Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum

Alice H. Wu*

December 2017

Abstract

Stereotyping, the process of ascribing characteristics based on group membership, can exaggerate the contrast between *in-group* and *out-group* and foster an unwelcoming atmosphere. This paper examines the existence and extent of gender stereotyping on Economics Job Market Rumors, an anonymous online forum with academic and professional purposes. First, I use a Lasso Logistic model to directly capture the gender stereotyped language. Discussions about women tend to focus more on physical appearance or family information, whereas discussions about men are more on their academic or professional aspects. The topic analysis provides further evidence on this finding from a more aggregate perspective. In addition, I develop an econometric framework to study gender stereotyping in the dynamics of a conversation. I find that there is a significantly stronger deviation from an *Academic/Professional* focus when there is a prior mention of women; in contrast, the deviation from a *Personal/Physical* topic is stronger if the prior post is about men rather than women. Last, female economists tend to receive more attention online than their male counterparts, a pattern that further emphasizes the need to reduce stereotyping and maintain an inclusive environment.

*I would like to thank my advisor David Card for his invaluable guidance and support. I also thank Janet Currie, David Romer, Pat Kline, Amanda Pallais, Jessie Rothstein, Ulrike Malmendier, Wei Jiang, Alessandra Fenizia, Sydnee Caldwell, Jonathan Tan, Justin Wolfers, Patrick Button, Dick Startz, seminar participants at Bowdoin College, UC Berkeley and Harvard Business School for very helpful comments and suggestions. All errors are my own. Contact information: alice.hw15@gmail.com.

Despite the remarkable gains in educational attainment in recent decades, women are still underrepresented in math-intensive fields like economics, engineering and computer science (Ceci et al. 2014; Bayer and Rouse 2016; Kahn and Ginther 2017). The persistent gender gap can consolidate the perception of *in-group* versus *out-group*, and social identity theory suggests that members of the well-represented *in-group* are likely to engage in stereotyping – the act of ascribing characteristics based on group membership – to emphasize "intragroup similarity" and "intergroup differences" (Tajafel and Turner 1986; Oakes et al. 1994).

Although there is a rising literature in economics formally modeling stereotype beliefs (Bordalo et al. 2016a) and testing for them in lab experiments (Bordalo et al. 2016b), it remains challenging to capture the stereotyping behavior in real world settings and evaluate its impact on the overall environment. One difficulty is that the day-to-day interactions between people are not easily observable. Another difficulty is that subjects who are concerned about their social or political correctness would not necessarily reveal their true attitudes in public.

This paper aims to fill in this gap of the literature by examining the existence and extent of gender stereotyping in everyday "conversations" that take place online between people in economics. I use text scraped from Economics Job Market Rumors[1] (EJMR), an online forum established to share information about job applications and results in each year's hiring cycle, though it is now active all year round. EJMR users post anonymously about economics-related or miscellaneous issues. Anonymity presumably eliminates any social pressure participants may feel to edit their speech, and thus creates a natural setting to capture what people believe but would not openly say. I focus on threads initiated or updated within the last four years, from October 2013 to October 2017. About 62% of the threads in my dataset include at least one post that directly addresses female(s) or male(s). Gender-related threads are also more popular: the mean number of posts per thread is 11 in the overall sample, but 14 in the gender sample. In particular, a thread starting with a title related to women contains about 2 more posts than one starting with a title related to men.

I start from the question whether women and men are portrayed differently on EJMR. Assuming an underlying causal relationship between the gender of the subject being discussed and the characteristics the poster would emphasize, I take an inversion step to infer gender from the

[1] More information about EJMR on https://www.econjobrumors.com/topic/about-ejmr

text, a strategy often used in the analysis of high-dimensional textual data (Taddy 2013; see Gentzkow 2017 for a summary). I train a Lasso-Logistic model on over four hundred thousand *Female* and *Male* posts, and it identifies the words with meaningful predictive power[2] on. The five words most uniquely associated with *Female* posts, in descending order of the marginal effect on $Pr(Female = 1|text)$, are: "hotter", "pregnant", "plow", "marry", and "hot", , while the top five associated with *Male* posts are: "homo", "testosterone", "chapters", "satisfaction", and "fieckers". The moderation of the forum is based on both automatic censoring and reports by users[3]. However, the terms captured by the Lasso Logistic model suggest that either the automatic system is not robust, or the EJMR users themselves do not find it necessary to report content of potential discrimination. A closer look into the contrast between top "female" vs. top "male" terms reveals that women are more likely to be characterized by their physical appearance or personal information, whereas men are more associated with academic or work-related content. To make inferences on the pervasiveness of the stereotyped language, I also consider the frequency of each word, and it gives a similar picture of the differential portrayal of women and men.

From a more aggregate perspective, I analyze the topics in gendered discussions. I measure the total occurrences of words under two topics of interest: *Academic/Professional* and *Personal/Physical*. The first topic is consistent with the original purposes of the forum to share job market information and discuss issues in economics, while the second topic includes descriptions of one's physical appearance or family information that can be inappropriate in a professional setting. At the post level, on average a *Male* includes 3 academic or professional terms, whereas a *Female* post contains about 1.35 terms less under this topic, a significant 45% decrease. The gender gap in *Academic/Professional* is robust under different sample restrictions by gender classifiers[4]. For the other topic, *Female* posts consistently includes about 1.1 *Personal/Physical* terms on average, more than double of what shows up in a typical *Male* post. At the thread level, I consider the mean number of terms under each topic. Relative to threads mostly centered on men, a thread with more *Female* posts than *Male* posts contain over 50% less academic terms, but significantly more words in *Personal/Physical*.

The findings in the static analysis above can reflect two forces in tandem. First, as women

[2] That is, the marginal effect of the occurrence of each word on $Pr(Female = 1|text)$ is nonzero.
[3] Moderation policy: https://www.econjobrumors.com/topic/request-a-thread-to-be-deleted-here
[4] Gender classifiers include words like "he" or "she", which I use to identify the gender of the subject of each post.

are underrepresented in economics, there are less mentions of women in an academic or professional discussion. Second, the theory of stereotyping in Bordalo et al. 2016a suggests that the contrast between women and men can be exaggerated as a result of representativeness-based discounting: that is, the *Academic/Professional* aspects of women are under-weighted, whereas the *Personal/Physical* aspects are over-weighted, relative to men.

To put stereotyping in a dynamic setting, I examine the flow of the conversation empirically, in particular, whether a thread is persistent in each topic, and how gender can potentially affect such persistence. I focus on gender-related threads that include at least one *Female* or *Male* post. Within each thread, a post can discuss about *Female*, *Male* or be *Neutral*, i.e. not directly related to gender. There is a mean reversion pattern on average: when the prior post talks about *Academic/Professional*, regardless of the gender in the prior, the next post is 5.0 ppt significantly less likely to stay on the same topic. Relative to the neutral group where the prior posts are genderless - *Neutral*, the deviation from an *Academic/Professional* focus is about 52% stronger in the *Female* group, and 32% stronger in the *Male* group. The effect of a mention of female(s) in the prior post is significantly different from that of a mention of male(s), with a p-value of 0.0001. I also break down the results by the initial conditions set up by the topic and the first post of each thread. The contrast between the effects of *Female* and *Male* on the persistence of an academic topic becomes even more salient under a thread that starts from an academic theme without any mention of women and men. From a behavioral perspective, a comment on the research by a female may contradict one's prior beliefs about women, resulting in an immediate deviation from the academic topic to protect the presumed stereotype. In contrast, the deviation from *Personal/Physical* is smaller in the *Female* group, which can reflect a confirmation bias.

Finally, I present a difference-in-difference analysis on the attention received by a comparable set of 190 female and 190 male high-profile economists who rank among the Top 5% of Authors on the RePEc ranking[5], and a second analysis of a cohort of 204 assistant professors (45 women, 159 men) from Top 20 economics departments[6] in the United States. I estimate the amount of attention each person receives by the number of results returned via a name search on EJMR.

[5] RePEc ranking of Top 5% Authors (Last 10 Years Publications), as of September 2016: `https://ideas.repec.org/top/top.person.all10.html`

[6] based on U.S. News ranking of best graduate programs in Economics as of 2013 and 2017, and RePEc ranking of top Economics Departments.

Among high-profile economists, women tend to get more attention than their male counterparts, and the difference is wider for relatively less prominent economists. Among junior faculty, women working at the top 5 departments are discussed more than men on the forum, but this trend is reversed for those at lower-ranked departments.

The rest of the paper is structured as follows. Section 1 provides an overview of the EJMR data and the construction of the gender sample, and discusses the popularity of threads in relation to gender. Section 2 presents the Lasso Logistic model I use to infer gender from the text and directly capture the gender stereotyped language online. Section 3 analyzes the topic differences in *Female* vs. *Male* posts, and how gender can affect the dynamics of the conversation. Section 4 uses an alternative design to analyze the attention comparable economists of different genders receive. Section 5 discusses the next steps and concludes.

1 EJMR Data and Sample Overview

As of October 28th, 2017, there were over three hundred thousand threads on the site of EJMR forum in a span of seven years. The threads are organized in reverse chronological order, by the time of each's latest update. I take the following two steps to create my dataset. First, I scrape the main pages of the forum, numbered from 1 to $8,750$. A typical page contains 35 threads, and it records each thread's title, the time of the latest update, the number of posts, the number of views, and the votes by users. I then scrape the posts on the first page and the last page (if a thread exceeds one page[7]) of each thread initiated or updated within the last four years, from October 2013 to October 2017. As a result, I obtain a dataset of $2,217,046$ posts across $223,475$ threads.

Without a pre-existing dictionary, I use the open-vocabulary strategy (Schwartz et al. 2011) to consider the most frequent $10,000$ words that emerge from the raw text. I record the word counts in a N-by-$10,000$ sparse matrix, where $N = 2,217,046$, the number of posts. In order to identify the gender-related posts, I extract a list of gender classifiers from the top $10,000$ words, which contain 57 words indicating females, and 236 words indicating males. The most straightforward classifiers are pronouns - “she”, “he” etc., while others can refer to a group or identity such as “women”,

[7] A typical thread contains at most 20 posts on each page.

"men", "wife", "husband". The imbalance in the total number of classifiers is mainly driven by the pattern that more male first names or male economists' last names emerge among the $10,000$ words than female ones. Based on the characteristics of the classifiers, I subsequently divide them into four groups, and define four increasingly restrictive levels as illustrated in Figure 1. Level 1 uses all classifiers, whereas Level 4 restricts to pronouns only. Such specifications are particularly useful for robustness checks in later sections. The more restrictive levels also help exclude cases where posters refer to themselves as "bros" or "guys" but the topic they are discussing is not gendered.

At each level, I define a post to be $\textit{Female} = 1$ ("female") if it includes any word indicating a female, $\textit{Female} = 0$ ("male") if it includes any word indicating a male, and NA ("neutral") otherwise. Under this classification rule, at Level 1, there arise $44,081$ "duplicate" posts that contain both female and male classifiers. To resolve this issue, I design a Lasso-Logistic model to infer gender from words other than the classifiers. This predictive model helps re-classify $14,028$ (31.8%) of the duplicate posts as "female", and the other $30,053$ posts as "male". Section 2.1 and Appendix A discuss the model and it training process in detail and display a list of words with the strongest predictive power for gender.

Table 1 provides a summary of the number of female and male posts identified at each level. Using all gender classifiers (Level 1), I find $444,810$ posts to be either about females or males, which make up over 20% of the posts in the entire dataset. The gender-related posts span across $138,477$ threads, about 62% of all threads in the past four years. I consider a thread to be related to gender if its title or at least one of its posts is discussing about females or males, i.e. $\textit{Female} \in \{0,1\}$. In later analysis, I examine the differences between "female" and "male" posts directly, and then extend to all $1,736,204$ posts within gender-related threads to study the flow of the conversation.

Popularity of Threads in relation to Gender

Threads in the gender sample tend to be more popular: the mean number of posts per thread is 11 in the overall sample, whereas a gender-related thread at Level 1 attracts 14 posts on average[8]. From a user's perspective, he or she first reads the title of a thread, and then decides whether to

[8] I use the number of posts shown on the main pages of the EJMR forum to calculate the means. The numbers are higher than $\frac{\text{No. Posts}}{\text{No. Threads}}$ in Table 1 because the dataset only preserves posts on the *First* page and the *Last* page (if more than one page) of each thread. About 2% of the threads span over 3 pages or more, and for them, posts in the middle pages are not scraped.

continue reading the posts under it and contribute to the discussion. Based on this observation, I further break down the popularity measure by gender in the title, which could be *Female*, *Male*, or *Neutral* (not related to gender). Table 2 shows that within the gender sample, a thread with a *Neutral* title contains 15 posts on average. A typical *Female* title attract about 12 posts, lower than *Neutral*, but about 2 more significantly than *Male*. The number of views per thread is an alternative measure of popularity in column (2). Gendered titles also get significantly less views than *Neutral* ones, but the difference between *Female* and *Male* is small and insignificant under this measure. In other words, *Female* titles initially get about the same amount of interests (measured by no. views) as *Male* ones, but there are some underlying incentives that motivate EJMR users to comment within a thread, resulting in a significant gap in the number of posts.

