

# Bot Versus Humans: Who Can Challenge Corporate Hypocrisy on Social Media?

Social Media + Society  
October–December 2024: 1–13  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20563051241292578  
journals.sagepub.com/home/sms



Serena Armstrong<sup>1</sup> , Caitlin Neal<sup>1</sup>, Rongwei Tang<sup>1</sup>,  
Hyejoon Rim<sup>2</sup>, and Emily K. Vraga<sup>1</sup>

## Abstract

Social media offer opportunities for companies to promote their image, but companies online also risk being denounced if their actions do not align with their words. The rise of social media bots amplifies this risk, as it becomes possible to automate such efforts to highlight corporate hypocrisy. Our experimental survey demonstrated that bots and human actors who confront a corporation touting their commitment to equality by calling out organizational pay gaps damage perceptions of the corporation, heighten anger toward them, and ultimately can elicit boycott intentions. These hypocrisy challenges are equally effective when they come from bots and user accounts. Challenges to hypocritical behavior on social media are consequential and require further exploration.

## Keywords

social media, bots, hypocrisy, corporate communication

In March of 2021, a bot account on Twitter (now X)<sup>1</sup> was created by Francesca Lawson and Ali Fensome to challenge corporations tweeting about their support for International Women’s Day. This account, *@PayGapApp*, automatically responds to companies listed on the UK government’s Gender Pay Gap Service website with their actual median gender pay differences, as all companies with 250 employees or more in the UK are mandated to report their gender pay gap (Gender Pay Gap Bot—About, n.d.). By the time of this study in March of 2023, this bot account had amassed over 247,000 followers, posted over 11,400 times on Twitter, and had a website dedicated to explaining the process behind its messages. Companies have responded to the bot in various ways, including blocking the account, removing their tweet from the feed, or deleting their initial tweet from public view (Breen, 2022).

The question of how bots are affecting the social media landscape is of great importance. It may be common to think that the social media landscape is composed of human users, but in the past decade, there has been an uptick in bot accounts across platforms (Ferrara et al., 2016; Hagen et al., 2022). These bots are defined as “automatic or semi-automatic computer programs that mimic humans and/or human behavior” (Wagner et al., 2012) and can fulfill various purposes, from the innocent like posting weather reports (U.S. Department of Commerce, n.d.) to the nefarious like spreading misinformation (Himelein-Wachowiak et al., 2021) or sowing discord (Broniatowski et al., 2018).

In this study, we consider the effects of a gender pay bot, calling attention to gaps between what companies profess to value on social media (i.e., support for International Women’s Day) and their actual practices (i.e., a sustained gender wage gap). We compare the ability of a bot to call out corporate hypocrisy with a human actor. In doing so, we first consider how companies use social media as a site for reputation management and describe the literature about how people respond to corporate hypocrisy. We then elaborate on how bot accounts are changing social media, and whether the source of information—specifically, whether a challenge to corporate hypocrisy comes from a bot versus a human user—may influence people’s responses to such efforts. Our experimental survey study suggests that bots and human accounts are equally effective in challenging corporate hypocrisy, damaging corporate reputation, provoking anger towards the corporation, and increasing people’s boycott intentions. This highlights the potential for bots to reshape the social media landscape as a source of interaction and confrontation, rather than a space for broadcasting corporate good deeds uncontested.

<sup>1</sup>University of Minnesota, USA

<sup>2</sup>The Chinese University of Hong Kong, Hong Kong

## Corresponding Author:

Emily K. Vraga, Hubbard School of Journalism and Mass Communication,  
University of Minnesota, 111 Murphy Hall, 206 Church Street SE,  
Minneapolis, MN 55455, USA.  
Email: ekvraga@umn.edu



## Social Media as Space for Corporate Reputation Management

Social media, as a platform for an organization to communicate with its stakeholders, plays a key role in engaging with them and managing its reputation (Briones et al., 2011). We define social media as an interactive, internet-based channel of mass personal communication that allows for two-way interaction and derives its value primarily from user-generated content (Carr & Hayes, 2014; Kent, 2010). Social media platforms enable organizations to bypass traditional communication channels and institutional media to interact directly with the public (Entman & Usher, 2018). This “flattening” of hierarchies of information control makes it possible for organizations to reach key stakeholders—but also for stakeholders to reach them.

Social media presents a potent channel for an organization’s reputation management efforts, but managing and controlling the message can be difficult (Macnamara & Zerfass, 2012). Corporations can burnish their reputation on social media by enhancing their ability to collect information, strengthen corporate identity, monitor public opinion, and engage with key publics (Y. Wang, 2015). They can also use social media to publicize their corporate social responsibility (CSR) practices, broadly defined as a corporation’s self-governing investment and involvement in its resources to support societal goals (Frederick, 1994).

But user-centered social media platforms are unlike traditional corporate-controlled media in that individual users become media gatekeepers and content-creators who decide how organization-related content is used and shared. This arrangement transfers “the power to define corporate images from corporate communicators to stakeholders’ online networks” (Y. Wang, 2015, p. 9). Therefore, the same tools corporations use on social media are also available to empower activist groups in commanding an organization’s attention (Coombs, 1998).

Research in crisis communication shows that information about an organization disseminated by a third party on social media activates publics’ emotions such as anger, contempt, and disgust (Jin et al., 2014), and that these emotions are contagious (Kowalski, 1996). A single social media user that is dissatisfied with an organization or challenges them publicly can therefore set off a firestorm or pile-on (Einwiller & Steilen, 2015). While this can constitute a viable threat for the organization that is charged with irresponsible or unethical behavior (Coombs & Holladay, 2012), it also offers an opportunity for social media users to have a real impact in altering the organization’s actions. We examine this specifically in the context of an account calling attention to potential hypocrisy between a corporation’s public message (supporting women) and its practices (gender pay inequality).

## Targeting Hypocrisy on Social Media

Corporate hypocrisy is defined as “the belief that a firm claims to be something that it is not” (Wagner et al., 2009),

which should apply to corporations claiming to support gender equality while also systematically paying women less than men. The mere use of CSR as a branding and marketing tool can be viewed as self-serving, potentially inducing the perception of corporate hypocrisy and backfiring on a firm’s reputation, if the firm does not live up to its claims or makes empty promises (Bae & Cameron, 2006; Wagner et al., 2009; Yoon et al., 2006). As such, CSR efforts are particularly susceptible to social media attacks designed to engender hypocrisy perceptions. Past research showed that perceptions of hypocritical action from corporations produce lower perceptions of trust and credibility (Bhatti et al., 2013; Cooper et al., 2019; von Sikorski & Herbst, 2020), higher negative emotions (Simonovits et al., 2022; von Sikorski & Herbst, 2020), and more intentions to boycott from consumers (Wagner et al., 2009). Because social reinforcement acts as the engine powering social media (Aral, 2020; Singh & Singh, 2021) and boycott intentions are often powerful driven by social norms (Delistavrou et al., 2020), posts that stand to damage a firm’s reputation can be critical to our understanding of how consumers are motivated to boycott—and how they motivate others to boycott—in an online environment.

Building from past research, we explore the mechanisms by which social media challenges designed to elicit perceptions of corporate hypocrisy affect corporate credibility and boycott intentions. This question has practical and theoretical value, given the ways in which algorithmically-reinforced social pressures on social media may heighten corporations’ vulnerability to reputation attacks, also referred to as paracrises in social-mediated crisis communication (Coombs & Holladay, 2012). We consider two potential pathways by which challenges may impact corporations. First, we test whether perceptions of corporate hypocrisy explain the harms to corporate credibility and boycott intentions that previous research has uncovered (Klein et al., 2004). We explicitly contrast this pathway with an alternative explanation: that challenges cause moral anger toward the company, and it is this anger (in addition to or in place of) that explains attitudes and behaviors.

Moral anger is characterized as an emotional response arising from the perceived violation of a moral norm (Lindebaum & Geddes, 2016). Those who feel angry about the target company have lower evaluations of the company (Grappi et al., 2013; Kim & Cameron, 2011; Xie & Bagozzi, 2019) and perceive the spokespeople as less trustworthy and less favorable (Clementson & Xie, 2020). Importantly, moral anger can trigger behaviors aimed at correcting the situation, even when these involve personal behavior such as boycott behaviors (Braunsberger & Buckler, 2011; Hino, 2023; Klein et al., 2002). In addition, anger could serve as a mediator between blame attribution and boycott intentions (Shim et al., 2021). Therefore, we compare these two potential mechanisms—perceptions of corporate hypocrisy and moral anger toward the company—for explaining assessments of corporate credibility and boycott intentions.

