(2), 337–355 (2010).
44. Bandura, A. The explanatory and predictive scope of the self-efficacy theory. *J. Soc. Clin. Psychol.* **4**(3), 359 (1986).
45. Stajkovic, A. D. & Luthans, F. Self-efficacy and work-related performance: A meta-analysis. *Psychol. Bull.* **124**(2), 240 (1998).
46. Huang, C. Gender differences in academic self-efficacy: Ameta-analysis. *Eur. J. Psychol. Educ.* **28**(1), 1–35 (2013).
47. Ashford, S., Edmunds, J. & French, D. What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. *Br. J. Health. Psychol.* **15**(2), 265–288 (2010).
48. Mahmood, K. Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Commun. Inf. Lit.* **10**(2), 3 (2016).
49. Mazor, M. & Fleming, S. M. The Dunning–Kruger effect revisited. *Nat. Hum. Behav.* **5**(6), 677–678 (2021).
50. Kruger, J. & Dunning, D. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* **77**(6), 1121 (1999).
51. Jansen, R. A., Rafferty, A. N. & Griffiths, T. L. A rational model of the Dunning–Kruger effect supports insensitivity to evidence in low performers. *Nat. Hum. Behav.* **5**(6), 756–763 (2021).
52. Sun, Y., Pan, Z. & Shen, L. Understanding the third-person perception: Evidence from a meta-analysis. *J. Commun.* **58**(2), 280–300 (2008).
53. Perloff, R. M. Third-person effect research 1983–1992: A review and synthesis. *Int. J. Pub. Opin. Res.* **5**(2), 167–184 (1993).
54. Paul, B., Salwen, M. B. & Dupagne, M. The third-person effect: A meta-analysis of the perceptual hypothesis. *Mass Commun. Soc.* **3**(1), 57–85 (2000).
55. Yan, H. Y. The rippled perceptions: The effects of LGBT-inclusive TV on own attitudes and perceived attitudes of peers toward lesbians and gays. *J. Mass Commun. Q.* **96**(3), 848–871 (2019).
56. Rosenthal, S., Detenber, B. H. & Rojas, H. Efficacy beliefs in third-person effects. *Commun. Res.* **45**(4), 554–576 (2018).
57. Tsay-Vogel, M. Me versus them: Third-person effects among Facebook users. *New Med. Soc.* **18**(9), 1956–1972 (2016).
58. Mo Jang, S. & Kim, J. K. Third person effects of fake news: Fake news regulation and media literacy interventions. *Comput. Hum. Behav.* **80**, 295–302 (2018).
59. Paek, H.-J. *et al.* The third-person perception as social judgment: An exploration of social distance and uncertainty in perceived effects of political attack ads. *Commun. Res.* **32**(2), 143–170 (2005).
60. Morgan, M. & Shanahan, J. Television and the cultivation of authoritarianism: A return visit from an unexpected friend. *J. Commun.* **67**(3), 424–444 (2017).
61. Nabi, R. L. & Sullivan, J. L. Does television viewing relate to engagement in protective action against crime? A cultivation analysis from a theory of reasoned action perspective. *Commun. Res.* **28**(6), 802–825 (2001).
62. Gunther, A. C. & Mundy, P. Biased optimism and the third-person effect. *J. Q.* **70**(1), 58–67 (1993).
63. Lyons, B. A. Why we should rethink the third-person effect: Disentangling bias and earned confidence using behavioral data. *J. Commun.* **72**(5), 565–577 (2022) (**ISSN: 0021-9916.**).
64. Flew, T. & Gillett, R. Platform policy: Evaluating different responses to the challenges of platform power. *J. Digit. Med. Policy* **12**(2), 231–246 (2021).
65. Lamo, M. & Calo, R. Regulating bot speech. *UCLA Law Rev.* **66**, 988 (2019).
66. Bak-Coleman, J. B. *et al.* Combining interventions to reduce the spread of viral misinformation. *Nat. Hum. Beh.* **6**(10), 1372–1380 (2022).
67. Hermann, E., Morgan, M. & Shanahan, J. Cultivation and social media: A meta-analysis. *New Med. Soc.* **25**(9), 2492–2511 (2023).
68. Tsay-Vogel, M., Shanahan, J. & Signorielli, N. Social media cultivating perceptions of privacy: A 5-year analysis of privacy attitudes and self-disclosure behaviors among Facebook users. *New Med. Soc.* **20**(1), 141–161 (2018).
69. Yang, K. C., & Menczer, F. Anatomy of an AI-powered malicious social botnet. arXiv preprint arXiv:2307.16336 (2023).
70. Berinsky, A. J., Huber, G. A. & Lenz, G. S. Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Polit. Anal.* **20**(3), 351–368 (2012).
71. Guess, A. M. & Munger, K. Digital literacy and online political behavior. *Polit. Sci. Res. Methods* **11**(1), 110–128 (2023).

Author contributions

H.Y. designed the study, H.Y. and K.-C.Y. performed the analyses, and all authors contributed to the writing and assisted in the revision of the manuscript and the refinement of its arguments.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to H.Y.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023