To further illustrate this point, Figure 2 plots the distribution of the no. posts under *Female* versus *Male* titles. For purposes of illustration, I "right-censor" the number of posts at 40 in the plot[9]. For threads with *Male* titles, the mass of the distribution is more highly concentrated on the left than that of threads with *Female* titles.

2 Capturing the Gender Stereotyped Language

I use a Lasso-logistic model to predict the gender a post discusses about by the counts of the most frequent $10,000$ words, excluding the gender classifiers and additional last names[10]. Assuming an underlying causal relationship between the gender of the subject and the language patterns, I take an inversion step[11] to infer gender from text and the estimated model identifies words most uniquely associated with each gender. At the meantime, the model serves as an alternative classification strategy to resolve "duplicate" posts that include both "female" and "male" classifiers.

[9]For censoring I code the no. posts as 40 if it is ≥ 40. Note there are only about 2% of all threads that contain more than 40 posts.

[10]The last names of celebrities (non-economists) and the names from which I cannot tell the gender are not used as gender classifiers. As a result, $9,545$ words remain as predictors.

[11]The inversion strategy that creates a map from high-dimensional text to lower dimensional attributes of interest is often used in logistic regression models (e.g., Taddy 2013; also see Gentzkow 2017 for a summary.)

2.1 Lasso-Logistic Model and Training Process

Given a post and a corresponding vector of token counts W_i, assume the posterior probability is:

$$
\begin{aligned}
P(Female_i = 1|W_i) &= \frac{exp(\theta_0 + W_i'\theta)}{1 + exp(\theta_0 + W_i'\theta)} \\
P(Female_i = 0|W_i) &= \frac{1}{1 + exp(\theta_0 + W_i'\theta)}
\end{aligned}
$$

Write the likelihood of each observation as:

$$
P(Female_i|W_i) = P(Female_i = 1|W_i)^{Female_i} \times P(Female_i = 0|W_i)^{(1-Female_i)}
$$

Assume the observations are independent, I estimate the coefficients on word counts that maximizes the log likelihood under a constraint on $\|\theta\|_1$ - the ℓ_1-norm as follows:

$$
\hat{\theta}_\lambda = argmin_\theta \; - log(\Pi_{i=1}^{N} P(Female_i|W_i)) + \lambda||\theta||_1 \tag{1}
$$

Lasso regularization, i.e. the ℓ_1-norm penalty, promotes sparsity as the estimator shrinks the coefficients on variables with little explanatory power to zero, and thus is particularly useful for variable selection in high dimensional data. Lasso has become a popular approach in computational linguistics (e.g. Eisenstein et al. 2011). Gentzkow et al. (2016) also use this strategy to identify the most partisan phrases in Congressional speeches. In this case, the Lasso-logistic model sorts out words with the strongest predictive power on gender. The estimator $\hat{\theta}_\lambda$ is biased, but the variance of the model is reduced, and tends to yield more accurate predictions.

There are $401,734$ *non-duplicate* posts that include only "female" words or only "male" words at Level 1. I use 75% of them, i.e. $300,788$ posts, to train the model and select an optimal tuning parameter λ^* through 5-fold cross validation. I select the best p-score threshold by the prediction accuracy on the remaining 25% as the test set ($p^* = 0.40$ according to Appendix Figure A1). Finally, if the predicted probability of a *duplicate* post discussing females is ≥ 0.40, I re-classify it to be a *Female* $= 1$ post, and a *Female* $= 0$ post otherwise. As a result, 31.8% of the duplicate posts that include both "female" and "male" classifiers are re-classified to *Female* $= 1$, and the rest to *Female* $= 0$.

2.2 Word Selection

As for the variable selection, the coefficients on $5,034$ words are shrunk to zero; that is, they are considered irrelevant to the classification of gender in each post. I sort the remaining words by each's marginal effect - the increase in the probability of the subject of a post being *Female* when a given word occurs once more.

The left half of Table 3 displays the top 30 words with the strongest predictive power for gender at Level 1. None of the most "female" words are related to economics or the job market. Instead, most of them are related to physical appearance or attributes of women. The words "hot", "attractive", and "beautiful" increases the predicted probability of a post discussing about *Female* by approximately $24.0\% - 27.1\%$. Although some of these words might seem positive by themselves, it is arguably inappropriate to discuss one's look in a professionally-oriented forum. For example, there is a thread titled "Cute, unmarried HRM AP is doing a seminar at my school. Can I ask her out?" [12], which judges a female economist based on her appearance instead of her research ability. Words about personal or family information such as "marry", "pregnancy", "dating" also emerge on this list.

In contrast, the words most uniquely associated with "male" posts are more academically and professionally oriented. Terms like "macroeconomics", "supervisor", "adviser", and "RFS" (The Review of Financial Studies) and names of institutions are among the top 30 most "male" words. The list still contains some very offensive terms, which might suggest an unwelcoming environment in a broader sense. However, the drastic differences in the gender stereotyped language at the word level do illustrate a differential treatment of *in-group* (men) and *out-group* (women).

To check the robustness of the words selected by Lasso, I train this predictive model on posts identified by Level 4 gender classifiers, and the results are shown in the right half of Table 3. Level 4 uses the most restrictive set of classifiers - "he", "she" etc. (see Figure 1). There is a 60% turnover rate among the top 30 "male" words at Level 4 relative to Level 1. Additional terms related to research or one's intellectual ability occur, e.g. "RePEc" and "genius". Academic terms such as

[12] This thread ($id = 143907$ in the final dataset) was initiated and last updated 2 years ago. It contains 20 posts and gets $1,238$ views. It lies in the top quintile of the popularity distribution as in Figure 2.

“adviser”, “supervisor” and “Nobel” show even stronger marginal effects on predicting posts about males[13]. On the one hand, using more restrictive gender classifiers does help identify “male” posts that are more academic or professionally oriented. On the other hand, the comparison between the top “female” words identified at Level 4 versus Level 1, with a mere 30% turnover rate, shows that the discussions related to women consistently tend to deviate from academic and professional topics, no matter how restrictive the sample selection is. Words like “nurse” or “humanities” emerge at Level 4, but they are not related to economics or the job market, which again reveals a strong tendency to promote gender stereotypes.

To make inferences on the pervasiveness of gender stereotyping, I consider the frequency of the words with the strongest association with gender (Appendix Table A1), and compare it with the most commonly used words that occur in $Female = 1$ and $Female = 0$ posts respectively (Appendix Table A2). It is true that Lasso picks up terms such as “hotter” and “chapters” that are mostly unique to one gender but that may not be frequent in the overall sample. However, words sorted by frequency reveal similar patterns: the five most frequent non-symbol words in *Female* posts are “life”, “work”, “hot”, “love”, “sex”, whereas the most frequent in *Male* ones are “work”, “paper”, “job”, “economics” and “great”.

To some extent, the analysis at the word level is similar to the idea of the Implicit Association Test in psychology, which capture one’s implicit bias by how fast he or she relates certain characteristics to different groups (Greenwald 1998). However, the patterns revealed here go beyond implicit biases, as the words occur in real online discussions among people in the economics community. The existence and extent of gender-stereotyped language deviates from the putative academic and professional purpose of this forum, and both illustrates and contributes to an unwelcoming atmosphere online.

3 Static and Dynamic Topic Analysis

As the word selection above reveals a divergence in themes between discussions about women and men, here I develop a more aggregate approach to study the topic differences at both the post

[13] At Level 1, an additional occurrence of “adviser”, “supervisor” and “Nobel” increases the chance of a post discussing about males by 14.1%, 14.6%, and 12.9% respectively. At Level 4, the marginal effects increase to 15.3%, 15.3% and 14.3% in the same order.

level and the thread level. In addition, I examine the flow of the conversation, in particular the persistence of a topic and its interaction with gender. I manually classify the top $10,000$ words into 15 categories. Table 4 explains how I group certain categories to consider two main topics of interest: (i) *Academic/Professional*; (ii) *Personal/Physical*.

3.1 Static Topic Analysis

A. Topics at the Post level

First, I restrict my analysis to gender-related posts ($Female \in \{0, 1\}$), and the sample size varies by the level of gender classifiers defined in Figure 1. For each post, I count the number of occurrences of words from each category, which provides an explicit representation of the post's association with a given topic. For example, a post that includes eight economics terms is considered more academic than a post with only three such terms. I use two benchmark models to estimate the gender differences in topics. The first model looks at the effects of gender on the sum of word frequencies in each topic, while the second uses an indicator for whether any word from a given topic occurs:

$$(i): \; Topic_i = \gamma_0 + \gamma_1 Female_i + e_i \tag{2}$$

$$(ii): \; D_i = \theta_0 + \theta_1 Female_i + u_i \tag{3}$$

$$Topic \in \{No.\ Academic/Professional\ terms, No.\ Personal/Physical\ terms\}$$

$$D_i := 1[Topic_i > 0]$$

Table 5 presents the estimates of model (i) on the *Academic/Professional* topic. At Level 1 where all gender classifiers are used to identify gender-related posts, it shows that on average there are 3.00 academic or job-related words in each post associated with a male, but 1.35 fewer (a significant 45.0% decrease) when the post is associated with a female. In terms of probabilities, as shown in Table 6, 58.8% of the "male" posts include at least one academic/work term, while 12.2% of "female" posts do.

One potential issue with using Level 1 gender classifiers is that they pick up a large number of posts talking about "girlfriend" or "boyfriend" etc. that are necessarily not academic/work oriented.

The higher the level of classifiers, the more likely it is that the post focuses on people within the Economics community, including professors, colleagues and candidates. The sample restriction through gender classifiers is not a perfect filter, but Level 4 (using pronouns only) does successfully reduce the sample size by over 50% relative to Level 1, and the comparison across levels provides an opportunity for a robustness check. I test the models on the gender sample identified by each level, and find that the null hypothesis $E[Academic_i|Female_i = 0] = E[Academic_i|Female_i = 1]$ is rejected at 0.1% significance level across all four levels. The relative percentage gap in the number of *Academic/Professional* terms is estimated to fall between 44.1% and 47.5%, with Level 3 and Level 4 showing larger differences. As the sample becomes more selective by gender classifiers, the average number of *Academic/Professional* terms increase for both genders, which helps illustrate the validity of the sample restrictions - that is, the posts identified are more centered on the Economics community.

For the other topic - *Personal/Physical*, I also estimate the benchmark models on posts identified by each level of gender classifiers. As shown in Table 7, at Level 1, a "female" post on average includes 1.12 terms related to personal information or physical attributes, almost three times of what occurs in an average "male" post. Even though the overall number of *Personal/Physical* terms seems smaller than the number of *Academic/Professional* ones, it is worth noting that this category includes a significant portion of words related to physical appearance or sexual content, which are arguably inappropriate in a forum for economists. In terms of probability (Table 8), 46.9% of "female" posts at Level 1 includes at least one term associated with this topic, more than double of the proportion of "male" posts with such terms. The gender difference shrinks as the sample becomes more restrictive, but the shrinkage is mainly driven by a small increase in the number of such terms in "male" posts, and on average a "female" post consistently has about 1.1 terms under this topic.

B. Topics at the Thread level

To capture a more complete picture of the gender-related discussions, I extend the static topic analysis to threads that contain at least one *Female* = 1 or *Female* = 0 post. Using Level 1 gender classifiers, I construct a panel dataset that contains $1,736,204$ individual posts[14] under

[14] $444,810$ posts in this panel dataset are *Female* = 1 or *Female* = 0 posts at Level 1 (see Table 1).

138,477 gender-related threads (see Table 1).

For each thread, I define $\%Female - \%Male = \frac{nFemale - nMale}{nPosts}$, the difference between the fraction of $Female = 1$ posts and that of $Female = 0$ ones, as an aggregate measure of the representation of "female" posts relative to "male" ones. I divide this measure into quartiles, where the first quartile$[-1, -0.333]$ corresponds to threads that most heavily center on men while the last quartile $[0, 1]$ refers to threads that include more posts related to females than to males.

The corresponding benchmark model at the thread level is:

$$\overline{Topic}_t = \gamma_0 + (\%Female - \%Male)'_t \gamma_1 + e_t \tag{4}$$

where the $\overline{Topic}_t$ refers to the mean of *Academic/Professional*; (ii) *Personal/Physical* terms across all posts within a thread t, and $(\%Female - \%Male)_t$ is a vector indicators for quartiles. Table 9 shows the outputs for both the OLS and the weighted version where I use the number of gender-related posts within a thread as its weight. Threads mostly centered on men (Quartile 1) on average have 4.00 *Academic/Professional* terms per post, or 2.47 when I put higher weights on more gender intensive threads. The more "female" posts a thread contains, the lower the mean number of *Academic/Professional* terms[15].

