

# Outnumbered But Well-Spoken: Female Commenters in the New York Times

Emma Pierson

Department of Statistics, Oxford University  
Oxford, United Kingdom  
emma.pierson@st-annes.ox.ac.uk

## ABSTRACT

Using eight months of online comments on *New York Times* articles, we find that only 28% of commenters of identifiable gender are female, but that their comments receive more recommendations from other readers. Comments from women are more common on forums about parenting, fashion, and health, and on articles written by women. The number of recommendations comments from women receive is positively correlated with the percentage of men on a forum, and the number of recommendations men receive is *negatively* correlated with the percentage of men on a forum. Female commenters are more likely to remain anonymous and anonymous commenters receive fewer recommendations. Male and female commenters differ in their choice of topics to emphasize, backgrounds, and language; we find three specific examples in responses to articles about sexual assault, contraception, and farm subsidies. We discuss the implications of these gender differences for democratic discourse and suggest ways to increase gender parity.

## Author Keywords

Gender; computer-mediated communication; online commenting.

## ACM Classification Keywords

H.5.3. Group and Organization Interfaces: Asynchronous interaction

## INTRODUCTION

Newspapers have long been male-dominated. 67% of newspaper reporters are men, as are 76% of people featured in newspaper stories [17]. These two disparities are probably entwined: stories by female reporters are more likely to feature female subjects and to challenge gender stereotypes [17, 63]. Among all categories of people featured in the news, only homemakers and students are majority female

[17]. Women outnumber men worldwide [17], so it is somewhat troubling that the news is primarily written by and about men.

The rise of the internet has added a new dimension to this dynamic: newspapers are increasingly accessed online [21], allowing readers to comment on articles and providing forums for discussion and argument. Previous researchers have considered the ways in which these forums interact with gender [28, 46, 27]. On the one hand, they may make it easier for women to participate, by allowing authors to choose anonymity and thus overcome social stereotypes, find like-minded individuals and potentially facilitate activism, and use a purely textual medium which may avoid many of the stereotypes which may bias face-to-face communication [28, 46]. On the other hand, the forums do not always fulfill this potential: men still post more and longer comments and are more likely to assert opinions strongly, begin and close discussions, and persist when their comments receive no response [28]. Women also are more likely to face harassment [46, 42], and perhaps as a consequence are more likely to worry about privacy [28, 46, 27].

Here we explore the gender dynamics of online comments on newspapers using 892,999 comments on *New York Times* articles made from June 2013 to January 2014. We infer the gender of the commenter from their name for 413,072 comments, yielding 9,809 unique female commenters and 25,607 unique male commenters. Using the URL of the article, we are also able to derive the article's forum (for example, "parenting.blogs.nytimes.com", or "sports/olympics") and its title. In total, our dataset contains 13,429 articles posted on 144 forums.

This dataset is of interest for two reasons. First, the *New York Times* is one of the most widely read and influential newspapers in the world: it has the largest digital readership (1.1 million) of any newspaper in the United States [21], and its website ranks second among all online newspapers in terms of daily visitors [50]. This dataset thus offers an opportunity to study an institution with large social influence. Second, *New York Times* comments are moderated: every comment is read by a Times staffer prior to publication and comments are rejected if they are obscene, insulting, or inappropriate. (On average, 17 articles, and all blog posts, are opened for comments every day, and articles are left open for comments for 24 hours [62].) This modera-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org). CSCW'15, March 14–18, 2015, Vancouver, Canada. Copyright © 2015 ACM 978-1-4503-2540-0/14/02...\$15.00.

tion, more stringent than that of other news sites like CNN, is of interest for two reasons. First, previous research has found that women may comment more actively if forums are moderated and insulting comments are prohibited [28]; one might expect the *New York Times* comment pages to show greater participation. (It is worth noting that this moderation itself exhibits interesting gender dynamics, because some studies suggest that most moderators are female [69]. Data on the *New York Times* moderators is not public, but we confirmed with staff at the *New York Times* that most moderators were female.) Second, as a result of the moderation, comments are more coherent and linguistically sophisticated than those on unmoderated forums like YouTube, offering a richer dataset for analysis. We are making the full dataset available at <http://cs.stanford.edu/~emmap1/data/nytimes/>. We also provide an interactive visualization online at <http://outnumberedbutwellspoken.herokuapp.com/>.

## DATA AND FRAMEWORK

### Data Collection and Inference of Commenter Gender

Comments on articles were collected using the *New York Times* Community API. Authors of articles were collected using the Article Search API [14]. We determined gender from commenter first name, a commonly used methodology [61, 22, 43, 51], by comparing the first name to a list containing 40,000 first names which were classified by native speakers [48]. To ensure clean data, we only assigned a commenter a gender if their name was exclusively used for men or women, and discarded commenters with names that could be used for both men and women (e.g., “Pat”), commenters who identified themselves with words that were not names (e.g., “A Reader”), and commenters who used only initials (e.g., “AJ”). We were able to infer the gender of the commenter for 46% of comments. To confirm the performance of this method on our dataset, we hand-classified 300 randomly selected commenter names and compared these classifications to the classifications of our algorithm. When the algorithm assigned a definite gender to a commenter, it agreed with our hand classifications 96% of the time. We were thus confident that when our algorithm classified gender, it did so correctly. We also confirmed that our algorithm was not better at identifying men’s or women’s names, which might produce illusory gender disparities: the algorithm identified as women’s names 67% of the names human labelers identified as women’s names, and identified as male 69% of the names human labelers identified as men’s names. The roughly equal percentages indicate a lack of bias. (The algorithm failed to identify all the names humans could identify either because, to ensure a clean dataset, it classified names as androgynous if there was any uncertainty, or because some names contained gender cues – e.g., “FemaleLawyerOf2Boys” which were not recognizable as names.) As a final confirmation, this algorithm was also used in a previous analysis of online education [51], in which its outputs were compared to students’ self-identified genders; the algorithm was not found to be better at identifying men’s or women’s names.

Throughout this paper, we refer to commenter “gender” as opposed to commenter “sex”. This follows previous work which has drawn inferences based on first name [61, 22, 43],

and is motivated by the fact that a name is a socially chosen attribute and therefore part of gender, not sex. (Sex is based on biological characteristics, while gender is fluid, socially constructed and refers to the cultural and behavioral characteristics which society attributes to the sexes [15].) When we use “male” and “female”, we refer only to gender, not sex, inferred from name; thus, “male” includes both cisgender and transgender men, and “female” includes both cisgender and transgender women. We note, as an important caveat, that this approach does not account for commenters whose gender does not fit a binary description or does not match that implied by their name.

### Gender in online discourse

Many paradigms for considering gender within human computer interaction exist [54]. We follow previous authors [10] in using the *gender as a variable* approach, which posits that gender is a variable that may be relevant to analysis. This perspective is well-substantiated by the previous literature, which details the myriad ways in which men and women approach the world differently: offline, in policy preferences, political behavior, moral reasoning, and language use [58, 24, 25, 64, 31]; online, in their behavior on forums like MySpace [63] and Twitter [13]; and, most saliently for our purposes, in their comments on online forums [28, 46, 27].

