FollowBias: Supporting Behavior Change Toward Gender Equality by Networked Gatekeepers on Social Media

J. Nathan Matias
MIT Center for Civic Media
MIT Media Lab
jnmatias@mit.edu

Sarah Szalavitz
7 Robot
dearsarah@gmail.com

Ethan Zuckerman
MIT Center for Civic Media
MIT Media Lab
ethanz@gmail.com

ABSTRACT

Networked gatekeepers on social media increasingly influence which people and groups receive media attention. Many unknowingly direct much greater attention to men than to women. Can technologies support these gatekeepers to follow their own values of equality? Theories of value consistency suggest that confronting people with inconsistencies between values and behavior can prompt behavior change. In this paper, we introduce FollowBias, a novel system that offers feedback on the percentage of women that users follow on Twitter. We conduct field deployments of FollowBias with 61 and 78 participants, exploring differences between their values and behavior, their explanations of those differences, and their changes in behavior. In the first, FollowBias users had a 45 percentage point greater chance of increasing the percentage of women followed over one week. In the second, we fail to find an effect. We also offer findings on political and ethical trade-offs in designing systems for behavior change toward equality.

Author Keywords

Feminist HCI; Social Justice; Transparency; Personal Behavior Change; Inductive Field Experiments; Value Consistency; Gender Equality; Social Media

ACM Classification Keywords

H.5.3 Group and Organizational Interfaces: Computer-supported cooperative work

INTRODUCTION

Information in the public sphere is increasingly mediated by online social networks, restructuring the challenges for women's equal representation in society. Where organizations once functioned as primary gatekeepers of who was most visible in the media, "networked gatekeepers" now direct considerable attention towards the people they interact with and amplify [6]. Advocates of networked journalism on the microblogging platform Twitter have argued that it has broadened access to diverse voices beyond institutional gatekeepers [57, 22, 33]. Organizing online, women have gained substantial visibility and power in areas including parenting and feminism [32, 5, 43]. Yet availability is not the same as attention. Collective preferences from individuals and their networks can limit their exposure to diverse viewpoints [4]. In the United States, internet users have discriminated against women and people of color in online classifieds [13], accommodation rentals [15], and charitable donations [51]. Women remain a minority of voices in print and online news media [36], and early evidence suggests that online news audiences may share women's writing less than men's writing [37]. Consequently, there is no guarantee that networked gatekeepers will behave any more equitably than institutions toward women.

Quantitative measures of women's presence in the media have long played a part in advocacy towards women's equality [62, 61]. Yet as networked gatekeepers take greater power in public attention, networked approaches need to join these institutional efforts toward equality.

In this paper, we introduce *FollowBias*, a system that supports networked gatekeepers to monitor and change the percentage of women they pay attention to and amplify on Twitter. FollowBias draws from theories of personal behavior change in social psychology, exposing people to inconsistencies between their beliefs about equality and their observed behavior toward women online. When a participant logs in, the system accesses information about the accounts they follow on Twitter and presents them with information about the percentage of women that they follow. FollowBias also offers collaborative recommendations of women to follow on Twitter.

We situate FollowBias in the history of quantitative advocacy for women's representation. We also explore the need for FollowBias in an analysis of the representation of women in the Twitter behavior of 3,656 journalists. Finally, in two mixed-methods, randomized field deployments of FollowBias with 61 and 78 participants on Twitter, we investigate gender differences in the accounts that participants follow, their explanations for their behavior, their values towards gender equality, and their guesses on the attention they offer women online. We offer experimental findings on the actions over time of participants exposed to FollowBias, followed by observational findings on the reported experience of users. Taken together, this paper offers findings on the design of systems to support personal behavior change towards equality in online social networks.

DESIGN MOTIVATIONS

Women's Representation in the Media

The visibility of women in public conversation is an important link toward widened participation and equality. While scholarship in communications and media studies consider details in "media representations" that shape cultural expectations about certain groups, the "symbolic representation" of simply including women in institutions and media may also affect outcomes for women [50]. Media coverage of women is linked with political participation; more women demonstrate political knowledge and vote in places where women run and are elected for public office [10]. In the long term, knowledge of women role models influences adolescents' career aspirations [63]. Globally, women journalists have been shown to represent a more diverse range of identities in their articles than men [36]. Among occasional media contributors, the exposure from opinion articles, interviews, and book reviews often leads to paid speaking opportunities, book contracts, book sales, grant funding, and job opportunities [17, 28]. Yet women continue to receive levels of public attention that are disproportionately low given their contributions [36].

Online publishers, from blogs to social media, have broadened the availability of diverse voices. Women have also strengthened solidarity and collective action using online platforms [53, 34]. For example, parenting blogs have occasionally gained national social and political influence [32, 5]. But greater connections among women may not amount to greater visibility within society. Writing about online feminism in 2011, cultural critic Emily Nussbaum remarked that online feminist blogs "emboldened readers to join in, to take risks in the safety of the shared spotlight." At the same time, she worried that these conversations might not actually broaden the visibility or influence of women: "who is going to hear your voice if you can't get their attention?" [43].

Overall, women receive no more coverage online than in other media. Women were no more likely to appear in online news media than print or broadcast globally in 2013, constituting 26% of people in print, broadcast, and online news, consistent with findings from previous years [35, 36]. Furthermore, early evidence suggests that for some topics where authorship is prominent, articles by women may be shared less by social media audiences than similar articles by similar men [37].

The Under-Representation of Women in Social Media By Networked Gatekeepers

Hope of broadened diversity through online social networks have rested in the work of journalists like Andy Carvin, who used social media during the Arab uprisings of 2011 to access "alternative actors" from usual news sources on the Middle East and North Africa [22, 33]. Analyses of information flows at such moments have shown that certain people emerge as "curators" [39] or "networked gatekeepers" [6] who work outside traditional news institutions to shape what their large audiences encounter.

These networked gatekeepers, whether individual, collective, or automated [4], are coming under greater scrutiny as social media becomes a primary means of distributing media. Formerly heralded as bringers of diversity, they are now seen as forces that may undermine the "careful curation of plurality" that is expected of media institutions [7] by professional journalism societies [2] and government regulators.

Parallel literature in computer science has explored the under-representation of women in a wide variety of peer production contexts, where online volunteers collaborate to create information resources. In these contexts, research has identified gender differences in online platforms including Wikipedia [29, 23], OpenStreetMap [60], Pinterest [45, 11], and news comment sections [49]. These differences may create feedback loops of marginalization; further research has documented under-representations of women in the knowledge created when women are a minority of contributors [52].

In this paper, we focus on systematic under-representation of women by network gatekeepers on Twitter. Plausible explanations for this under-representation are numerous, including cultural attitudes towards women's equality [24], social factors including the available pool of notable women [59, 52], implicit biases against women [20], and social preferences among men [38]. The FollowBias system is designed to address the case where the behavior of a networked gatekeeper differs from their personal values or their beliefs about what is acceptable in society (e.g. social norms).

Gender Differences in Twitter Behavior of Journalists

To explore the need for a system to expand women's representation on social media, we conducted an analysis of tweets and follow networks of 3,656 US and UK journalists. This analysis is offered as documentation of our design process rather than a finding with precise estimates of gender differences in journalist behavior on Twitter. We collected journalist twitter accounts in December 2013 from official, public twitter lists by 21 print and digital-first media organizations. For example, the "Seattle Times Staff" Twitter list includes current staff at the Seattle Times.[1] We supplemented this archive with public lists maintained by a journalist who has been recognized publicly for Twitter lists of media organization staff [44].

We classified the name sex of each account in this list using an automated software library [56], supplemented with manual coding in cases of uncertainty. Accounts that were not people or whose identity was not identifiably male or female were labeled as brands/bots/other. One account remained unlabeled. We also accessed approximately 6.3 million tweets from these accounts on Jan 3, 2014[2] and labeled the name sex of accounts they mentioned in their tweets. We counted more than 1.5 million individual mentions of other accounts. The 1,000 top mentioned accounts whose name sex was unclassified (65% of total interactions) were manually coded for the gender presented in profile images.[3] Within our sample, 28% of journalist Twitter accounts appeared to be by women, ten percent less than the percentage of women employed as

[1] We sent tweets to those accounts asking if the accounts were up-to-date. Two media organizations confirmed the accuracy of the list, and one organization asked to be removed from the dataset.