## Bots as Emerging Actors on Social Media

Recently, more scholarly and public attention has been focused on the growing role bots play on social media platforms (Assenmacher et al., 2020; Gorwa & Guilbeault, 2020). Bots are automated accounts based on algorithms to generate content and interact with other users (Howard & Kollanyi, 2016). Although bots attempt to mimic human users (Oberer et al., 2019), bots' interactions on social media, such as sharing, sharing, and responding, often lack the responsiveness and variability that are inherent in human interactions, and their activities are more predictable (Cai et al., 2022; Chu et al., 2012). In addition, compared with human users, content generated by bots is less likely to be aligned with the overall mood of an event (e.g., bots share negative posts for a positive event) (Kusen & Strembeck, 2019).

The presence and activity of bots on social media can have significant societal implications. Bots have been criticized for manipulating public opinion (Weng & Lin, 2022), as a weapon for hate speech and m/disinformation (Hameleers et al., 2022; Shao et al., 2018; Uyheng et al., 2022; Vosoughi et al., 2018), and for influencing the public agenda (Zhang et al., 2024). Even when bots represent only a relatively small percentage of discussion participants, they can activate a spiral of silence (Cheng et al., 2020; Ross et al., 2019), wherein people avoid speaking out of fear of being isolated when they perceive themselves to be in the minority (Noelle-Neumann, 1974).

Although bots are often perceived negatively for their roles in spreading misinformation and manipulating public discourse, they can also have positive impacts. For example, they can also be used as tools for good, such as supporting online activism (Chen et al., 2021; Savage et al., 2016), responding to reduce racial harassment (Munger, 2017), or combatting problematic information on Wikipedia (Jiang & Vetter, 2020; Zheng et al., 2019). In this study, the bot we examined has this salutary intention: to broadcast the disconnect between a corporation's public words (declaring support on International Women's Day) and behaviors (pay inequality).

## Bots Versus Humans as Sources

Given the increasingly important role bots play on social media (Cheng et al., 2020; Ross et al., 2019; Zhang et al., 2024), we need to further understand how people respond to algorithms-based bots and human users as information sources. Existing literature presents mixed evidence of how people perceive bots and human actors, with most existing research focusing on the perceived credibility of algorithmic versus human-created news (Graefe & Bohlken, 2020; Jia & Liu, 2021; Liu & Wei, 2019; Tandoc et al., 2020; Waddell, 2018; Wölker & Powell, 2021). Some researchers

found that there was no significant difference in perceived credibility between a bot and a human user for both Twitter pages (Edwards et al., 2014) and news sources (Tandoc et al., 2020; Wölker & Powell, 2021). In contrast, other researchers found people rated content attributed to automated algorithms as either more objective (Liu & Wei, 2019) or less credible (Jia & Liu, 2021; Waddell, 2018) than human authors. One meta-analysis found news purportedly written by a human source would make the participants perceive the news content as more credible than news written by an algorithm (Graefe & Bohlken, 2020). However, this question of bot versus human sources has not been studied in the context of challenges to corporate hypocrisy in CSR efforts online.

Bots, due to their algorithm-based machine nature, might be perceived as less human-like, potentially more impartial, and could elicit less emotional engagement than human authors (Liu & Wei, 2019; Wischnewski et al., 2022). However, it is also possible that no significant difference exists between social media bots and human authors, as research also indicates that bots were perceived as equally credible, competent, fair, and objective compared with human authors, and there were no different interactional intentions between content attributed to humans or bots (Edwards et al., 2014). Therefore, we contribute to the literature comparing bots and human sources by extending it to a new space: in their ability to challenge corporate hypocrisy on social media.

## Research Questions and Hypotheses

This study tests several specific hypotheses and research questions based on the existing literature. Building from previous research on corporate hypocrisy (Klein et al., 2004), we propose our first hypothesis:

*Hypothesis 1 (H1). A corporation that is challenged for hypocrisy will have (a) higher perceived corporate hypocrisy, (b) lower corporate credibility ratings, (c) higher anger toward the company, and (d) higher boycott intentions than in the control condition.*

We go beyond existing research in our next hypotheses to examine how these challenges to corporate hypocrisy are uniquely responsive to the affordances of social media. In particular, we ask whether an (unknown) user or a clearly labeled bot will differ in public responses to challenges to corporate hypocrisy. Given the contradictory findings in existing literature in terms of perceptions of bots versus human actors on social media (Jia & Liu, 2021; Liu & Wei, 2019; Waddell, 2018), we explore whether these two actors differ in their effects on response toward the corporation (RQ1) and evaluations of the *actor* making the response itself (RQ2):

*Research Question 1 (RQ1). Will a bot versus a user challenging the corporation for hypocrisy differ in terms of (a)*

higher perceived corporate hypocrisy, (b) lower corporate credibility ratings, (c) higher anger toward the company, and (d) higher boycott intentions?

*Research Question 2 (RQ2).* Will a bot versus a user challenging a corporation for hypocrisy differ in terms of the challenger credibility ratings?

Finally, we offer a theoretical model by which social media challenges to corporate hypocrisy from bots versus humans affect perceptions of corporate credibility and boycott intentions. Specifically, we follow Shim and Yang (2016) to propose that perceptions of corporate hypocrisy will mediate the effects of the challenge on corporate credibility and boycott intentions. Likewise, we expect that moral anger elicited in response to the challenge of corporate hypocrisy (e.g., gender discrimination) will mediate the effects of the challenge on consumers' attitudes and behaviors toward the target company (Krishna et al., 2021; Z. Wang et al., 2020; Z. Wang & Zhu, 2020). We add to this literature by exploring which of these processes serve as a better explanation for attitudinal and behavioral outcomes (RQ3) as well as whether these mechanisms for explaining possible effects differ depending on whether the challenge comes from a bot versus a human actor (RQ4):

*Hypothesis 2 (H2).* The effects of the public challenge on (a) corporate credibility and (b) boycott intentions will be mediated by heightened anger toward the corporation.

*Hypothesis 3 (H3).* The effects of the public challenge on (a) corporate credibility and (b) boycott intentions will be mediated by heightened perceived hypocrisy of the corporation.

*Research Question 3 (RQ3).* Will anger or perceived corporate hypocrisy serve as a better explanation for the mediation effects of the hypocrisy challenge on (a) corporate credibility or (b) boycott intentions?

*Research Question 4 (RQ4).* Will the bot versus user challenger differ in terms of the mediating pathways predicting (a) corporate credibility or (b) boycott intentions?

## Methods

To test these research questions and hypotheses, we used a pre-registered experimental design<sup>2</sup> to precisely manipulate the source (bot versus human) of a challenge to corporate hypocrisy to explore its effects on our outcomes of interest. We used Prolific to recruit 600 participants from the United Kingdom in March of 2023, paying them £1.25 for taking our 7-minute survey. Our participants skewed female (65.5%), educated (54% had a Bachelor's degree or higher), White (87.2%), and younger ( $M=37.87$ ,  $SD=12.58$ ) than the UK population.

After a short pre-test questionnaire, participants were shown a simulated Twitter feed. In our control condition, they saw four filler tweets, unrelated to the topic of the study.

In our two experimental conditions, one of the filler tweets was replaced with a tweet from Procter & Gamble (P&G) promoting their efforts to support equality in honor of International Women's Day in addition to three of the filler tweets. In both experimental conditions, this P&G tweet appears as a quote tweet, with the message emphasizing that P&G paid women 21% less than men (see Figure 1). The challenging messages are identical except for the source of the message: either from a human user (e.g., Taylor Jacobsen) or from a bot account (e.g., the Gender Pay Gap Bot). Therefore, we are able to precisely determine the effect of this shift in source cue in the ability of human versus bot accounts to challenge corporate hypocrisy on social media.

After seeing the simulated social media feed, participants first answered a series of manipulation checks about the tweets they saw on the feed (all), what the quoted tweet said about the pay gap at P&G, and the source of the quoted tweet (experimental conditions only). Then, they rated their agreement with statements to measure the five key outcomes: perceptions of corporate hypocrisy, corporate credibility, anger toward the corporation, bot/user credibility, and boycott intentions. The participants who failed the attention check ( $n=3$ ) were excluded from the analysis. All analyses match the pre-registration unless otherwise noted.