We briefly review previous theoretical and empirical research on gender in online discourse to provide context for our results. When online forums first became popular, there was hope that they would increase gender equality by allowing authors to choose anonymity and thus overcome social stereotypes or use a purely textual medium which might avoid many of the stereotypes which bias face-to-face communication [28, 46, 39, 70, 29]. Unfortunately, this ideal has not been achieved for several reasons. First, women initially had lesser access to the internet [29]. But even in situations where women have achieved equal access, gender equality has sometimes remained elusive. Women are often driven away from forums by aggressive tactics or harassment [29, 46, 42]. Even when they are not, gender norms often persist online: women tend to be more polite, to assert opinions less strongly, and to “like” products which are consistent with gender stereotypes [27, 29]. Consequently, online forums often end up reproducing the same power imbalances observed in offline interactions. One study of an online forum for the Linguistic Society of America found that women were underrepresented relative to their membership in the society, posted shorter messages, and received fewer replies [28, 26]. Another study found that men and high-status individuals continued to dominate online forums even when participants were given the option to conceal gender [57]. Overall, while online forums offer promise for increasing gender equality, gender differences can persist and these differences often work to the disadvantage of women [16, 29]. Our study provides a large-scale analysis of the extent to which early hopes for increased gender equality have been realized.

### Reader demographics

Readers of the *New York Times* are 56% male, 63% younger than 50, and 80% at least partially college-educated [9].

These statistics are for all *New York Times* readers, not just readers of the digital edition: in general, readers of the digital editions of newspapers are 5% more likely to be male than readers of the print edition, and they are also younger and less wealthy [32]. An analysis of the political affiliations of online news readers found that 45% of online *New York Times* readers described themselves as liberal, 25% as moderate, and 30% as conservative [23].

## RESEARCH QUESTIONS

Among commenters whose gender we could identify (whom we refer to henceforth as “gender-identified” commenters), only 27.7% were women. The average woman also made fewer comments than the average man (10.4 versus 12.1), so in total women made 24.8% of comments. This is considerably lower than the fraction of *New York Times* readers who are women (44%) even accounting for the fact that readers of the digital version of a newspaper tend to be slightly more male than readers overall [32, 9]. To better understand the disparities between male and female commenters, we investigated gender differences in four areas. First, for each comment, our dataset included the number of recommendations the comment received from other readers, and whether the comment was chosen as a “Times pick” by *New York Times* staffers (indicating particularly noteworthy comments).

**RQ1:** Is there a difference in how positively comments by male and female commenters are received (both by other readers and by *New York Times* staffers) and does this vary by forum or article?

Second, previous researchers have found that men and women are interested in news in different areas: women report greater interest in news relating to lifestyle topics, education, and health, while men report greater interest in sports and international issues [32]. We searched for evidence of this in the *New York Times* dataset.

**RQ2:** Do male and female commenters comment on different forums, and are there other factors which affect the gender distribution on an forum?

Third, 54% of comments were made by commenters who chose anonymous pseudonyms with no identifiable gender (we refer to these commenters as “anonymous commenters”). To provide a more complete view of our dataset, we sought to identify the gender distribution of these anonymous commenters. Women face a greater threat of harassment on the internet and display greater concern for privacy than do men; they are, for example, less likely to display their phone numbers and addresses online [46, 42, 20, 3]. Nonetheless, we might expect the moderation of the *New York Times* comment pages to mitigate this effect.

**RQ3:** Can we detect gender differences in the extent to which commenters choose to remain anonymous?

Finally, the underrepresentation of women is of potential concern because comments on online forums can play an important role in democratic deliberation [67, 20]. If men and women express different views, as substantial previous research indicates that they do [58, 24, 25, 64, 31], discourse

will not contain an equal representation of all views, to the detriment of representative democracy.

**RQ4:** Do male and female commenters express different views on the same article?

## RQ1: WOMEN RECEIVE MORE RECOMMENDATIONS

Women’s comments received significantly more recommendations from other readers ( $M = 13.0, SD = 41.4$ ) than did men’s ( $M = 11.3, SD = 37.1$ ),  $t(413,011) = 12.6$ ,  $p < 10^{-5}$ . (The large standard deviations occur because the number of recommendations comments receive is right-skewed, suggesting that analysis of the logarithm of the number of recommendations may be appropriate. We confirmed that we observed the results in this section regardless of whether or not we took the logarithm of the number of recommendations.) However, women’s comments no more likely to be selected by *New York Times* editors as “Times picks” (2.62% vs. 2.59%,  $p > .05$ ). Consequently, we examined the hypothesis that women were not actually making comments that appealed more to other readers, but were merely commenting more than men on more popular forums and articles. We found two signs, however, that this was not the case. First, when we regressed the number of recommendations a comment received on the gender of the commenter and the number of comments on the article (as a way of controlling for article popularity) the coefficient for gender remained highly significant ( $p < 10^{-5}$ ) with women receiving 1.9 more recommendations than men on average (Table 1). Second, women received more recommendations than men on individual forums and articles. Women’s comments received more recommendations than men’s on 67/100 forums with more than 100 comments (67%, 2-tailed binomial  $p = 8.7 \cdot 10^{-4}$ ) and 636/1153 articles with more than 100 comments (55%, 2-tailed binomial  $p = 5.1 \cdot 10^{-4}$ ).

Variable	Coefficient	$p$
Commenter Female	$1.9 \pm 0.1$	$< 10^{-5}$
Comments on Article	$0.0071 \pm 0.0002$	$< 10^{-5}$

**Table 1.** Regression of the number of recommendations a comment received on the gender of the commenter and the number of comments made on the article.  $N = 413,011$

Women’s comments received more recommendations on more male forums; men’s comments received *fewer* recommendations. For each comment made by women, we regressed the number of recommendations the comment received on the female fraction of the forum (the percentage of comments on that forum made by women). The lower the female fraction, the more recommendations women’s comments received ( $p < 10^{-5}$ ). This was not because forums with few women tended to give more recommendations to *everyone*. On the contrary, we found a highly significant relationship for men in the opposite direction ( $p < 10^{-5}$ ): the lower the female fraction of the forum, the *fewer* recommendations men’s comments received. These correlations endured when we looked at the female fraction of an article as opposed to the female fraction of a forum. As a final check, we ran a regression on both genders controlling for article popularity as described above; the interaction between

gender and the female fraction of an article was significant ( $p < 10^{-5}$ ; Table 2). For male commenters, a 10% increase in the female fraction of an article corresponded to 1.1 more recommendations on average; for female commenters, a 10% increase in the female fraction of an article corresponded to 1.0 fewer recommendations. In short, the less frequently a gender commented on a forum or article, the more positively their comments were received when they did.

Variable	Coefficient	$p$
Commenter Female	$7.7 \pm 0.3$	$< 10^{-5}$
Article Female Fraction	$11.4 \pm 0.5$	$< 10^{-5}$
Commenter Female * Article Female Fraction	$-20.9 \pm 0.8$	$< 10^{-5}$
Comments on Article	$0.0067 \pm 0.0002$	$< 10^{-5}$

Table 2. Regression of the number of recommendations a comment received on the gender of the commenter, the female fraction of the article, the interaction between the two, and the number of comments made on the article. The interaction term is negative and highly significant: as the female fraction of the article increases, men receive more recommendations, and women receive fewer recommendations.  $N = 413,011$

### Frequent commenters receive more recommendations

Male commenters on average made 12.1 comments; female commenters made 10.4 comments. But most commenters fell below this average: 49% of men, and 54% of women, made only one comment during the eight months we examined, and 86% of men and 87% of women made fewer than 10. Commenters who made more comments received more recommendations per comment (Figure 1), and this effect was particularly pronounced for female commenters. Comments made by women who made only one comment averaged only .7 more recommendations than comments made by men who made only one comment; comments made by women who made more than 100 comments averaged 2.6 more than those by men who made more than 100 comments. We confirmed that the difference in effect sizes was statistically significant by regressing the number of recommendations a comment received on the gender of the commenter, the number of comments the commenter made, and the number of comments the article received (Table 3). There was a significant interaction term between the commenter being female and the number of comments the commenter made, indicating that the number of comments a commenter made had a greater effect on recommendations for women. (We filtered out commenters who made more than 100 comments to avoid unduly weighting the dataset, but confirmed that the interaction remained significant –  $p < 10^{-5}$  – when we did not.) We note that including the number of comments the commenter made and the interaction term with gender makes the coefficient for gender much smaller: this does not imply that gender is not significant, but that the greater number of recommendations that women receive is due primarily to women who make many comments.