Relative to Q1, threads in Q4 where the number of "female" posts exceed that of "male" posts contain about 53.3% - 67.4% less *Academic/Professional* terms. This relative gap is even wider than the estimates at the post level ($44.1\% - 47.5\%$ in Table 5). Threads mainly discussing about men might be more persistent in an *Academic/Professional* topic, whereas those more intensively about women might not start as an academic discussion, or deviate from its original academic focus as the conversation evolves. These potential explanations require a dynamic analysis that I will discuss in Section 3.2

As for the *Personal/Physical* topic, the unweighted model shows that threads in Quartile 4 contain about 16.7% significantly more terms about personal information and physical attributes. In contrast, the relative increase becomes much more drastic, rising to 123.1%, when I use the number of gender-related posts as weights. The weights seem to have a larger influence on results

[15]The unweighted version in (1) of Table 9 shows that Quartile 3 on average contains slightly more *Academic/Professional* terms than Quartile 2, but that might be driven by threads that actually contain very few "female" or "male" posts, resulting in a measure of (*%Female-%Male*) near 0. Using the number of gender-related posts addresses this concern, and it does show a monotonically decreasing trend in (2).

for this topic, which is potentially because the words under *Personal/Physical* are more directly associated with gender discussions than the *Academic/Professional* words. Also, note that the weights lead to a shrinkage of the differences between Q2/Q3 and Q1, and this finding is in line with the observation that when $\%Female - \%Male$ is close to 0, it is either because a thread is very balanced in the number of posts related to "female" or "male" or because the discussion overall is not really related to gender. Therefore, it is important to put more weight on threads that contain more gender-related posts.

To summarize, the static analyses at both the post level and the thread level show that discussions about women are significantly less academically or professionally oriented on average, and significantly more about personal information or physical appearance. This conclusion is consistent with the gender-stereotyped language captured by the Lasso-logistic model at the word level.

3.2 Dynamic Topic Analysis

Moving beyond the static analysis of topic differences, I develop a dynamic approach to study the flow of the conversation in gender-related threads[16]. Intuitively speaking, gender stereotyping can be examined in a dynamic setting, as an extension of the stereotype model in Bordalo et al. (2016a): subjects might react to new information about *in-group* (males) vs. *out-group* (females) differently, in particular when the information contradicts their prior beliefs in certain characteristics or threatens a preexisting contrast between groups.

A. Econometric Framework

I define *persistence* as the tendency to stick with the same topic within a thread. In theory, the current post can be a reaction to both the initial topic (in the title and the first post) and one or more prior posts within the thread. For purposes of illustration, I focus on the persistence between adjacent posts. I test for two hypotheses: (1) whether the topic of the current post (p) depends on its prior one ($p-1$), i.e. AR1 process; (2) whether the *persistence* becomes stronger or weaker when the prior post is directly discussing about women or men ($Female_{t,p-1} \in \{0,1\}$), under different initial conditions set up by the *Title* & the *First* post of each thread.

[16] A thread is "gender-related" if its title or at least one of its post is discussing women or men.

First, I assume the data-generating process (DGP) in topics between adjacent posts to be

$$D_{t,p} = \beta_0 + \beta_1 D_{t,p-1} + \alpha_t + u_{t,p} \tag{5}$$

where $D_{t,p}$ is a dummy for containing any *Academic/Professional* or *Personal/Physical* respectively in the p-th post of thread t, and α_t is an unobserved component determining the underlying theme of a thread that influences all posts within it.

A user is prompted to click on a thread based on its *Title* shown on the main pages of the forum. The *First* post, written by the same person who started the thread in most cases, also plays an important role in shaping the theme of the thread. All other posts within the thread are presumably equally informative of the unobserved "thread" effect α_t. Therefore, I assume α_t can be absorbed linearly by the initial topic in its title and the first post, and the mean topic across all posts[17]. Formally,

$$\alpha_t = \phi_0 + \phi_1 D_{t,0} + \phi_2 D_{t,1} + \phi_3 \overline{D}_t + \epsilon_t \tag{6}$$

where the residual ϵ_t is uncorrelated with the remaining observable characteristics on topics. Therefore, I estimate the reduced form model as follows.

$$D_{t,p} = \beta_0 + \beta_1 D_{t,p-1} + (\phi_1 D_{t,0} + \phi_2 D_{t,1} + \phi_3 \overline{D}_t) + \theta\, 1[\textit{last page}] + \nu_{t,p} \tag{7}$$

where in an abuse of notation β_0 also absorbs the constant ϕ_0 in (6), and $\nu_{t,p} = u_{t,p} + \epsilon_t$. Since for each thread I scrape the first page and the last page (if over one), I add $1[\textit{last page}]$, an indicator for posts on the last page to control for potentially systematic differences between posts toward the end of the discussion and those in the beginning.

To examine whether gender in the prior post, denoted by $\textit{Gender}_{t,p-1} \in \{\textit{Female}, \textit{Male}, \textit{Neutral}\}$[18], shifts the topic directly or affects the persistence between posts, I revise the reduced form model by adding dummies for $\textit{Gender}_{t,p-1}$ and their interaction with $D_{t,p-1}$. Each post in the base group

[17] The mean topic is taken across All posts including the first one. The estimated coefficient ($\hat{\phi_2}$)on $D_{t,1}$ shall be interpreted as an additional weight on the *First* post relative to the following posts.

[18] A post is "neutral" if it contains neither female nor male classifier.

follows a genderless ("neutral") post and it occurs on the first page of the thread it belongs to.

$$\begin{aligned} D_{t,p} = \beta_0 + \beta_1 D_{t,p-1} + Gender_{t,p-1}\lambda' + (D_{t,p-1} \times Gender_{t,p-1})\eta' \\ + (\phi_1 D_{t,0} + \phi_2 D_{t,1} + \phi_3 \overline{D}_t) + \theta\, 1[last\ page] + \nu_{t,p} \end{aligned} \tag{8}$$

The model above yields two ways to consider the effects of gender on the flow of the conversation. First, *within* each gender (female, male, or neutral), the coefficient on the lagged topic captures the relationship between adjacent post:

$$\beta_1 = E[D_{t,p} \,|\, D_{t,p-1} = 1,\, Gender_{t,p-1} = Neutral,\, X] - E[D_{t,p} \,|\, D_{t,p-1} = 0,\, Gender_{t,p-1} = Neutral,\, X]$$

$$\beta_1 + \eta_F = E[D_{t,p} \,|\, D_{t,p-1} = 1,\, Gender_{t,p-1} = Female,\, X] - E[D_{t,p} \,|\, D_{t,p-1} = 0,\, Gender_{t,p-1} = Female,\, X]$$

$$\beta_1 + \eta_M = E[D_{t,p} \,|\, D_{t,p-1} = 1,\, Gender_{t,p-1} = Male,\, X] - E[D_{t,p} \,|\, D_{t,p-1} = 0,\, Gender_{t,p-1} = Male,\, X]$$

where X includes all regressors other than the lagged variables. If β_1 or $\beta_1 + \eta_F$ or $\beta_1 + \eta_M$ are negative, there is a reversion effect relative to the prior post *within* the corresponding gender group. η_F and η_M are the difference-in-difference estimators capturing whether the potential reversion becomes stronger or weaker in the gendered cases relative to the neutral group.

The other way is to directly compare the probability of the current post staying on the same topic as its prior one *between* genders, conditional on $D_{t,p-1} = 1$.

$$\lambda_F + \eta_F = E[D_{t,p} \,|\, D_{t,p-1} = 1,\, Gender_{t,p-1} = Female,\, X] - E[D_{t,p} \,|\, D_{t,p-1} = 1,\, Gender_{t,p-1} = Neutral,\, X]$$

$$\lambda_M + \eta_M = E[D_{t,p} \,|\, D_{t,p-1} = 1,\, Gender_{t,p-1} = Male,\, X] - E[D_{t,p} \,|\, D_{t,p-1} = 1,\, Gender_{t,p-1} = Neutral,\, X]$$

From the comparison between $\lambda_F + \eta_F$ and $\lambda_F + \eta_F$, I can make inference on whether the topic is more likely to degenerate due to the mention of a female or male in the prior post, especially when a group is not representative in a type of discussion, e.g. mention of women in an academic thread or mention of men in a thread about physical appearance.

B. The Basic Results

I estimate the models above on posts within all gender-related threads. Standard errors are clustered at the thread level to take into account the potential correlation between posts within the same thread. Table 10 displays the results with and without the effects of gender ($Gender_{t,p-1}$), for the *Academic/Professional* topic (i.e. $D_{t,p} = 1$ if the post p contains any academic term) and the *Personal/Physical* topic respectively.

The average reversion effect across all genders in the *Academic/Professional* topic is about 5.0 percentage point (ppt) as shown in column (1). That is, if the prior post has an academic focus ($D_{t,p-1} = 1$), there is a mean reversion pattern that the next post is about 5.0 ppt less likely to stay on the same topic conditional on thread characteristics. Column (2) breaks down the reversion effect by gender: when the prior post is neutral, the reversion is about 4.4 ppt, but it becomes 2.6 ppt and 1.9 ppt stronger in magnitude, each significant at 0.1% level, for the female and male groups respectively.

Suppose the prior post contains at least one *Academic/Professional* term ($D_{t,p-1} = 1$), then relative to the neutral group, the female group (i.e. posts whose priors are discussing about women) is 2.3 ppt less likely to stay on the academic focus, while the male group is 1.4 ppt less likely. The effects of gender here are mostly driven by the interaction between gender and topic in the prior. I conduct a F-test on the null hypothesis that female and male in the prior post have equal effects: $\lambda_F + \eta_F = \lambda_M + \eta_M$, and it gives a p-value 0.0001.

For the *Personal/Physical* topic, the mean reversion pattern also holds: if the prior contains an term about personal information or physical appearance, the next post is 5.5 ppt less likely to be on the same topic in the neutral group, and the counterpart is 7.3 ppt in the female group, and 7.6 ppt in the male group. Relative to the neutral group, conditional on the prior related to *Personal/Physical* ($D_{t,p-1} = 1$), the female group is 1.1 ppt less likely to be persistent in topic, whereas the male group is 1.8 ppt less likely. The p-value from the F-test on equal gender effects: $\lambda_F + \eta_F = \lambda_M + \eta_M$ is 0.003. It is worth noting that $\hat{\lambda}_F$, $\hat{\eta}_F$ are very similar to $\hat{\lambda}_M$, $\hat{\eta}_M$ in column (2) for the *Academic/Professional* topic, in terms of both the estimates and their standard errors. This "symmetry" suggests that *Academic/Professional* to men is like *Personal/Physical* to women, which to some extent reflect the stereotype beliefs held by the posters.

The main limitation of using dummy variables $D := 1$ if including an *Academic/Professional* term, or 1 if including a *Personal/Physical* term, is that it cannot capture the subtle deviation from each topic. For example, the prior post contains five academic terms, but the next one only contains one. Both posts are labeled as $D = 1$. In this sense, using dummies may have underestimated the actual differences in the effects of gender on the persistence in each topic. To address this concern, in Appendix B, I replace D by *Topic* - the number of *Academic/Professional* terms and the number of *Personal/Physical* terms, and re-do the estimations as above. The robustness checks provide a more complete picture, and yield the same conclusions that there is a significantly higher deviation from an academic focus and a significantly lower deviation from a personal topic when the prior post is female rather than male or neutral.

C. Further Discussion under Different Initial Conditions

Since the title of each thread and the first post in most cases by the same poster set up the theme of the following discussion, I describe the initial conditions through 16 mutually exclusive combinations of gender, if the initial topic is *Academic/Professional*, if the initial topic is *Personal/Physical*:

$$\{Female, Male, Both, Neither\} \times \{0, 1\} \times \{0, 1\}$$

A thread starts off as *Female* if its title or its first post contains any female classifier[19] but none of the male classifiers, and vice versa for *Male*. The additional *Both* category refers to threads that include both female and male classifiers initially. I do not force the title and the first post to be about the same gender. Last, the *Neither* category consists of threads that are not gender related in the beginning.

I consider the initial topic to be *Academic/Professional* if (1) both the title and the first post include at least one academic or professional term, AND (2) the fraction of academic/professional terms in the title and the first post as a whole is $\geq$ the median % across all threads in the sample[20]. The initial indicator for *Personal/Physical* is defined likewise, but since the median fraction of terms in this category is 0, condition (2) is not binding.

I split the sample by initial condition and the resulting estimates for model (8) are summarized in Table 11 and Table 12 for the two topics respectively. For the *Academic/Professional* topic, the null hypothesis of equal effects of female or male relative to the neutral group is rejected at 5% significance level under the following initial conditions: (Female, 1, 0), (Neither, 1, 0), (Female, 0, 1), (Male, 0, 1), (Female, 0, 0) and (Male, 0, 0). The ratio between the effects of gender relative

[19] I use all gender classifiers at Level 1.

[20] The sample consists of all gender-related threads that include at least one "female" or "male" post.

to the neutral group – $\frac{\hat{\lambda}_F+\hat{\eta}_F}{\hat{\lambda}_M+\hat{\eta}_M}$ also provides some insights on the potentially drastic differences in the persistence in the academic topic, which might move in opposite directions after a mention of female versus male. In seven out of the sixteen cases, $\hat{\lambda}_M+\hat{\eta}_M$ turn out to be positive, which means the probability of the current post staying on the same academic focus as its prior post is higher in the male group relative to the neutral group. In contrast, the estimated $\hat{\lambda}_F+\hat{\eta}_F$ are positive in only three cases with either a *Personal* initial topic or *Male* as the initial gender[21].