[2] up to 3,200 historical tweets from each account

[3] Gender labeling by three coders did not overlap, so inter-rater reliability measures are not available.

reporters in U.S. newsrooms [2] and four percent greater than the global average of female reporters [36].

Journalists in our sample followed and interacted with far fewer women than men. For the median account in our sample, women constituted 21% of the accounts they followed. The median percentage of women retweeted was 22%. While prior research has found that men retweet mostly men, and that women retweet mostly women [64], this was not the case among the journalists in our sample. Among women, the median percentage of mentions or retweets about women was 27%, while the median among male journalists was 17%. Women also constituted a smaller percentage of who journalists followed than men. Among women journalists, the median percentage of women followed was 23.9%, while among men, the median was 19.5%.

This analysis is not a representative sample of online journalism and our classification methods may be biased [42]. Yet as a design motivation activity, our analysis convinced us that many networked gatekeepers draw attention to women in smaller proportions than other parts of the media. These findings motivate our work to create systems to support greater inclusion of women by networked gatekeepers on Twitter.

Changing the Representation of Women In the Media

The FollowBias system participates in a forty year history of quantitative research and advocacy for women's representation in the media. Early research in this tradition combined academic scholarship with advocacy by counting the number of women and girls appearing in media. In the 1970s and 80s, scholar-advocates Lauren Weitzman and Dorothy Jurney counted the absence of women in U.S. children's books and political news to show how women's interests were poorly served when women were excluded from public conversations [62]. Jurney took her findings to news editors, pointing out gaps in their coverage and urging them to change [61]. Today, advocates continue to use quantitative findings to pressure institutions towards equality. Since the 1990s, the Global Media Monitoring Project has coordinated a large-scale longitudinal analysis of women's representation in print and broadcast media across 114 countries, supporting global advocacy for women's equality in the news [36, 35]. VIDA Women in Literary Arts publicizes charts about the limited coverage of books by women in specific literary publications [28]. The Op Ed Project collects data on women's presence in opinion sections to inform their advocacy and capacity development work among women opinion makers [17].

Editors and journalists often reply to advocates that they can't find women to publish or include in their writing. To address this response, several organizations offer recommendations of eligible women, manually curating lists of notable women and their topics of expertise [8, 1]. Feminist bloggers also participate in list-making to encourage their communities to follow women's voices [40]. Some of these efforts involve systems for computer-supported cooperative work: the Public Insight network coordinates a database of potential sources across reporters at public radio stations [48].

Within journalism, there is no evidence in any direction that publishing quantitative measures of news diversity has any effect on institutional behavior. Longitudinal statistical models have found that societal changes in U.S. women's employment and political involvement, not internal leadership factors, explain the slow growth in visibility for women in the news [59]. Public pressure can also prompt a backlash, as VIDA board member Amy King worried in her introduction to the advocacy group's 2013 release of data on book reviews about women's writing [28]:

> *"I fear the attention we've already given them has either motivated their editors to disdain the mirrors we've held up to further neglect or encouraged them to actively turn those mirrors into funhouse parodies at costs to women."*

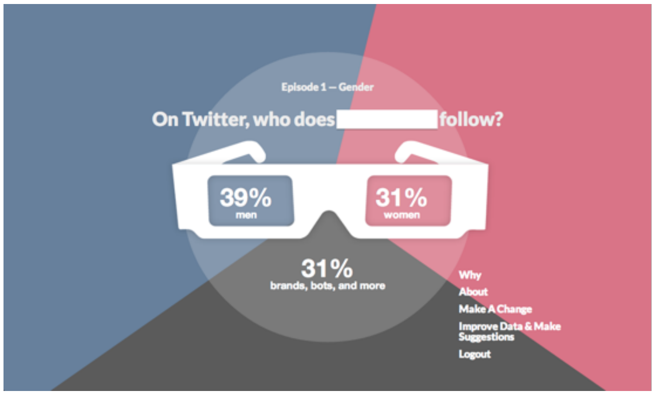
The FollowBias system is situated in the tradition of quantitative advocacy for women's representation. By creating a system for personal behavior change, we have also drawn lessons from the uncertain effects of transparency and public pressure. Unlike efforts to pressure reluctant institutions, FollowBias supports people who are already supportive of women's equality to adopt those values in their own behavior.

Personal Behavior Change

FollowBias draws from theories of behavior change from social psychology, participating in a wider set of systems in Computer Supported Cooperative Work (CSCW) and Human Computer Interaction (HCI) for personal behavior change.

Among the many theory-driven field experiments on prejudice reduction in social psychology, research on "value consistency" has tested the effect of exposing participants to the inconsistencies between their values and their behavior [47]. In the Rokeach value confrontation experiment, researchers instructed college students that people who value equality tend to sympathize with civil rights for black Americans. In followups after 17 and 21 months, Rokeach found that students who received the lecture experienced greater increases in support for civil rights than the control group. Twice as many treatment students enrolled in ethnic studies courses and 2.8 times more treatment students responded to a mailing from the National Association for the Advancement of Colored People [54, 55]. Subsequent research has attributed these effects to a tendency to reduce "cognitive dissonance" between values and behavior, inspiring similar interventions to encourage water conservation and condom use [3, 12]. The FollowBias system uses a similar approach by revealing to users the percentage of women they follow on Twitter, making visible the discrepancies between their values and their behavior.

Systems for behavior change towards diversity have focused on encouraging notions of "balance" in the political and geographical diversity of reading and commenting. Some of these systems offer a visual display of diversity: illustrating "balance" of a user's reading behavior [41] or displaying the ideological positions of other commenters [31]. Other systems offer spatial illustrations of reading behavior [27] and complex visualisations of an "opinion space" [16]. Some of these systems also recommend news that might expand participants' information diversity [27, 41]. While one field deployment



Episode 1 — Gender

On Twitter, who does [redacted] follow?

Figure 1. Presenting The FollowBias Result

showed a modest effect on political diversity of reading behavior [41], exposure to these systems has not consistently been associated with significant changes in reading behavior. This inconsistency may derive from the difficulty of defining change; many differing values may underly a participant's expressed interest in information diversity [31].

Rather than relate complex values to complex measures of political or geographic information diversity, FollowBias addresses a single-dimension scale of gender on social media.

THE FOLLOWBIAS SYSTEM

In the design of FollowBias, we set out to develop a system that could support individual networked gatekeepers to change their behavior towards women on Twitter. FollowBias shows users the percentage of women and men that they choose to follow on Twitter. Inspired by value consistency research, the software reveals to users the inconsistencies between their values and their behavior toward women on Twitter.

Users log into the service using their Twitter account. After completing a survey, users are shown the gender ratio of who they follow: women, men, and brands/bots/more (Figure 1). They are then offered opportunities to make corrections to the automated results, review suggestions from other users on who to follow, and make suggestions of their own. Records on participants are updated every six hours from Twitter; users can log in at any time to see their current result.

System Architecture

As a system design, FollowBias is a "successor system," a work of critical design that monitors an existing platform to support social critique and social change [18]. For example, the Turkopticon system monitors work requests on Amazon's Mechanical Turk platform and collects ratings on work requesters. These ratings offer a critique of labor practices on the platform and support workers to make informed decisions about what work to accept [26]. The Snuggle system monitors socio-technical vandalism response systems on Wikipedia. Snuggle offers alternative measures to Wikipedia's vandalism detection processes and mobilizes Wikipedians for social change in the socialization of newcomers [21]. Like these other successor systems, FollowBias monitors user behavior on Twitter and offers an alternative metric that supports critique and behavior change.

The FollowBias interface is a Javascript web application for desktop and mobile devices. A server application manages permissions, aggregates and serves data used by the browser layer, and responds to user corrections. A multi-server queueing system manages processes to query the Twitter API for participant behavior, monitor changes in who users follow, and archive participant results. Name sex inferences are provided by Open Gender Tracker [56].

Design Considerations

During the design process, we considered decisions about (a) personal feedback versus public pressure, (b) the quality of the gender measures (c) trade-offs between identity and privacy, and (d) the social constraints of behavior change.