## Measures

### Corporate Hypocrisy

Perceptions of corporate hypocrisy is measured by asking the participants to rate the following statements for P&G on a 7-point scale from "Strongly Agree" to "Strongly Disagree": "P&G acts hypocritically," "What P&G says and does are two different things," "P&G pretends to be something it is not," "P&G does exactly what it says" (reversed), "P&G keeps its promises" (reversed), and "P&G puts its words into action" (reversed). This scale was adapted from the study by Wagner et al. (2009). Factor analysis confirmed a single factor seven-point scale ( $\alpha=.94$ ,  $M=4.40$ ,  $SD=0.97$ ).

### Moral Anger

Moral anger toward the corporation is measured by asking the participants to rate the following: "Please indicate your response using the scale provided. While you read the tweet about P&G to what extent did you experience these emotions toward P&G?" (on a 7-point scale from "Not at all" to "Very much"). This measure is adapted from the scale developed by Harmon-Jones et al. (2016). Differing from our pre-registration, we only include three items (anger, rage, and disgust) in our analysis as recent research indicates that moral emotions, such as moral anger and disgust, would be elicited by the manipulation of hypocrisy (Laurent et al., 2014; Shim et al., 2021; Z. Wang et al., 2020; Z. Wang & Zhu, 2020) ( $\alpha=.93$ ,  $M=2.47$ ,  $SD=1.55$ ).<sup>3</sup>

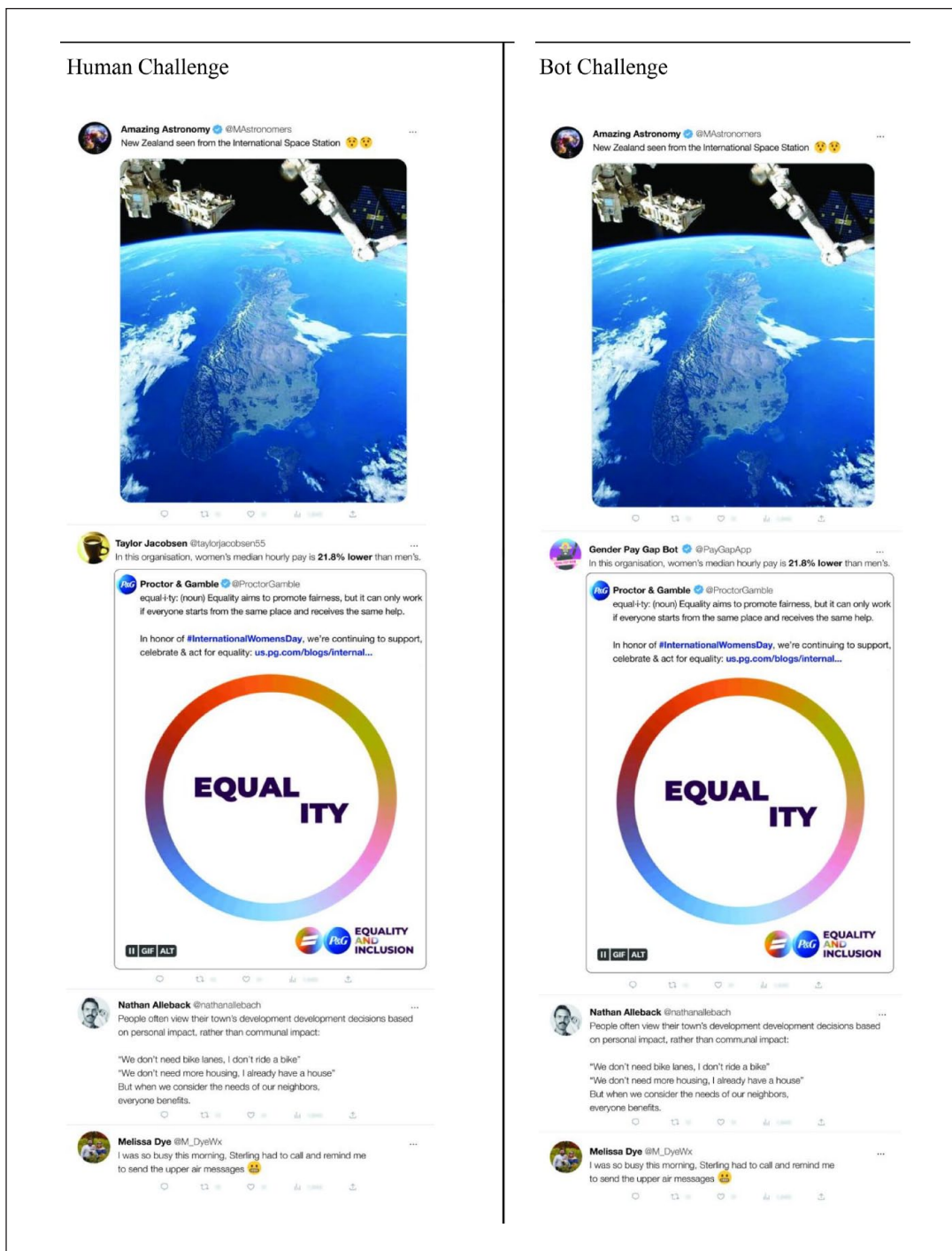


Figure 1. Example stimuli.

**Corporate Credibility**

Corporate credibility is measured by asking the participants to rate eight statements for P&G on a 7-point scale from “Strongly Agree” to “Strongly Disagree.” Example statements include “P&G has a great amount of experience” to “I do not believe what P&G is telling me” (reversed) (Newell & Goldsmith, 2001) ( $\alpha = .85, M = 4.27, SD = .79$ ).

**Boycott Intentions**

Boycott intentions are measured by asking the participants to rate the three statements for P&G: Statements range from “I would recommend others to avoid P&G products” to “I will purchase products made by P&G” (reversed) to “I would feel guilty if I bought a P&G product” on a 7-point scale from “Strongly Agree” to “Strongly Disagree.” This measure is

adapted from the work by Shim et al. (2021) ( $\alpha = .73$ ,  $M = 3.49$ ,  $SD = 1.07$ ).

### Challenger Credibility

Bot/User credibility is measured by asking the participants to rate the following to evaluate the account that quote retweeted the P&G tweet on a 7-point scale: the account is “Informed/Uninformed” (reversed), “Incompetent/Competent,” “Inexpert/Expert,” “Cares about me/Doesn’t care about me” (reversed), “Concerned with me/Unconcerned with me” (reversed), “Has my interests at heart/Doesn’t have my interests at heart” (reversed), “Untrustworthy/Trustworthy,” “Honest/Dishonest” (reversed), “Unethical/Ethical,” and “Moral/Immoral” (reversed). This measure is adapted from the study by McCroskey and Teven (1999) ( $\alpha = .88$ ,  $M = 4.11$ ,  $SD = .75$ ).

## Results

To test H1, we used a series of analyses of variance (ANOVAs) with the two experimental conditions combined to compare against the control condition, run separately for each dependent variable as pre-registered. H1 is supported across three of the four dependent variables (see Table 1), as exposure to a challenge produced significantly higher perceptions of corporate hypocrisy, higher anger toward the company, and lower credibility ratings for the company—but did not significantly affect boycott intentions.

RQ1 asked whether there would be differences in the effects of the hypocrisy challenge depending on whether the source was a bot versus an anonymous social media user. To test this question, we again ran a series of ANOVAs, with the three experimental conditions as the independent variable (see Table 2). The results reinforce H1: both the user and bot challenge influenced corporate hypocrisy, anger, and credibility, but not boycott intentions. In no case did the bot versus the user challenges differ in these outcomes, even using the more permissive Least Significant Difference test. In addition, the bot versus user did not significantly differ in affecting perceptions of or feelings toward the corporation, per RQ2.

Our next set of hypotheses proposed that perceptions of corporate hypocrisy and anger toward the corporation would mediate the effects of the challenge on perceptions of

corporate credibility and boycott intentions. To test these hypotheses, we used the PROCESS macro version 3.3 (Hayes, 2017), Model 4, with a heteroskedasticity estimator. We find support for H2: the effects of the challenge (either from the bot or a Twitter user, per RQ4) on perceptions of corporate credibility are fully mediated by both perceptions of corporate hypocrisy and anger toward the corporation (see Figure 2, Table 3), with the direct pathway between the challenge and perceptions of corporate credibility reduced to non-significance. For corporate credibility, perceptions of corporate hypocrisy appear to be a stronger pathway, as indicated by the lack of overlap between the confidence intervals for the indirect pathways for anger versus perceptions of corporate hypocrisy (per RQ3).