The most interesting case to examine stereotyping in a dynamic setting is **(Neither, 1, 0)**, which consists of 31,294 (25%) threads starting off as a purely academic or professional discussion and not directly related to either women or men. Presumably the posts under this type of threads should stay on the same academic topic, and the mention of a female or a male should not make a difference on the persistence if there were no stereotyping invovled. Bordalo et al. (2016a) models stereotype belief as a representativeness-based heuristic. Women are traditionally underrepresented in academia; therefore, the stereotype model in Bordalo et al. (2016a) would suggest the academic aspect of women are down-weighted relative to other aspects, which could be physical appearance. In a dynamic setting, the posters might "overreact" to a post emphasizing the academic performance of a female and thus deviates from the academic focus and converges back to his or her own prior beliefs about gender characteristics. The empirical results (Table 11) show that the mention of women in the prior post decreases the probability of staying on the academic topic by 2.1 ppt (significantly negative at 0.1% level) relative to the neutral group, whereas there is a slight increase such probability in the male group. The F-test on $\hat{\lambda}_F+\hat{\eta}_F=\hat{\lambda}_M+\hat{\eta}_M$ yields a p-value of 0.0005.

The results on the *Personal/Physical* topic (Table 12) again show some "symmetry": in eight out of the sixteen cases, the effects of female ($\hat{\lambda}_F+\hat{\eta}_F$) are positive, while the effects of male are only positive in four cases when a thread starts off discussing about women or a personal topic. Among threads in **(Neither, 0, 1)**, there is a significant increase in the probability of staying on the personal topic when the prior post mentions women, while the effects of male is significantly negative, in contrast with the case of **(Neither, 1, 0)** for the *Academic/Professional* topic. The main issue of analyzing the *Personal/Physical*, however, is the under-identification of threads under this topic, due to a relatively small list of such terms in the overall lexicon and the strict definitions of the initial topic, which require the title and the first post to be consistent in themes. About

[21]The three cases are (Neither, 1, 1), (Male, 0, 1), and (Male, 0, 0).

11.75% of all gender-related threads are considered to start off with a *Personal/Physical* focus. It would be helpful to expand the vocabulary and phrases to capture similar threads from the last four cases.

In summary, the dynamic topic analysis reveals a significantly stronger tendency to deviate from an academic or professional focus when the prior post mentions a female rather than being neutral, whereas the mention of a male shows smaller or even opposite effects. It is particularly interesting to examine the stereotyping behavior under the (Neither, 1, 0) initial conditions for the *Academic/Professional* topic, and (Neither, 0, 1) for the *Personal/Physical*. Appendix Table B5 provides some stylized examples that illustrate the effects of gender on the persistence in topics. The stereotype model developed in Bordalo et al. (2016a) can be extended to a dynamic setting to explain the patterns I find in the flow of the conversation. In future analysis, it would also be important to consider the movement between these two topics, which can be understood as characteristics of women and men, instead of analyzing them independently.

4 Alternative Design: Attention Received by Economists

While the previous sections study the patterns in all gender-related discussions, this final part of the paper examines whether gender plays a role in determining how much attention an economist receives on EJMR. In this alternative design, I select two cohorts of economists: (1) 380 high-profile economists who ranked among the Top 5% Authors on RePEc[22]; (2) 204 assistant professors in Top 20 U.S. Economics Departments[23]. Using a difference-in-difference approach, I find that high-profile female economists tend to receive more attention than their male counterparts, and the gap is wider for relatively lower-ranked economists. The junior cohort shows different patterns when I group economists by the ranking of their current institutions.

[22]RePEc ranking of Top 5% Authors (Last 10 Years Publications), as of September 2016: https://ideas.repec.org/top/top.person.all10.html. To identify the gender of each economist, I match the overall ranking with a separate RePEc ranking on female economists: https://ideas.repec.org/top/top.women.html.

[23]based on U.S. News ranking of best graduate programs in Economics as of 2013 and 2017, and RePEc ranking of top Economics Departments.

4.1 Selection of Economists

The economists most likely to be discussed on EJMR are either prominent senior faculty, or tenure-track junior economists who have been through the job market recently. Based on this observation, I select two cohorts of economists: (1) 380 high-profile economists who ranked among the Top 5% Authors on RePEc[24]; (2) 204 assistant professors in Top 20 U.S. Economics Departments[25].

For the senior cohort, I generate a balanced set of female and male economists, who are comparable according to the RePEc ranking of the Top 5% Authors. I find 190 female economists among the top $2,422$ authors. For each of them, a coin is tossed to decide whether the male economist who ranks 1 above or 1 below will be included in the control group. I use each economist's rank as a proxy for his or her prominence in the field of economics. Hence I have a sample of 190 female and 190 male high-profile economists. For the junior cohort, I select all assistant professors in Top 20 U.S. Economics Departments. I find 45 female and 159 male junior faculty among these schools as of January 2017.

Given the 584 economists in total, I search by each person's full name within EJMR forum and then preserve as many threads in which he or she is mentioned as possible[26]. Then I keep all the posts on a given page of a thread. As a result, I construct a data set of $3,299$ unique threads. There is no restriction on the years of the discussions in this data set. Among 380 senior economists, there are 278 economists (145 women, 133 men) mentioned at least once in EJMR. Among 204 junior faculty, 187 economists (38 women, 149 men) were mentioned at least once. Seniority increases the attention one receives significantly. On average, a high-profile economist is discussed in 20.5 threads, whereas an assistant professor occurs in 14.8 threads.

4.2 Difference-in-Difference Analysis of Gender on Attention

Given the number of search results- N_i on each economist[27], I define A_i, a metric that represents the amount of attention person i receives as $A_i = asinh(N_i) = log(N_i + \sqrt{1 + N_i^2})$.

[24] RePEc ranking of Top 5% Authors (Last 10 Years Publications), as of September 2016: https://ideas.repec.org/top/top.person.all10.html. To identify the gender of each economist, I match the overall ranking with a separate RePEc ranking on female economists: https://ideas.repec.org/top/top.women.html.

[25] based on U.S. News ranking of best graduate programs in Economics as of 2013 and 2017, and RePEc ranking of top Economics Departments.

[26] In each query, I maximize the number of results Google display, but if there are over 20 results, the amount of URLs I can successfully scrape is shrunk by 25% on average.

[27] The number of threads in the final dataset is considered as an alternative measure, and it gives consistent results.

I estimate the following difference-in-difference specification:

$$A_i = \gamma_0 + \gamma_1 Female_i + \Gamma' Group_i + \Lambda'(Female_i \times Group_i) + \varepsilon_i$$

where Λ are the coefficients of interest. For the high-profile cohort, each “Group” contains 10 female economists and 10 male economists, based on their RePEc ranking. Figure 3 shows that the higher ranked an economist is, the more attention he or she receives on EJMR. Female economists tend to receive more attention than their male counterparts, and this gap, though insignificant, widens as the economist’s ranking goes down. This finding is in line with the hypothesis that women as the minority group are more “visible” (Kanter 1997).

For the junior cohort, since I do not have a measure of prominence at the individual level, I split them into 6 groups by the ranking of their current departments. Figure 4 reveals that junior faculty in higher ranked institutions receive significantly more attention. Female assistant professors receive more attention than their male counterparts in the first two groups (top 5 economics departments), but this trend is reversed for people in relatively lower ranked departments. In other words, for women the amount of attention one gets is more sensitive to the prestige of the institutions. However, note that the junior cohort is imbalanced in gender: 45 women and 159 men. The gender differences can be exaggerated if there are outliers among men who receive much more attention than their peers. For a more careful analysis on the junior cohort, it would be helpful to use the publication information of each economist as a measure of individual achievement in lieu of the institutional ranking.

For both the high-profile and the junior cohorts, the selection is limited as I focus on the best people in the field in terms of their academic and professional achievements. A more informative analysis would require expanding the sample of economists to be more representative of the overall academic community. It is also worth mentioning that within these samples, there is no clear relationship between the prestige of the department one works at and an economist’s own prominence. In particular, junior faculty within the same department ranking group are not as comparable as the high-profile economists within the same RePEc ranking group based on individual performance. Therefore, the results for high-profile and junior cohorts should be viewed separately.

5 Discussions

Gender stereotyping can take a subtle or implicit form that makes it difficult to measure and analyze in economics. In addition, people tend not to reveal their true beliefs about gender if they care about political and social correctness in public. The anonymity of the Economics Job Market Rumors forum, however, removes such barriers, and thus provides a natural setting to study the existence and extent of gender stereotyping in this academic community online.

There are mainly three contributions of this paper. First, it provides a systematic evaluation of the gender stereotyped language on EJMR, which can create an unwelcoming atmosphere online. Second, the topic analysis provides an empirical framework to test for the stereotyping model developed in Bordalo et al. 2016a and also extend it to a dynamic setting. It reveals that women to *Personal/Physical* is like men to *Academic/Professional*, and there is a stronger tendency to deviate from an academic focus when women are mentioned previously. Third, in terms of methodology, this study illustrates the use of text analytic techniques in combination with econometric methods to draw meaningful insights from the textual data.

The release of the earlier version of this study and the review by Justin Wolfers on New York Times[28] in August 2017 give a shock to the forum itself. Appendix C provides a trend analysis. For theads initiated before August 2017, *Female* posts consistently contain about 45% less *Academic/Professional* terms than *Male* posts on average, but the gap shrinks almost by half as shown in Table C1 from August to October 2017. In particular, the month-to-month variation in Figure C1 shows that among threads started in August 2017, there is stronger link between women and *Academic/Professional* for the first time. The intervention is effective in the sense that the academic aspects of women are discussed more intensively than before and might help shrink the contrast between *in-group* and *out-group*. However, it is not clear whether this trend will persist at this point.

A missing dimension in the current topic analysis is sentiment. In the dynamic setting, I show that there is a significantly stronger immediate deviation from the *Academic/Professional* focus after a mention of female(s). It will be more informative if I can differentiate between positive and negative comments about the research work by men and women. The examples in Table B5

[28] Wolfers, Justin. 2017. "Evidence of a Toxic Environment in Economics". New York Times. 18 August.

suggest that the deviation might be stronger when the comment on a woman's academic aspects is positive, but less so if it is negative. In other words, the *Academic/Professional* characteristics shall be considered in two dimensions and allowed different weights in stereotyping. It will also be helpful to formally extend the representativeness-based discounting model of stereotype in Bordal et al. 2016 to a dynamic process.

In conclusion, my results suggest the need for changes to maintain an inclusive online environment for everyone in the academic community. The casual setting of this online forum cannot be an excuse for gender stereotyped conversations, and the freedom to express one's opinions anonymously should not be abused to create a sense of isolation, which can be discouraging and harmful to the academic and professional development of all genders.

References

[1] Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." Quarterly Journal of Economics 131 (4): 1753-1794.

[2] Bordalo, Pedro, Katie Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. Working Paper. "Beliefs about Gender".

[3] Ceci, S.J., Ginther, D.K., Kahn, S. and Williams, W.M., 2014. "Women in academic science A changing landscape". Psychological Science in the Public Interest, 15(3), pp.75-141.

[4] Eisenstein J, Smith NA, Xing EP. 2011. "Discovering sociolinguistic associations with structured sparsity". Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, pp. 1365–1374.

[5] Ginther, D.K. and Kahn, S., 2006. "Does science promote women? Evidence from academia 1973-2001" (No. w12691). National Bureau of Economic Research.

[6] Gentzkow, M., Shapiro, J.M. and Taddy, M., 2016. "Measuring polarization in high-dimensional data: Method and application to congressional speech" (No. w22423). National Bureau of Economic Research.

[7] Gentzkow, Matthew and Kelly, Bryan T. and Taddy, Matt. "Text As Data". February 15, 2017.

[8] Greenwald AG, McGhee DE, Schwartz JL. 1998. "Measuring individual differences in implicit cognition: The implicit association test". J Pers Soc Psychol 74(6):1464–1480.

[9] Hutto, C.J. and Gilbert, E.E. 2014. "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text". Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[10] Kahn, S. and Ginther, D.K. 2017. "Women and STEM." Prepared for The Oxford Handbook on the Economics of Women. Averett, S., Argys, A. and Hoffman, S.D. eds.

[11] Kanter, Rosabeth Moss. 1997. "Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women." American Journal of Sociology, vol. 82, no. 5, 1977, pp. 965–990. www.jstor.org/stable/2777808.

[12] Lee, Peggy and James, Erika. (2007). "SHE-E-OS: Gender effects and investor reactions to the announcements of top executive appointments". Strategic Management Journal. 28. 227 - 241. 10.1002/smj.575.

[13] Oakes, P.J., Haslam, S.A. and Turner, J.C., 1994. Stereotyping and social reality. Blackwell Publishing.

[14] Pennebaker, J.W., Booth, R.J., Boyd, R.L., and Francis, M.E. 2015. "Linguistic Inquiry and Word Count: LIWC2015". Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

[15] RePEc, 2016. "Top 10% Authors (Last 10 Years Publications), as of October 2016". https://ideas.repec.org/top/top.women10.html (accessed October 2016.)

