Applied Network Science

RESEARCH Open Access

Bots influence opinion dynamics without direct human-bot interaction: the mediating role of recommender systems

N. Pescetelli[1,2*], D. Barkoczi[2,3] and M. Cebrian[2,4,5]

*Correspondence: niccolo.pescetelli@njit.edu

[1] New Jersey Institute of Technology, 323 Dr. Martin Luther King Jr Blvd, Newark, NJ 07102, USA
[2] Center for Humans and Machines, Max Planck Institute for Human Development, 94 Lentzeallee, 14195 Berlin, Germany
[3] Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, China
[4] Statistics Department, Universidad Carlos III de Madrid, Marid, Spain
[5] UC3M-Santander Big Data Institute, Universidad Carlos III de Madrid, Madrid, Spain

Abstract

Bots' ability to influence public discourse is difficult to estimate. Recent studies found that hyperpartisan bots are unlikely to influence public opinion because bots often interact with already highly polarized users. However, previous studies focused on direct human-bot interactions (e.g., retweets, at-mentions, and likes). The present study suggests that political bots, zealots, and trolls may indirectly affect people's views via a platform's content recommendation system's mediating role, thus influencing opinions without direct human-bot interaction. Using an agent-based opinion dynamics simulation, we isolated the effect of a single bot—representing 1% of nodes in a network—on the opinion of rational Bayesian agents when a simple recommendation system mediates the agents' content consumption. We compare this experimental condition with an identical baseline condition where such a bot is absent. Across conditions, we use the same random seed and a psychologically realistic Bayesian opinion update rule so that conditions remain identical except for the bot presence. Results show that, even with limited direct interactions, the mere presence of the bot is sufficient to shift the average population's opinion. Virtually all nodes—not only nodes directly interacting with the bot—shifted towards more extreme opinions. Furthermore, the mere bot's presence significantly affected the internal representation of the recommender system. Overall, these findings offer a proof of concept that bots and hyperpartisan accounts can influence population opinions not only by directly interacting with humans but also by secondary effects, such as shifting platforms' recommendation engines' internal representations. The mediating role of recommender systems creates indirect causal pathways of algorithmic opinion manipulation.

Keywords: Bots, Opinion dynamics, Bayesian belief update, Recommender systems, Social influence

Introduction

Bots and recommender systems

Bots are becoming pervasive in our social media. From Twitter to Reddit, bots can interact with humans without detection, influencing opinions, and creating artificial narratives (Hurtado et al. 2019; Yanardag et al. 2021). This study uses an agent-based

© The Author(s) 2022. **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/.

simulation to explore the interaction between bots and content recommendation algorithms.

Recommender systems such as collaborative filtering can provide hyper-personalized content recommendations. However, they partly rely on average population characteristics and shared features between nodes to produce their recommendations. We test the hypothesis that recommender systems mediating information access can mediate bot influence. We hypothesize that bots can affect a population's mean opinion not just by direct interactions with other nodes but via skewing the training sample fed to the recommender system during training (i.e., indirect interactions). Thus, a bot may influence content recommendation at the population level by subtly affecting how a centralized recommender system represents a population's preferences and patterns of content engagement. This indirect social influence may be more pervasive than direct social influence because it occurs without direct bot-human interaction.

The potential of algorithmic agents, commonly called bots, to influence public opinion has recently been under closer scrutiny. Special attention has been given to social and political bots that operate under human disguise on social media. Early studies documented the potential effects of bots on skewing opinion distributions on social media users and voters (Bessi and Ferrara 2016). Bots can inflate the perception of the popularity of particular views (Lerman et al. 2016), polarise opinions around divisive issues (Broniatowski et al. 2018; Stewart et al. 2018; Carley 2020), contribute to the spread of misinformation, conspiratory theories or hyper-partisan content (Paul and Matthews 2016; Shao et al. 2018), and promote harmful or inflammatory content (Stella et al. 2018). These generalized concerns have mobilized platforms to improve algorithmic agents' automatic detection and removal (Howard 2018; Ferrara et al. 2016; Ledford 2020; Beskow and Carley 2018). Bot influence often acts on public opinion in concert with human trolls, fake accounts, pink-slime newspapers, and "fake news" (Hurtado et al. 2019; Linvill and Warren 2018; Aral and Eckles 2019; Tucker et al. 2018). Researchers have started to untangle this complex web of interactions. The content spread by this class of agents spreads faster due to its emotional or sensationalist features (Vosoughi et al. 2018; Lazer et al. 2018). While often studied together, misinformation and extreme views can be orthogonal dimensions. Empirical evidence shows that misinformation and extreme online views tend to be more engaging than accurate or moderate content (Edelson et al. 2021). Recommender systems seem critical in promoting extreme content over moderate ones (Whittaker et al. 2021). Partisan content tends to remain confined in insulated clusters of users, thus reducing the opportunity to encounter cross-cutting content (Bakshy et al. 2015). Although algorithmic agents represent only a small part of general media manipulation tactics (Kakutani 2019; Sunstein 2018), they pose a problem for online platforms. Their ease of implementation, low cost, and scalability hurt the overall media environment. In this paper, we estimate the lower bound of algorithmic influence by focusing on the effect of a single algorithmic agent on a population. Our findings can be generalized to other 'pre-programmed' or 'stubborn' agents of media manipulation, such as partisan accounts and human trolls (Hegselmann and Krause 2015). Pre-programmed agents share several features, such as pre-set opinions and pushing political agendas while being scarcely influenced by others' beliefs.

The effect of bots and troll factories on public opinion is hard to estimate. Several researchers have recently attempted to measure hyper-partisan content's effect by looking at social media data from the 2016 USA presidential election (Guess et al. 2019; Allen et al. 2020). These studies suggest that sharing and consuming fake or hyper-partisan content was relatively rare relative to the total volume of content consumed. One study (Bail et al. 2020) attempted to measure the effect of exposure to Russia's Internet Research Agency (IRA) content on people's opinions. The authors found that interactions with highly partisan accounts were most common among respondents with already strong ideological alignment with those opinions. The researchers interpreted these findings as suggesting that hyper-partisan accounts might fail to change beliefs because they primarily interact with already highly polarised individuals. This phenomenon, also named "minimal effect", is not specific to social media platforms but can also be found in offline political advertisement and canvassing practices (Zaller 1992; Endres and Panagopoulos 2019; Kalla and Broockman 2018). In other words, changing political attitudes tends to be less effective than one imagines. A recent study found that human accounts are significantly more visible during political events than unverified accounts (González-Bailón and De Domenico 2021). This finding casts doubt on the centrality and impact of bot activity on political mobilizations' coverage (Ferreira et al. 2021). Overall, these findings show that, notwithstanding the well-documented spread of bots and troll factories on social media, their effect on influencing opinions may be limited.

The studies reviewed above were primarily concerned with direct influence among agents, namely direct interactions between algorithmic and human accounts (e.g., likes, retweets, and comments). Although common in many social settings, we argue that direct social influence does not consider the complexity of the digital influence landscape. Direct social influence has long been studied outside the domain of social media platforms, e.g., opinion change in social psychology (Yaniv 2004; Bonaccio and Dalal 2006; Sherif et al. 1965; Festinger and Carlsmith 1959; Rader et al. 2017) and in opinion dynamics in sociology (Flache et al. 2017; Deffuant et al. 2000; DeGroot 1974; Friedkin and Johnsen 1990). Direct influence assumes the unfiltered exposure to another person's belief (e.g., an advisor) changes a privately held belief. However, this simple social influence model may be outdated in the modern digital environment.