Likewise, although we did not find a main effect of our challenge on boycott intention toward the corporation, our mediation model provides some evidence for why this occurred (see Figure 3, Table 4). There is a significant indirect effect of the bot and user challenge (per RQ4) on boycott intentions via both perceptions of corporate hypocrisy and anger toward the corporation. These effects are roughly equivalent in size, as indicated by the overlapping confidence intervals for each indirect pathway (per RQ3). However, these positive indirect pathways on increasing boycott intentions are offset by the negative direct pathway between the challenge and boycott intentions that remains, an unanticipated result.

### Additional Analysis: Recall

However, one difficulty with social media manipulations is that attention and recall for social media stimuli tend to be quite low (Ellison et al., 2011; Lang, 2000; Smith & Duggan, 2016; Vraga et al., 2016). This is also the case in our study. In our two experimental conditions, only two-thirds (65.0% bot, 66.3% user) correctly identified that the tweet claimed P&G pays men more than women. Even fewer recognized the source of the challenge: only 27.4% recognized that it came from a bot in the bot condition; 48.2% reported it originated from a user in the user condition.

Therefore, as indicated in our pre-registration, we replicate our analyses among those who recalled the position of the challenge tweet (that P&G pays men more than women). This strengthened but did not change our findings reported above (see Supplemental Table A1). Likewise,

**Table 1.** Experimental Effects on Dependent Variables Comparing Experimental Conditions to Control.

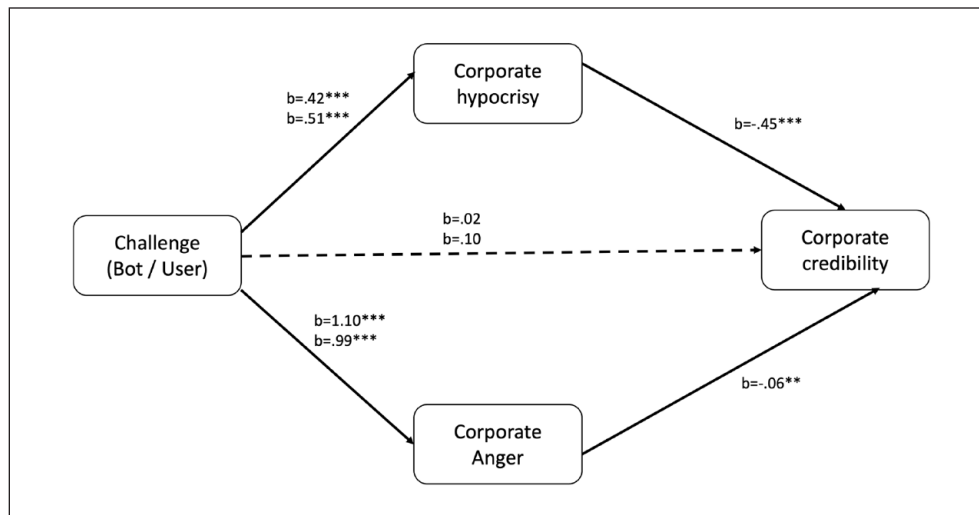
DV	F-test	$\eta_p^2$	Control mean	Challenge mean
Corporate hypocrisy	32.35***	.052	4.09 <sub>a</sub>	4.56 <sub>b</sub>
Corporate anger	67.55***	.102	1.77 <sub>a</sub>	2.82 <sub>b</sub>
Corporate credibility	9.73**	.016	4.42 <sub>a</sub>	4.20 <sub>b</sub>
General boycott intentions	2.00	.003	3.41 <sub>a</sub>	3.54 <sub>a</sub>

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ ; Different subscripts indicate significant differences between conditions for that DV,  $p < .05$ .

**Table 2.** Experimental Effects on Dependent Variables Among Full Sample.

DV	F-test	$\eta_p^2$	Control mean	Bot mean	User mean
Corporate hypocrisy	16.55***	.053	4.09 <sub>a</sub>	4.51 <sub>b</sub>	4.60 <sub>b</sub>
Corporate anger	34.05***	.103	1.77 <sub>a</sub>	2.88 <sub>b</sub>	2.76 <sub>b</sub>
Corporate credibility	5.08**	.017	4.42 <sub>a</sub>	4.18 <sub>b</sub>	4.23 <sub>b</sub>
General boycott intentions	1.26	.004	3.41 <sub>a</sub>	3.58 <sub>a</sub>	3.50 <sub>a</sub>
Challenger credibility	.39	.001	—	4.13 <sub>a</sub>	4.08 <sub>a</sub>

\*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ ; Different subscripts indicate significant differences between conditions for that DV,  $p < .05$ .



**Figure 2.** Mediation effects on corporate credibility.

Note. \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ ; Bot as compared to control condition listed above, user as compared to control condition listed below.

**Table 3.** Indirect Effects on Corporate Credibility.

	B	SE	LLCI	ULCI
Bot → Hypocrisy → Corporate credibility	<b>-.19</b>	<b>.04</b>	<b>-.28</b>	<b>-.11</b>
Bot → Anger → Corporate credibility	<b>-.06</b>	<b>.02</b>	<b>-.10</b>	<b>-.02</b>
User → Hypocrisy → Corporate credibility	<b>-.23</b>	<b>.05</b>	<b>-.32</b>	<b>-.14</b>
User → Anger → Corporate credibility	<b>-.06</b>	<b>.02</b>	<b>-.11</b>	<b>-.02</b>

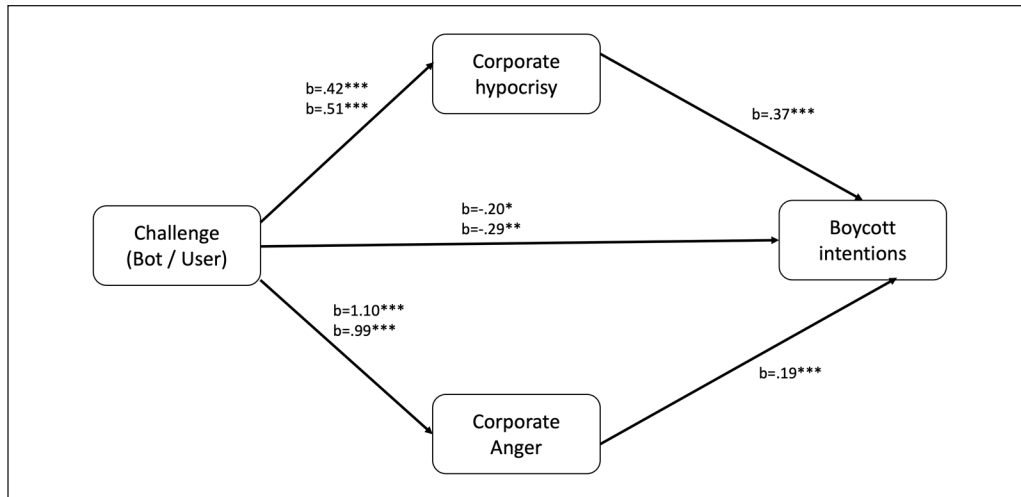
Bolded rows signal significant indirect effects, as indicated by the confidence interval that does not include 0. SE = standard error; CI = confidence interval.

looking only at those who correctly recalled the content and source of the challenge largely produced the same results: either challenge was (equally) effective in changing perceptions of corporate hypocrisy and anger toward the corporation (see Supplemental Table A2); both the bot and the user challenge were seen as equally credible. While they both appeared somewhat more successful in lowering corporate credibility and increasing boycott intentions as compared to the user, these results must be interpreted with caution given the small sample size and the likely differences between those who can recall the challenger’s position and source, as compared to those who cannot.

## Discussion

This article addresses the question of whether challenges to corporate hypocrisy on social media are effective in changing attitudes and behavioral intentions toward the company. We pay special attention to the *source* of such challenges, exploring the question of whether bots are equally effective as human actors in challenging corporate hypocrisy. Our experiment found that bots and human actors largely function similarly: when a bot or a human challenge a corporation for failing to pay men and women equitably, participants not only view the corporation as hypocritical, but they also experience a sense of moral anger toward the corporation and view the corporation as less credible. Importantly, we also provide a mechanism for downstream effects: Participants who expressed moral anger and viewed the corporation as hypocritical saw the corporation as less credible and were more likely to say they would boycott.