Variable	Coefficient	$p$
Commenter Female	$0.31 \pm 0.26$	0.23
Comments Made by Commenter	$0.02 \pm 0.003$	$< 10^{-5}$
Comments Made by Commenter * Commenter Female	$0.02 \pm 0.006$	0.002
Comments on Article	$0.01 \pm 0.001$	$< 10^{-5}$

Table 3. Regression of the number of recommendations a comment received on the gender of the commenter, the number of total comments the commenter made, the interaction, and the number of comments made on the article. The significant positive interaction indicates that, for female commenters, making more comments correlated with a greater increase in recommendations.  $N = 182,237$ . The smaller  $N$  occurs because we filter out commenters who made more than 100 comments to avoid unduly weighting the dataset.

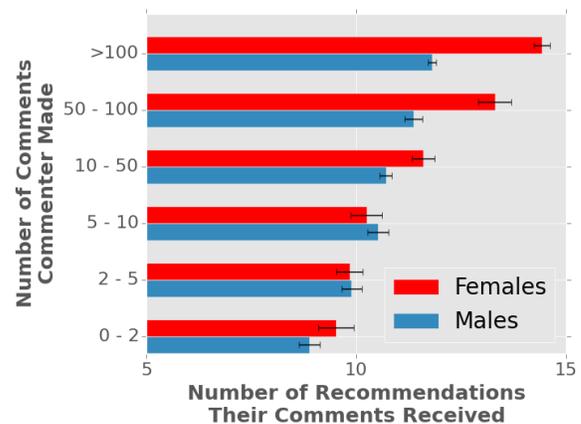


Figure 1. Commenters who make more comments receive more recommendations on average, and this effect is especially pronounced for female commenters.

### RQ2: MEN AND WOMEN FAVOR DIFFERENT FORUMS

In Tables 4 and 5, we provide the articles and forums with the highest percentages of comments made by men and women (filtering out articles and forums with fewer than 100 comments). Women constituted the majority in only 5 of the 144 forums. Topics relating to sports and science tended to be dominated by men; topics relating to parenting, fashion, and health tended to be dominated by women.

Three of the ten most male-dominated articles related to global warming, all of them posted on Andrew Revkin’s Dot Earth blog. This is curious given that previous research has found that if anything, women have more knowledge and concern about global warming and the environment than do men [41, 44]. This discrepancy may be partially due to a small number of men who returned to the forum frequently and dominated the discussion. (As one irritated commenter wrote to the blog’s author, “Have you noticed what an utter disaster your blog threads have become? Basically, you have the same 5 cranks authoring thousands and thousands of repetitive words”). The Dot Earth blog was one of the forums with the largest number of comments per user (7.4

comments, 4th out of 101 forums with more than 100 comments). To more systematically explore this “power user” effect, we compared the fraction of women among the 20 “power users” who commented most frequently on each forum to the fraction of women on the forum overall. On most forums, women were less represented among the power users than they were among users overall (68% of forums, 2-tailed binomial  $p = 3 \cdot 10^{-4}$ ). However, the power users also tended to be more gender-skewed than the users on the forum overall: on the most female forums, power users tended to be more female than the users on the forum overall, and on the most male forums, they tended to be more male ( $p = .006$ , logistic regression).

To examine connections between forums, we computed the forum-commenter matrix  $F$ , where  $F_{ij} = 1$  if commenter  $j$  commented on forum  $i$  and  $F_{ij} = 0$  otherwise. We then computed  $C$ , the correlation matrix of  $F$ , where  $C_{ij}$  was the correlation between forums  $i$  and  $j$  (in other words, two forums were correlated if commenters on one tended to comment on the other). We clustered this matrix using spectral clustering [49], dividing forums into ten groups. We provide an interactive visualization online at <http://outnumberedbutwellspoken.herokuapp.com/> and a screenshot in Figure 2. Several specific-interest clusters were apparent: a cluster of sports-related forums (basketball, football, olympics, etc); a cluster related to arts and music; an international-events consisting of news pages (world, world/asia, world/middleeast), and so on. Notably, all five forums in which women constituted the majority occurred in the same cluster: in addition to these forums, the cluster contained forums related to travel, real estate, gardening, health, dining, and arts.

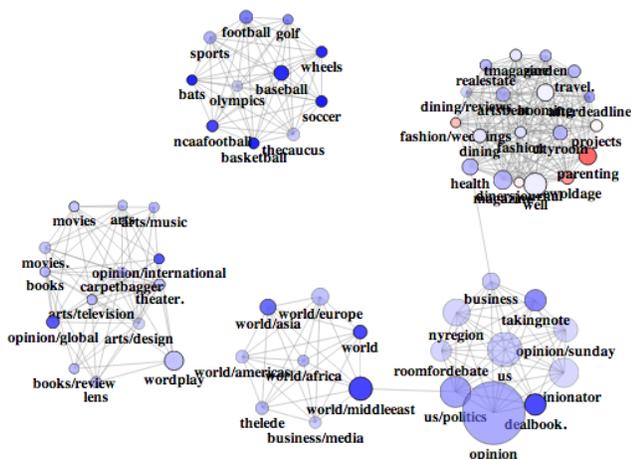


Figure 2. Five of the clusters of forums portrayed in our online visualization. Blue clusters are majority male; red clusters are majority female. All five majority female forums occur in the top right cluster.

Most Male Articles	Most Female Articles
Can We Respond To Problems Like Global Warming Where There’s No Simple Villain? (0.0%)	How Much Would You Pay To Have A Baby? (90.6%)
An Update On Risks Of Abrupt Jolts From Global Warming (0.0%)	Fertility Diary: I’m Old (85.8%)
That Unskewed Feeling (1.8%)	When The Bully Is A Sibling (81.5%)
Milton Friedman, Unperson (1.8%)	A Balanced School Needs A Balance Of Volunteers (81.2%)
A Nuclear Submariner Challenges A Pro Nuclear Film (1.9%)	“Maxed Out” Author Thought Readers Would Critique Her Ideas. Instead, They Judged Her Choices. (80.8%)
A Radio Chat On Cold Snaps And Global Warming (2.0%)	My Half-Jewish Child Wants Christmas. I Don’t, Anymore (80.4%)
On The Symmetry Between Microsoft And Apple (2.5%)	Bounced From Hospice (77.2%)
A Closer Look At Tornadoes And Global Warming (2.7%)	Widening Ripples Of Grief In Adoptee’s Death (76.4%)
Iran Nuclear Talks – Unfinished, But Alive (3.6%)	“Mixed Kids Are Always So Beautiful” (76.1%)
What Janet Yellen And Everyone Else Got Wrong (3.9%)	Blaming The Patient, Then Asking Forgiveness (70.4%)

Table 4. The article titles with the highest percentage of comments made by men (two left columns) and women (two right columns). All articles have at least 100 gender-identified comments.