Although direct interactions on most online platforms do occur (e.g., friends exchanging messages and users tweeting their views), information exchange is also mediated by algorithmic procedures that sort, rank, and disseminate or throttle information. The algorithmic ranking of content can affect exposure to specific views (Bakshy et al. 2015). Recommender systems can learn population averages and trends, forming accurate representations of individual preferences from collective news consumption patterns (Das et al. 2007; Analytis et al. 2020). One crucial difference between traditional social influence and machine-mediated social influence is that in the latter case, single users can influence not only other people's beliefs but the "belief" of the content curation algorithm (i.e., its internal model). This paper investigates a previously unexplored indirect causal pathway connecting social bots and individuals via a simple recommendation algorithm (Fig. 1a). We test the hypothesis that algorithmic agents, like bots and troll factories, can disproportionately influence the entire population by biasing the training sample of recommender

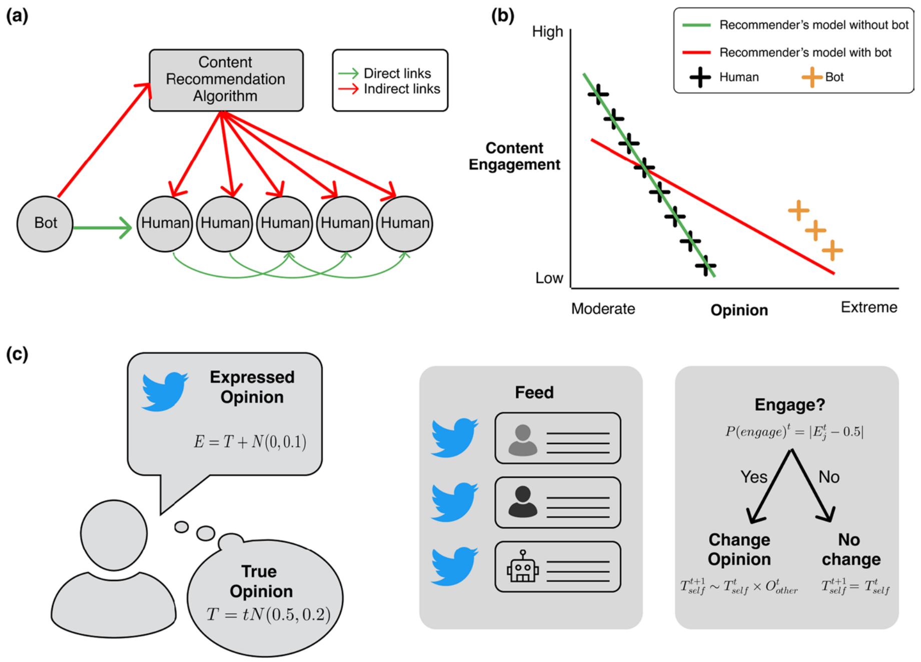

Fig. 1 The indirect influence of bots on social information networks. **(a)** Representation of opinion dynamics network mediated by a content recommender system (grey box on top). Bot and Human agents (circles) consume and share content. A bot agent can influence human opinions via direct interaction with human agents (e.g., retweets, at-mentions, likes, and comments) or indirectly via affecting the internal representation of the content recommendation algorithm. **(b)** Schematic representation of the effect of bot presence on the internal representation learned by a simple recommender system trained to predict a user's engagement with various types of content. Including the bot behavior in the training set skews the model to think that engagement with extreme content is more likely than it would be without the bot presence. **(c)** Agents in the simulation were modeled to include a true private opinion and an expressed public opinion. Agents were presented with one of their neighbors' public opinions on every round based on the recommender's predicted content engagement. Then the agents decided whether to engage with this content or not according to their engagement function (Eq. 3). Opinion change took place only if the agent decided to engage with the recommended piece of content

algorithms predicting user engagement and user opinions (Fig. 1b). This disproportionate influence may be facilitated by their resistance to persuasion and greater content engagement and sharing activity (Yildiz et al. 2013; Hunter and Zaman 2018). Affecting recommender systems' internal representations would be a more effective influence strategy that can influence a network's nodes in parallel rather than serially.

We call this type of influence *machine-mediated indirect influence*, as opposed to indirect influence occurring via intermediary nodes (a bot may directly influence one human but indirectly influence all the humans to whom the first human is connected). Recent research in opinion dynamics has already shown the importance of weak ties and the indirect influence of bots on the rest of the network (Keijzer and Mäs 2021; Aldayel and Magdy 2022). Here, however, we are especially interested in the influence of social bots on network opinion dynamics when platform-wide algorithmic content recommendation mediates information sharing.

A cognitive model of Bayesian opinion update

Several opinion dynamics models represent belief updates as linear combinations of opinions, such as weighted averages (Friedkin and Johnsen 2011; DeGroot 1974). Linear models, however, need to explain the non-linear dynamics of belief escalations often observed in online and lab settings (Bail et al. 2018; Pescetelli and Yeung 2020b). Models that try to account for these effects—e.g., similarity bias and repulsive influence (Flache et al. 2017)—often use parameters that are difficult to match with the well-known cognitive processes underlying opinion change (Resulaj et al. 2009; Fleming et al. 2018; Yaniv 2004).

Opinion change has been the focus of an active investigation in cognitive neuroscience and social psychology (Bonaccio and Dalal 2006). This research shows that people update their opinions based on subjective estimates of uncertainty in their beliefs: more confident opinions are more influential in group settings (Price and Stone 2004; Penrod and Cutler 1995; Sniezek and Van Swol 2001), and confident individuals show smaller opinions shifts (Soll and Mannes 2011; Yaniv 2004; Becker et al. 2017). Opinion dynamics models have used confidence to model the susceptibility to persuasion—or vice versa the influence of an opinion (Hegselmann and Krause 2015). However, while this literature models confidence as a free parameter, we build on recent theoretical and empirical work on the neurocognitive bases of confidence (Ma et al. 2006; Fleming et al. 2018; Fleming and Daw 2017). According to this framework, confidence behaves and is mathematically described as a probability estimate. The probabilistic framework has two direct benefits. First, it grounds opinion dynamics models in cognitive psychology and empirical behavioral findings of decisions, beliefs, and changes of mind. Second, it allows for modeling a wide range of opinion dynamics (e.g., belief escalation, risky shift, polarization, convergence to consensus, bias assimilation, similarity bias) within the well-understood mathematical framework of probability (Hahn and Oaksford 2006, 2007).

In this paper, we use a binary choice (0, 1), where opinion and confidence are represented as the sign (opinion) and the magnitude (confidence) of the difference from 0.5, respectively. Thus, zero represents extreme confidence in one opinion, 0.5 represents a moderate or uncertain opinion, and 1 represents extreme confidence in the opposite opinion. We use a Bayesian opinion updating rule used in experimental psychology to model opinion change (Pescetelli and Yeung 2020a, b; Harris et al. 2016; Pescetelli et al. 2016). The Bayesian update offers a natural way to consider all aspects of beliefs, including opinion direction, belief conviction, and resistance to changes of mind or new information (Sun and Müller 2013; Hegselmann and Krause 2015). This opinion update function produces non-linear dynamics mirrored by belief updates in laboratory experiments (Pescetelli and Yeung 2020b; Pescetelli et al. 2016). Agreeing people tend to reinforce each other's beliefs and move to more confident positions. In comparison, people who disagree tend to converge to more uncertain positions (see (Bail et al. 2018) for an exception). These non-linear dynamics—often called biased assimilation in opinion dynamics—naturally emerge when using the Bayesian theorem to update opinions.