These findings add to our understanding of how bots and human-presenting accounts are interpreted in social media settings. This is an active area of inquiry; the research remains unclear how people rate the credibility, neutrality, and authority of bot versus human actors (e.g., Graefe & Bohlken, 2020; Jia & Liu, 2021; Liu & Wei, 2019; Waddell, 2018). Here, we find that bots serve as equally effective challenges to corporate hypocrisy. Theoretically, this suggests that work remains in



**Figure 3.** Mediation effects on boycott intentions.

Note. \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ ; Bot as compared to control condition listed above, user as compared to control condition listed below.

**Table 4.** Indirect Effects on Corporate Boycott Intentions.

	<i>b</i>	<i>SE</i>	LLCI	ULCI
Bot → Hypocrisy → Boycott intentions	<b>.16</b>	<b>.04</b>	<b>.09</b>	<b>.24</b>
Bot → Anger → Boycott intentions	<b>.21</b>	<b>.04</b>	<b>.13</b>	<b>.30</b>
User → Hypocrisy → Boycott intentions	<b>.19</b>	<b>.04</b>	<b>.11</b>	<b>.28</b>
User → Anger → Boycott intentions	<b>.19</b>	<b>.04</b>	<b>.12</b>	<b>.27</b>

Bolded rows signal significant indirect effects, as indicated by the confidence interval that does not include 0. *SE* = standard error; *CI* = confidence interval.

considering how people respond to bot activity on social media, and this work is urgent beyond examining their effects for news consumption behaviors (Graefe & Bohlken, 2020; Jia & Liu, 2021; Liu & Wei, 2019; Tandoc et al., 2020; Waddell, 2018; Wölker & Powell, 2021). For corporate communication strategies, our current results suggest that advocacy individuals and organizations can utilize bot technology to challenge hypocritical behavior of corporations, rather than devoting manpower to this work. This produces a more scalable method to draw public attention to the misdeeds of corporations.

Of course, the social media environment is constantly changing. Among other major shifts that made global headlines, in February 2023, Twitter announced that it would begin charging third-party developers, such as bots, for access to their API data, a service that was previously free. Access to the API enables programmatic access to Twitter's data, making it possible for bots to autonomously post and respond to tweets (Barnes, 2023). While the crackdown was intended to reduce bots, this policy change was later partially reversed, with the announcement that a "light" version would still be available for free (Binder, 2023). More recent research suggests that bots are an increasing problem on Twitter as the company shrinks its content moderation efforts (Henriksen & Wang, 2022; Taylor, 2023; Yang &

Menczer, 2023). Bots that serve as "watchdog" accounts could be shut down at any moment, as Twitter CEO Elon Musk (2023) tweeted that Twitter will allow bots continued free access to the API provided they post "good content." What Twitter and Musk deem "good content" remains an open question—and a potential threat to said watchdog accounts. The instability and uncertainty surrounding Twitter's bot policies highlight the importance of scholarly focus not only on malicious bots but also on bots designed to bring attention to social issues and the societal implications of attempting to categorize the two.

At a more theoretical level, we also must recognize not just the growth of bot accounts but also accelerating interest in artificial intelligence (AI) in general. The development of AI has important implications for bots on social media. AI can strengthen the automated communication done by bots (Hepp, 2020). As AI makes bots more sophisticated (e.g., more responsive and human-like) (Chang & Ferrara, 2022), it will be harder to distinguish the differences between bots and human users on social media (Ferrara, 2023). In addition, with further advancement of Artificial General Intelligence (AGI), a form of future stronger AI that is smarter than human intelligence (McLean et al., 2023), bots could evolve to become the most active and influential actors on social media. Bots' growing sophistication has raised many concerns about their negative outcomes (Hameleers et al., 2022; Shao et al., 2018; Uyheng et al., 2022; Vosoughi et al., 2018; Weng & Lin, 2022), but these AI-driven tools may also help the public to counter bad social bots (e.g., Botometer) (Yang et al., 2019), enhance the detection of malicious social bots (Zago et al., 2019), and advance the quality of interactions and user experience on social platforms (Hepp, 2020). They may also enable users to challenge hypocritical corporations more efficiently, producing the kinds of consequences our study describes.

We also should consider the ethical implications of using bots—even for goodwill—such as using them to challenge corporate wrongdoing. For example, human users might not be aware that they are interacting with bots (Marechal, 2016), even for the bots have clear labels because these automation tools can garner “unearned social capital” and potentially be disrespectful for other human users on social media (Coleman, 2018, p.124.) Moreover, bots more generally can hijack social media hashtags, disrupt online conversations (Marechal, 2016), manipulate user behavior (Guilbeault, 2016), affect public opinion (Bastos & Mercea, 2018).

Given this evolving landscape and ethical concerns, it is important to recognize that users can continue to perform this work of challenging corporate hypocrisy themselves across many social media platforms, whose affordances may not allow for bots (or can change rapidly, as in the case of Twitter), and which can minimize some of the ethical questions of using bots. However, these challenges still require access to high-quality information—in this case, from the UK government—to offer these challenges.

This research also extends previous work into corporate hypocrisy into a new space: social media. Much like in other spaces (Bhatti et al., 2013; Cooper et al., 2019; Simonovits et al., 2022; von Sikorski & Herbst, 2020), once a corporation is challenged on Twitter for hypocrisy, people perceived it to be more hypocritical, less credible, and experienced more intense moral anger toward the company. Corporations using social media to promote their corporate reputations must be aware of this risk and ensure they are matching what they say they value and how they behave to avoid unintended negative consequences (Rim et al., 2020; Xu & Chang, 2023).

However, our results also suggest that these challenges on social media may go unnoticed by many. The low recall for what the bot said (65.0% bot, 66.0% user) still likely over-estimates real-world attention; while individuals can gain information through incidental exposure in an online media environment, they also engage in information filtering (Prior, 2007). Our effects are even stronger among those who recalled what the bot said, signaling that efforts to understand what makes some content more easily recalled than others are an important ancillary aspect of this research. Regardless of the reason for the low recall, this presents an additional difficulty for the visibility of social media challenges. They can be effective—if they are seen.

We also do not know if those users who follow these kinds of accounts in the real world are different from the participants of our study. It is possible that users who are motivated to follow these types of watchdog or activist accounts hold higher levels of moral anger to activate against corporations. If this is the case, we might expect challenges to have even stronger effects on these populations primed for anger, given its role as a mechanism for affecting the credibility and boycott intentions toward the company. Future research should replicate this experimental design using other methods and approaches to more closely approximate real-world experiences with social media platforms.

In addition, we studied this question in the context of a corporation not upholding its stated ideals in the form of gender equality. It remains an open question how challenges to hypocritical behavior on social media may function for other issues or when targeting other prominent actors outside of corporations. The fact that our sample was not representative—and especially in over-representing women—may have thus strengthened our findings regarding responses to corporate hypocrisy in this context. Moreover, in our case, the bot was producing accurate information about corporate hypocrisy, but that would not have to be the case. Future research should continue to explore how challenges function on social media in contexts outside of corporate hypocrisy, how corporations (or others) can respond if unfairly criticized, and how this process happens on other platforms outside of Twitter.

Ultimately, this study offers important theoretical and practical insights into how bots and humans challenging corporate hypocrisy on Twitter are perceived by individuals. On Twitter, an emotional reaction can be triggered by a challenge to corporate hypocrisy from either a human or a bot, which can translate into perceptions of the company overall and intentions to boycott their products. This insight provides a window into how social media firestorms and pile-ons, particularly those aimed at corporations, have the potential to gain speed and momentum in practice, as a single user or bot can activate an emotional response from viewers of their challenging tweet. This leads us to the central question of our study: Who can challenge corporations on social media? For now, the answer seems to be anyone.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

Funding for this article was provided by the Don and Carole Larson Endowed Professorship through the University of Minnesota.

### ORCID iD

Serena Armstrong  <https://orcid.org/0009-0007-9833-4322>

### Supplemental Material

Supplemental material for this article is available online.