Public Pressure Versus Private Feedback

When designing FollowBias, we decided to keep a person's results private. Quantitative measures of gender disparities are often published widely to create public pressure upon powerful institutions [28]. These efforts have yielded uncertain results in the case of institutions, and applying public pressure to individuals on social media can introduce substantial risks. Advocacy campaigns sometimes fruitfully pressure powerful individuals by highlighting their shortcomings. However, online transparency creates differential risks for marginalized populations [9]. In our analysis of journalists on Twitter, we found that most of the women journalists in our sample followed more men than women. If we had designed a system that automatically applied public pressure on its users, we might have done substantial harm to the voice and careers of the very people we set out to support.

When designing FollowBias, we chose against a design that would apply public pressure to Twitter users at scale. FollowBias minimizes risk by only showing participants their own results. Private, personal results also align the design with findings from social psychology on the effect of value confrontation on behavior change.

Quality of Gender Measures

FollowBias is capable of offering personalized feedback at greater scales and frequencies than the human-coded gender counts published by media representation advocates. As an automated service, FollowBias achieves this scale through a trade-off in the quality of its measures. Since the automated name sex coding of Twitter accounts is not always correct, FollowBias invites users to correct these judgments for the accounts they follow. All judgments are added to a collective datastore that improves the accuracy of everyone's results (Figure 2). The FollowBias result is automatically updated on screen every time the participant makes a correction. Where more than one user has made a correction to a given account, the system shows individual users the correction they made while internally adopting the most common judgment.

Gender Identity and Privacy

Any advocacy or research that focuses solely on gender binaries offers a limited view of gender. Yet in FollowBias, we

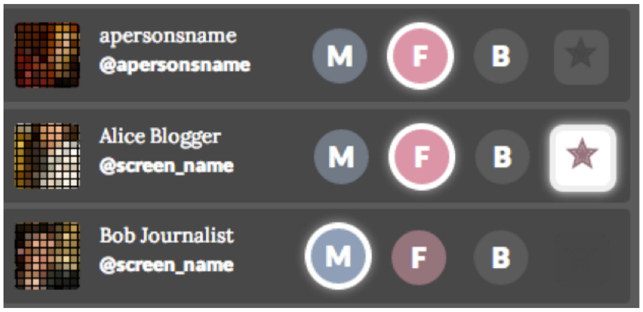


Figure 2. Making gender corrections and suggestions. M=male F=female B=Brands/Bots/More. Users can click on an entry to view an account's Twitter feed

chose to limit our system design to a gender binary to balance competing trade-offs of personalization, quality, and privacy. To achieve scalable personalization of the system, we adopted an automated measure of Twitter account gender presentation, supplemented with collaborative corrections. If we had invited collaborative corrections of more nuanced gender identities, our system could potentially disclose private information about the gender of third parties without their consent.

Since disclosure of gender represents a risk for many people, our system only requests that users record a male or female identity that the account owner has made public. FollowBias prominently reminds participants of the privacy risk when inviting corrections.

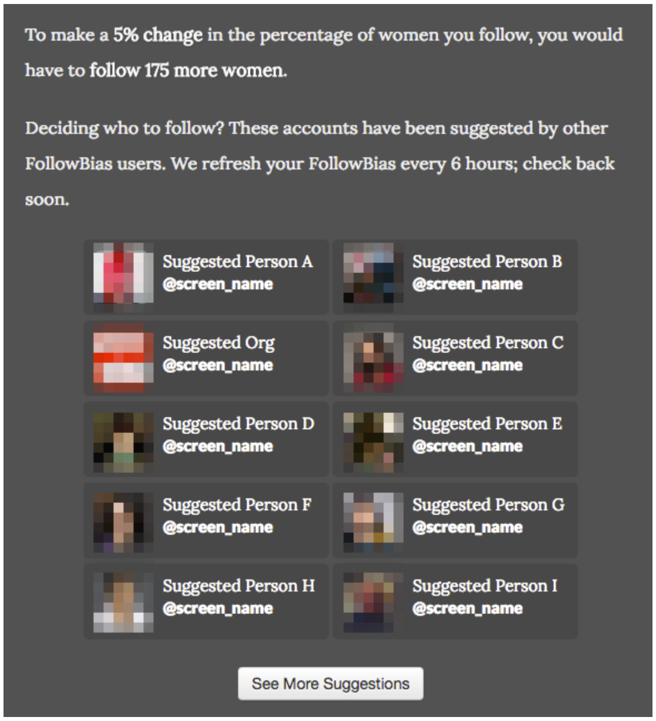
Although privacy concerns lead us to adopt binary gender categories in FollowBias, we have also incorporated critiques of that binary into the design of the system. By presenting the result through stereotypically pink and blue 3D glasses, FollowBias offers an implicit argument for equality while also foregrounding the constructed nature of the information it collects and shares with users. The metaphor of tinted glasses implicitly calls into question the automated metrics of gender binaries shown through the system's lenses. The feedback that FollowBias offers is a work of artifice, like colored lenses, an imperfect filter to support users to evaluate the relationship between their values and their behavior.

Social Constraints of Behavior Change

Especially in matters of representation, a person's capacity for behavior change is bounded by their social context. If a the social context of a FollowBias user includes very few women, it may be difficult for them to expand the percentage of women they follow.

In the second deployment, we introduced peer recommendations for who to follow. As participants corrected the automated judgments, they could select a star next to an account to recommend the account to others (Figure 2). Only accounts labeled as women were permissible as recommendations.

FollowBias shows users a random sample of recommendations immediately below the FollowBias result. We chose a random sample to maximize the diversity of our recommendations, attempting to avoid the winner-takes-all outcomes from popularity-based recommendation systems [58]. We also judged that the participants in our pilot deployments were in similar enough fields that a random selection of suggestions would not represent a relevance problem for participants.



To make a **5% change** in the percentage of women you follow, you would have to follow 175 more women.

Deciding who to follow? These accounts have been suggested by other FollowBias users. We refresh your FollowBias every 6 hours; check back soon.

Figure 3. Browsing recommendations. Users can click on an entry to view an account's Twitter feed

In addition to showing recommendations, the system indicated how many women's accounts a user would have to follow in order to increase the percentage of women they followed by five percent (Figure 3).

METHODS: STRUCTURING INDUCTIVE RESEARCH THROUGH FIELD EXPERIMENTS

In this paper, we offer findings from two pilot deployments of the FollowBias system with networked gatekeepers on Twitter. We investigate the values and behavior of these gatekeepers and evaluate design of FollowBias itself. To do so, we carry out mixed methods survey research, email interviews, observational analysis, and statistical analysis of participant behavior.

In both pilot deployments, we use field experiments to structure our quantitative and qualitative investigations. The contribution of this paper should not be seen as a purely deductive endeavour leading to a single hypothesis test. Instead, we adopt qualitative field experiment methods from social psychology and political science [46], where field experiments offer a way to structure findings from multiple methods.

Experiment Procedures

We evaluated the FollowBias system in two pilot deployments, one in April 2013 and another in December 2013 - January 2014. In both cases, we recruited Twitter users from a convenience sample of accounts that we considered to be networked gatekeepers because their position as professional journalists, their practice of regular blogging, or their standing in their

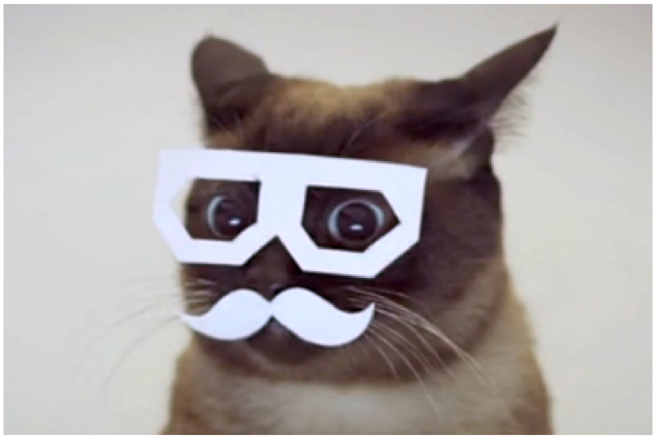


Figure 4. FollowBias Placebo Condition for Control Group

profession. The pool of recruits was drawn from a list of people who followed us on Twitter, who had attended workshops hosted by our institution, or who were otherwise personally known to the authors. In both cases, participants in our recruitment pool were randomly assigned to treatment and placebo groups before recruitment. Participants were recruited through an email that offered participants an opportunity to try software that "analyzes who you follow on Twitter." No mention was made of gender in the recruitment materials.