Most Male Forums	Female	Most Female Forums	Female
soccer	5.4%	parenting	79.3%
bats.blogs	6.9%	newoldage	70.0%
basketball	7.2%	weddings	62.6%
wheels	8.0%	dinersjournal	53.0%
baseball	8.1%	projects	51.8%
dotearth	9.6%	booming	47.6%
keller	10.4%	well	46.7%
energy-environment	10.6%	tmagazine	45.3%
krugman.blogs	10.9%	dining	44.6%
science/space	13.4%	india	44.4%

Table 5. The forums with the highest percentage of comments made by men (two left columns) and women (two right columns). All forums have at least 100 gender-identified comments

### Women comment more on articles written by women

Women wrote only 34% of all articles, but the articles they wrote had a higher percentage of female commenters. 32% of commenters on articles by women were female, as opposed to only 21% on articles by men. This is at least partially because female authors and commenters frequented the same forums:

85% of articles on the Parenting blog, for example, were written by women. But we found the relationship endured when we controlled for this by including the forum on which the article was written as a categorical variable: an article written by a woman will tend to have more female commenters than an article *on the same forum* by a man.

### RQ3: WOMEN ARE MORE PRONE TO ANONYMITY

To provide a more complete view of the dataset, we sought to infer the gender distribution of the anonymous commenters. Since this is obviously a somewhat difficult task, we used two different estimation methods which yielded the same conclusion: women are disproportionately likely to remain anonymous.

As a first method of estimating the relative strength of preferences for anonymity, we looked at the relative probability that commenters we could identify as male or female from their first names would also include their last names. Male commenters were more likely to use their full names (78% of male commenters versus 63% of female commenters;  $p < 10^{-5}$ , logistic regression). One explanation for this is the greater threat of harassment that women face online [46, 42]; one might expect this threat to both reduce someone’s probability of using their full name, and to reduce their probability of even using their first name.

As a second method, we examined anonymous commenters who identified themselves only by first initial (e.g., “W. Bostock” or “CG”) and examined the proportion of names beginning with each letter (ignoring “A” and “I” which frequently corresponded to spurious names like “A Reader”). We then compared this distribution  $p_{anon}$  to the distribution of female commenter names beginning with each letter  $p_f$ , and the distribution of male commenter names beginning with each letter  $p_m$  by fitting the following model:

$$p_{anon} = \beta_f p_f + \beta_m p_m + (1 - \beta_m - \beta_f) p_{uniform}$$

where  $p_{uniform}$  was the uniform probability vector and  $\beta_f$  and  $\beta_m$  were the coefficients of the model ( $\beta_m, \beta_f \geq 0$ ,  $\beta_m + \beta_f \leq 1$ ). Essentially, this model says that anonymous commenters arise from a mixture of three possibilities: women abbreviating their names, men abbreviating their names, and commenters who choose their anonymous pseudonym via some other process (for which we use the uniform prior  $p_{uniform}$ ). We fit this model using maximum likelihood; the fit to the data was very good ( $r = .97$ , Figure 3) with  $\beta_f = .312$ ,  $\beta_m = .494$ ,  $\beta_{uniform} = 1 - \beta_m - \beta_f = .194$ . Taking the ratio  $\frac{\beta_f}{\beta_m + \beta_f}$  yields an estimate of the proportion of anonymous commenters who are female: 39%, considerably higher than the 28% for non-anonymous commenters. (We confirmed that our model better fit the data than did the models with only female commenters, only male commenters, or only female and male commenters with no uniform term;  $p < 10^{-5}$  for all comparisons, likelihood ratio test. The model with only female and male commenters yielded a similar estimate of the proportion of women: 41%).

Anonymous comments received fewer recommendations than gender-identified comments even when we controlled for article popularity (Table 6). They were also less likely to be chosen as editors’ selections (2.2% versus 2.6%;  $p < 10^{-5}$ , t-test). Similarly, comments made by authors who used their full name as opposed to just their first name received 2.8 more recommendations when we controlled for gender and article popularity (Table 7). (The effect remained significant when we did not control for gender).

Variable	Coefficient	$p$
Commenter Anonymous	$-1.4 \pm 0.1$	$< 10^{-5}$
Comments on Article	$0.002 \pm 0.0001$	$< 10^{-5}$

Table 6. Anonymous commenters get fewer recommendations when controlling for article popularity.  $N = 892,548$

Variable	Coefficient	$p$
Commenter Full Name Uses	$2.8 \pm 0.1$	$< 10^{-5}$
Commenter Female	$2.3 \pm 0.1$	$< 10^{-5}$
Comments on Article	$0.007 \pm 0.0002$	$< 10^{-5}$

Table 7. Commenters who use their full name get more recommendations when controlling for author gender and article popularity.  $N = 413,011$

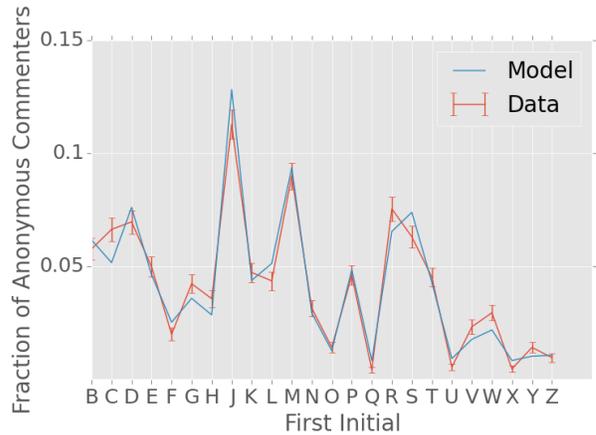


Figure 3. Comparison between modeled and actual proportions of anonymous usernames beginning with each letter.

### RQ4: MEN AND WOMEN EXPRESS DIFFERENT VIEWS

These gender skews would be of less concern if men and women said the same things; however, this is not, of course, the case, as numerous previous authors have found [58, 24, 25, 64, 31]. We searched for evidence of this in the *New York Times* comments. We initially explored standard natural language techniques like sentiment analysis [60] and topic modeling [5] but found, by comparing the results of these algorithms with hand-coded results for a small subset of comments, that the comments were too linguistically complex for the algorithms to yield reliable results. Nonetheless, we found clear evidence that men and women differed in terms of their choice of topics to emphasize, backgrounds, and language.

We began by analyzing the frequencies with which men and women used different words. Men and women use different words, of course, simply because they comment on different types of articles – but this is less interesting than gender differences in reactions to the *same* article. To isolate the latter effect, we looked for words that men used more frequently than women (or vice versa) in commenting on the same article as follows. For each article, we computed the word count vectors for both men and women (sampling the same number of comments for both men and women to avoid frequency effects). Then, for each word, we computed the fraction of articles in which men used the word more frequently in responding to the article. We report the single words and 2-grams whose fractions showed the largest discrepancies from one half in Table 9, after filtering out stop words and words that appeared fewer than 10 times in the dataset as a whole. (For conciseness, we also filtered out phrases that are semantically uninteresting: for example, we filtered out “my husband”, which is the 2-gram that indicates most strongly that a commenter is a woman, and “my wife”, which indicates most strongly that a commenter is a man).