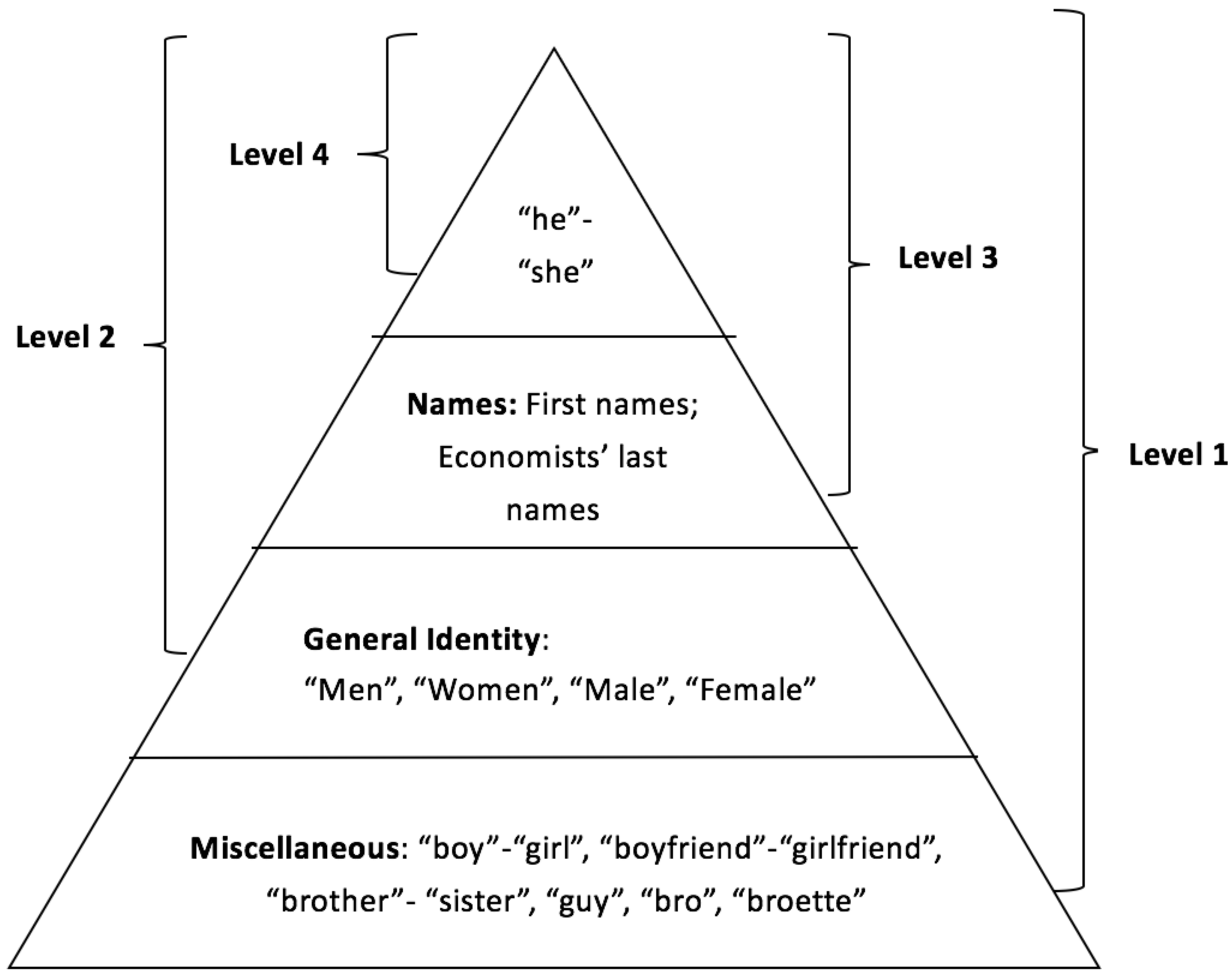
[16] RePEc, 2016. "Top 10% Female Economists (Last 10 Years Publications), as of October 2016". https://ideas.repec.org/top/top.women10.html (accessed October 2016.)

[17] Reuben, E., Sapienza, P. and Zingales, L., 2014. "How stereotypes impair women's careers in science". Proceedings of the National Academy of Sciences, 111(12), pp.4403-4408.

[18] Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al., 2013. "Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach". PLoS ONE 8(9): e73791. doi:10.1371/journal.pone.0073791.

[19] Tajfel, H. and Turner, J. C., and 1986. "The Social Identity Theory of Intergroup Behavior". Psychology of intergroup relations, 7-24.

Figure 1: Levels of Gender Classifiers



Notes: the words displayed are examples. See Appendix for the complete list of gender classifiers.

Table 1: Summary of the EJMR data

	No. Threads	No. Posts	No. *Female*	No. *Male*	No. *Neutral*
All	223,475	2,217,046			
Gender Sample					
Level 1	138,477	1,736,204	103,584	341,226	1,292,394
			(23.29%)	(76.71%)	
Level 2	110,933	1,467,949	77,405	248,530	1,142,014
			(23.75%)	(76.25%)	
Level 3	101,052	1,362,091	54,944	228,613	1,078,534
			(19.38%)	(80.62%)	
Level 4	76,325	1,122,782	50,435	144,940	927,407
			(25.81%)	(74.19%)	

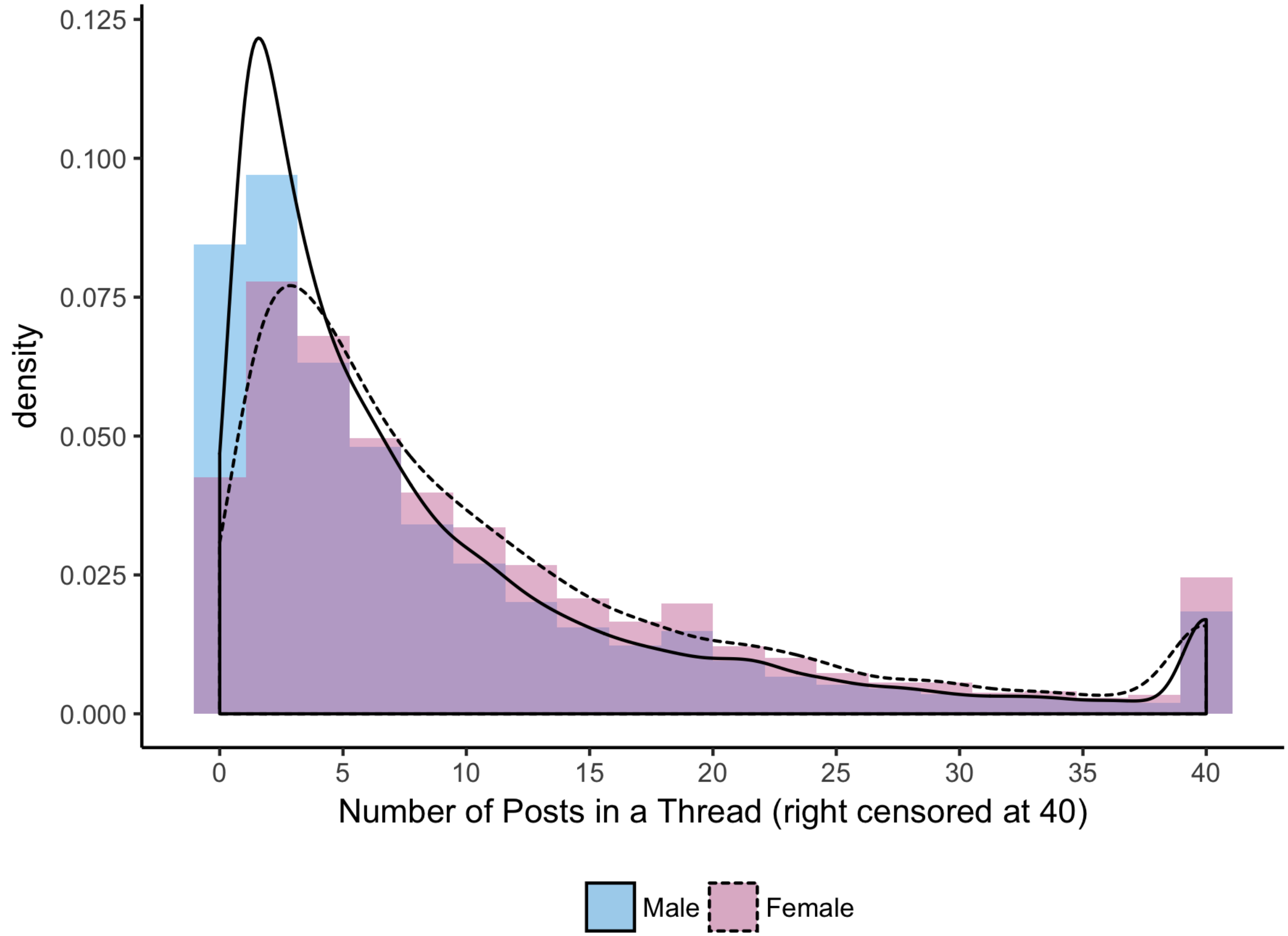
Notes: "All" refers to the entire dataset of threads created/updated from Oct 2013 to Oct 2017. Level 1 to Level 4 refer to the increasingly restrictive subsets of gender classifiers I use to identify gender-related posts. At each level, "Gender Sample" preserves all posts within threads that contain at least one gender-related post. The %s in parentheses refer to the percentage of *Female* posts among all gender-related posts, and that of *Male* posts respectively. Duplicate observations that contain both female and male classifiers have been resolved by the Lasso-Logistic model in Section 2.1.

Table 2: Popularity on Gender in Titles

	(1) No. Posts	(2) No. Views
Gender in Titles		
$Female_{t,0} = 1$	-3.347	-384.406
	(1.007)	(126.940)
$Female_{t,0} = 0$	-5.585	-385.420
	(0.629)	(79.324)
Constant	15.626	1,076.197
	(0.272)	(34.231)
No. Threads	138,477	138,477
R^2	0.001	0.0002

Notes: Standard Errors are in parentheses. No. Posts and No. Views of each thread are reported the main pages of EJMR. I did a simple check on the no. posts reported on EJMR: the no. posts I could scrape from the *First* and the *Last* pages of each thread should be $\leq$ the number reported at the thread level. I found about 0.7% misreported threads, for which I replace the value by the no. posts I scraped successfully.

Figure 2: Popularity by Gender in Titles



Notes: I restrict to threads with *Female* or *Male* titles, i.e. $Female_{t,0} \in \{0, 1\}$. The number of posts are reported on the main sites of EJMR forum at the thread level, except for 0.7% misreported cases I found and corrected as in the notes under Table 2. For purposes of illustration, I "right-censor" the data that I code no. posts as 40 if it is ≥ 40.

Table 3: Top 30 Words with the strongest predictive power for $Female_i = 1$

Level 1				Level 4			
Most "female"		Most "male"		Most "female"		Most "male"	
Word	Marginal Effect	Word	Marginal Effect	Word	Marginal Effect	Word	Marginal Effect
hotter	0.422	homo	-0.303	pregnant	0.358	testosterone	-0.271
pregnant	0.323	testosterone	-0.195	sexism	0.353	handsome	-0.250
plow	0.277	chapters	-0.189	breast	0.316	homo	-0.218
marry	0.275	satisfaction	-0.187	hotter	0.307	dictator	-0.199
hot	0.271	fieckers	-0.181	marry	0.286	blog	-0.185
marrying	0.260	macroeconomics	-0.180	feminist	0.285	gray	-0.184
pregnancy	0.254	cuny	-0.180	plow	0.268	hateukbro	-0.172
attractive	0.245	thrust	-0.169	attractive	0.262	hero	-0.170
beautiful	0.240	nk	-0.165	hot	0.237	irate	-0.167
breast	0.227	macro	-0.163	hp	0.237	knocking	-0.163
dumped	0.225	fenance	-0.162	vagina	0.234	gay	-0.159
kissed	0.224	founding	-0.160	pregnancy	0.233	fieckers	-0.158
misogynistic	0.222	blog	-0.157	marrying	0.223	adviser	-0.153
feminist	0.218	mountains	-0.156	divorce	0.219	supervisor	-0.153
sexism	0.210	grown	-0.156	blonde	0.215	ferguson	-0.146
dated	0.209	frat	-0.155	dated	0.214	nobel	-0.143
whore	0.208	handsome	-0.154	whore	0.212	repec	-0.141
sexy	0.202	nba	-0.151	classified	0.212	mirror	-0.141
raped	0.200	lyrics	-0.151	shopping	0.206	register	-0.141
attracted	0.198	ferguson	-0.150	dumped	0.199	deadwood	-0.138
slept	0.195	wasn	-0.147	gorgeous	0.199	genius	-0.137
blonde	0.193	supervisor	-0.146	beautiful	0.199	gop	-0.134
unattractive	0.193	rfs	-0.145	date	0.197	fans	-0.133
gorgeous	0.192	adviser	-0.141	tinder	0.187	pulled	-0.131
assaulted	0.191	minnesota	-0.140	cute	0.184	player	-0.130
cute	0.185	hero	-0.136	nurse	0.182	spell	-0.130
vagina	0.184	gay	-0.135	dump	0.182	bowl	-0.125
date	0.181	puerto	-0.134	humanities	0.180	minnesota	-0.124
dating	0.181	nobel	-0.129	gender	0.180	retard	-0.123
ugly	0.181	keynesian	-0.128	sexy	0.177	players	-0.123

Notes: the marginal effect of word w is the change in probability of a post being classified as *female*, i.e. 1 if it is discussing women, when it contains one more word w.