We created two identical, fully connected networks of 100 agents to test our hypothesis. The two networks differed only in whether the bot was present or absent. We initialized the two simulations using the same random seed, which allowed us to directly test the counterfactual of introducing a single bot in the network while holding all other conditions constant. The bot could influence other users via the recommendation algorithm (Fig. 1a). Contrary to previous studies (Friedkin and Johnsen 1990; DeGroot 1974), we distinguish between internally held beliefs and externally observable behavior. We assume that observable behavior represents a noisy reading of true internal beliefs. This assumption captures the fact that people on several online platforms, such as fora and social media, can form beliefs and change opinions simply by consuming content and never posting or sharing their own (Lazer 2020; Muller 2012). One does not need to tweet about climate change to form an opinion on climate change. The distinction between internally held and publicly displayed beliefs allows us to train the recommender algorithm with externally observable behavior. The recommender algorithm does not make the unrealistic assumption that it can access a user's private opinions. We call 'engagement' all externally observable behaviors such as tweets, likes, and reactions. Thus, the recommender algorithm and agents must infer other agents' underlying opinions from engagement behaviors.

Across a series of simulations, we quantify the effect of adding a single bot to a network of fully connected agents. We show that the bot can influence human agents even though few direct interactions exist between human agents and the bot. We conclude that in an information system where algorithmic models control who sees what, bots and hyper-partisan agents can influence the entire users' population by influencing the internal representation learned by the recommender algorithm. In other words, the recommender belief might be as crucial as people's beliefs in determining the outcome of network opinion dynamics. We discuss these findings in light of the contemporary debate on social media regulation.

Methods

Overview

We simulate a simplified social network model where a recommender system learns and presents a personalized content feed to agents in the network. This feed contains the expressed opinions of other agents in the network. Each agent can observe and interact with other agents' opinions by updating and expressing their own opinions. We manipulate whether a single bot is also part of the potential pool of agents that the recommender system draws upon to create the feeds in two separate but identical conditions. We study whether this bot can infiltrate the feed controlled by the recommender system by influencing the statistical relationships it learns.

Simulation procedure

All simulations were run using R version 4.1.2. We simulate N = 100 fully connected agents. We run the model for 100 steps. The limit of 100 steps was chosen because pilots converged long before this time. We test 100 replications per condition.

Agents

Each agent is represented by a private *opinion* in the range [0.01, 0.99] drawn from a truncated Normal distribution.

$$T_i = tN(0.5,\ 0.2) \tag{1}$$

and by an *expressed opinion* representing a noisy observation of their true opinions:

$$E_i = T_i + N(0,\ 0.1) \tag{2}$$

On each time step, agents go through a two-step process:

Engagement First, agents decide whether or not to *engage* with content in their feed (see below). Content is the expressed opinions of other agents ranked by the recommender system for each agent, based on the model's predicted engagement for a given piece of content. Agents decide whether to engage with the content based on two well-documented biases, namely the bias to engage with similar content (similarity bias or homophily bias) (Mäs and Flache 2013; Dandekar et al. 2013) and the bias to engage with extreme content and confident opinions (Penrod and Cutler 1995; Price and Stone 2004; Hegselmann and Krause 2015; Edelson et al. 2021; Whittaker et al. 2021). An engagement function is defined as:

$$P(engage_j)^{t+1} = \alpha\left(\left|E_j^t - T_i^t\right|\right) + 2(1-\alpha) \times \left|E_j^t - 0.5\right| \tag{3}$$

where E is the expressed opinion of another agent *j*. We represent engagement as a binary decision to engage or not engage, drawn from a binomial distribution with probability P(engage). The right-hand side of the equation represents a weighted sum of the similarity bias and the extremity bias. The importance of each bias is controlled by the weight parameter alpha, set to 0.2. We explore in Supplementary Material various values of alpha (Additional file 1: Fig. S2). Opinion similarity is represented by the absolute distance between E and T (the first term). In contrast, the extremity bias is represented by the absolute distance between E and the midpoint of 0.5 (the second term). The engagement function in Eq. 3 makes it more likely that agents engage with content that is (a) similar to their initially held opinion and (b) more distant from moderate opinions (represented by the mid point 0.5) their own opinions. This behavior represents people's online tendency to engage with shocking or count-intuitive content more than moderate content (Vosoughi et al. 2018; Lazer et al. 2018; Edelson et al. 2021).

Opinion update If agents decide to *engage,* they update their own opinions using a Bayesian opinion update function:

$$T_i^{t+1} = \frac{T_i^t O_i^t}{(T_i^t O_i^t) + (1 - T_i^t)(1 - O_j^t)} \tag{4}$$

where O is the expressed opinion of another agent *j*, discounted by a trust factor θ between 0 (no trust at all) and 1 (complete trust).

$$O_i^t = \theta(E_j^t - 0.5) + 0.5 \tag{5}$$

The addition of a discount factor helps stabilize the model. It avoids that agents who engage with distant opinions update their initial opinion with large shifts in the opinion space (an unrealistic behavior). We set $\theta = 0.2$.

If agents decide not to engage, they keep their opinion from the previous timestep, time *t*.

Feed

Each agent is presented with a feed consisting of the *expressed opinion* of another agent among their neighbors. The recommended piece of content (Feed) was created as a simple logistic recommender system. We chose this simple logistic model to provide a minimal proof of concept of our hypothesis. Based on past observations, the feed aims to provide the content that the agent is most likely to engage with (see *Engagement* above). To achieve this, we train a simple logistic regression using all agents' binary engagement history as a dependent variable and the absolute distance between the agent's public opinion at time t-1 and the opinion they observed in their feed as the independent variable.

$$H \sim L(D) \tag{6}$$

where L is a logistic regression model, H is the history of binary engagement events (0 = did not engage; 1 = did engage), and D is the absolute difference between an agent and its neighbors' public opinions. In other words, the model aims to learn the agents' engagement function by observing their prior engagement history and the content they observed in their feeds. To provide sufficient training data for the recommender system, we start the simulation's first ten steps by randomly presenting content in the feeds. The logistic regression was implemented using the *glm* package in R version 4.1.2.

Bot

The bot is represented as a stubborn agent that does not change its opinion but sticks to the same opinion throughout the simulation (Stewart et al. 2019; Karan et al. 2018; Yildiz et al. 2013). In different conditions, we manipulate the degree to which this opinion is extreme (i.e., the distance from the mean opinion of the agents).

We initialize the simulation with the following parameters: N(0.5, 0.2) and bot opinion = 0.8. This setting represents a situation where agents hold a moderate opinion and are not polarized. In probabilistic terms, the average population opinion is uncertain (i.e., around 0.5). The bot's opinion is more confident than the average opinion, but the mean difference between agents and the bot is not very large. On each timestep (starting from t = 10 onwards), agents are presented with a unique feed based on which they decide whether to engage and update their opinions. Once each agent has made a decision, the simulation proceeds to the next timestep. We repeat the procedure for t = 100 timesteps and r = 100 replications. We record each agent's opinion on each timestep and the cases where the bot gets recommended to an agent. We simulate two conditions, one where the bot is present and one where it is absent. We initialize both simulation conditions with the same random seeds, thereby producing virtually identical simulation conditions except for the presence of the bot. This manipulation allows for precise measurements regarding the influence of the bot on the network.