As this analysis makes clear, men and women tended to emphasize different things and use different language *even in responding to the same articles*. Women were more likely to use female-specific words and phrases (“breast”, “misogyny”, “pregnancy”, “woman”, “female”, “rights [for/of] women”) as well as phrases relating to issues which disproportionately affect women (“birth control”, “domestic violence”). They were more likely to use words related to family (“daughter”, “babies”, “children”, “sister”, “mother”, “son”, “parents”, “family”). (Note that “father” does not appear on the list.) They were also more likely to use phrases related to education (“critical thinking”, “my students”, “public schools”), healthcare (“health issues”, “existing condition”, “cancer”, “nurse”, “my doctor”, “mental health”, “blood pressure”), and labor (“time jobs”, “poor working”, “living wage”).

Men, on the other hand, were more likely to use numerical words: “trillion”, “billion”, “zero”. They were also more likely to use words related to economics – “currency”, “economic growth”, “wealth” – and government – “liberals”, “constitutional”, “federal government”, “administration”. The fact that men and women used different words in responding to the same articles may indicate that they found different topics worth emphasizing even when reading the same article.

In addition to emphasizing different topics, men and women also used different language. Women were more likely to use positive phrases like “thank [you so] much”. (47% of 7770 “thank you”s came from female commenters, as opposed to 28% of comments overall). This is consistent with prior findings that women on online forums tend to be more polite [46, 27]. Men and women also differed in how they referred to Barack Obama: men preferred “Barack Obama”, “Mr. Obama”, or “Obama administration”, while women preferred “Pres[.] Obama”. The arguably more positive and less formal female epithet may be due in part to the fact that Barack Obama has enjoyed higher approval ratings among

women than men [52]. Women were also more likely to use the more respectful “affordable care act” or “ACA” rather than “Obamacare” (70% of comments from women versus 63% of comments from men,  $p < 10^{-5}$ ).

To explore whether male and female commenters came from different backgrounds, we looked at discrepancies in commenter self-descriptions by examining phrases following “I am a” or “I’m a” (Table 8). While most commenters did not provide detailed self-descriptions, we did find phrases which showed significant gender skews: all 13 mathematicians, for example, were men, while 14/15 feminists were women.

Self-Description	Female commenters/ Total
canadian	0% (0/14)
gay	0% (0/15)
libertarian	0% (0/14)
mathematician	0% (0/13)
conservative	6% (2/35)
republican	6% (2/34)
gun	10% (3/29)
scientist	11% (3/28)
fan	24% (53/220)
teacher	53% (49/93)
school	62% (21/34)
single	64% (21/33)
family	67% (12/18)
working	68% (17/25)
therapist	73% (8/11)
read	75% (9/12)
trained	75% (9/12)
child	76% (13/17)
public school	80% (8/10)
social worker	83% (10/12)
nurse	90% (26/29)
feminist	93% (14/15)

**Table 8. Self-description words and phrases that showed the most pronounced gender skews; 37% of self-descriptions came from female commenters overall. All gender skews are significant (binomial  $p$ ) after multiple hypothesis correction for number of n-grams (false discovery rate  $< .05$ ).**

We supplemented this automated analysis by using human coding to find specific articles on which men and women expressed different views. One of the most commented-on articles was an open letter from Dylan Farrow [18] which accused Woody Allen of sexual abuse. In the first 48 hours after the letter was posted, more than 3,000 comments were made, and of those with identifiable author gender, 53% were from women. Based on previous findings that women exhibit greater sympathy towards victims of sexual assault [47], we investigated whether there was a significant difference in how sympathetic men and women were to Farrow’s allegations. We used both two methods to quantify sympathy. First, we marked comments that included a link to a widely read rebuttal entitled “The Woody Allen Allegations: Not So Fast” [68], which cast doubt on Dylan Farrow’s claims. Women

were much less likely to cite this rebuttal than men: only 38% of those citing the rebuttal were women, as opposed to 53% overall, a statistically significant discrepancy (odds ratio = .52,  $p = .046$ , logistic regression). Because few comments cited this rebuttal, however, we supplemented this automated analysis with human-aided coding. A randomly selected, gender-balanced subset of comments ( $n = 200$ ) were coded by a researcher blinded as to the gender of the commenter as to whether they expressed support for Dylan Farrow, expressed opposition to Dylan Farrow, or provided no definite view. Men were far more likely to express opposition: 37% of men expressed opposition, as opposed to 11% of women ( $p < 10^{-5}$ ,  $\chi^2$ ). (38% of men expressed support, as opposed to 71% of women; the remainder took no definite position.)

Since analyzing every article in this depth would have introduced multiple-hypothesis problems, we filtered for articles on which men and women seemed likely to express different views by sorting all articles by the cosine distance between the word frequency vector for comments from men and the word frequency vector for comments from women. Among the most divergent articles, we selected ones where the divergences appeared to be semantically significant based on the title of the article or the words that showed divergence (for example, an article in which the divergence was primarily due to the word “probable” was less interesting than one in which it was due to “contraception”). A gender-balanced subset of comments was then classified by a researcher blinded as to the gender of the commenter based on a pre-specified criterion. (We note a potential confound: while researchers were not aware of the true gender of the commenter, they might be biased in their coding by expectations about the commenter’s gender. The effect of this bias is likely small, however, because the criteria for coding were unequivocal in the vast majority of cases. We also confirmed that the coding done by researchers was consistent with that done by a volunteer who was not involved in the study.) We examined three divergent articles: “Your Submissions for Most Neglected Topic”, “Trouble for the Contraception Mandate”, and “The Insanity of Our Food Policy”. The latter two showed statistically significant gender differences; the first did not.

“Your Submissions for Most Neglected Topic” invited readers to submit topics they believed had not received sufficient coverage. Based on the high proportion of male commenters on the global warming articles discussed above, we hypothesized that men might be more likely to submit topics related to global warming; however, we found no statistically significant discrepancy. This may reinforce the idea, discussed above, that female commenters may not actually be less interested in global warming, but are rare on the Dot Earth blog for other reasons.

“Trouble For The Contraception Mandate” argued that employers should not be granted religious exemptions to the Affordable Care Act’s contraception mandate. We hypothesized that female commenters would be more likely than male commenters to support this view, based on previous findings that women are more likely to express support for mandated cov-

erage of birth control medication [45]. Comments were rated by a gender-blinded researcher on whether they expressed support for employer-provided contraception, expressed opposition to employer-provided contraception, or expressed no definite view ( $n = 73$ ; the small  $n$  is due to the smaller number of comments on the article). Comments made by women were significantly more likely than comments made by men to support employer-provided contraception: 0% of comments made by women opposed employer-provided contraception, as opposed to 26% of comments made by men (70% of comments by women, and 53% of comments by men, supported contraception;  $p = .01$ ,  $\chi^2$ ).

“The Insanity of Our Food Policy” argued against giving large farm subsidies to wealthy commercial operations while many Americans struggled to survive on food stamps. This was a long and complex article that commenters reacted to in a variety of ways: some discussed the economic details of subsidies, some focused on the details of providing food to people. Women were more likely to use food-related words like “corn” and “food”; based on this, comments were classified based on whether they focused on food (healthy diets, good or bad eating habits, specific foods, or nutritional information). Comments from women were significantly more likely to be food-related (30% of comments from women as opposed to 6% of comments from men;  $p = .006$ ,  $\chi^2$ ).