### Notes

1. We have opted to continue the use of the name Twitter for this research as that was the platform’s name at the time of data collection, although during the course of this research, Twitter announced it would be changing its name to X (Ivanova, 2023).
2. In this study, we address a subset of the pre-registered hypotheses. Specifically, we include all pre-registered hypotheses regarding the main effects and mediated effects of the challenge on the specified outcomes but do not address the potential moderating role that the legitimacy of gender hierarchy

(LGH) may play in this relationship given the space limitations of this journal. The hypothesis and research question associated with LGH will be addressed in a subsequent paper.

3. As previous research indicates, factor analysis of emotional data typically produces two different factors: positive emotion and negative emotion. Theoretically, different discrete emotions could have different effects on the outcomes (Dillard & Seo, 2012), and recent research indicates that moral emotions, such as moral anger and disgust, would be elicited by the manipulation of hypocrisy (Laurent et al., 2014; Shim et al., 2021; Z. Wang et al., 2020; Z. Wang & Zhu, 2020).

## References

- Aral, S. (2020). *The hype machine how social media disrupts our elections, Our economy, and our health*. Crown Currency.
- Assenmacher, D., Clever, L., Frischlich, L., Quandt, T., Trautmann, H., & Grimme, C. (2020). Demystifying social bots: On the intelligence of automated social media actors. *Social Media + Society*, 6(3), Article 2093926. <https://doi.org/10/gg9fts>
- Bae, J., & Cameron, G. T. (2006). Conditioning effect of prior reputation on perception of corporate giving. *Public Relations Review*, 32(2), 144–150. <https://doi.org/10.1016/j.pubrev.2006.02.007>
- Barnes, J. (2023, February 3). Twitter ends its free, API: Here's who will be affected. *Forbes*. <https://www.forbes.com/sites/jenaebarnes/2023/02/03/twitter-ends-its-free-api-heres-who-will-be-affected/>
- Bastos, M., & Mercea, D. (2018). The public accountability of social platforms: Lessons from a study on bots and trolls in the Brexit campaign. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), Article 20180003. <https://doi.org/10.1098/rsta.2018.0003>
- Bhatti, Y., Hansen, K. M., & Leth Olsen, A. (2013). Political hypocrisy: The effect of political scandals on candidate evaluations. *Acta Politica*, 48(4), 408–428. <https://doi.org/10.1057/ap.2013.6>
- Binder, M. (2023, May 2). Public services will get free API access again, Twitter says. *Mashable*. <https://mashable.com/article/twitter-reverses-api-decision-for-emergency-weather-alerts-public-services>
- Braunsberger, K., & Buckler, B. (2011). What motivates consumers to participate in boycotts: Lessons from the ongoing Canadian seafood boycott. *Journal of Business Research*, 64(1), 96–102. <https://doi.org/10.1016/j.jbusres.2009.12.008>
- Breen, A. (2022, March 15). It's equal pay day this Twitter bot is calling out companies that pay men more than women. *Entrepreneur*. <https://www.entrepreneur.com/business-news/its-equal-pay-day-and-this-twitter-bot-is-calling-out/422096>
- Briones, R. L., Kuch, B., Liu, B. F., & Jin, Y. (2011). Keeping up with the digital age: How the American Red Cross uses social media to build relationships. *Public Relations Review*, 37(1), 37–43.
- Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health*, 108(10), 1378–1384. <https://doi.org/10.2105/AJPH.2018.304567>
- Cai, M., Luo, H., Meng, X., & Cui, Y. (2022). Differences in behavioral characteristics and diffusion mechanisms: A comparative analysis based on social bots and human users. *Frontiers in Physics*, 10, Article 875574. <https://doi.org/10.3389/fphy.2022.875574>
- Carr, C. T., & Hayes, R. A. (2014). The effect of disclosure of third-party influence on an opinion leader's credibility and electronic word of mouth in two-step flow. *Journal of Interactive Advertising*, 14(1), 38–50.
- Chang, H. C. H., & Ferrara, E. (2022). Comparative analysis of social bots and humans during the COVID-19 pandemic. *Journal of Computational Social Science*, 5(2), 1409–1425. <https://doi.org/10.1007/s42001-022-00173-9>
- Chen, C. F., Shi, W., Yang, J., & Fu, H. H. (2021). Social bots' role in climate change discussion on Twitter: Measuring standpoints, topics, and interaction strategies. *Advances in Climate Change Research*, 12(6), 913–923. <https://doi.org/10.1016/j.accre.2021.09.011>
- Cheng, C., Luo, Y., & Yu, C. (2020). Dynamic mechanism of social bots interfering with public opinion in network. *Physica A: Statistical Mechanics and Its Applications*, 551, Article 124163. <https://doi.org/10.1016/j.physa.2020.124163>
- Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing*, 9(6), 811–824. <https://doi.org/10.1109/TDSC.2012.75>
- Clementson, D., & Xie, T. (2020). Narrative storytelling and anger in crisis communication. *Communication Research Reports*, 37(4), 212–221. <https://doi.org/10.1080/08824096.2020.1811660>
- Coleman, M. C. (2018). Bots, social capital, and the need for civility. *Journal of Media Ethics*, 33(3), 120–132. <https://doi.org/10.1080/23736992.2018.1476149>
- Coombs, W. T. (1998). An analytic framework for crisis situations: Better responses from a better understanding of the situation. *Journal of Public Relations Research*, 10(3), 177–191. [https://doi.org/10.1207/s1532754xjpr1003\\_02](https://doi.org/10.1207/s1532754xjpr1003_02)
- Coombs, W. T., & Holladay, J. S. (2012). The paracrisis: The challenges created by publicly managing crisis prevention. *Public Relations Review*, 38(3), 408–415. <https://doi.org/10.1016/j.pubrev.2012.04.004>
- Cooper, J., Feldman, L. A., & Blackman, S. F. (2019). Influencing republicans' and democrats' attitudes toward Obamacare: Effects of imagined vicarious cognitive dissonance on political attitudes. *The Journal of Social Psychology*, 159(1), 112–117. <https://doi.org/10.1080/00224545.2018.1465023>
- Delistavrou, A., Krystallis, A., & Tilikidou, I. (2020). Consumers' decision to boycott "unethical" products: The role of materialism/post materialism. *International Journal of Retail & Distribution Management*, 48(10), 1121–1138.
- Dillard, J., & Seo, K. (2012). Affect and persuasion. In J. P. Dillard & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (pp. 150–166). SAGE. <https://doi.org/10.4135/9781452218410>
- Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior*, 33, 372–376. <https://doi.org/10.1016/j.chb.2013.08.013>
- Einwiller, S. A., & Steilen, S. (2015). Handling complaints on social network sites: An analysis of complaints and complaint responses on Facebook and Twitter pages of large U.S. companies. *Public Relations Review*, 41(2), 195–204. <https://doi.org/10.1016/j.pubrev.2014.11.012>
- Ellison, N. B., Steinfield, C., & Lampe, C. (2011). Connection strategies: Social capital implications of Facebook: Enabled com-