Every participant who responded was administered a pre-survey with questions about their values and behavior on Twitter. Since the primary goal of both deployments was to evaluate the design and experience of the system, we allocated 80% of recruitment pool to the treatment group. The treatment group was offered access to the FollowBias system, followed by a post-survey, several minutes after exposure to their FollowBias results. Several minutes after exposure to their FollowBias results, respondents in the treatment group were issued a post-survey. We also conducted semi-structured email interviews with over a dozen treatment group participants across both deployments.

The control group was offered a placebo in the form of a photograph of a cat wearing 3D-classes similar to the glasses in the main FollowBias display (Figure 4).[4]

In the second deployment, we introduced a peer suggestions system and implemented logging systems to monitor participant actions more closely. In particular, we monitored the action of clicking on a link to view a Twitter profile from the corrections or suggestions interface.

Estimation Variables

We use a placebo design to estimate the average treatment effect among compliers, the participants who responded to our recruitment materials [19]. We include the following variables:

Main Outcome: Did a User Increase the Percentage of Women They Followed?

The main outcome variable for experimental analysis is the binary outcome of whether the participant's difference in the percentage of women followed in the experiment period is greater than zero. In the first deployment, we observed the difference over one week. In the second deployment, we observed this difference over three weeks.

Like the results shown to participants, the dependent variable is constructed from the percentage of women within the total number of followed accounts. The dependent variable takes into account all corrections made in the history of the system up to the conclusion of the study period, for treatment and placebo participants alike.

Changes in this percentage over time come from changes in the relationships of three variables, each of which could increase or decrease. For example, one participant in a treatment group reduced who they followed by 63 accounts (18% of who they followed), in the process increasing the percentage of women they follow by 2.6 percentage points. Another participant followed 96 additional Twitter accounts in the experiment period, increasing the number of accounts they followed by 10% while only increasing the percentage of women they follow by less than one percentage point. A third participant followed more new women on Twitter than men, but they also followed a large number of accounts with gender unspecified. In consequence, the percentage of women they followed stayed the same, even though they followed 68 new accounts. This participant may have been deliberately following people with gender identities not included in our measure.

Changing the FollowBias result by one percentage point required different amounts of activity from different participants. One participant in a placebo group who was following no women added one female account the next week, resulting in a 5.3 point increase in the percentage of women they followed. In contrast, a treatment group participant following over 4,700 accounts would have needed to follow 262 new women in order to achieve the same percentage point increase.

Regression Adjustment Variables

To improve estimation of standard errors, we conduct regression adjustment on the following variables: the reported gender of the user (male, female, unreported), the number of accounts followed by the user before the experiment, and the percentage of women they followed before the experiment.

Estimation Procedures

In this study, we test the following main hypothesis:

H1: Among participants who respond (compliers), FollowBias has a positive effect on the chance of increasing the percentage of women followed on Twitter.

The complier average treatment effect of FollowBias exposure on the probability of following more women after a period of time is estimated using a logistic regression model ($\alpha = 0.05$).

We also consider a secondary, non-causal hypothesis:

H2: Within the treatment group of the second deployment, use of the recommendation system is associated with larger differences in the percentage of women followed on Twitter over time.

[4] Image from a music video by Denis Borisovich

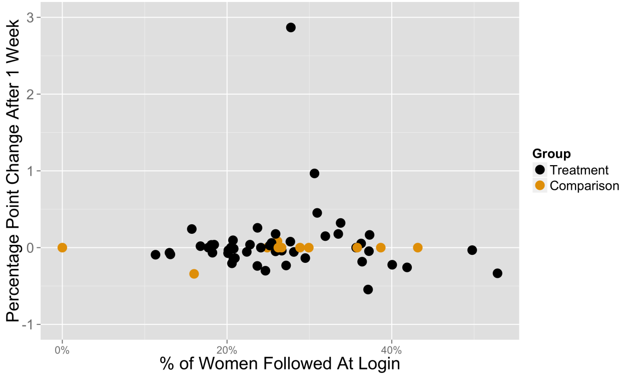


Figure 5. Percentage Point Changes in % Women Followed after One Week, First Deployment

The relationship between use of the recommendation system and increases in the percentage of women followed is estimated with a linear regression model ($\alpha = 0.05$).

EXPERIMENT RESULTS

In two randomized pilot deployments of FollowBias, we test the complier average treatment effect of using FollowBias on the chance of a participant increasing the percentage of women they follow, compared to the placebo condition. The first deployment observed the dependent variable change after one week; the second observed it after three weeks.

For the first field deployment of FollowBias, we recruited 227 journalists and bloggers who also had Twitter accounts. Of these, 61 completed the study. In this first deployment, 50 treatment participants and 11 placebo group participants completed the study. In total, 32 women and 29 men completed the study. The number of Twitter accounts they followed ranged from 5 to 4804, with a mean of 1000 accounts.

In the first deployment, the percentage point difference in women followed after one week ranged from -0.55 to 2.87, with a mean of 0.04. After one week, 9% of participants in the placebo group increased the percentage of women that they followed, and 40% of participants in the treatment group increased percent of women they followed. To our surprise, only one participant increased the women they followed by more than one percentage point; many in the treatment group reduced the percentage of women they followed (Figure 5).

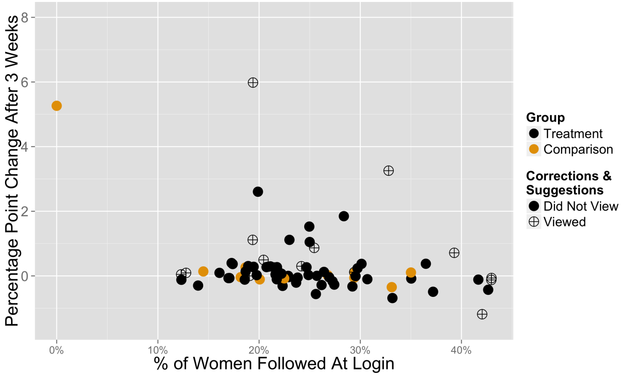


Figure 6. Percentage Point Changes in % Women Followed After Three Weeks, Second Deployment

Logistic Regression Predicting Opting Out

	Base	Adjusted
(Intercept)	-18.57	-16.19
	(1966.65)	(1836.42)
Treatment	16.44	17.01
	(1966.65)	(1836.42)
Male		-1.34
		(0.93)
% Women Initially Followed		-14.63
		(8.13)
Total Followed Initially		0.00
		(0.00)
Log Likelihood	-25.46	-20.74
Deviance	50.92	41.49
Num. obs.	86	86

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1. Estimating Opt-Out Imbalances, Second Deployment

In the second deployment of FollowBias, we recruited 332 journalists, academics, and bloggers who also had Twitter accounts. Of these, 86 out of the recruitment pool of 332 began the study. Twitter data was not available for 8 respondents, either due to software failures or participants choosing to opt out by revoking access to their Twitter data. In a logistic regression of participants who opted out, we find no difference between treatment and placebo groups on observable characteristics. Since participants with missing data do not influence the balance of the treatment and placebo groups, we drop them from the analysis (Table 1) [19]. Accounting for non-completers, the second experiment included 67 complying treatment group participants and 11 placebo group participants. In total, 43 women and 35 men completed the study. The number of Twitter accounts they followed ranged from 18 to 4761, with a mean of 1030 accounts.

Among second deployment compliers, the difference in women followed after three weeks ranged from -1.2 to 8.8, with a mean of 0.5. In that period, 54.5% of participants in the placebo group increased the percentage of women that they followed, and 59.7% of participants in the treatment group changed the number and percentage of women they followed. Over three weeks, 11 participants increased the women they followed by more than one percentage point (Figure 6).

H1: The Effect of FollowBias Use on An Increase in the Percentage of Women Followed on Twitter

In two randomized pilot deployments, we find mixed support for Hypothesis 1.