Results

Population-level influence of the bot on the average opinion

We start by looking at the population-level influence of the bot on agents' opinions. Figure 2a shows the mean opinion in the entire group over time for the two conditions (bot vs. no bot). Note that for the first t = 10 timesteps, there is no change in opinion since those trials serve as training samples for the recommender system and, therefore, present content in the feed randomly. From t = 10 onwards, we see a significant difference between the two conditions, with the bot shifting the average opinion of the population by an average of more than ten percentage points. This effect is also reflected by the average engagement levels in the population, as depicted in Fig. 2b. This effect holds across different initial opinion distributions (Additional file 1: Fig. S3) and different bot opinions (Fig. 5). From t = 10 onwards, we observe a significant jump in engagements. The recommender system becomes increasingly efficient at recommending content. The increase in engagement towards the end of the plot indicates that, as consensus emerges, interactions between agreeing agents lead to greater confidence (biased assimilation) and thus greater engagement. The plot shows an average between simulations where the average opinion converged to 0 and simulations where the average opinion converged to 1 (Additional file 1: Fig. S1).

We let the simulation run for 300 timesteps (Additional file 1: Fig. S4). Surprisingly, Figure S4b shows lower engagement when the bot is present (red line). In some simulations, the bot's pull toward its opinion (0.8) while agents try to converge toward 0 keeps agents in the uncertainty region (around 0.5) for longer (Additional file 1: Fig. S1d). In turn, this reduces average opinion extremity and engagement. However, the bot can also increase engagement, as seen in Additional file 1: Fig. S3. When agents start with moderate levels of consensus (0.4), the bot temporarily increases engagement as it holds the most confident opinion. As agents become more confident themselves, engagement in the bot-absent condition increases again.

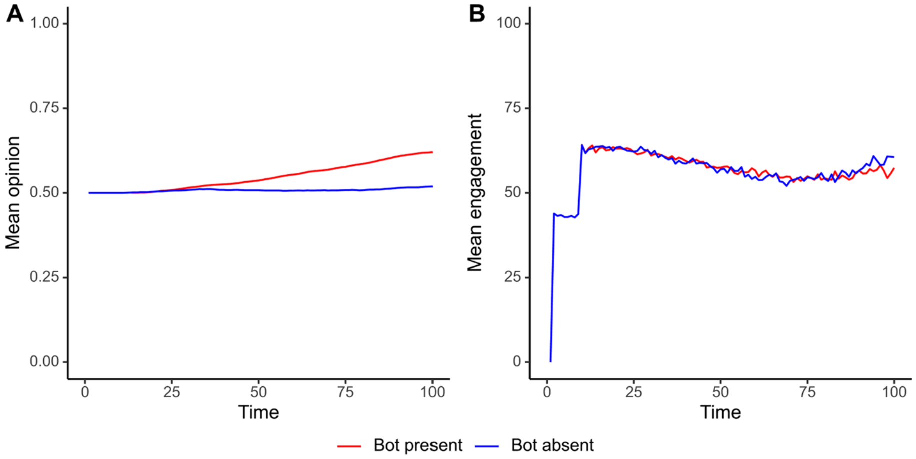

Fig. 2 Mean opinion and mean engagement in networks with and without bot influence. **(a)** Mean opinion of the agents over time. **(b)** Mean number of agents engaging with content in each timestep. Red: Treatment condition, the bot is part of the social network. Blue: Control condition, the bot is not part of the social network. A single bot can produce substantial changes in the mean opinion and mean engagement levels in the network

The magnitude of direct bot influence on the individual agents

So far, we have seen that a single bot can affect the population's average opinion and engagement levels. Here, we investigate the reasons underlying this effect more directly. Figure 3a shows the number of agents directly influenced by the bot on each timestep. By direct influence, we mean that the bot's content was recommended to an agent via the feed. The agent decided to engage with the bot's content (and thus update its private opinion based on its content). On average, 2 agents engage with and change their opinions after observing the bot on any timestep. Yet, we observe an average opinion change of 26% (Fig. 3b). Opinion change is the absolute difference between opinions pre and post-engagement. The finding of low engagement and opinion shift is consistent with the existing literature on the "minimal effect", which suggests that both online (Bail et al. 2020) and offline (Zaller 1992; Endres and Panagopoulos 2019; Kalla and Broockman 2018) efforts at persuasion are rarely effective. It suggests that direct influence (e.g., direct bot interaction or political advertisement and canvassing practices) is often ineffective at shifting population averages. Our finding only captures the direct influence from bot to agent but does not measure the bot's indirect influence by influencing an agent that will influence further agents. We believe that indirect influence may be more pervasive and more pronounced, especially in online contexts where recommender systems facilitate information spread. To measure this indirect n-th order influence of bots on agents, in the next paragraph, we compare the two simulation conditions (bot vs. no bot) while using the same random seed and holding all other conditions constant.

The individual-level shift in opinion as a result of direct or indirect bot influence

We then looked at the difference in opinion between the same agent across the two simulation conditions, holding all other aspects of the simulation constant (Fig. 4). Initializing the two simulations with identical parameters and random seed allowed us to isolate the effect of the bot. Estimating the within-agents effect improves our estimation of the

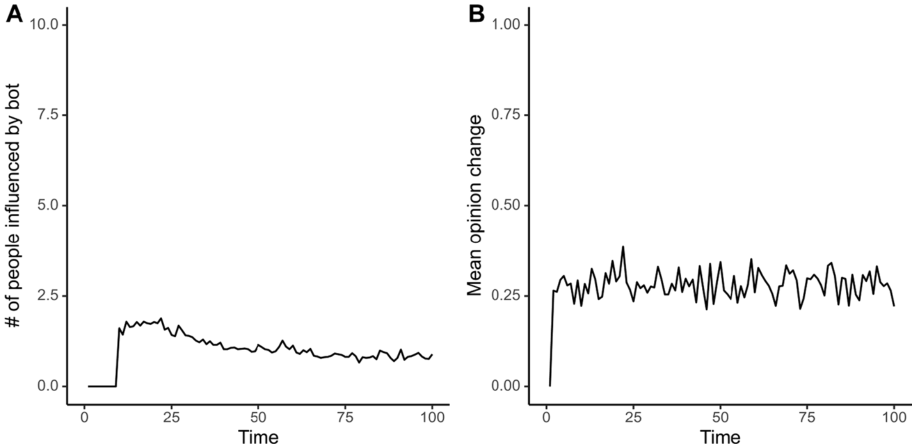

Fig. 3 Direct bot influence. **(a)** The average number of nodes influenced by the bot on each timestep. Influence is defined as when an agent is presented with content produced by the bot, engages with this content, and shifts its own opinion. **(b)** Mean opinion change for agents influenced by the bot on each timestep. A single bot can influence multiple people on each timestep and produce substantial opinion change

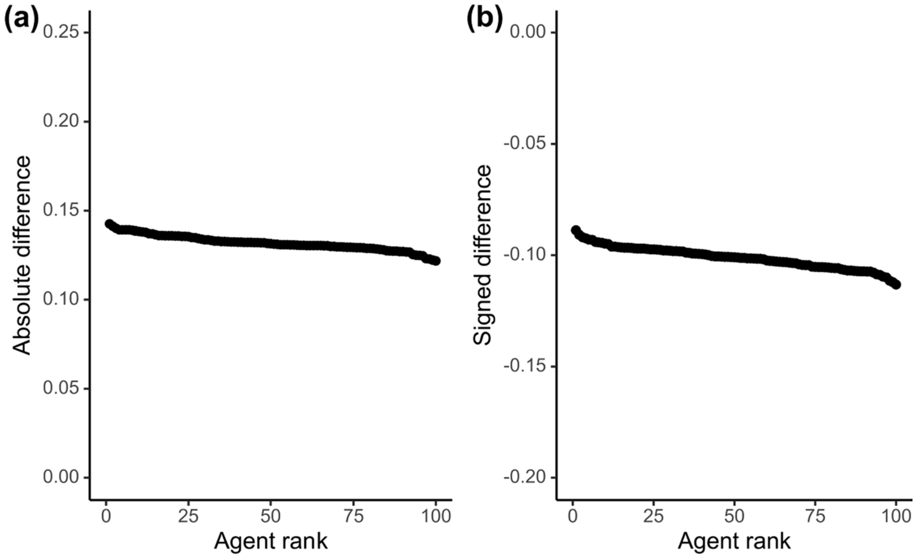

Fig. 4 Within-agents effect of bot across the two simulations. **(a)** The absolute difference between each agent's opinion at time *t* = *300* between the two simulation conditions (bot vs. no bot). This analysis measures the bot's total impact on the opinions of the same agents in the network. **(b)** The signed difference between each agent's opinion at time *t* = *300* between the two simulation conditions (bot vs. no bot). This analysis shows the direction of the social influence of the bot on individuals' opinions