- munication practices. *New Media & Society*, 13(6), 873–892. <https://doi.org/10.1177/1461444810385389>
- Entman, R. M., & Usher, N. (2018). Framing in a fractured democracy: Impacts of digital technology on ideology, power and cascading network activation. *Journal of Communication*, 68(2), 298–308.
- Ferrara, E. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities. *First Monday*, 28, Article 13185. <https://doi.org/10.5210/fm.v28i6.13185>
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. <https://doi.org/10.1145/2818717>
- Frederick, W. C. (1994). From CSRI to CSR2: The maturing of business-and-society thought. *Business & Society*, 33(2), 150–164. <https://doi.org/10.1177/000765039403300202>
- Gender Pay Gap Bot —About. (n.d.). <https://genderpaygap.app/>
- Gorwa, R., & Guilbeault, D. (2020). Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*, 12(2), 225–248. <https://doi.org/10.1002/poi3.184>
- Graefe, A., & Bohlken, N. (2020). Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news. *Media and Communication*, 8(3), 50–59. <https://doi.org/10.17645/mac.v8i3.3019>
- Grappi, S., Romani, S., & Bagozzi, R. P. (2013). The effects of company offshoring strategies on consumer responses. *Journal of the Academy of Marketing Science*, 41(6), 683–704. <https://doi.org/10.1007/s11747-013-0340-y>
- Guilbeault, D. (2016). Automation, algorithms, and politics| growing bot security: An ecological view of bot agency. *International Journal of Communication*, 10, 5003–5012.
- Hagen, L., Neely, S., Keller, T. E., Scharf, R., & Vasquez, F. E. (2022). Rise of the machines? Examining the influence of social bots on a political discussion network. *Social Science Computer Review*, 40(2), 264–287. <https://doi.org/10.1177/0894439320908190>
- Hameleers, M., van der Meer, T. G. L. A., & Dobber, T. (2022). You won't believe what they just said! The effects of political deepfakes embedded as vox populi on social media. *Social Media + Society*, 8(3), Article 1116346. <https://doi.org/10.1177/20563051221116346>
- Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The Discrete Emotions Questionnaire: A new tool for measuring state self-reported emotions. *PLOS ONE*, 11(8), Article e0159915. <https://doi.org/10.1371/journal.pone.0159915>
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.
- Henriksen, L., & Wang, C. (2022). Hello, Twitter bot!: Towards a bot ethics of response and responsibility. *Catalyst: Feminism, Theory, Technoscience*, 8(1), 1–22.
- Hepp, A. (2020). Artificial companions, social bots and work bots: Communicative robots as research objects of media and communication studies. *Media, Culture & Society*, 42(7–8), 1410–1426. <https://doi.org/10.1177/0163443720916412>
- Himelein-Wachowiak, M., Giorgi, S., Devoto, A., Rahman, M., Ungar, L., Schwartz, H. A., Epstein, D. H., Leggio, L., & Curtis, B. (2021). Bots and misinformation spread on social media: Implications for COVID-19. *Journal of Medical Internet Research*, 23(5), Article e26933. <https://doi.org/10.2196/26933>
- Hino, H. (2023). More than just empathy: The influence of moral emotions on boycott participation regarding products sourced from politically contentious regions. *International Business Review*, 32(1), Article 102034. <https://doi.org/10.1016/j.ibusrev.2022.102034>
- Howard, P., & Kollanyi, B. (2016). *Bots, #StrongerIn, and #Brexit: Computational propaganda during the UK-EU referendum* (Arxiv 160606356). <https://doi.org/10.2139/SSRN.2798311>
- Ivanova, I. (2023, July 31). Twitter is now X: Here's what that means. *CBS News*. <https://www.cbsnews.com/news/twitter-rebrand-x-name-change-elon-musk-what-it-means/>
- Jia, C., & Liu, R. (2021). Algorithmic or human source? Examining relative hostile media effect with a transformer-based framework. *Media and Communication*, 9(4), 170–181. <https://doi.org/10.17645/mac.v9i4.4164>
- Jiang, J., & Vetter, M. A. (2020). The good, the bot, and the ugly: Problematic information and critical media literacy in the post-digital era. *Postdigital Science and Education*, 2(1), 78–94. <https://doi.org/10.1007/s42438-019-00069-4>
- Jin, Y., Liu, B. F., & Austin, L. L. (2014). Examining the role of social media in effective crisis management: The effects of crisis origin, information form, and source on publics' crisis responses. *Communication Research*, 4(1), 174–194. <https://doi.org/10.1177/0093650211423918>
- Kent, M. L. (2010). Directions in social media for professionals and scholars. *The SAGE Handbook of Public Relations*, 2, 643–656.
- Kim, H. J., & Cameron, G. T. (2011). Emotions matter in crisis: The role of anger and sadness in the publics' response to crisis news framing and corporate crisis response. *Communication Research*, 38(6), 826–855. <https://doi.org/10.1177/0093650210385813>
- Klein, J. G., Smith, N. C., & John, A. (2002). *Exploring motivations for participation in a consumer boycott* (ACR North American advances NA-29). <https://www.acrwebsite.org/volumes/8678/volumes/v29/NA-29/full>
- Klein, J. G., Smith, N. C., & John, A. (2004). Why we boycott: Consumer motivations for boycott participation. *Journal of Marketing*, 68(3), 92–109. <https://doi.org/10.1509/jmkg.68.3.92.34770>
- Kowalski, R. M. (1996). Complaints and complaining: Functions, antecedents, and consequences. *Psychological Bulletin*, 119(2), 179–196. <https://doi.org/10.1037/0033-2909.119.2.179>
- Krishna, A., Kim, S., & Shim, K. (2021). Unpacking the effects of alleged gender discrimination in the corporate workplace on consumers' affective responses and relational perceptions. *Communication Research*, 48(3), 426–453. <https://doi.org/10.1177/0093650218784483>
- Kusen, E., & Strembeck, M. (2019). Something draws near, I can feel it: An analysis of human and bot emotion-exchange motifs on Twitter. *Online Soc. Networks Media*, 10–11, 1–17. <https://doi.org/10.1016/J.OSNEM.2019.04.001>
- Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, 50(1), 46–70. <https://doi.org/10.1111/j.1460-2466.2000.tb02833.x>
- Laurent, S. M., Clark, B. A. M., Walker, S., & Wiseman, K. D. (2014). Punishing hypocrisy: The roles of hypocrisy and moral emotions in deciding culpability and punishment of criminal and civil moral transgressors. *Cognition and Emotion*, 28(1), 59–83. <https://doi.org/10.1080/02699931.2013.801339>
- Lindebaum, D., & Geddes, D. (2016). The place and role of (moral) anger in organizational behavior studies. *Journal of Organizational Behavior*, 37(5), 738–757. <https://doi.org/10.1002/job.2065>