In the first study, after adjusting for covariates, we find that using FollowBias has a positive effect on the chance that a user increases the percentage of women that they follow over one week. On average, among compliers in our sample, a person who used FollowBias had a 45 percentage point greater chance of increasing the percentage of women they followed, when accounting for other factors (Table 2). In the second

	1st Experiment After 1 Week		*2nd Experiment After 3 Weeks*	
	Base	Adjusted	Base	Adjusted
(Intercept)	-2.30^{*}	-1.91	0.18	2.39^{*}
	(1.05)	(1.49)	(0.61)	(1.19)
Treatment	1.90	2.24^{*}	0.21	0.61
	(1.09)	(1.13)	(0.65)	(0.73)
Male		0.14		-0.55
		(0.59)		(0.54)
Total Followed Initially		-0.00		0.00
		(0.00)		(0.00)
% Women Initially Followed		-1.10		-9.72^{**}
		(3.33)		(3.63)
Log Likelihood	-37.00	-36.23	-52.75	-48.54
Deviance	74.00	72.47	105.50	97.09
Num. obs.	61	61	78	78

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2. H1: Logistic Regression Results for the Complier Average Treatment Effect of FollowBias on a Binary Outcome of an Increase in the % of Women Followed in the Experiment Period

study, after adjusting for covariates, we fail to reject the null hypothesis that using FollowBias has no effect on average over three weeks, among compliers in our sample.

These competing results are likely the result of our small sample sizes. In a follow-up power analysis, we simulated 10^4 field experiments with a complier average treatment effect at the size observed in the first experiment. If the population effect is indeed a 45 percentage point increase in the chance of following a greater percentage of women, we would expect to observe a statistically-significant result 50.1% in the first study and 51.7% of the time in the second study. Our findings are consistent with the results of this power analysis.

H2: The Relationship Between Recommendation System Use and Increases in the Percentage of Women Followed

In the second deployment, we find support for the second, non-causal hypothesis. On average, among complying participants who used FollowBias, clicking at least once on a Twitter account in the recommendation system was associated with a 1.18 percentage point increase in the percentage of women followed after three weeks, holding all else equal (Table 3).

Experiment Threats to Validity

Missing Observations in the Second Deployment

The internal validity of our analysis on the second deployment may be affected by missing observations by opted-out participants. Even though there is no statistically significant difference between treatment and placebo in the chance of opting out, a greater proportion of treatment group participants did opt out, and participants who opted out tended to follow smaller percentages of women (Table 1). It is possible that participants with low FollowBias scores may have been so worried about the implications that they opted out of the study upon seeing their result. We explored this possibility by applying upper and lower Lee bounds [30] in followup analyses and found no meaningful differences from the result of dropping observation presented in our main findings.

3 Week Percentage Point Difference in Women Followed		
	Base	Adjusted
(Intercept)	0.18	1.43^{*}
	(0.19)	(0.72)
Viewed Accounts	1.18^{**}	1.18^{**}
	(0.40)	(0.41)
Male		-0.18
		(0.36)
Total Followed Initially		-0.00
		(0.00)
% Women Initially Followed		-4.06
		(2.31)
R^2	0.12	0.17
Adj. R^2	0.11	0.11
Num. obs.	67	67
RMSE	1.35	1.35

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3. H2: Associations Between System Use And Changes in the % of Women Followed Within the Second Deployment Treatment Group

Effect of the Survey Instrument

The internal validity of our analysis of exposure to FollowBias may be confounded by a possible effect from the pre-survey we issued to participants. Although our recruitment materials did not not mention gender, the survey did. While our placebo design avoids the most common biases from pre-tests that do not survey control group participants [19], our survey still introduces a threat to validity. Prior experiments have found that drawing attention to values may be sufficient for influencing behavior [54, 55]. Indeed, several participants in the first deployment remarked in an open-ended form field that the survey had led them to rethink how they relate to women on Twitter. If the survey had an effect on placebo and treatment participants alike, our models may have under-estimated the treatment effect of exposure to FollowBias alone.

For the second deployment, we test the hypothesis of a possible survey effect in a logistic regression model among the set of all recruited accounts that did not opt out. On average, recruits who complied with the study by taking the survey had a 17.6 percentage point greater likelihood that they would increase the percentage of women they followed compared to non-compliers, holding all else equal (Table 4). Since compliers do not differ from non-compliers on observable characteristics, this result offers strong, if non-causal, evidence that our estimates on the effect of FollowBias may have been confounded by the survey.

Interference from Network Spillover

Interference effects in networks are a further possible confounding factor in estimating the effect of exposure to FollowBias. Since the measured outcome involves changes in the social network of participants, placebo group participants might be affected by changes in the behavior of treatment group participants who they follow [14]. Due to the approach we took in our convenience sample to recruit people known to us, many of our participants also followed each other. Consequently, these spillover effects represent a substantial threat to internal validity. In the second deployment, only one placebo group

3 Week Increase In the % of Women Followed		
	Base	Adjusted
(Intercept)	-0.39^{**}	-1.02^{*}
	(0.13)	(0.43)
Participated	0.75^{**}	0.78^{**}
	(0.27)	(0.27)
Male		0.39
		(0.25)
Unknown Gender		-0.06
		(0.64)
% Women Followed at start		1.51
		(1.41)
Total Followed at start		0.00
		(0.00)
Log Likelihood	-213.30	-210.79
Deviance	426.60	421.59
Num. obs.	316	316

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4. Estimating Possible Survey Effects, Second Deployment

participant followed no one in the treatment group. Placebo group participants followed between 0 and 54 accounts in the treatment group, with a mean of 19.5. Sensitivity analyses that included a covariate counting second-degree exposure were not substantively different from the primary models for the second deployment.

QUALITATIVE AND OBSERVATIONAL FINDINGS

Values Toward Women's Equality In Twitter Use

In the first deployment, when asked on a five-point likert scale if the gender of who they follow on Twitter is important in their professional work, 32 reported that it was sometimes, often, or always important (79% of women and 31% of men). Many participants (25%) reported already paying attention to the gender of who they follow on Twitter.

We asked similar questions of the second deployment. When asked on a five-point likert scale if the gender of who they follow on Twitter is important in their professional work, 45 reported that it was sometimes, often, or always important (89% of women and 40% of men). Among participants, 23% reported tracking the demographics of their Twitter audience, and 19% reported already paying attention to the gender of who they follow on Twitter.

Over-Estimating the Percentage of Women Followed

In both studies, we asked participants to guess the percentage of men, women, and brand/bots/other accounts they followed on Twitter. In both studies, most participants over-estimated the percentage.

In the first deployment, 10 participants declined to guess, answering 0%. Among the rest, 88.2% participants over-estimated the percentage of women they follow on Twitter. These over-estimates were not consistent. We found no significant relationship between guesses and observed behavior.[5]

In the second deployment, 18 guessed 0%. Among those who guessed, 87.1% over-estimated the percentage of women they follow on Twitter. On average, participants made guesses that were 1.6 times higher than their observed behavior.[6]

High Expectations on the Effect of FollowBias

In responses to the pre-survey, most expected that receiving a report on the gender of who they followed would influence their behavior on Twitter, with 64% in the first deployment and 76% in the second deployment expecting an effect.

Although most participants in both deployments expected an effect, the participants who explained their answer in a free text question were less certain. Several drew attention to their social context, mentioning their role in the "male dominated" fields of journalism and technology. For some, our survey represented the first time they had been asked to consider gender equality in their own behavior. One argued for the importance of "gender/sex blind" behavior, and others argued that it was more important to "judge people on merit" rather than follow someone based on gender. Yet others worried about having a "pro-men bias," and expressed interest in quantifying gender differences in other parts of their behavior, including whose articles they read online.

Questioning the Constructions of Identity in FollowBias

Some participants in the second deployment also questioned the constructions of identity used in this research, encouraging us expand the data our system monitors. For example, one offered a "SERIOUS criticism" that the survey didn't consider how the participant monitored "my engagement with my trans network." Some expressed disappointment that we didn't include race. Other participants requested support for recording personal data of pseudonymous Twitter accounts, even where the owners of those accounts declined to make that information visible on Twitter.