bot effect. Differences between the two counterfactual worlds reflect the direct and indirect effects caused by introducing the bot. Notwithstanding the little direct influence (Fig. 3), we found that, compared to a counterfactual simulation, the bot had an indirect effect on the entire population, with the magnitude of influence on opinion varying considerably, from 11 to 15 percentage points (Fig. 4a). This effect is explained by agents observing other agents who might have interacted with the bot, leading to a trickle-down effect of the bot's opinion on other agents who might not have interacted with the bot. Figure 4b shows the signed difference between agents' opinions in the control and bot conditions ($d = T_{nobot} - T_{bot}$). Notice that most points are negative, indicating that nodes' opinions shifted toward the bot's opinion. Even though using the same random seed initialization does not ensure that all interaction events are the same, the fact that signed differences are systematically skewed towards the bot's opinion (Fig. 4a) shows that the bot influenced all individual agents. Our model shows that bots' influence is magnified when we account for indirect influence via the recommender system or other intermediary agents. This striking result indicates that a single bot can have a much more robust and lasting effect beyond individuals it directly interacts with. This finding seems to suggest that studies focusing only on direct influence (bots' influence on people they directly interacted with) might have underestimated the actual capacity of a bot to bias population opinion dynamics.

An exploration of the parameter space for bot opinion and population average

Finally, the above results assumed that the average opinion in the population is N(0.5, 0.2) and the bot opinion is 0.8. The results are specific to this parametrization of our model. To test the generalisability of our conclusion, we explore the sensitivity of our results to different values of agent and bot opinion. Figure 5 shows a heatmap where the

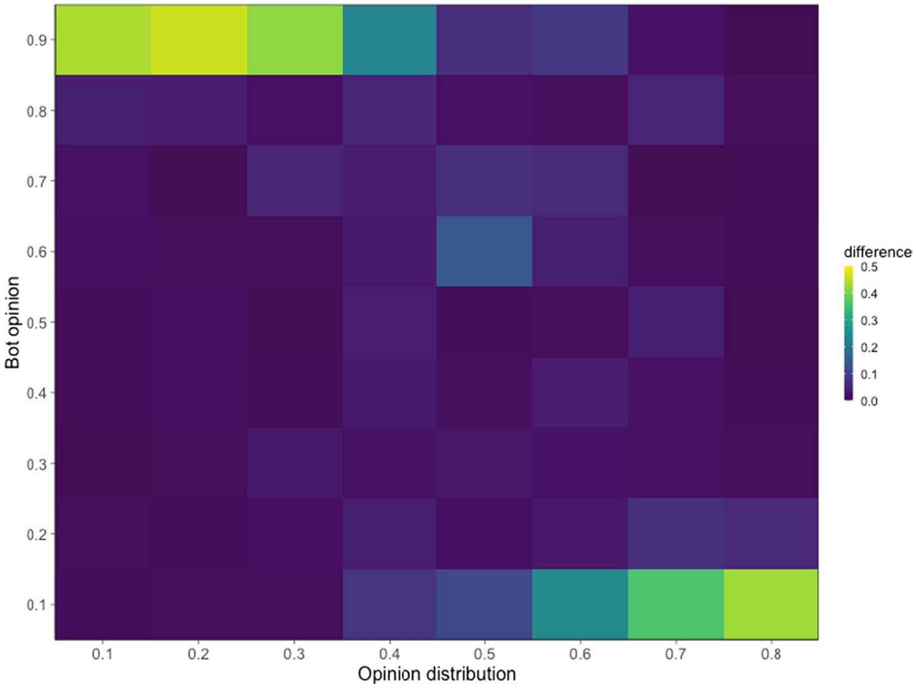

Fig. 5 Heatmap of different initial opinion distributions. The principal analysis assumed that the average opinion in the population is N(0.5, 0.2) and the bot opinion is 0.8. Here, we explore the sensitivity of our results to different values of agent and bot opinion. This figure shows a heatmap where the x-axis shows the bot's different opinion values and the y-axis shows different population average opinion values. The results remain qualitatively similar to those presented in the main text. The bot has a more substantial effect on the population when its opinion is more distant from the average opinion of the population

x-axis shows different values of the bot opinion and the y-axis shows the mean opinion in the population. The results remain qualitatively similar to those presented in the main text. The bot has a more substantial effect on the population when its opinion is more distant from the average opinion of the population. The results further support the conclusion that a bot (representing 1% of the total population) can have a disproportionate effect on population-level dynamics when we consider indirect influence.

Effect on the recommendation system's internal representations

We conducted a last set of analyses to detect differences in the model's internal representations. The recommender system used in this study was a simple logistic regression. The model was trained using agents' binary engagement history as a dependent variable and the absolute difference between the agent's public opinion at time $t-1$ and other agents' opinions as to the independent variable. After model fitting, the logistic's slope beta coefficients were used to compare the model's internal representations across conditions on the last timestep of each simulation. The distributions' mean values were negative in both conditions (No bot $= -4.86$; Bot $= -4.08$), suggesting that opinion distance between an agent and its neighbor negatively predicted engagement. This difference is expected given the influence of the similarity bias in Eq. 3. We then ran a Welch two-sample t-test on the distributions of beta coefficients on the last time step of the simulation across the 100 repetitions. We found a significant difference between the two conditions (t(173.11) = -3.96, $p < 0.001$), suggesting that the bot significantly reduced the negative effect of the opinion distance on engagement (Cf. Fig. 1b).

Discussion

This paper investigated the indirect influence that programmed media manipulators, such as bots, trolls, and zealots, can have on population opinion dynamics via recommender systems. We posited that even without direct exposure to bots' content, bots could influence population-wide content ranking by providing unduly training evidence to recommender algorithms. For instance, bots' greater activity, content engagement and production, and resilience to persuasion may contribute to bots skewing the training sample algorithms use to infer population preferences, averages, and typical content consumption patterns.

Using an opinion dynamics simulation on a 100-node network, we find that a single bot can substantially shift the mean opinion and engagement compared to a control condition without a bot. Even though only a minority of 'human' nodes (2%) directly engaged with the bot's content, the bot disproportionately affected the average shift in opinion observed in the population. Notably, virtually all nodes in the population were influenced by the bot presence, with opinion shifts ranging from 11 to 15 percentage points. The results are robust across different initialization parameters and different opinion update functions. We tested the effect of removing the bot after 40 timesteps. The number of agents directly influenced by the bot and the mean opinion change is shown in Additional file 1: Fig. S5. We find a sudden drop in direct influence after the bot is removed from the network. Comparing within-node opinion shifts across conditions, we find that all nodes showed shifted opinions at t = 100 (Additional file 1: Fig. S6). However, compared to the main results shown in Fig. 4, the magnitude of the shift is vastly reduced. Finally, we find that the internal representations of our simple recommendation model (beta logistic coefficients) were significantly impacted by the bot's presence. The coefficients were significantly larger in the bot-present than in the bot-absent condition.