Table 4: Categories of Words

Category	No. Words	Examples
Gender Classifiers (All - Level 1)		
Female	44	“she”, “female”
Male	134	“he”, “male”
Academic/Professional		
Economics	177	“economics”, “macro”, “empirical”,“QJE”, “Keynesian”
Academic-General	1,515	“research”, “papers”,“tenure”, “teaching”, “professor”
Professional	138	“career”, “interview”, “payrolls”, “placement”, “recruit”
Personal/Physical		
Personal Information	118	“family”,“married”, “kids”, “relationship”,“lifestyle”
Physical Attributes	125	“beautiful”, “handsome”, “attractive”, “body”,“fat”
Gender related	86	“gender”, “feminine”, “masculine”, “sexist”, “sexual”
Swear Words		
Swear	78	“shit”, “wtf”, “asshole”
Intellectual		
Intellectual-Positive	115	“intelligent”, “creative”,“competent”
Intellectual-Neutral	29	“brain”, “iq”, “ability”
Intellectual-Negative	134	“dumb”,“ignorant”,“incompetent”
Miscellaneous		
Emotion/Feelings	74	“happy”,“depressing”
Emojis	11	“:)”, “;)”, “:p”
Others	7,222	“years”, “places”, “everything”
Total	10,000	

Notes: “Gender related” category under *Personal/Physical* are not used as gender classifiers.

Table 5: Academic/Professional - counts

	Number of *Academic/Professional* Words			
	Level 1	Level 2	Level 3	Level 4
$Female_i$	-1.349	-1.535	-1.675	-1.514
	(0.020)	(0.025)	(0.029)	(0.032)
Constant	3.000	3.368	3.526	3.434
	(0.014)	(0.018)	(0.019)	(0.022)
R^2	0.008	0.009	0.009	0.010
F Stat.	4645.115	3754.923	3363.119	2239.054
N	$435,617$	$318,289$	$276,310$	$194,583$

Notes: Standard errors in parentheses are clustered at the thread level. Sample restricted to posts with ≥ 3 and ≤ 300 words, roughly 98% of each sample. "Level 1" to "Level 4" refer to increasingly restrictive levels of gender classifiers to identify gender-related posts.

Table 6: Academic/Professional - $1[counts > 0]$

	1 if includes *Academic/Professional* words			
	Level 1	Level 2	Level 3	Level 4
$Female_i$	-0.122	-0.127	-0.144	-0.164
	(0.002)	(0.003)	(0.003)	(0.003)
Constant	0.588	0.620	0.633	0.660
	(0.001)	(0.001)	(0.001)	(0.002)
R^2	0.011	0.012	0.014	0.022
F Stat.	2993.244	2359.953	2195.627	2449.906
N	$435,617$	$318,289$	$276,310$	$194,583$

Notes: Standard errors in parentheses are clustered at the thread level. Sample restricted to posts with ≥ 3 and ≤ 300 words, roughly 98% of each sample. "Level 1" to "Level 4" refer to increasingly restrictive levels of gender classifiers to identify gender-related posts.

Table 7: Personal/Physical - counts

	Number of *Personal/Physical* Words			
	Level 1	Level 2	Level 3	Level 4
$Female_i$	0.710	0.688	0.603	0.592
	(0.009)	(0.011)	(0.012)	(0.013)
Constant	0.408	0.452	0.442	0.521
	(0.002)	(0.003)	(0.003)	(0.004)
R^2	0.041	0.033	0.024	0.023
F Stat.	6327.566	4135.636	2492.290	2040.092
N	$435,617$	$318,289$	$276,310$	$194,583$

Notes: Standard errors in parentheses are clustered at the thread level. Sample restricted to posts with ≥ 3 and ≤ 300 words, roughly 98% of each sample. "Level 1" to "Level 4" refer to increasingly restrictive levels of gender classifiers to identify gender-related posts.

Table 8: Personal/Physical - $1[counts > 0]$

	1 if includes *Personal/Physical* words			
	Level 1	Level 2	Level 3	Level 4
$Female_i$	0.243	0.223	0.197	0.184
	(0.002)	(0.002)	(0.003)	(0.003)
Constant	0.226	0.236	0.226	0.263
	(0.001)	(0.001)	(0.001)	(0.001)
R^2	0.052	0.044	0.031	0.030
F Stat.	13115.743	8082.688	4714.754	3555.060
N	$435,617$	$318,289$	$276,310$	$194,583$

Notes: Standard errors in parentheses are clustered at the thread level. Sample restricted to posts with ≥ 3 and ≤ 300 words, roughly 98% of each sample. "Level 1" to "Level 4" refer to increasingly restrictive levels of gender classifiers to identify gender-related posts.

Table 9: Mean Frequencies of Words by Topic

	$\overline{Academic_t}$		$\overline{Personal_t}$	
	(1)	(2)	(3)	(4)
Posts: $(\%Female - \%Male)$				
Quartile 1: $[-1, -0.333)$ (base)				
Quartile 2: $[-0.333, -0.157)$	-2.110	-0.532	-0.203	-0.028
	(0.034)	(0.023)	(0.007)	(0.005)
Quartile 3: $[-0.157, 0)$	-2.022	-0.550	-0.232	-0.004
	(0.035)	(0.025)	(0.007)	(0.005)
Quartile 4: $[0, 1]$	-2.694	-1.316	0.074	0.357
	(0.035)	(0.023)	(0.007)	(0.005)
Constant	3.999	2.468	0.444	0.290
	(0.025)	(0.015)	(0.005)	(0.003)
Weighted		X		X
N	138,468	138,468	138,468	138,468
Adjusted R^2	0.046	0.024	0.021	0.059
F Statistic	2,222.896	1,093.083	1,008.088	2,728.695

Notes: Standard errors are in parentheses. Each title can be classfied as $Female_{t,0} = 1$, $Female_{t,0} = 1$ or not related to gender. Columns (2) and (4) use #gender-related posts in each thread as the weight.

Table 10: Persistence in Topics (Any Thread; dummies)

	Academic/Professional		Personal/Physical	
	(1)	(2)	(3)	(4)
$D_{t,p-1}$	-0.050	-0.044	-0.062	-0.055
	(0.001)	(0.001)	(0.001)	(0.001)
Gender in the Prior Post				
Neutral (base)				
Female		0.003		0.007
		(0.002)		(0.002)
Male		0.006		0.003
		(0.001)		(0.001)
$Female \times D_{t,p-1}$		-0.026		-0.018
		(0.003)		(0.003)
$Male \times D_{t,p-1}$		-0.019		-0.021
		(0.002)		(0.002)
$\overline{D}_t$	1.141	1.140	1.145	1.145
	(0.001)	(0.001)	(0.001)	(0.001)
$D_{t,0}$ (Titles)	0.005	0.004	0.005	0.004
	(0.000)	(0.000)	(0.000)	(0.000)
$D_{t,1}$ (*First* posts)	-0.079	-0.078	-0.081	-0.081
	(0.000)	(0.000)	(0.000)	(0.000)
1[last page]	0.024	0.024	0.017	0.017
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-0.002	-0.002	0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Adj. R^2	0.268	0.268	0.182	0.182
F	$964,952.73$	$536,543.02$	$314,652.98$	$175,564.92$
N	$1,333,515$	$1,333,515$	$1,333,515$	$1,333,515$

Notes: Standard errors in parentheses are clustered at the thread level.

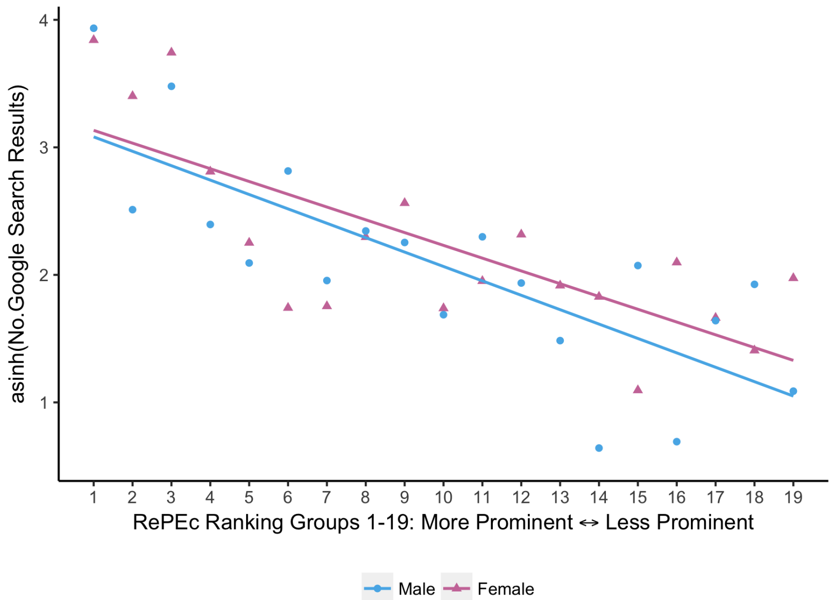
Table 11: Gender on Persistence in *Academic/Professional* under 16 Initial Conditions

			Neutral	Female		Male		Ratio	H0: $\lambda_F + \eta_F = \lambda_M + \eta_M$	
Titles	No. Titles (%)	No. Posts (%)	β_1	η_F	$\lambda_F + \eta_F$	η_M	$\lambda_M + \eta_M$	$\frac{\lambda_F+\eta_F}{\lambda_M+\eta_M}$	F stat	p-value
Any title	127029 (100%)	1333515 (100%)	-0.044	-0.026	-0.023	-0.019	-0.014	1.706	15.700	0.0001
Initial: (Gender, 1 if Academic, 1 if Personal)										
(Female,1,0)	2,117 (1.67%)	19,470 (1.46%)	-0.049	-0.024	-0.034	0.016	0.013	-2.531	11.950	0.001
(Male,1,0)	15,215 (11.98%)	141,659 (10.62%)	-0.056	-0.033	-0.025	-0.019	-0.021	1.207	0.120	0.732
(Both,1,0)	1,929 (1.52%)	19,377 (1.45%)	-0.031	-0.044	-0.044	-0.029	-0.030	1.461	1.120	0.290
(Neither,1,0)	31,294 (24.64%)	403,977 (30.29%)	-0.030	-0.033	-0.021	-0.013	0.0004	-56.096	12.130	0.0005
(Female,1,1)	434 (0.34%)	4,270 (0.32%)	-0.059	-0.016	-0.014	0.011	0.0005	-28.791	0.190	0.664
(Male,1,1)	908 (0.71%)	8,081 (0.61%)	-0.040	-0.014	-0.029	-0.051	-0.042	0.698	0.130	0.721
(Both,1,1)	446 (0.35%)	4,727 (0.35%)	-0.064	-0.032	-0.020	-0.027	0.001	-19.712	0.840	0.361
(Neither,1,1)	1,815 (1.43%)	22,101 (1.66%)	-0.053	0.026	0.011	-0.006	-0.001	-8.219	0.400	0.528
(Female,0,1)	2,394 (1.88%)	23,168 (1.74%)	-0.051	-0.004	-0.014	0.058	0.041	-0.354	8.960	0.003
(Male,0,1)	2,838 (2.23%)	24,970 (1.87%)	-0.063	0.012	0.014	-0.012	-0.030	-0.475	5.030	0.025
(Both,0,1)	2,050 (1.61%)	22,024 (1.65%)	-0.051	-0.023	-0.022	-0.018	-0.007	2.992	1.240	0.266
(Neither,0,1)	4,065 (3.2%)	45,476 (3.41%)	-0.044	-0.020	-0.005	-0.010	0.003	-1.537	0.300	0.585
(Female,0,0)	6,020 (4.74%)	51,126 (3.83%)	-0.062	-0.023	-0.026	0.013	0.008	-3.078	10.090	0.002
(Male,0,0)	22,285 (17.54%)	180,255 (13.52%)	-0.061	0.022	0.018	-0.015	-0.018	-0.976	10.520	0.001
(Both,0,0)	5,006 (3.94%)	45,617 (3.42%)	-0.051	-0.034	-0.033	-0.032	-0.036	0.934	0.070	0.790
(Neither,0,0)	28,213 (22.21%)	317,217 (23.79%)	-0.046	-0.017	-0.012	-0.014	-0.001	12.928	2.170	0.141

Table 12: Gender on Persistence in *Personal/Physical* under 16 Initial Conditions