Trusting FollowBias Results & Making Corrections

As designers, we expected that participants would express a variety of views about the trust they put in FollowBias upon seeing their result. In the post-survey and email interviews, most participants reported trusting the FollowBias result and reflected on the personal and social reasons for their result.

In email interviews and survey results, many participants expressed appreciation for the ability to make corrections. In the first deployment, the corrections interface was used by 23% of participants, who made a total of 3,245 corrections. Users tended to carry out corrections in a single session taking between 5 and 25 minutes. Five users carried out their corrections over a period longer than an hour. One participant noted that some women speak pseudonymously on Twitter, which might have influenced their result.

Over one fourth (28%) of second deployment participants used the corrections interface, making a total of 1807 corrections.

[5] Two-sided t-test for Pearson correlation $p = 0.17$

[6] The Pearson correlation between guesses of the percentage of women they followed and their observed percentage was 0.61 (Two-sided t-test: $p = 2.5 \times 10^{-7}$). The correlation between guesses and the percentage of women among gender-identified accounts was 0.61 (Two-sided t-test: $p = 2.7 \times 10^{-7}$).

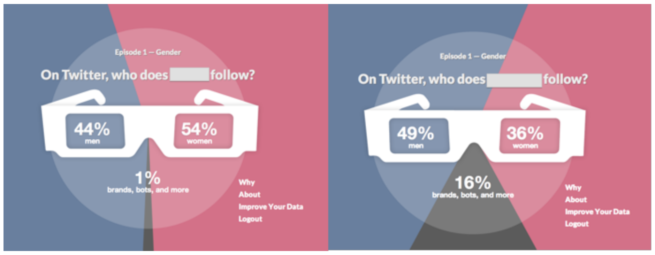


Figure 7. FollowBias results varied between separate personal (left) and professional (right) accounts for one participant

Among those who made corrections, participants made corrections on up to 38% of the total number of accounts they followed, with a mean of 13%.[7] As before, a small number of users were very active in making corrections, with 10% of participants making over 100 corrections each. Twelve percent clicked on at least one screen name in the corrections system, perhaps to check the person's gender or to consider unfollowing or recommending that account.

Sharing Suggestions With Others

In the second deployment, participants interacted with recommendations less than corrections, although the recommendations were more prominent. Suggestions were made by 15% of participants, while 12% of them clicked on suggestions. Seven percent viewed more than one page of suggestions.

Some participants subverted our design intent when making suggestions. Since our system only showed the recommendation button for accounts labeled as women, several men's accounts and organizational Twitter accounts were labeled as women by users and added to the list of recommendations.

Explaining The Reasons for FollowBias Results

Most participants declined to offer an explanation of the factors that influenced the outcome of their FollowBias results. Among those who did, some saw professional pressures in tension with preferred personal behavior, "Who I follow, in part, is a function of my professional networks," wrote one, citing a male-dominated environment. Another participant emailed us screenshots of the difference be-tween a personal and professional account. The professional account followed 36% women, while the personal account followed 54% women (Figure 7). That participant explained:

> *"The performance of people I need to follow for politeness and algorithmic stuff is more male. The people that I actually follow and read every day is more female. [....] I work in the tech sector and that there's a lot of politeness involved with following people so as to not offend. And since the tech sector is predominantly male, this bias is visible."*

This distinction between personal and professional Twitter accounts was common for many. In the opening survey, 40% of users reported distinctions between professional and personal Twitter uses, and a quarter of responders kept both a professional and personal account. For one "male public figure," the professional/personal nature of social media led him to follow few women, concerned that following more women might be seen as flirting. "I often stop and wonder if the 'relationship' gesture is appropriate," he wrote.

Journalist social media activity is often monitored as part of employee evaluation. One respondent felt pressured by metrics systems to follow accounts that might not be actual interests: "the problem with my followers on this account above all else is that it's a performance for other algorithmic analyses [from employers], not actually indicative of who I pay attention to."

In the second deployment, participants appreciated the feature describing how much activity would be required to make a substantial change. In a typical response, one participant expressed surprise and appreciation together:

> *I was baffled by the result—so far from what I perceived about my own account!... I *loved* the idea of being told how many women I would have to follow to bring up my percentage of female 'followees' by 5%*

Privacy Concerns About FollowBias Results

Several participants contacted us with privacy concerns. A journalist wrote, "I'm slightly nervous. The organization I work for prides itself on being objective and I take that value seriously in my work." Acknowledging that FollowBias uses entirely public information, they remarked that "if you're making the process really easy and calling it "FollowBias", it might be a bit uncomfortable if that then got published with my name attached to it." Another participant, who operates a feminism-oriented Twitter account, also encouraged us to keep our results private. "We already receive a lot of abuse from men," this participant emailed us: "revealing the demographic of who we follow on Twitter (probably mostly women) would be likely to increase that and open us up to further criticism and accusations."

DISCUSSION

As networked gatekeepers take a greater role in shaping attention in the public sphere, networked approaches to monitoring and establishing media representation will grow in importance. In this paper, we have outlined the issues at stake for women's representation in online media and have demonstrated consistent under-representation of women among thousands of influential media-makers on Twitter. Motivated by those findings, we have introduced FollowBias, a system to support personal behavior change towards gender equality among network gatekeepers on Twitter. Our findings from two field deployments of FollowBias offer implications for the design of technologies to support equality and social justice online.

As intended, FollowBias succeeded at drawing attention to disparities between values and behavior. Participants consistently underestimated the percentage of women they followed, expressing surprise in their observed behavior. Most participants voiced a desire to pay attention to more women online.

Our users praised FollowBias and had high expectations for its effects on their behavior. We find mixed results on the effect of

[7] Correction figures may be inflated since they include multiple corrections of the same account. The user account making the greatest number of corrections followed many accounts outside the countries whose records were used for automated name sex classifier

FollowBias in two randomized trials. In the first deployment, compared to the placebo group, an experiment complier who used FollowBias had a 45 percentage point greater chance of increasing the percentage of women they followed on average, when accounting for other factors. We observed no statistically significant difference between treatment and placebo in the second deployment, a finding that a power analysis leads us to expect for our small sample sizes.

Our results may also be confounded by factors outside FollowBias, including our pre-survey and network interference. We note that all of the threats to validity that we explored tend towards false negatives rather than false positives. Based on these findings, we are optimistic about the possibility of observing positive effects in future research with larger samples.

Despite this somewhat promising result, it is not easy for anyone to substantially change the percentage of women they follow. Only a minority of participants in either deployment increased their result by more than than one percentage point. As we found in qualitative analysis, wider factors from a person's workplace, profession, and social surround can constrain individual behavior. When women are underrepresented in a particular profession or community, the available pool to follow may be small. Furthermore, employer metrics and obligations of reciprocity may require people to follow more people from a dominant group than they would prefer. A quarter of participants reported keeping different personal and professional accounts, and at least one participant had a secret personal Twitter account where they followed more women.

Social constraints may be overcome through recommendations that reach beyond a participant's network, as we found in a non-causal analysis on the use of the recommendation system. Use of our recommendation system was associated with a 1.18 percentage point increase in the percentage of women followed by our participants on average, holding all else constant.

Even a single percentage point chance can require dozens or hundreds of actions. Some FollowBias users made large changes in their Twitter networks yet reduced the percentage of women they follow. These users may have lost track of the overall balance of the change in their behavior, accidentally reducing the percentage of women they follow. Prior research has tested an effect of value confrontation on a single commitment (joining the NAACP, enrolling in an ethnic studies course) [54]. In this study, we observe outcomes that derive from a large number of ongoing, sustained activities. It is possible that a single moment of surprise about one's own behavior cannot alone support an ongoing effort of substantial change to one's network. Further designs towards network equality might offer participants a chance to join a support group, track their progress, or subscribe to a publication after being shown information about their behavior.

Our findings offer guidance on future research on mediating factors in the effect of value confrontation on behavior change. In our survey results and email interviews, we encountered participants who do not hold women's equality or media representation as meaningful guides to their behavior on Twitter. In these cases, we might not expect FollowBias to have the an effect. Future experiments with larger samples could randomize on participant values to identify the mediating role of values on the effect of value confrontation. Secondly, participants varied in the magnitude of the difference between their values and behavior. In our pilot studies, FollowBias users who made the greatest changes tended to be in the middle of the distribution of FollowBias results (Figures 5 and 6). Future studies could randomize blocks of participants on the magnitude of the difference between values and behavior to identify the influence of that difference on value confrontation effects. A third line of enquiry could explore the mediating role of social context on value confrontation effects, randomizing treatments on variables derived from participants' network contexts.