## DISCUSSION

These results are troubling. Identifiably female commenters are rare, and the effects of this gender skew are exacerbated by the fact that men and women comment on different forums: the average man will see that only 21% of other comments are made by women (the average woman will see 36%). While it is possible that this gender skew is purely the product of individual volition, we find two lines of evidence that it is not. The first is that women can be rare even on forums discussing topics in which external evidence indicates they have as much interest as men, the Dot Earth blog being a representative example. The second is that women are disproportionately likely to favor anonymity – to not include their last name, or to use only initials. This may indicate lesser comfort on the forums: either with being identified as women, or with having their full identity known. We also found that anonymous commenters received fewer recommendations from other readers. Anonymity may offer benefits to women: the equalization hypothesis predicts that minorities with less power in traditional forums may use anonymity to free themselves from social stereotypes and express unpopular views. Some previous research has indicated that anonymity can increase equalization [59], although other studies have challenged this view [44, 53]. Women may favor anonymity to a greater degree than men in order to increase their influence [19].

Strikingly, we also find that women receive more recommendations from other readers, as do people commenting on forums dominated by the opposite gender. It is somewhat remarkable that in spite of societal prejudices against women offering opinions about football, and men offering opinions about parenting, women who comment on both football fo-

forums (sports/football and sports/ncaafootball) do in fact receive more recommendations than do men, and men who comment on the Parenting Blog receive more recommendations than do women. There are at least two possible explanations for this. One is that men and women might not comment on issues usually dominated by the other gender unless they have something unusually substantive to say. A second might be that when they do choose to comment, they offer a novel perspective which is attractive to other readers.

Our results imply that those with the most to offer to other readers (as measured by their recommendations) are those most often silenced. We see this in two ways: women comment infrequently, though their comments are more positively received, and commenters rarely venture to forums dominated by the opposite gender, though their comments are also more positively received. So views that would be valued go unvoiced, to the detriment of democratic discourse.

Further, the discourse that does occur is not equally representative of all views. As we illustrate both with automated and human-aided analysis, men and women differ in their choice of topics to emphasize, their backgrounds, their language, and their opinions on sexual assault, food policy, and contraception. Given the richness of our dataset, we suspect we have only scratched the surface of the ways in which male and female commenters differ, and that the evolution of natural language techniques will only bring them into sharper focus. The fact that commenters on the *New York Times* are predominantly liberal [23] may make the gender gap smaller in views smaller than it would be on a more diverse forum.

These gender discrepancies in views, combined with our finding that men are overrepresented on the forums, have a simple implication: the views that are heard do not fairly reflect the views that are held. This has profound implications for a democratic society [67], which strives for fair representation of all opinions. For example, the overrepresentation of men has direct consequences for the level of sympathy expressed towards sexual assault victims; this in turn makes it less likely that victims of sexual assault will speak out [1]. Since people are influenced by reading online comments [30, 37], these discrepancies may propagate throughout society. Readers of comments about intimate partner violence, for example, were influenced by the views expressed in those comments [8]. Worse, these discrepancies may be exacerbated because if people falsely believe their own views are unpopular, they may be persuaded not to express them at all, creating a “spiral of silence” [66].

Four things ought to be kept in mind in interpreting our results. First, while the *New York Times* is a large and influential newspaper, it is only a single newspaper. Future analyses should test our findings on other newspapers or online forums to determine how broadly they apply. Any discrepancies are themselves interesting grounds for future study. Second, online discourse is full of complexities and nuances which quantitative analysis may not always capture. For example, our regression models accounted for only a small fraction of the variance ( $r^2 < .1$ ) in the number of recommendations a comment received, indicating that, while our results are highly

statistically significant, there are many factors which determine how people react to a comment which are not captured in our analysis. We have supplemented our quantitative methods with qualitative hand-coding of the comments in order to better capture the nuances in our dataset.

Third, one might question to what extent our results on gender-identified commenters extend to the anonymous commenters. This is certainly a question worthy of future study, but we believe our results about gender-identified commenters are of interest for three reasons. The first is that, while anonymity can certainly influence behavior [44], gender norms run deep enough that it seems likely that what is learned about gender dynamics under non-anonymous conditions extends at least partially to anonymous ones. (The persistence of gender norms explains, for example, why some studies have failed to find that anonymity exerts an equalizing effect [44, 53].) Second, comments made by gender-identified commenters will impact societal gender stereotypes as comments by anonymous commenters cannot: a woman who anonymously makes an insightful comment on an article with mostly male commenters, for example, will not increase the perception that women can offer insight on this topic, but an identifiably female commenter may. Third, many forums do not easily allow anonymity – writing for the *New York Times* rather than reading it, using Facebook, running for political office – and for these forums our results on non-anonymous commenters may be especially salient.

Finally and most importantly, while our quantitative analysis reveals important gender gaps, it cannot always explain the reasons these gaps occur. For example, we do not know if women comment more on articles by women because they are influenced by the author’s gender, or because women are interested in writing and reading about similar topics. In particular, the importance of social forces in producing gender gaps must be kept in mind. For example, our findings that women are more likely to comment on articles related to parenting, and to use words related to family, should not be taken to indicate that women have a greater inherent interest in these topics. Women spend far more time on child care, and do so in part because traditional gender ideologies define child care as feminine; this dynamic is often influenced by government policies that influence how child rearing tasks are allocated [33]. Similarly, our finding that women rarely comment on sports-related articles, while consistent with previous findings [32], must be considered in light of the stigma that confronts women who are interested in sports [55, 6]. Our findings that women are rare in general should be considered in light of the factors which encourage women to stay silent. Men tend to be more confident [36, 38], especially in areas perceived as masculine [4], and are therefore more eager to offer opinions [34]. This may occur in part because women who behave assertively fear social backlash [65, 2]. Previous literature on online discourse makes it clear that social forces that create inequality offline persist online as well [29, 27].

The importance of social forces can be viewed positively, because it means the gender gaps we observe may not be im-

mutable. While the roots of these problems run deep, we offer three potential solutions:

- *Increase the number of female authors.* We find that women are more likely to comment on articles written by women even when we control for the forum on which the article is posted. It is unclear if this is a causal effect of the author's gender, or whether female authors simply choose topics (or favor writing styles) that appeal to female readers. A simple means of discriminating between the hypotheses would be to run an AB test varying whether the gender of an author was displayed on an article, and to examine whether the gender composition of the readership was affected. But even if commenters do not care at all about the author's gender and are responding only to the topic of the article, getting more women to write articles might increase the number of articles written on topics of interest to women, and thereby the number of women who comment. Previous authors have found that female journalists are more likely to feature women in their stories and to challenge gender stereotypes [17, 63].
- *Increase awareness.* Making people aware of the gender skew in forums they read might increase the extent to which they visit forums favored by the opposite gender. This idea has been investigated in other online forums: for example, a project from the MIT Media Lab seeks to decrease gender bias on Twitter by making people aware of the gender split of those they follow [40]. One might also encourage comments from women, and people on opposite-gender forums, by making them aware of the fact that their comments tend to be better-received. This may be especially important for women given our finding that there is a stronger correlation for women between the number of recommendations they receive and how many comments they make, and in light of previous findings that women are more deterred when their comments receive no response [28].
- *Limit collaborative filtering.* The *New York Times*, like many other websites, recommends new content based on what its readers have already consumed. Many websites do this using collaborative filtering algorithms [56], where readers are directed to articles that other "similar" readers have consumed. While this often produces good recommendations, it also reinforces gender segregation, because women are referred to articles that many other women read, and similarly for men. While it seems extreme to discard collaborative filtering entirely, one might experiment with interventions where people were occasionally referred to articles primarily read by the opposite gender.