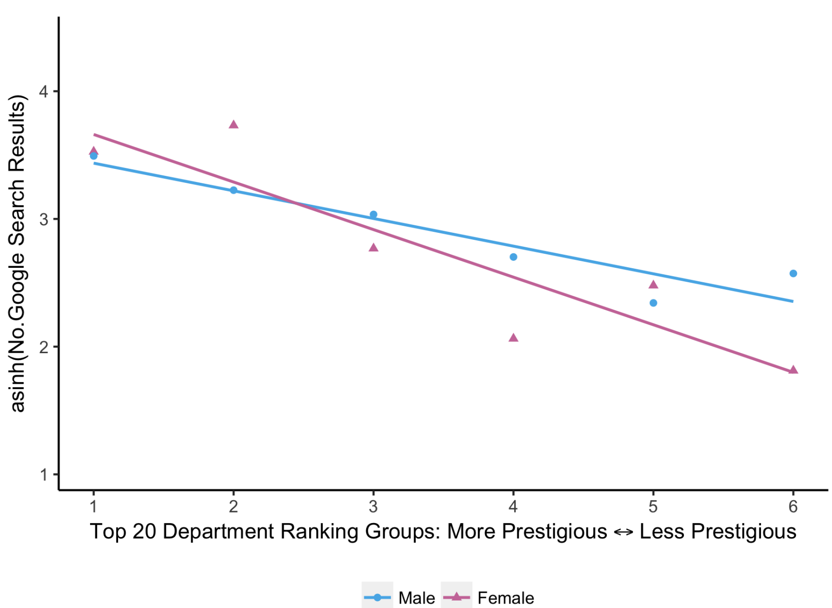
			Neutral	Female		Male		Ratio	H0: $\lambda_F + \eta_F = \lambda_M + \eta_M$	
Titles	No. Titles (%)	No. Posts (%)	β_1	η_F	$\lambda_F + \eta_F$	η_M	$\lambda_M + \eta_M$	$\frac{\lambda_F+\eta_F}{\lambda_M+\eta_M}$	F stat	p-value
Any title	127,029 (100%)	1,333,515 (100%)	-0.055	-0.018	-0.011	-0.021	-0.018	0.584	8.650	0.003
Initial: (Gender, 1 if Academic, 1 if Personal)										
(Female,1,0)	2,117 (1.67%)	19,470 (1.46%)	-0.040	-0.031	-0.033	-0.012	-0.008	3.933	1.520	0.218
(Male,1,0)	15,215 (11.98%)	141,659 (10.62%)	-0.060	-0.023	0.001	-0.019	-0.021	-0.025	2.360	0.124
(Both,1,0)	1,929 (1.52%)	19,377 (1.45%)	-0.048	-0.011	-0.002	-0.006	-0.028	0.065	2.480	0.115
(Neither,1,0)	31,294 (24.64%)	403,977 (30.29%)	-0.045	-0.009	0.005	-0.010	-0.004	-1.331	1.250	0.263
(Female,1,1)	434 (0.34%)	4,270 (0.32%)	-0.040	-0.085	-0.047	-0.059	0.009	-5.524	1.850	0.174
(Male,1,1)	908 (0.71%)	8,081 (0.61%)	-0.100	0.088	0.024	0.030	0.015	1.630	0.080	0.778
(Both,1,1)	446 (0.35%)	4,727 (0.35%)	-0.088	0.041	0.027	0.003	-0.003	-9.628	1.060	0.304
(Neither,1,1)	1,815 (1.43%)	22,101 (1.66%)	-0.044	-0.040	-0.032	-0.006	-0.002	15.613	1.970	0.160
(Female,0,1)	2,394 (1.88%)	23,168 (1.74%)	-0.071	-0.013	-0.006	0.022	0.002	-2.861	0.260	0.612
(Male,0,1)	2,838 (2.23%)	24,970 (1.87%)	-0.065	0.012	0.005	-0.021	-0.025	-0.178	3.460	0.063
(Both,0,1)	2,050 (1.61%)	22,024 (1.65%)	-0.068	0.004	-0.00001	0.008	-0.00002	0.605	0	1.000
(Neither,0,1)	4,065 (3.2%)	45,476 (3.41%)	-0.048	-0.001	0.013	-0.013	-0.008	-1.594	2.530	0.112
(Female,0,0)	6,020 (4.74%)	51,126 (3.83%)	-0.066	-0.017	-0.019	-0.004	0.002	-8.647	3.550	0.060
(Male,0,0)	22,285 (17.54%)	180,255 (13.52%)	-0.067	-0.005	0.005	-0.023	-0.024	-0.231	8.780	0.003
(Both,0,0)	5,006 (3.94%)	45,617 (3.42%)	-0.056	-0.026	-0.024	-0.040	-0.035	0.684	1.500	0.220
(Neither,0,0)	28,213 (22.21%)	317,217 (23.79%)	-0.058	-0.006	0.003	-0.014	-0.008	-0.333	2.240	0.135

Figure 3: 380 High-profile Economists (190 female,190 male)



Notes: 190 economists of each gender are assigned to 19 groups based on their ranking. Each plotted point represents the mean attention measure for a group of 10 economist of the given gender. The lines show the linear trends of attention measure on ranking groups ranging from 1 to 19.

Figure 4: 204 Assistant Professors (45 female, 159 male)



Notes: 204 junior economists are assigned to 5 groups based on the ranking of their current departments. Each plotted point represents the mean attention measure for economists of a given gender within the same group. The lines show the linear trends of attention measure on ranking groups ranging from 1 to 5.

APPENDIX

A. Lasso-logistic Model for Gender Prediction and Word Selection

The objective of the Lasso-Logistic model (Section 2.1) is to estimate $P(Female = 1|Text)$ - the probability of the subject of a post being female conditional on characteristics emphasized in the text, which are in the format of individual words in this case. I exclude gender classifiers and the last names of celebrities (non economists) from the most frequent $10,000$ words that emerge from the raw data (over 1.1 million posts). As a result, I have $9,545$ words as predictors for gender. The model is constructed as follows:

Let W_i denotes the vector of word frequencies for post i, and assume the posterior probability is:

$$P(Female_i = 1|W_i) = \frac{exp(\theta_0 + W_i'\theta)}{1 + exp(\theta_0 + W_i'\theta)}$$
$$P(Female_i = 0|W_i) = \frac{1}{1 + exp(\theta_0 + W_i'\theta)}$$

Write the likelihood of each observation as:

$$P(Female_i|W_i) = P(Female_i = 1|W_i)^{Female_i} \times P(Female_i = 0|W_i)^{(1-Female_i)}$$

Assume the observations are independent, the log likelihood of N observations is

$$\begin{aligned} l_N(\theta) &= log(\Pi_{i=1}^{N} P(Female_i|W_i)) \\ &= \sum_{i=1}^{N} Female_i * (\theta_0 + W_i'\theta) - log(1 + exp(\theta_0 + W_i'\theta)) \end{aligned}$$

I estimate θ on words through the following objective function:

$$\begin{aligned} \hat{\theta}_\lambda &= argmin_\theta \, (-l_N(\theta)) + \lambda||\theta||_1 \\ &= argmin_\theta \, \frac{1}{N}\Sigma_i[log(1 + exp(W_i'\theta)) - Female_i(W_i'\theta)] + \lambda\|\theta\|_1 \\ where \; \|\theta\|_1 &= \sum_{j\geq 1} |\theta^j|. \end{aligned}$$

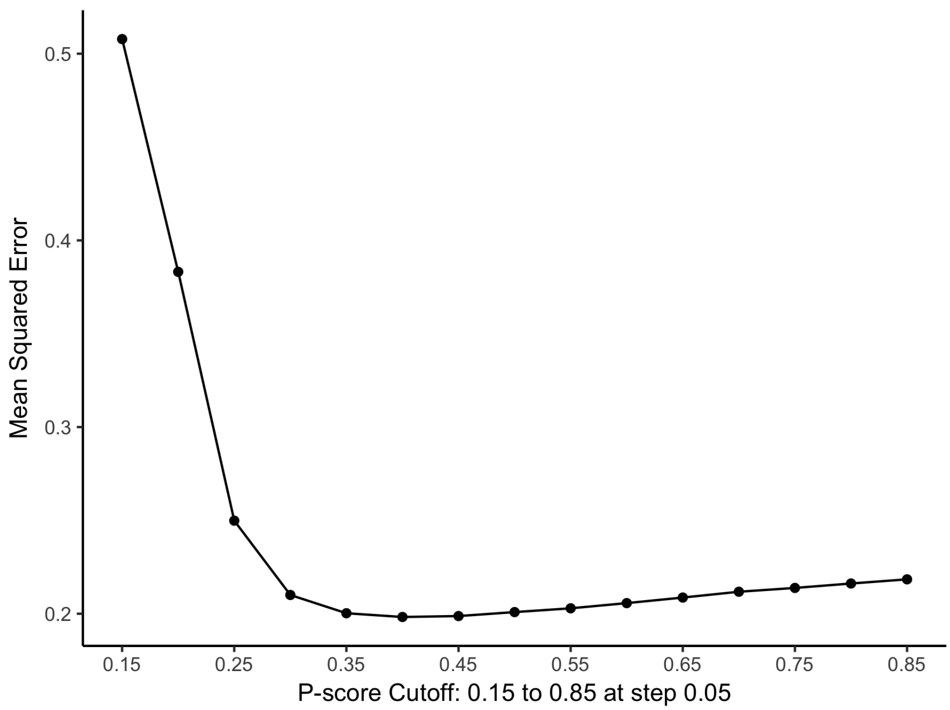
In this case, each W_i is a $9,545$-by-1 vector of word counts. Due to the penalization, the estimator $\hat{\theta}_\lambda$ is biased, but the variance of the model is reduced, and tends to yield more accurae estimates of $P(Female|Language)$.

There are $400,729$ posts that include only "female" words or only "male" words at Level 1. I use 75% of them, i.e. $300,788$ posts, to train the model and select an optimal tuning parameter λ through 5-fold cross validation. The remaining 25% - $99,941$ posts are assigned to the test set to select the best cutoff on p-scores, which turns out to be 0.40 (Figure A1). Finally, I apply the model to the $44,081$ duplicates, and reclassify $14,028$ of them to $Female = 1$ and the rest to $Female = 0$. As for the variable selection, the coefficients of $5,034$ words are shrunk to zero; that is, they are considered irrelevant to the gender identification of a post. The average marginal effect of word k is estimated by:

$$\begin{aligned} \textit{Word k's marginal effect} &= P(Female_i = 1|W_{i,(-k)}, W_{ik} + 1) - P(Female_i = 1|W_{i,(-k)}, W_{ik}) \\ &= \frac{1}{N} \sum_i P(Female_i = 1|Wi) \times (1 - P(Female_i = 1|Wi))\, \widehat{\theta^k_\lambda} \end{aligned}$$

where W_{ik} is the frequency of word k in post i, and $W_{i,(-k)}$ is the vector of frequencies of words other than k in post i.

Figure A1: Selection of Optimal P-score Cutoff by Mean Squared Error



Note: "P-score" refers to the predicted probablity of the subject of a post being female, i.e. $P(Female = 1 \| Words)$. The MSEs are calcuated on the test set ($99,941$ posts) - 25% of all posts that include only "female" or only "male" words at Level 1. $p = 0.40$ is selected as the optimal cutoff to make a call on gender among the ultimate test set - $44,081$ duplicate posts.

Table A1: Number of Posts containing the Most *Predictive* Words selected by Lasso

Level 1						Level 4					
Most "female"			Most "male"			Most "female"			Most "male"		
Word	*#Female*	*#Male*	Word	*#Female*	#Male	Word	*#Female*	*#Male*	Word	*#Female*	*#Male*
hotter	307	31	homo	48	715	pregnant	280	88	testosterone	16	30
pregnant	564	120	testosterone	51	102	sexism	88	75	handsome	45	166
plow	274	83	chapters	9	361	breast	66	26	homo	29	164
marry	1,287	258	satisfaction	59	145	hotter	120	31	dictator	6	167
hot	3,613	1,053	fieckers	49	604	marry	564	184	blog	89	1,244
marrying	262	49	macroeconomics	19	850	feminist	164	126	gray	14	69
pregnancy	202	61	cuny	8	248	plow	151	55	hateukbro	0	70
attractive	1,578	417	thrust	6	47	attractive	559	234	hero	32	412
beautiful	1,419	610	nk	3	260	hot	1,322	645	irate	25	234
breast	134	48	macro	178	4,282	hp	26	14	knocking	7	81
dumped	361	100	fenance	46	640	vagina	138	40	gay	167	733
kissed	218	50	founding	6	186	pregnancy	105	28	fieckers	20	201
misogynistic	66	48	blog	109	1,839	marrying	118	33	adviser	66	591
feminist	422	234	mountains	14	90	divorce	381	142	supervisor	36	197
sexism	269	171	grown	69	394	blonde	156	50	ferguson	10	126
dated	362	148	frat	59	290	dated	195	85	nobel	124	1,945
whore	239	148	handsome	103	323	whore	130	105	repec	8	176
sexy	430	207	nba	16	301	classified	33	56	mirror	28	92
raped	297	155	lyrics	17	111	shopping	100	68	register	20	110
attracted	415	182	ferguson	10	221	dumped	244	84	deadwood	39	425
slept	368	85	wasn	32	171	gorgeous	110	41	genius	51	649
blonde	292	79	supervisor	40	273	beautiful	541	329	gop	28	320
unattractive	172	32	rfs	7	284	date	908	471	fans	27	221
gorgeous	213	78	adviser	78	712	tinder	107	33	pulled	80	328
assaulted	98	52	minnesota	35	703	cute	467	294	player	82	706
cute	912	488	hero	47	579	nurse	51	26	spell	27	126
vagina	199	68	gay	406	1,755	dump	369	215	bowl	14	104
date	1,729	835	puerto	7	101	humanities	53	146	minnesota	20	283
dating	1,423	399	nobel	204	3,379	gender	296	324	retard	44	328
ugly	1,046	404	keynesian	8	567	sexy	175	105	players	26	501

Notes: The words above are in the same order as the Top 30 most "female" and most "male" words selected by Lasso-logistic model (see Table 3). Level 1 includes all possible gender classifiers, while Level 4 includes only pronouns like "he" or "she". "*#Female*" and "*#Male*" refers to the no. of $Female = 1$ posts and $Female = 0$ posts each word occurs in resepectively.