FollowBias participates in a long tradition of quantitative monitoring and advocacy for equality, but as we discovered, personalizing those methods introduces substantial ethical and political design challenges. Quantitative systems for accountability and social change involve dynamics of risk and power that are very different when power is distributed across networks of individuals rather than institutions. Our privacy, safety, and reliability decisions created complicated trade-offs for users by omitting race and adopting gender binaries. Consequently, our system may have under-estimated gender representation for participants who pay substantial attention to people from a variety of gender identities. Yet our users also thanked us for attending to privacy risks. By foregrounding limitations in the system's construction of gender in the visual design of FollowBias, we successfully prompted critical reflection on these issues from users.

As designers motivated by social justice, our use of field experiments to structure our work created a difficult political decision for us: should we release FollowBias more widely without conclusive evidence on its average effects? The goals of critical design tend be focused on ideological critique, and designers have been caught off-guard when large numbers of people rely on their work [25]. Since creating FollowBias in 2013, we have faced substantial pressure to make the system widely available. As a critical design, our system successfully draws attention to an important inequality. From that perspective, widespread use of FollowBias would draw meaningful attention toward the under-representation of women by networked gatekeepers. Yet while we have promising evidence on a possible effect from FollowBias use, we cannot yet offer conclusive evidence on the average treatment effect. If FollowBias has no effect on behavior on average, a wide-scale deployment might misdirect valuable energy away from more effective efforts for women's equality. For now, we have chosen to keep the system private and conduct further research.

The equality of women in society is a fundamentally important challenge. As social media systems restructure the brokers of attention and opportunity, we must ensure that the incomplete gains of women are not rolled back. In this paper, we address one area where this may be happening. Based on findings with FollowBias, we argue that personalized behavior change systems offer a promising direction for design and research toward equality among networked gatekeepers.

ACKNOWLEDGMENTS

This research was partly made possible by the Knight Foundation, who funded Open Gender Tracker. We are grateful to all our participants who generously shared their personal data with us and thoughtfully explained their experiences. James Home generously implemented the system's beautiful design from our sketches. We wish to thank Tom Steinberg and Kate Crawford their role on the committee for the master's thesis that started this project, as well as Ben Goldacre for suggesting that we employ a randomized trial. We are grateful to Irene Ros and Adam Hyland for their work on Open Gender Tracker, and for Adam's early involvement in data analysis. Ali Hashmi helped code of our sample of journalists. Elizabeth Levy Paluck provided key advice during the literature review portion of the project. We also extend thanks to the reviewers of this paper, whose careful attention to detail and thoughtful suggestions have substantially improved this article.

REFERENCES

1. 2013. BBC publishes list for first '100 Women' event - Media Centre. *BBC* (Oct. 2013). http://www.bbc.co.uk/mediacentre/latestnews/2013/100-women-names
2. 2015. *2015 newsroom census results*. Technical Report. American Society of Newspaper Editors. http://asne.org/blog_home.asp?Display=1948
3. Eliot Aronson, Carrie Fried, and Jeff Stone. 1991. Overcoming denial and increasing the intention to use condoms through the induction of hypocrisy. *American Journal of Public Health* 81, 12 (1991), 1636–1638. http://ajph.aphapublications.org/doi/abs/10.2105/AJPH.81.12.1636
4. Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015). http://science.sciencemag.org/content/348/6239/1130.short
5. Rory Barnett. 2010. Politicians woo 'Mumsnet' generation. *BBC* (Feb. 2010). http://news.bbc.co.uk/local/london/hi/people_and_places/newsid_8522000/8522841.stm
6. Karine Barzilai-Nahon. 2008. Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American society for information science and technology* 59, 9 (2008), 1493–1512. http://onlinelibrary.wiley.com/doi/10.1002/asi.20857/full
7. Emily Bell. 2016. Facebook is eating the world. *Columbia Journalism Review* (March 2016). http://www.cjr.org/analysis/facebook_and_media.php
8. Sarah Bostock. 2013. 20 Women You Should Follow On Twitter (PICTURES). *The Huffington Post* (Aug. 2013). http://www.huffingtonpost.co.uk/2013/08/02/top-20-women-to-follow-on-twitter_n_3695708.html#slide=2765999
9. Danah Boyd. 2012. The power of fear in networked publics, Vol. 10. Austin, TX. http://www.danah.org/papers/talks/2012/SXSW2012.html
10. Nancy Burns and Kay Lehman Schlozman. 2001. *The private roots of public action*. Harvard University Press.
11. Shuo Chang, Vikas Kumar, Eric Gilbert, and Loren G. Terveen. 2014. Specialization, Homophily, and Gender in a Social Curation Site: Findings from Pinterest. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 674–686. DOI: http://dx.doi.org/10.1145/2531602.2531660
12. Chris Ann Dickerson, Ruth Thibodeau, Elliot Aronson, and Dayna Miller. 1992. Using cognitive dissonance to encourage water conservation1. *Journal of Applied Social Psychology* 22, 11 (1992), 841–854. http://onlinelibrary.wiley.com/doi/10.1111/j.1559-1816.1992.tb00928.x/full
13. Jennifer L. Doleac and Luke CD Stein. 2013. The visible hand: Race and online market outcomes. *The Economic Journal* 123, 572 (2013), F469–F492. http://onlinelibrary.wiley.com/doi/10.1111/ecoj.12082/full
14. Dean Eckles, Brian Karrer, and Johan Ugander. 2014. Design and analysis of experiments in networks: Reducing bias from interference. *arXiv preprint arXiv:1404.7530* (2014). http://arxiv.org/abs/1404.7530
15. Benjamin G. Edelman, Michael Luca, and Dan Svirsky. 2016. *Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment*. SSRN Scholarly Paper ID 2701902. Social Science Research Network, Rochester, NY. http://papers.ssrn.com/abstract=2701902
16. Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1175–1184. http://dl.acm.org/citation.cfm?id=1753502
17. Erika Fry. 2012. It's 2012 already: why is opinion writing still mostly male? *Columbia Journalism Review* (2012).
18. R. Stuart Geiger. 2014. Successor Systems: The Role of Reflexive Algorithms in Enacting Ideological Critique. *Selected Papers of Internet Research* 4 (2014). http://spir.aoir.org/index.php/spir/article/view/942
19. Alan S. Gerber and Donald P. Green. 2012. *Field experiments: Design, analysis, and interpretation*. WW Norton.
20. Anthony G. Greenwald, Debbie E. McGhee, and Jordan LK Schwartz. 1998. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology* 74, 6 (1998), 1464. http://psycnet.apa.org/psycinfo/1998-02892-004
21. Aaron Halfaker, R. Stuart Geiger, and Loren G. Terveen. 2014. Snuggle: Designing for efficient socialization and ideological critique. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 311–320. http://dl.acm.org/citation.cfm?id=2557313

22. Alfred Hermida, Seth C. Lewis, and Rodrigo Zamith. 2014. Sourcing the arab spring: a case study of Andy Carvin's sources on twitter during the Tunisian and Egyptian revolutions. *Journal of Computer-Mediated Communication* 19, 3 (2014), 479–499. http://onlinelibrary.wiley.com/doi/10.1111/jcc4.12074/full

23. Benjamin Mako Hill and Aaron Shaw. 2013. The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation. *PLoS ONE* 8, 6 (June 2013), e65782. DOI: http://dx.doi.org/10.1371/journal.pone.0065782

24. Ronald Inglehart, Pippa Norris, and Christian Welzel. 2002. Gender equality and democracy. *Comparative Sociology* 1, 3 (2002).