- Liu, B., & Wei, L. (2019). Machine authorship—In situ: Effect of news organization and news genre on news credibility. *Digital Journalism*, 7(5), 635–657. <https://doi.org/10.1080/21670811.2018.1510740>
- Macnamara, J., & Zeffass, A. (2012). Social media communication in organizations: The challenges of balancing openness, strategy, and management. *International Journal of Strategic Communication*, 6(4), 287–308. <https://doi.org/10.1080/1553118X.2012.711402>
- Marechal, N. (2016). Automation, algorithms, and politics| when bots tweet: Toward a normative framework for bots on social networking sites (feature). *International Journal of Communication*, 10. <https://consensus.app/papers/automation-algorithms-politics-when-bots-tweet-toward-marechal/42a4bf8fb434577384d834d6fb2c6869/>
- McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communication Monographs*, 66(1), 90–103. <https://doi.org/10.1080/03637759909376464>
- McLean, S., Read, G. J. M., Thompson, J., Baber, C., Stanton, N. A., & Salmon, P. M. (2023). The risks associated with Artificial General Intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence*, 35(5), 649–663. <https://doi.org/10.1080/0952813X.2021.1964003>
- Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, 39(3), 629–649. <https://doi.org/10.1007/s11109-016-9373-5>
- Musk, E. [@elonmusk]. (2023, February 23). Responding to feedback, Twitter will enable a light, write-only API for bots providing good content that is free [Tweet]. *Twitter*. <https://twitter.com/elonmusk/status/1622082025166442505?s=20&t=XMiz3fSEEt9UKwVcx1C4zw>
- Newell, S. J., & Goldsmith, R. E. (2001). The development of a scale to measure perceived corporate credibility. *Journal of Business Research*, 52(3), 235–247. [https://doi.org/10.1016/S0148-2963\(99\)00104-6](https://doi.org/10.1016/S0148-2963(99)00104-6)
- Noelle-Neumann, E. (1974). The spiral of silence a theory of public opinion. *Journal of Communication*, 24(2), 43–51. <https://doi.org/10.1111/j.1460-2466.1974.tb00367.x>
- Oberer, B. J., Erkollar, A., & Stein, A. (2019). Social bots: Act like a human, think like a bot. *Digitalisierung Und Kommunikation*. [https://doi.org/10.1007/978-3-658-26113-9\\_19](https://doi.org/10.1007/978-3-658-26113-9_19)
- Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.
- Rim, H., Park, Y. E., & Song, D. (2020). Watch out when expectancy is violated: An experiment of inconsistent CSR message cueing. *Journal of Marketing Communications*, 26(4), 343–361.
- Ross, B., Pilz, L., Cabrera, B., Brachten, F., Neubaum, G., & Stieglitz, S. (2019). Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks. *European Journal of Information Systems*, 28(4), 394–412. <https://doi.org/10.1080/0960085X.2018.1560920>
- Savage, S., Monroy-Hernandez, A., & Höllerer, T. (2016). Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM conference on computer supported cooperative work & social computing* (pp. 813–822). Association for Computing Machinery. <https://doi.org/10.1145/2818048.2819985>
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, 9(1), 1–9. <https://doi.org/10.1038/s41467-018-06930-7>
- Shim, K., Cho, H., Kim, S., & Yeo, S. L. (2021). Impact of moral ethics on consumers' boycott intentions: A cross-cultural study of crisis perceptions and responses in the United States, South Korea, and Singapore. *Communication Research*, 48(3), 401–425. <https://doi.org/10.1177/0093650218793565>
- Shim, K., & Yang, S.-U. (2016). The effect of bad reputation: The occurrence of crisis, corporate social responsibility, and perceptions of hypocrisy and attitudes toward a company. *Public Relations Review*, 42(1), 68–78. <https://doi.org/10.1016/j.pubrev.2015.11.009>
- Simonovits, G., McCoy, J., & Littvay, L. (2022). Democratic hypocrisy and out-group threat: Explaining citizen support for democratic erosion. *The Journal of Politics*, 84(3), 1806–1811. <https://doi.org/10.1086/719009>
- Singh, I., & Singh, S. (2021). The hype machine: How social media disrupts our elections, our economy and our health- and how we must adapt. *Business and Society Review*, 126(1), 101–104. <https://doi.org/10.1111/basr.12225>
- Smith, A., & Duggan, M. (2016, October 25). *The political environment on social media*. Pew Research Center: Internet, Science & Tech. <https://www.pewresearch.org/internet/2016/10/25/the-political-environment-on-social-media/>
- Tandoc, E. C., Jr., Yao, L. J., & Wu, S. (2020). Man vs. machine? The impact of algorithm authorship on news credibility. *Digital Journalism*, 8(4), 548–562. <https://doi.org/10.1080/21670811.2020.1762102>
- Taylor, J. (2023, September 9). Bots on X worse than ever according to analysis of 1m tweets during first Republican primary debate. *The Guardian*. <https://www.theguardian.com/technology/2023/sep/09/x-twitter-bots-republican-primary-debate-tweets-increase>
- U.S. Department of Commerce. (n.d.). *Social media feeds: NOAA's national weather service*. <https://www.weather.gov/mob/social>
- Uyheng, J., Bellutta, D., & Carley, K. M. (2022). Bots amplify and redirect hate speech in online discourse about racism during the COVID-19 pandemic. *Social Media + Society*, 8(3), Article 2211047. <https://doi.org/10/gr5ksv>
- von Sikorski, C., & Herbst, C. (2020). Not practicing what they preached! Exploring negative spillover effects of news about ex-politicians' hypocrisy on party attitudes, voting intentions, and political trust. *Media Psychology*, 23(3), 436–460. <https://doi.org/10.1080/15213269.2019.1604237>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Vraga, E., Bode, L., & Troller-Renfree, S. (2016). Beyond self-reports: Using eye tracking to measure topic and style differences in attention to social media content. *Communication Methods and Measures*, 10(2–3), 149–164. <https://doi.org/10.1080/19312458.2016.1150443>
- Waddell, F. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism*, 6(2), 236–255. <https://doi.org/10.1080/21670811.2017.1384319>
- Wagner, C., Mitter, S., & Körner, C. (2012). When social bots attack: Modeling susceptibility of users in online social networks. *#MSM2012 Workshop Proceedings*, 838, 41–48.
- Wagner, T., Lutz, R. J., & Weitz, B. A. (2009). Corporate hypocrisy: Overcoming the threat of inconsistent corporate social responsibility perceptions. *Journal of Marketing*, 7(6), 377–391. <https://doi.org/10.1509/jmkg.73.6.77>

- Wang, Y. (2015). Incorporating social media in public relations: A synthesis of social media-related public relations research. *Public Relations Journal*, 9(3). [https://scholars.cityu.edu.hk/en/publications/incorporating-social-media-in-public-relations\(06703df3-4960-4827-b460-a356e7a3c9ba\).html](https://scholars.cityu.edu.hk/en/publications/incorporating-social-media-in-public-relations(06703df3-4960-4827-b460-a356e7a3c9ba).html)
- Wang, Z., Zhang, L., & Liu, X. (2020). Consumer response to corporate hypocrisy from the perspective of expectation confirmation theory. *Frontiers in Psychology*, 11, Article 580114. <https://doi.org/10.3389/fpsyg.2020.580114>
- Wang, Z., & Zhu, H. (2020). Consumer response to perceived hypocrisy in corporate social responsibility activities. *SAGE Open*, 10(2), Article 2092287. <https://doi.org/10.1177/2158244020922876>
- Weng, Z., & Lin, A. (2022). Public opinion manipulation on social media: Social network analysis of Twitter bots during the COVID-19 pandemic. *International Journal of Environmental Research and Public Health*, 19(24), Article 24. <https://doi.org/10.3390/ijerph192416376>
- Wischniewski, M., Ngo, T., Bernemann, R., Jansen, M., & Krämer, N. (2022). "I agree with you, bot!": How users (dis)engage with social bots on Twitter. *New Media & Society*, 26, Article 72307. <https://doi.org/10.1177/14614448211072307>
- Wölker, A., & Powell, T. E. (2021). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, 22(1), 86–103. <https://doi.org/10.1177/1464884918757072>
- Xie, C., & Bagozzi, R. P. (2019). Consumer responses to corporate social irresponsibility: The role of moral emotions, evaluations, and social cognitions. *Psychology & Marketing*, 36(6), 565–586. <https://doi.org/10.1002/mar.21197>
- Xu, H., & Chang, B. (2023). Goodwill or just for show? The effects of different corporate social justice statements and the role of perceived authenticity. *Journal of Communication Management*, 27, 493–521. <https://doi.org/10.1108/JCOM-09-2022-0105>
- Yang, K.-C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, 1(1), 48–61. <https://doi.org/10.1002/hbe2.115>
- Yang, K.-C., & Menczer, F. (2023). *Anatomy of an AI-powered malicious social botnet* (arXiv:2307.16336). arXiv. <https://doi.org/10.48550/arXiv.2307.16336>
- Yoon, Y., Gürhan-Canli, Z., & Schwarz, N. (2006). The effect of corporate social responsibility (CSR) activities on companies with bad reputations. *Journal of Consumer Psychology*, 16(4), 377–390.
- Zago, M., Nespoli, P., Papamartzivanos, D., Perez, M. G., Marmol, F. G., Kambourakis, G., & Perez, G. M. (2019). Screening out social bots interference: Are there any silver bullets? *IEEE Communications Magazine*, 57(8), 98–104. <https://doi.org/10.1109/MCOM.2019.1800520>
- Zhang, M., Chen, Z., Liu, X., & Liu, J. (2024). Who leads? Who follows? Exploring agenda setting by media, social bots and public in the discussion of 2022 South Korea presidential election. *SAGE Open*. Advance online publication. <https://doi.org/10.21203/rs.3.rs-3023846/v1>
- Zheng, L. N., Albano, C. M., Vora, N. M., Mai, F., & Nickerson, J. V. (2019). The roles bots play in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction*, 3, Article 215. <https://doi.org/10.1145/3359317>

### Author Biographies

**Serena Armstrong** (M.A. Mass Communication, University of Minnesota) is a recent graduate of the Hubbard School of Journalism and Mass Communication M.A. program at the University of Minnesota. Their research interests include misinformation mitigation, political communication, and understanding and impacts of algorithmic content.

**Caitlin Neal** (M.A. Strategic Communication, University of Minnesota) is a Mass Communication M.A. student at the Hubbard School of Journalism and Mass Communication at the University of Minnesota. Her research interests are in public opinion, partisan identity, and public policy.

**Rongwei Tang** (M.A. Teachers College, Columbia University) is a Ph.D. student at the Hubbard School of Journalism and Mass Communication at the University of Minnesota. Her research interests include strategic communication and how best to counter misinformation on social media.

**Hyejoon Rim** (Ph.D., University of Florida) is an Associate Professor in the School of Journalism and Communication at the Chinese University of Hong Kong. Her research focuses on corporate social responsibility, consumer skepticism, and social media activism.

**Emily K. Vraga** (Ph.D. University of Wisconsin-Madison) is the Don and Carole Larson Professor of Health Communication in the Hubbard School of Journalism and Mass Communication at the University of Minnesota. Her research tests methods to identify and correct misinformation on social media, to build news literacy to improve audience resilience, and to encourage attention to more diverse content online.