These results would be unlikely if bots could influence human agents only via direct exposure. As bots represent only a minority of the population of agents (1% in our simulation), it is unlikely that they can interact with and directly influence all other agents. Our findings show that a simple recommender system (a logistic regression in our simulation) dramatically increases the influence of a bot on the population. Our first contribution is advancing the debate around bots' influence and media manipulation. Our study highlights a previously unexplored phenomenon and draws attention to a subtle yet potentially pervasive phenomenon.

Contrary to previous studies investigating social media bots, our work does not model direct interactions between bots and human agents (arguably representing a minority of interactions) but focuses on indirect effects via recommendation systems. Agents-based simulations have shown how bots can have a long-range, pervasive, and most critically stealthy influence on the network even without direct social influence (Keijzer and Mäs 2021). Our findings highlight that malicious agents, such as bots and trolls factories, can further increase their influence by infiltrating the internal representations of trained models tasked with content filtering. Our setup allows us to compare counterfactual worlds, thus strengthening causal inference. We initialized control (without-bot) and treatment (with-bot) simulations with the same parameters and random seed. Furthermore, effects on opinion shifts and engagement were calculated at the individual node level, thus measuring the effect of our treatment (bot presence) on the opinion dynamics and engagement of virtually identical 'human' agents.

Although our agent-based model provides valuable insights into machine-mediated information systems, it is limited by the ecological validity of simulation studies. Testing the same hypotheses in real-world contexts may be problematic due to the difficulty in conducting randomized control trials on social media platforms and the proprietary nature of natural recommender systems. Although it may be challenging to study these systems, researchers have recently successfully inferred the hidden mechanisms underlying several proprietary algorithms by systematically prompting them (Ali et al. 2019; Hannak et al. 2013; Robertson et al. 2018). Furthermore, real-world opinion dynamics are arguably more complex than the simple simulated world. Complex dynamics may be elicited by bots operating on media platforms not captured by our simulation (Mønsted et al. 2017). Nevertheless, our findings show that one component of such a complex network of influence may occur not via direct interactions between nodes but by subtly skewing the recommender systems' training set. We invite future researchers in computational social science to investigate this indirect causal pathway linking bots and human social media accounts through recommender systems (Fig. 1a).

The second contribution of our paper lies in using a Bayesian update function that bridges classical opinion dynamics findings (e.g., opinion averaging, biased assimilation) with behavioral observations of opinion change from psychology and cognitive science. Although several other opinion models exist that reproduce these dynamics (Flache et al. 2017), not many use parameters that can be associated with explicit psychological constructs. Several classical models assume that opinion change results from a linear combination of neighboring nodes' observed opinions, such as averages and weighted means (DeGroot 1974; Friedkin and Johnsen 1990; Deffuant et al. 2000). However, experimental evidence suggests that non-linear multiplicative dynamics often govern opinion change (Bail et al. 2018; Pescetelli and Yeung 2020b; Pescetelli et al. 2016; Moscovici and Zavalloni 1969). Here, we used a Bayesian opinion update model that captures the dynamics of belief conviction, uncertainty, and probabilistic judgments (Pescetelli and Yeung 2020a, b; Harris et al. 2016). We selected this Bayesian update model because opinion shifts can be interpreted as shifts in confidence estimates. We argue that this opinion update model has several advantages. It represents opinions in the well-known language of probability. It can be seen as a normative rational model of opinion update (Harris et al. 2016). Using the Bayesian theorem to model belief updates allows us to quantify a best-case scenario, namely, the impact of bots if people were rational.

Similarly, it represents opinion dynamics as shifts in subjective probability estimates (e.g., the probability of being "right"), given the perceived evidence from other agents' opinions. More confident agents (i.e., with a stronger prior) are less influenced by other agents and more influential than uncertain agents. This approach explains several social phenomena (e.g., polarization, hyper-partisanship, escalation, and averaging) without requiring arbitrary free parameters. Encounters with agreeing agents tend to increase one's belief conviction (biased assimilation), while encounters with disagreeing agents increase uncertainty (assimilative social influence). Modeling trust and perceived expertise could potentially explain recent evidence suggesting that disagreement may sometimes entrench people further in their decisions (Bail et al. 2018; Harris et al. 2016). Bayesian update represents opinion escalation dynamics better than linear aggregation models. While linear updates may better model estimation tasks, Bayesian updates may better represent belief convictions and partisan affiliations, i.e., cases where interaction with like-minded individuals makes people more extreme.

We also acknowledge that our findings are specific to our choice of parameters and may not generalize well to other scenarios. Contrary to previous work, we do not explore the effect of different network sizes and structures on the effect under consideration. We acknowledge this as a limitation of our study and invite future studies to test the robustness of our results to alternative network architectures. Our study used a simple logistic model to predict engagement scores to provide recommended content. One limitation is that existing recommender systems are more complex than the simple logistic regression employed in this study. For instance, recommender systems can consider many more features and provide greater personalization thanks to highly granular information about users and user similarity. However, the effects highlighted in our findings are likely to affect, at least to some degree, any content filtering algorithm trying to extrapolate the behavior of one user to another. We speculate that more complex recommendation systems may still be affected by the same dynamics highlighted here as long as they use population averages to predict individual preferences. By biasing the estimation of a population mean, algorithmic agents can change the model's expectation for a given cluster of users or the whole population. Extrapolating a user's behavior to another represents the standard in many recommender systems (Ricci et al. 2011), e.g., collaborative filtering algorithms (Das et al. 2007; Koren and Bell 2015; Ricci et al. 2011). Recently, researchers have shown that individual social influence can be affected by an individual's position in the population distribution and similarity with others (Analytis et al. 2018, 2020). Thus we may also expect to observe our findings with more realistic content recommendation algorithms.

Furthermore, the complexity of realistic recommender systems makes the findings of this work even more significant. Indeed, our findings suggest that bots and troll factories' influence may be subtle but highly pervasive. The opacity and complexity of natural recommender systems suggest that such pervasive effects may continue to operate undetected. The potential consequences are difficult to imagine but should prompt further investigation.

Finally, some of our findings may depend on the specific opinion update model that we used here. We explore in the Supplementary material different values of alpha in Eq. 3 (Additional file 1: Fig. S2) and different average opinion distributions (Additional file 1: Fig. S3). A caveat in our simulation pertains to the modeling opinion and opinion change and operationalizing bots as stubborn agents (Hunter and Zaman 2018; Yildiz et al. 2013). In the present study, we represent beliefs along a single opinion dimension. People's beliefs outside the lab are often more complex and multifaceted than our model.

Nevertheless, using beliefs spanning a single dimension represents a necessary first step in many opinion dynamic models and advice-taking paradigms (Bonaccio and Dalal 2006; Deffuant et al. 2000; Flache et al. 2017; Friedkin and Johnsen 1990). Similarly, political polarisation and beliefs across several domains, especially divisive issues, may be well described by a single belief dimension (Navajas et al. 2019). Future modeling efforts could generalize our findings to multi-dimensional attitude spaces.

We suggest possible ways to reduce the risk of public opinion manipulation. First, improving the detection and removal of automated accounts can reduce bots' impact on population-wide behaviors (Additional file 1: Fig. S5). However, uniquely relying on this strategy is not sustainable in the long run as automated detection becomes outdated and new and more sophisticated bots are developed. Detection and removal tend to be more effective with relatively simple bots, thus creating a selective pressure for bots to develop more human-like features that are more likely to remain undetected (like the Red Queen hypothesis in biology). A more valuable strategy might be to regulate recommender and filtering algorithms to make them more transparent. Knowledge of the features used to make content recommendations can help academics and practitioners to monitor features that ill-willing entities can exploit. Open auditing of recommender systems and open-source software can go a long way in preventing some types of bots from doing harm and minimizing algorithmic tampering with public opinions.