25. Lilly Irani and M. Silberman. 2014. From critical design to critical infrastructure: Lessons from Turkopticon. *interactions* 21, 4 (2014), 32–35. http://dl.acm.org/citation.cfm?id=2627392

26. Lilly C. Irani and M. Silberman. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 611–620. http://dl.acm.org/citation.cfm?id=2470742

27. Kanarinka. 2014. *Engineering serendipity : Terra Incognita and other strange encounters with global news*. Thesis. Massachusetts Institute of Technology. http://dspace.mit.edu/handle/1721.1/95597

28. Amy King. 2013. Vida Count 2012: Mic Check, Redux. *Vidaweb. org* (2013).

29. Shyong Tony K. Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R. Musicant, Loren Terveen, and John Riedl. 2011. WP: clubhouse?: an exploration of Wikipedia's gender imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. ACM, 1–10. http://dl.acm.org/citation.cfm?id=2038560

30. David S. Lee. 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies* 76, 3 (2009), 1071–1102. http://restud.oxfordjournals.org/content/76/3/1071.short

31. Q. Vera Liao and Wai-Tat Fu. 2014. Can you hear me now?: mitigating the echo chamber effect by source position indicators. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 184–196. http://dl.acm.org/citation.cfm?id=2531711

32. Lori Kido Lopez. 2009. The radical act of'mommy blogging': redefining motherhood through the blogosphere. *New media & society* 11, 5 (2009), 729–747. http://nms.sagepub.com/content/11/5/729.short

33. Gilad Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, and others. 2011. The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication* 5 (2011), 31. http://ijoc.org/index.php/ijoc/article/viewArticle/1246

34. Susana Loza and others. 2014. Hashtag feminism,# SolidarityIsForWhiteWomen, and the other# FemFuture. *Ada: A Journal of Gender, New Media, and Technology* 5 (2014).

35. Sarah Macharia, Lilian Ndangam, Mina Saboor, Esther Franke, Sara Parr, and Eugene Opoku. 2015. *Who Makes the News?: Global Media Monitoring Project 2015*. World Association for Christian Communication.

36. Sarah Macharia, Dermot O'Connor, and Lilian Ndangam. 2010. *Who Makes the News?: Global Media Monitoring Project 2010*. World Association for Christian Communication.

37. J. Nathan Matias and Hanna Wallach. 2015. Working Paper: Gender Discrimination by Audiences of Online News. *Computation + Journalism* (2015). http://cj2015.brown.columbia.edu/papers/gender-discrimination.pdf

38. Miller McPherson, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* (2001), 415–444. http://www.jstor.org/stable/2678628

39. Andrs Monroy-Hernndez, Emre Kiciman, Munmun De Choudhury, Scott Counts, and others. 2013. The new war correspondents: The rise of civic media curation in urban warfare. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1443–1452. http://dl.acm.org/citation.cfm?id=2441938

40. Claire Moshenberg. 2016. The Problem is Not a Lack of Women Experts; It's a Lack of Effort. (May 2016). http://www.genderavenger.com/blog/the-problem-is-not-a-lack-of-women-experts

41. Sean A. Munson, Stephanie Y. Lee, and Paul Resnick. 2013. Encouraging Reading of Diverse Political Viewpoints with a Browser Widget.. In *ICWSM*. http://dub.uw.edu/djangosite/media/papers/balancer-icwsm-v4.pdf

42. Dong-Phuong Nguyen, R. B. Trieschnigg, A. S. Doruz, Rilana Gravel, Marit Theune, Theo Meder, and F. M. G. de Jong. 2014. Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment. (2014). http://eprints.eemcs.utwente.nl/25496/

43. Emily Nussbaum. 2011. The rebirth of the feminist manifesto. *New York Magazine* 30 (Oct. 2011). http://nymag.com/news/features/feminist-blogs-2011-11/

44. Michael O'Connell. 2013. Freelance journalist Nina L. Diamond finds her home on Twitter. (March 2013). http://itsalljournalism.com/nina-l-diamond-finds-her-home-on-twitter/

45. Raphael Ottoni, Joo Paulo Pesce, Diego Las Casas, Geraldo Franciscani Jr., Wagner Meira Jr., Ponnurangam Kumaraguru, and Virgilio Almeida. 2013. Ladies First: Analyzing Gender Roles and Behaviors in Pinterest. In *Seventh International AAAI Conference on Weblogs and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6133

46. Elizabeth Levy Paluck. 2010. The promising integration of qualitative methods and field experiments. *The ANNALS of the American Academy of Political and Social Science* 628, 1 (2010), 59–71. http://ann.sagepub.com/content/628/1/59.short

47. Elizabeth Levy Paluck and Donald P. Green. 2009. Prejudice reduction: What works? A review and assessment of research and practice. *Annual review of psychology* 60 (2009), 339–367. http://www.annualreviews.org/doi/abs/10.1146/annurev.psych.60.110707.163607

48. Andrew Phelps. 2012. The Public Insight Network, now swimming in data, launches its own reporting unit. (Jan. 2012). http://www.niemanlab.org/2012/01/the-public-insight-network-now-swimming-in-data-launches-its-own-reporting-unit/

49. Emma Pierson. 2015. Outnumbered but Well-Spoken: Female Commenters in the New York Times. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 1201–1213. DOI:http://dx.doi.org/10.1145/2675133.2675134

50. Hanna Fenichel Pitkin. 1967. *The concept of representation*. Univ of California Press.

51. Jason Radford. 2014. Architectures of Virtual Decision-Making: The Emergence of Gender Discrimination on a Crowdfunding Website. *arXiv preprint arXiv:1406.7550* (2014). http://arxiv.org/abs/1406.7550

52. Joseph Reagle and Lauren Rhue. 2011. Gender bias in Wikipedia and Britannica. *International Journal of Communication* 5 (2011). http://ijoc.org/index.php/ijoc/article/viewArticle/777

53. Carrie A. Rentschler and Samantha C. Thrift. 2015. Doing feminism in the network: Networked laughter and the Binders Full of Womenmeme. *Feminist Theory* (2015), 1464700115604136. http://fty.sagepub.com/content/early/2015/09/17/1464700115604136.abstract

54. Milton Rokeach. 1971. Long-range experimental modification of values, attitudes, and behavior. *American psychologist* 26, 5 (1971), 453. http://psycnet.apa.org/journals/amp/26/5/453/

55. Milton Rokeach and others. 1973. *The nature of human values*. Vol. 438. Free press New York. https://www.uzh.ch/cmsssl/suz/albert/lehre/wertewandel2011/B01_Rokeach1973.pdf

56. Irene Ros, J. Nathan Matias, and Adam Hyland. 2013. Open Gender Tracker. (2013). http://opengendertracking.github.io/

57. Jay Rosen. 2006. *The people formerly known as the audience*. PressThink. http://archive.pressthink.org/2006/06/27/ppl_frmr.html

58. Matthew J. Salganik and Duncan J. Watts. 2008. Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market. *Social Psychology Quarterly* 71, 4 (Dec. 2008), 338–355. DOI:http://dx.doi.org/10.1177/019027250807100404

59. Eran Shor, Arnout van de Rijt, Alex Miltsov, Vivek Kulkarni, and Steven Skiena. 2015. A Paper Ceiling Explaining the Persistent Underrepresentation of Women in Printed News. *American Sociological Review* 80, 5 (2015), 960–984. http://asr.sagepub.com/content/80/5/960.short

60. Monica Stephens. 2013. Gender and the GeoWeb: divisions in the production of user-generated cartographic information. *GeoJournal* 78, 6 (Aug. 2013), 981–996. DOI:http://dx.doi.org/10.1007/s10708-013-9492-z

61. Kimberly Wilmot Voss. 2010. Dorothy Jurney: A National Advocate for Women's Pages as They Evolved and Then Disappeared. *Journalism History* 36, 1 (2010).

62. Lenore J. Weitzman, Deborah Eifler, Elizabeth Hokada, and Catherine Ross. 1972. Sex-Role Socialization in Picture Books for Preschool Children. *Amer. J. Sociology* 77, 6 (1972), 1125–1150. http://www.jstor.org.libproxy.mit.edu/stable/2776222

63. Christina Wolbrecht and David E. Campbell. 2007. Leading by example: Female members of parliament as political role models. *American Journal of Political Science* 51, 4 (2007). http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2007.00289.x/full

64. Chunjing Xiao, Ling Su, Juan Bi, Yuxia Xue, and Aleksandar Kuzmanovic. 2012. Selective behavior in online social networks. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on*, Vol. 1. IEEE, 206–213. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6511886