## CONCLUSION

Our analysis indicates that there is still much work to be done to realize the early hopes that online forums would allow gender equity [27, 29]. Yet there is cause for hope: the finding that women's comments receive more recommendations, particularly on more male forums, is not one we expected, and speaks to the possibility of greater gender equity.

Further, the timing of our analysis is serendipitous: even as we completed our study, we learned that the *New York Times* was going to build a new platform for its comments pages [35], in collaboration with Mozilla and the Knight Foundation. We shared the conclusions in this paper with the team at the *New York Times* in charge of the comments pages and found that they shared our interest in drawing out women's voices. For example, the team has experimented in the past with allowing readers to separate comments by gender to see how men and women offer different perspectives [7]. We hope that some combination of the measures we have discussed, along with continued study, will decrease the polarization of discourse and draw out the silenced who, as the data indicate, have something to say.

## ACKNOWLEDGMENTS

We thank Suzanna Fritzberg, Nathan Matias, Nat Roth, and Jacob Steinhardt for helpful contributions, and the staff at the *New York Times* for their cooperation and guidance.

Male Words	Female	Female Words	Female
hoover	13.2%	huffingtonpost	83.1%
phenomena	19.3%	tariffs	78.5%
embraced	23.0%	breast	78.5%
currency	24.7%	horrified	77.8%
trillion	25.8%	misogyny	77.4%
substantially	26.5%	pregnancy	72.6%
roughly	27.5%	grateful	71.5%
worlds	28.3%	disgusted	71.4%
generate	30.5%	nurse	70.2%
liberals	32.0%	daughter	69.7%
principle	32.7%	thank	69.6%
constitutional	33.4%	woman	69.3%
academic	33.5%	lovely	68.3%
establishment	33.7%	scary	68.1%
virtually	34.2%	babies	67.8%
competitive	34.5%	children	67.4%
collapse	35.0%	sister	67.0%
conflict	35.3%	mom	66.9%
measure	35.9%	female	66.7%
billion	35.9%	struggling	64.7%
rational	37.2%	child	63.9%
zero	37.3%	son	63.6%
failure	37.3%	feel	63.2%
reduce	37.5%	beautiful	63.0%
central	37.6%	kids	63.0%
federal	38.0%	cancer	62.9%
political	38.4%	loved	62.2%
administration	38.4%	parents	61.6%
solution	38.6%	think	61.4%
usa	38.7%	school	61.0%
market	39.4%	tired	61.0%
society	39.5%	family	60.6%
wealth	39.7%	home	60.4%

Male 2-grams	Female	Female 2-grams	Female
mr krugman	23.6%	existing condition	78.4%
government run	25.4%	big corporations	74.2%
american exception- alism	25.5%	public transportation	73.7%
century ago	25.9%	poor working	73.4%
bill rights	28.3%	critical thinking	72.3%
create new	28.3%	rights women	71.9%
national debt	28.8%	oh please	71.9%
tax cuts	29.0%	my doctor	71.9%
election day	29.4%	free trade	70.5%
welfare state	29.4%	pres obama	70.4%
developed countries	30.9%	thank much	70.2%
us constitution	31.6%	time jobs	70.0%
nuclear weapons	31.9%	self employed	69.8%
richard nixon	32.2%	hard earned	69.8%
economic growth	33.3%	living wage	69.6%
executive branch	33.3%	big oil	69.5%
second amendment	34.4%	blood pressure	68.6%
mr obama	35.5%	public schools	68.3%
great depression	35.9%	my heart	68.3%
obama administra- tion	36.8%	domestic violence	67.9%
global warming	37.7%	china india	67.9%
gun control	37.7%	my students	67.9%
barack obama	38.2%	health issues	67.7%
federal government	39.3%	big business	67.1%
		bernie sanders	66.7%
		support family	66.7%
		campaign finance	66.7%
		birth control	66.0%
		chris christie	65.8%
		big pharma	64.7%
		mental health	63.4%

**Table 9. A selection of the words and 2-grams used more often by men and women. Percentages are the percent of articles on which women use the word more frequently than men. All percentages are statistically significantly different from 50% (binomial  $p$ ) after multiple hypothesis correction (false discovery rate  $< .05$ )**

## REFERENCES

1. Ahrens, C. E. Being silenced: The impact of negative social reactions on the disclosure of rape. *American Journal of Community Psychology* 38, 3-4 (2006), 263–274.
2. Amanatullah, E. T., and Morris, M. W. Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of Personality and Social Psychology* 98, 2 (2010), 256.
3. Bartel Sheehan, K. An investigation of gender differences in on-line privacy concerns and resultant behaviors. *Journal of Interactive Marketing* 13, 4 (1999), 24–38.
4. Beyer, S., and Bowden, E. M. Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin* 23, 2 (1997), 157–172.
5. Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent Dirichlet Allocation. *The Journal of Machine Learning Research* 3 (2003), 993–1022.
6. Blinde, E. M., and Taub, D. E. Women athletes as falsely accused deviants: Managing the lesbian stigma. *The Sociological Quarterly* 33, 4 (1992), 521–533.
7. Board, T. E. Who Says Math Has To Be Boring? *The New York Times* (2014).
8. Brossoie, N., Roberto, K. A., and Barrow, K. M. Making sense of intimate partner violence in late life: comments from online news readers. *The Gerontologist* 52, 6 (2012), 792–801.
9. Pew Research Center. Demographics and Political Views of News Audiences, 2012.
10. Chang, S., Kumar, V., Gilbert, E., and Terveen, L. Specialization, homophily, and gender in a social curation site: Findings from Pinterest. *CSCW 2014* (2014).
11. Chesney, T., and Su, D. K. The impact of anonymity on weblog credibility. *International Journal of Human-Computer Studies* 68, 10 (2010), 710–718.
12. Christopherson, K. M. The positive and negative implications of anonymity in Internet social interactions: “On The Internet, Nobody Knows You’re a Dog”. *Computers in Human Behavior* 23, 6 (2007), 3038–3056.
13. Cunha, E., Magno, G., Almeida, V., Gonçalves, M. A., and Benevenuto, F. A gender based study of tagging behavior in Twitter. In *Proceedings of the 23rd ACM conference on Hypertext and social media*, ACM (2012), 323–324.
14. New York Times Developers. APIs. Available online: <http://developer.nytimes.com/docs> (2014).
15. Espfen, E., and Jolly, S. Gender and sex: A sample of definitions. *Bridge (December 2006)* 10 (2010).
16. Ess, C. A plea for understanding—beyond false dilemmas on the net. *Philosophical Perspectives on Free Speech and CMC* (1996).
17. Gallagher, M., et al. Who Makes the News? *Global Media Monitoring Project* (2010).
18. Farrow, D. An open letter from Dylan Farrow. *The New York Times* (2014).
19. Flanagin, A. J., Tiyaamornwong, V., O'Connor, J., and Seibold, D. R. Computer-mediated group work: The interaction of sex and anonymity. *Communication Research* 29, 1 (2002), 66–93.
20. Fogel, J., and Nehmad, E. Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior* 25, 1 (2009), 153–160.
21. Alliance for Audited Media. Top 25 US Newspapers for March 2013 (2013).
22. Garcia, D., Weber, I., and Garimella, V. R. K. Gender asymmetries in reality and fiction: The Bechdel test of social media. *arXiv preprint arXiv:1404.0163* (2014).
23. Gentzkow, M., and Shapiro, J. M. Ideological segregation online and offline. *The Quarterly Journal of Economics* 126, 4 (2011), 1799–1839.
24. Gidengil, E., Blais, A., Nadeau, R., and Nevitte, N. *Women to the Left?: Gender Differences in Political Beliefs and Policy Preferences*. Université de Montréal, 2001.
25. Gilligan, C., and Attanucci, J. Two moral orientations: Gender differences and similarities. *Merrill-Palmer Quarterly* (1988), 223–237.
26. Herring, S. C. Gender and participation in computer-mediated linguistic discourse. *ERIC* (1992).
27. Herring, S. C. Gender differences in CMC: Findings and implications. *Computer Professionals for Social Responsibility Journal* 18, 1 (2000).
28. Herring, S. C. *The Handbook of Language and Gender: Chapter 9, Gender and Power in Online Communication*. John Wiley and Sons, 2008.
29. Herring, S. C., and Stoerger, S. Gender and (a) nonymity in computer-mediated communication. *The Handbook of Language, Gender, and Sexuality* (2013), 567–586.
30. Houston, J. B., Hansen, G. J., and Nisbett, G. S. Influence of user comments on perceptions of media bias and third-person effect in online news. *Electronic News* 5, 2 (2011), 79–92.
31. Huffaker, D. A., and Calvert, S. L. Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication* 10, 2 (2005).
32. American Press Institute. Social and Demographic Differences in News Habits and Attitudes (2014).