Conclusions

In this paper, we explored the hypothesis that algorithmic agents may have stealthy, undue influence on online social networks by biasing the internal representations of recommender systems. Bots' more extreme views, greater activity frequency, and content generation might distort content recommendation for the entire network. Researchers and watchdogs should be aware of the indirect causal pathways of bot influence.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s41109-022-00488-6.

Additional file 1: Supplementary Figures S1–S6.

Acknowledgements
We thank two anonymous reviewers for their time, feedback, and detailed, invaluable insights.

Author contributions

NP and MC conceptualized the study. NP and DB designed the formal analysis. DB curated and visualized the data and ran the formal analyses. NP and MC acquired funding. All authors were responsible for the investigation and methodology. NP curated the project administration. MC supervised the project. NP and DB wrote the original draft. All authors reviewed and edited the final draft.

Funding
The study was funded by the Max Planck Institute for Human Development. D.B. was partly funded by a research grant from the Institute of Psychology at the Chinese Academy of Sciences.

Availability of data and material

Barkoczi, D., & Pescetelli, N. (2021, August 21). Indirect causal influence of social bots through a simple recommendation algorithm. Retrieved from osf.io/7s83x.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 28 February 2022 Accepted: 13 June 2022
Published online: 06 July 2022

References

Aldayel A, Magdy W (2022) Characterizing the role of bots in polarized stance on social media. Soc Netw Anal Min 12(1):30

Ali M, Sapiezynski P, Bogen M, Korolova A, Mislove A, Rieke A (2019) Discrimination through optimization: how Facebook's Ad delivery can lead to biased outcomes. Proc ACM Hum Comput Interact 199(3):1–30

Allen J, Howland B, Mobius M, Rothschild D, Watts DJ (2020) Evaluating the fake news problem at the scale of the information ecosystem. Sci Adv 6(14):eaay3539

Analytis PP, Barkoczi D, Herzog SM (2018) Social learning strategies for matters of taste. Nat Hum Behav 2(6):415–424

Analytis PP, Barkoczi D, Lorenz-Spreen P, Herzog S (2020) The structure of social influence in recommender networks. In: Proceedings of the web conference 2020, 2655–61. WWW '20. New York, NY, USA: association for computing machinery

Aral S, Eckles D (2019) Protecting elections from social media manipulation. Science 365(6456):858–861

Bail CA, Argyle LP, Brown TW, Bumpus JP, Haohan Chen MB, Hunzaker F, Lee J, Mann M, Merhout F, Volfovsky A (2018) Exposure to opposing views on social media can increase political polarization. Proc Natl Acad Sci 115(37):9216–9221

Bail CA, Guay B, Maloney E, Aidan Combs D, Hillygus S, Merhout F, Freelon D, Volfovsky A (2020) Assessing the Russian internet research agency's impact on the political attitudes and behaviors of American twitter users in late 2017. Proc Natl Acad Sci 117(1):243–250

Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news and opinion on facebook. Science. https://science.sciencemag.org/content/348/6239/1130.abstract?casa_token=93SGKMyFHO4AAAAA:NLLn7cnwU-dniTFvSJ5wC7XUJ30w5AFKxPLDLfWyijbh8Z-NWk0vsYB2zgXtq7EyGRLUhHdYX2fBfQ

Becker J, Brackbill D, Centola D (2017) Network dynamics of social influence in the wisdom of crowds. Proc Natl Acad Sci USA 114(26):E5070–E5076

Beskow DM, Carley KM (2018) Bot conversations are different: leveraging network metrics for bot detection in twitter. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), 825–32. ieeexplore.ieee.org

Bessi A and Ferrara E (2016) Social bots distort the 2016 US presidential election online discussion. SSRN 21(11). https://ssrn.com/abstract=2982233

Bonaccio S, Dalal RS (2006) Advice taking and decision-making: an integrative literature review, and implications for the organizational sciences. Organ Behav Hum Decis Process 101(2):127–151

Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, Quinn SC, Dredze M (2018) Weaponized health communication: twitter bots and russian trolls amplify the vaccine debate. Am J Public Health 108(10):1378–1384

Carley KM (2020) Social cybersecurity: an emerging science. Comput Math Organ Theory 26(4):365–381

Dandekar P, Goel A, Lee DT (2013) Biased assimilation, homophily, and the dynamics of polarization. Proc Natl Acad Sci USA 110(15):5791–5796

Das A, Datar M, Garg A and Rajaram S (2007) Google news personalization: scalable online collaborative filtering. In: Proc of the 16th Int Conf on World Wide Web, 271–80

Deffuant G, Neau D, Amblard F, Weisbuch G (2000) Mixing beliefs among interacting agents. Adv Compl Syst A Multidis J 03(4):87–98

DeGroot MH (1974) Reaching a consensus. J Am Stat Assoc 69(345):118

Edelson L, Nguyen M-K, Goldstein I, Goga O, Mccoy D, et al. Understanding engagement with U.S. (mis)information news sources on Facebook. IMC '21: ACM Internet Measurement Conference, Nov 2021, Virtual Event, France. pp. 444–463. Link available here: https://hal.archives-ouvertes.fr/hal-03440083/file/news-interactions-imc2021.pdf

Endres K, Panagopoulos C (2019) Cross-pressure and voting behavior: evidence from randomized experiments. The J Polit 81(3):1090–1095

Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. Commun ACM 59(7):96–104

Ferreira LN, Hong I, Rutherford A, Cebrian M (2021) The small-world network of global protests. Sci Rep 11(1):19215

Festinger L, Carlsmith JM (1959) Cognitive consequences of forced compliance. J Abnorm Psychol 58(2):203–210

Flache A, Mäs M, Feliciani T, Chattoe-Brown E, Deffuant G, Huet S, and Lorenz J (2017) Models of social influence: towards the next frontiers. J Artif Soc Soc Simul. https://doi.org/10.18564/jasss.3521.

Fleming SM, Daw ND (2017) Self-evaluation of decision performance: a general bayesian framework for metacognitive computation. Psychol Rev 124(1):1–59

Fleming SM, van der Putten EJ, Daw ND (2018) Neural mediators of changes of mind about perceptual decisions. Nat Neurosci 21(4):617–624

Friedkin NE, Johnsen EC (1990) Social influence and opinions. The J Math Sociol 15(3–4):193–206

Friedkin NE and Johnsen EC (2011) Social influence network theory: a sociological examination of small group dynamics. Cambridge University Press

González-Bailón S, De Domenico M (2021) Bots are less central than verified accounts during contentious political events. Proc Natl Acad Sci USA. https://doi.org/10.1073/pnas.2013443118

Guess A, Nagler J, Tucker J (2019) Less than you think: prevalence and predictors of fake news dissemination on facebook. Sci Adv 5(1):4586

Hahn U, Oaksford M (2006) A Bayesian approach to informal argument fallacies. Synthese 152(2):207–236

Hahn U, Oaksford M (2007) The rationality of informal argumentation: a Bayesian approach to reasoning fallacies. Psychol Rev 114(3):704–732

Hannak A, Sapiezynski P, Kakhki AM, Krishnamurthy B, Lazer D, Mislove A, Wilson C (2013) Measuring personalization of web search. In: Proceedings of the 22nd international conference on world wide web, 527–38. WWW '13. New York, NY, USA: Association for Computing Machinery

Harris AJL, Hahn U, Madsen JK, Hsu AS (2016) The appeal to expert opinion: quantitative support for a Bayesian network approach. Cogn Sci 40(6):1496–1533

Hegselmann R, Krause U (2015) Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: a simple unifying model. Netw Heterog Media 10(3):477–509

Howard P (2018) How political campaigns weaponize social media bots. IEEE Spectrum Oct

Hunter SD, and Zaman T (2018) Optimizing opinions with stubborn agents under time-varying dynamics. arXiv [cs.SI]. arXiv. http://arxiv.org/abs/1806.11253.

Hurtado S, Ray P and Marculescu R (2019) Bot detection in reddit political discussion. In: Proceedings of the fourth international workshop on social sensing, 30–35. SocialSense'19. New York, NY, USA: Association for Computing Machinery

Kakutani M (2019) The death of truth. Tim Duggan Books

Kalla JL, Broockman DE (2018) The minimal persuasive effects of campaign contact in general elections: evidence from 49 field experiments. The Am Polit Sci Rev 112(1):148–166

Karan N, Salimi F, Chakraborty S (2018) Effect of zealots on the opinion dynamics of rational agents with bounded confidence. Acta Phys Pol, B 49(1):73

Keijzer MA, Mäs M (2021) The strength of weak bots. Online Social Networks and Media 21(January):100106

Koren Y and Bell R (2015) Advances in collaborative filtering. In: Recommender systems handbook, edited by Francesco Ricci, Lior Rokach, and Bracha Shapira, 77–118. Boston, MA: Springer US

Lazer D (2020) Studying human attention on the internet. Proceedings of the National Academy of Sciences of the United States of America

Lazer D, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Metzger MJ et al (2018) The science of fake news. Science 359(6380):1094–1096

Ledford H (2020) Social scientists battle bots to glean insights from online chatter. Nature 578(7793):17–17

Lerman K, Yan X, Xin-Zeng Wu (2016) The 'Majority Illusion' in social networks. PLoS ONE 11(2):e0147617

Linvill DL, Warren PL (2018) Troll factories: the internet research agency and state-sponsored agenda building. Resource Centre on Media Freedom in Europe. https://scholar.google.com/scholar?hl=en&q=Brandon+C+Boatwright%2C+Darren+L+Linvill%2C+and+Patrick+L+Warren.+2018.+Troll+factories%3A+The+internet+research+agency+and+statesponsored+agenda+building.+Resource+Centre+on+Media+Freedom+in+Europe+%282018%29

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nat Neurosci 9(11):1432–1438

Mäs M, Flache A (2013) Differentiation without distancing. Explaining Bi-polarization of opinions without negative influence. PLoS ONE 8(11):e74516

Mønsted B, Sapieżyński P, Ferrara E, Lehmann S (2017) Evidence of complex contagion of information in social media: an experiment using twitter bots. PLoS ONE 12(9):e0184148

Moscovici S, Zavalloni M (1969) The group as a polarizer of attitudes. J Pers Soc Psychol 12(2):125–135

Muller M (2012) Lurking as personal trait or situational disposition: lurking and contributing in enterprise social media. In: Proceedings of the ACM 2012 conference on computer supported cooperative work, 253–56. CSCW '12. New York, NY, USA: Association for Computing Machinery

Navajas J, Heduan FÁ, Garrido JM, Gonzalez PA, Garbulsky G, Ariely D, Sigman M (2019) Reaching consensus in polarized moral debates. Curr Biol: CB 29(23):4124–29.e6

Paul, Christopher, and Miriam Matthews. 2016. "The Russian 'firehose of Falsehood' Propaganda Model." *Rand Corporation*, 2–7.

Penrod SD, Cutler BL (1995) Witness confidence and witness accuracy: assessing their forensic relation. Psychol, Publ Pol, Law: an off Law Rev Univ Arizona College Law Univf Miami School Law 1:817–845

Pescetelli N, Yeung N (2020a) The role of decision confidence in advice-taking and trust formation. J Exp Psychol Gen. https://doi.org/10.1037/xge0000960

Pescetelli N, Yeung N (2020b) The effects of recursive communication dynamics on belief updating. Proc Royal Soc b: Biol Sci 287(1931):20200025

Pescetelli N, Rees G, Bahrami B (2016) The perceptual and social components of metacognition. J Exp Psychol Gen 145(8):949–965

Price PC, Stone ER (2004) Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic. J Behav Decis Mak 17(1):39–57

Rader CA, Larrick RP, Soll JB (2017) Advice as a form of social influence: informational motives and the consequences for accuracy. Soc Pers Psychol Compass 11(8):e12329

Resulaj A, Kiani R, Wolpert DM, Shadlen MN (2009) Changes of mind in decision-making. Nature 461:263–266

Ricci F, Rokach L and Shapira B (2011) Introduction to recommender systems handbook. In: Recommender systems handbook, edited by Ricci F, Rokach L, Shapira B, and Kantor PB, 1–35. Boston, MA: Springer US

Robertson RE, Lazer D, and Wilson C (2018) Auditing the personalization and composition of politically-related search engine results pages. In: Proceedings of the 2018 world wide web conference on World Wide Web - WWW '18, 955–65. New York, New York, USA: ACM Press

Shao C, Ciampaglia GL, Varol O, Yang K-C, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. Nat Commun 9(1):4787

Sherif CW, Sherif MS, Nebergall RE (1965) Attitude and attitude change. W.B. Saunders Company, Philadelphia

Sniezek JA, Van Swol LM (2001) Trust, confidence, and expertise in a judge-advisor system. Organ Behav Hum Decis Process 84(2):288–307

Soll JB, Mannes AE (2011) Judgmental aggregation strategies depend on whether the self is involved. Int J Forecast 27(1):81–102

Stella M, Ferrara E, De Domenico M (2018) Bots increase exposure to negative and inflammatory content in online social systems. Proc Natl Acad Sci USA 115(49):12435–12440

Stewart LG, Arif A, and Starbird K (2018) Examining trolls and polarization with a retweet network. In: Proc ACM WSDM, workshop on misinformation and misbehavior mining on the web. http://faculty.washington.edu/kstarbi/examining-trolls-polarization.pdf

Stewart AJ, Mosleh M, Diakonova M, Arechar AA, Rand DG, Plotkin JB (2019) Information gerrymandering and undemocratic decisions. Nature 573(7772):117–121

Sun Z, Müller D (2013) A framework for modeling payments for ecosystem services with agent-based models, Bayesian belief networks and opinion dynamics models. Environ Model Softw 45(July):15–28

Sunstein CR (2018) #Republic: divided democracy in the age of social media. Princeton University Press

Tucker JA, Guess A, Barbera P, Vaccari Cr, Siegel A, Sanovich S, Stukal D, Nyhan B (2018) Social media, political polarization, and political disinformation: a review of the scientific literature. SSRN J. https://doi.org/10.2139/ssrn.3144139

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151

Whittaker J, Looney S, Reed A, Votta F (2021) Recommender systems and the amplification of extremist content. Internet Policy Rev. https://doi.org/10.14763/2021.2.1565

Yanardag P, Cebrian M, Rahwan I (2021) Shelley: a crowd-sourced collaborative horror writer. Creat Cognit. https://doi.org/10.1145/3450741.3465251

Yaniv I (2004) Receiving other people's advice: influence and benefit. Organ Behav Hum Decis Process 93(1):1–13

Yildiz E, Ozdaglar A, Acemoglu D, Saberi A, Scaglione A (2013) Binary opinion dynamics with stubborn agents. ACM Trans Econ Comput 19,1(4):1–30

Zaller JR (1992) The nature and origins of mass opinion. Cambridge University Press

Publisher's Note

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

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com