33. Kan, M. Y., Sullivan, O., and Gershuny, J. Gender convergence in domestic work: Discerning the effects of interactional and institutional barriers from large-scale data. *Sociology* 45, 2 (2011), 234–251.
34. Kantor, J. Harvard Business School Case Study: Gender Equity. *The New York Times* (2013).
35. Kaufman, L. Mozilla to Develop Comments Platform with New York Times and Washington Post. *New York Times* (2014).
36. Kay, K., and Shipman, C. The Confidence Gap. *The Atlantic* (2014).
37. Lee, E.-J., and Jang, Y. J. What do others' reactions to news on internet portal sites tell us? Effects of presentation format and readers' need for cognition on reality perception. *Communication Research* 37, 6 (2010), 825–846.
38. Lenney, E. Women's self-confidence in achievement settings. *Psychological Bulletin* 84, 1 (1977), 1.
39. Li, Q. Computer-mediated communication and gender difference: A meta-analysis. Presented at the Annual Meeting of AERA (2005).
40. Matias, N., and Szalavitz, S. Followbias. Available online: <http://followbias.com/> (2014).
41. McCright, A. M. The effects of gender on climate change knowledge and concern in the American public. *Population and Environment* 32, 1 (2010), 66–87.
42. Meyer, R., and Cukier, M. Assessing the attack threat due to IRC channels. In *International Conference on Dependable Systems and Networks, 2006*, IEEE (2006), 467–472.
43. Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., and Rosenquist, J. N. Understanding the demographics of Twitter users. *ICWSM 11* (2011), 5th.
44. Mohai, P. Gender differences in the perception of most important environmental problems. *Race, Gender & Class* (1997), 153–169.
45. Moniz, M. H., Davis, M. M., and Chang, T. Attitudes about mandated coverage of birth control medication and other health benefits in a US national sample. *JAMA* (2014).
46. Morahan-Martin, J. Women and the internet: Promise and perils. *Cyber Psychology & Behavior* 3, 5 (2000), 683–691.
47. Mori, L., Bernat, J. A., Glenn, P. A., Selle, L. L., and Zarate, M. G. Attitudes toward rape: Gender and ethnic differences across Asian and Caucasian college students. *Sex Roles* 32, 7-8 (1995), 457–467.
48. Muller, B. A Python library to determine gender based on first name. Available online: <https://pypi.python.org/pypi/SexMachine/> (2014).
49. Ng, A. Y., Jordan, M. I., Weiss, Y., et al. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems 2* (2002), 849–856.
50. Pew Research Center, State of the News Media. Top 10 Online Newspapers Worldwide, (2012).
51. Pierson, E., and Do, C. What About the Women? Available online: <http://blog.coursera.org/post/79073490420/what-about-the-women> (2014).
52. Gallup Politics. Gender Gap in Obama Approval Constant Since Term Began (2012).
53. Postmes, T., and Spears, R. Behavior online: Does anonymous computer communication reduce gender inequality? *Personality and Social Psychology Bulletin* 28, 8 (2002), 1073–1083.
54. Rode, J. A. A theoretical agenda for feminist HCI. *Interacting with Computers* 23, 5 (2011), 393–400.
55. Sartore, M. L., and Cunningham, G. B. The lesbian stigma in the sport context: Implications for women of every sexual orientation. *Quest* 61, 3 (2009), 289–305.
56. Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, ACM (2001), 285–295.
57. Selfe, C. L., and Meyer, P. R. Testing claims for online conferences. *Written Communication* 8, 2 (1991), 163–192.
58. Shapiro, R. Y., and Mahajan, H. Gender differences in policy preferences: A summary of trends from the 1960s to the 1980s. *Public Opinion Quarterly* 50, 1 (1986), 42–61.
59. Siegel, J., Dubrovsky, V., Kiesler, S., and McGuire, T. W. Group processes in computer-mediated communication. *Organizational Behavior and Human Decision Processes* 37, 2 (1986), 157–187.
60. Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., and Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2013), 1631–1642.
61. Larivier, V., Ni, C., Gingras, Y., Cronin, B., Sugimoto, C. Bibliometrics: Global Gender Disparities in Science. *Nature* 504, (2013), 211 – 213.
62. Sullivan, M. Questions and Answers On How The Times Handles Online Comments From Readers. *The New York Times* (2012).
63. Thelwall, M. Social networks, gender, and friending: An analysis of Myspace member profiles. *Journal of the American Society for Information Science and Technology* 59, 8 (2008), 1321–1330.

64. Thomson, R., and Murachver, T. Predicting gender from electronic discourse. *British Journal of Social Psychology* 40, 2 (2001), 193–208.
65. Tinsley, C. H., Cheldelin, S. I., Schneider, A. K., and Amanatullah, E. T. Women at the bargaining table: Pitfalls and prospects. *Negotiation Journal* 25, 2 (2009), 233–248.
66. Walther, J. B., and Jang, J.-w. Communication processes in participatory websites. *Journal of Computer-Mediated Communication* 18, 1 (2012), 2–15.
67. Weber, P. Discussions in the comments section: Factors influencing participation and interactivity in online newspapers' reader comments. *New Media & Society* (2013).
68. Weide, R. B. The Woody Allen Allegations: Not So Fast. *The Daily Beast* (2014).
69. Wilson, J. Hate Sinks. *The New Inquiry* (2014).
70. Yates, S. J. Gender, identity and CMC. *Journal of Computer Assisted Learning* 13, 4 (1997), 281–290.