Table A2: Most Frequent Words in *Female* and *Male* Posts

Level 1						Level 4					
Most common in *Female*			Most common in *Male*			Most common in *Female*			Most common in *Male*		
Word	#*Female*	#*Male*	Word	#*Female*	#Male	Word	#*Female*	#*Male*	Word	#*Female*	#*Male*
*	5,253	14,525	*	5,253	14,525	*	2,747	7,355	work	2,259	7,986
life	4,034	7,644	work	3,800	13,989	work	2,259	7,986	*	2,747	7,355
work	3,800	13,989	paper	1,503	11,727	life	2,058	4,092	paper	1,035	6,495
hot	3,613	1,053	job	3,091	10,313	love	1,778	2,039	job	1,624	5,502
love	3,297	4,274	economics	1,120	9,808	job	1,624	5,502	great	1,382	4,829
sex	3,103	1,535	great	2,323	9,181	feel	1,529	2,333	economics	646	4,690
job	3,091	10,313	best	2,558	8,552	sex	1,414	794	best	1,336	4,407
feel	2,574	5,167	research	1,407	8,238	great	1,382	4,829	school	1,351	4,297
best	2,558	8,552	school	2,446	8,228	school	1,351	4,297	research	831	4,267
school	2,446	8,228	market	1,750	7,954	best	1,336	4,407	papers	592	4,194
kids	2,441	2,200	life	4,034	7,644	hot	1,322	645	life	2,058	4,092
great	2,323	9,181	phd	1,751	7,295	married	1,130	664	students	792	3,841
married	2,231	1,207	papers	854	7,177	student	1,128	3,762	phd	980	3,825
friends	2,048	2,504	econ	1,133	6,950	friends	1,117	1,430	student	1,128	3,762
nice	1,978	4,590	students	1,474	6,889	nice	1,067	2,400	market	714	3,694
money	1,951	6,011	theory	415	6,347	kids	1,043	1,202	economist	545	3,342
home	1,778	2,734	money	1,951	6,011	paper	1,035	6,495	money	992	3,290
phd	1,751	7,295	data	729	5,648	home	1,028	1,523	course	778	3,137
market	1,750	7,954	student	1,560	5,607	friend	1,001	1,924	wrong	835	3,136
date	1,729	835	economist	855	5,539	money	992	3,290	idea	714	2,997
family	1,653	2,685	wrong	1,344	5,487	phd	980	3,825	department	638	2,908
attractive	1,578	417	economists	697	5,461	date	908	471	econ	588	2,819
student	1,560	5,607	course	1,320	5,416	family	896	1,568	theory	256	2,789
relationship	1,506	1,169	question	1,109	5,257	relationship	893	631	question	641	2,695
paper	1,503	11,727	idea	1,158	5,184	happy	853	1,331	professor	486	2,577
students	1,474	6,889	feel	2,574	5,167	wrong	835	3,136	university	637	2,536
happy	1,452	2,536	economic	466	5,152	research	831	4,267	economists	340	2,480
dating	1,423	399	department	935	4,985	students	792	3,841	tenure	633	2,447
beautiful	1,419	610	university	955	4,970	course	778	3,137	working	719	2,432
friend	1,412	2,423	r	682	4,774	working	719	2,432	nice	1,067	2,400

Notes: The words above come from non-"0" categories (see Table 4) and are sorted by the number of *Female* and *Male* posts they occur in respectively. Level 1 includes all possible gender classifiers, while Level 4 includes only pronouns like "he" or "she". "#*Female*" and "#*Male*" refers to the no. of *Female* = 1 posts and *Female* = 0 posts each word occurs in resepectively.

B. Dynamic Topic Analysis - use counts

In Section 3.2, I use dummy variables to measure whether a post has an *Academic/Professional* or *Personal/Physical* focus, i.e. $D \coloneqq 1$ if it includes at least one term from a given topic. It is relatively easier to interpret the estimated coefficients on lagged variables as a change in probablity of staying on the same topic. However, the dummy variables cannot capture the subtle deviation from each topic. For example, the prior post contains five academic terms, but the next one only contains one. Both posts are labeled as $D = 1$. In this sense, using dummies may have underestimated the actual differences in the effects of gender on the persistence in each topic. Here I replace D by *Topic* - the number of *Academic/Professional* terms and the number of *Personal/Physical* terms, and re-do the estimations as above.

In parallel with model (7) in Section 3.2, I estimate the following reduced form model:

$$Topic_{t,p} = \beta_0 + \beta_1 Topic_{t,p-1} + (\phi_1 Topic_{t,0} + \phi_2 Topic_{t,1} + \phi_3 \overline{Topic_t}) + \theta\, 1[last\ page] + \nu_{t,p} \tag{9}$$

To specify the effects of gender in the prior post, in parallel with model (8), I estimate:

$$\begin{aligned} Topic_{t,p} = \beta_0 + \beta_1 Topic_{t,p-1} + Gender_{t,p-1}\lambda' + (Topic_{t,p-1} \times Gender_{t,p-1})\eta' \\ + (\phi_1 Topic_{t,0} + \phi_2 Topic_{t,1} + \phi_3 \overline{Topic_t}) + \theta\, 1[last\ page] + \nu_{t,p} \end{aligned} \tag{10}$$

where each post in the base group follows a genderless ("neutral") post and it occurs on the first page of the thread it belongs to.

Table B1 shows the regression outputs for model (9) and model (10), under each topic. Table B2 restrict to academic-oriented threads where the mean no. *Academic/Professional* terms across all posts within the same thread is $\geq$ the median 1. The estimates are comparable. In Table B3, I compare the effects of gender on persistence under 16 initial conditions based on each thread's title and its first post. Note under **(Neither, 1, 0)** - threads starting with an academic focus only and not related to gender initially, conditional on the prior post containing exactly 1 academic term ($Topic_{t,p-1} = 1$), the mention of male **increases** the number of academic terms in the next post by 0.12 relative to the neutral group, whereas the mention of female **decreases** the no. academic terms more than twice than the neutral group. The F-test on equal gender effects of *Female* vs.

Male gives a p-score around 0.003.

In summary, the robustness checks provide a more complete picture, and yield the same conclusions that there is a significantly higher deviation from an academic focus and a significantly lower deviation from a personal topic when the prior post is female rather than male or neutral.

Table B1. Persistence in Topics (Any Thread)

	Academic/Professional		Personal/Physical	
	(1)	(2)	(3)	(4)
$Topic_{t,p-1}$	-0.046	-0.010	-0.051	-0.031
	(0.003)	(0.003)	(0.004)	(0.003)
<u>Gender in the Prior Post</u>				
Neutral (base)				
Female		0.004		0.039
		(0.012)		(0.006)
Male		0.058		0.003
		(0.013)		(0.004)
$Female \times Topic_{t,p-1}$		-0.063		-0.034
		(0.007)		(0.006)
$Male \times Topic_{t,p-1}$		-0.058		-0.025
		(0.005)		(0.009)
$\overline{Topic_t}$	1.003	0.995	1.009	1.006
	(0.009)	(0.009)	(0.009)	(0.009)
$Topic_{t,0}$ (Titles)	0.079	0.068	0.046	0.044
	(0.005)	(0.005)	(0.003)	(0.003)
$Topic_{t,1}$ (*First* posts)	-0.058	-0.057	-0.070	-0.069
	(0.001)	(0.001)	(0.002)	(0.002)
1[last page]	0.382	0.373	0.052	0.051
	(0.013)	(0.013)	(0.003)	(0.003)
Constant	0.053	0.038	0.020	0.017
	(0.007)	(0.008)	(0.002)	(0.002)
Adj. R^2	0.191	0.193	0.168	0.168
F	28,901.235	17,004.891	10,661.330	6,549.256
N	$1,333,515$	$1,333,515$	$1,333,515$	$1,333,515$

Notes: Standard errors in parentheses are clustered at the thread level.

Table B2. (Threads s.t. $\overline{Academic_t} \geq 1$)

	Academic/Professional		*Personal/Physical*	
	(1)	(2)	(3)	(4)
$Topic_{t,p-1}$	-0.046	-0.010	-0.049	-0.021
	(0.003)	(0.004)	(0.006)	(0.005)
<u>Gender in the Prior Post</u>				
Neutral (base)				
Female		0.002		0.057
		(0.030)		(0.009)
Male		0.076		0.001
		(0.023)		(0.005)
$Female \times Topic_{t,p-1}$		-0.066		-0.050
		(0.008)		(0.008)
$Male \times Topic_{t,p-1}$		-0.058		-0.030
		(0.006)		(0.013)
$\overline{Topic_t}$	0.973	0.968	0.958	0.955
	(0.011)	(0.011)	(0.014)	(0.014)
$Topic_{t,0}$ (Titles)	0.071	0.061	0.073	0.069
	(0.005)	(0.004)	(0.007)	(0.007)
$Topic_{t,1}$ (*First* posts)	-0.058	-0.057	-0.070	-0.069
	(0.001)	(0.001)	(0.002)	(0.002)
1[last page]	0.528	0.517	0.046	0.046
	(0.020)	(0.019)	(0.003)	(0.003)
Constant	0.181	0.151	0.029	0.025
	(0.020)	(0.021)	(0.003)	(0.003)
Adj. R^2	0.124	0.126	0.160	0.160
F	5928.837	3520.076	3007.530	1873.375
N	$772,873$	$772,873$	$772,873$	$772,873$

Notes: Standard errors in parentheses are clustered at the thread level. Restrict to threads where the mean no. *Academic/Professional* across all posts is $\geq$ the median, which equals to 1. The idea is to check whether the state dependence results are robust among threads that are more academically oriented.

Table B3. Gender on Persistence in *Academic/Professional* under 16 Initial Conditions

			Neutral	Female		Male		Ratio	H0: $\lambda_F + \eta_F = \lambda_M + \eta_M$	
Titles	No. Titles (%)	No. Posts (%)	β_1	η_F	$\lambda_F + \eta_F$	η_M	$\lambda_M + \eta_M$	$\frac{\lambda_F+\eta_F}{\lambda_M+\eta_M}$	F stat	p-value
Any title	127029 (100%)	1333515 (100%)	-0.010	-0.063	-0.060	-0.058	-0.0002	339.824	26.240	0.000
Initial: (Gender, 1 if Academic, 1 if Personal)										
(Female,1,0)	2117 (1.67%)	19470 (1.46%)	-0.001	-0.095	-0.034	-0.037	0.093	-0.368	1.710	0.191
(Male,1,0)	15215 (11.98%)	141659 (10.62%)	0.017	-0.041	-0.104	-0.080	-0.077	1.359	0.100	0.755
(Both,1,0)	1929 (1.52%)	19377 (1.45%)	0.083	-0.103	-0.338	-0.112	-0.177	1.911	2.180	0.140
(Neither,1,0)	31294 (24.64%)	403977 (30.29%)	-0.032	-0.025	-0.043	-0.028	0.119	-0.364	9.090	0.003
(Female,1,1)	434 (0.34%)	4270 (0.32%)	0.012	-0.160	0.153	0.128	-0.039	-3.948	1.660	0.199
(Male,1,1)	908 (0.71%)	8081 (0.61%)	0.052	-0.112	-0.035	-0.097	-0.155	0.228	0.370	0.541
(Both,1,1)	446 (0.35%)	4727 (0.35%)	0.148	-0.157	-0.174	-0.177	-0.113	1.535	0.180	0.669
(Neither,1,1)	1815 (1.43%)	22101 (1.66%)	-0.057	-0.079	0.031	-0.173	0.349	0.090	6.760	0.009
(Female,0,1)	2394 (1.88%)	23168 (1.74%)	-0.037	-0.028	-0.021	-0.031	0.057	-0.371	4.500	0.034
(Male,0,1)	2838 (2.23%)	24970 (1.87%)	0.031	-0.115	-0.043	-0.086	-0.088	0.493	1.230	0.268
(Both,0,1)	2050 (1.61%)	22024 (1.65%)	0.065	-0.099	-0.064	-0.075	-0.083	0.773	0.440	0.505
(Neither,0,1)	4065 (3.2%)	45476 (3.41%)	-0.031	0.017	0.018	-0.003	0.026	0.668	0.090	0.770
(Female,0,0)	6020 (4.74%)	51126 (3.83%)	0.148	-0.225	-0.119	-0.156	-0.027	4.347	4.490	0.034
(Male,0,0)	22285 (17.54%)	180255 (13.52%)	-0.017	0.077	0.018	-0.054	-0.052	-0.348	2.790	0.095
(Both,0,0)	5006 (3.94%)	45617 (3.42%)	0.060	-0.090	-0.051	-0.084	-0.028	1.818	0.360	0.548
(Neither,0,0)	28213 (22.21%)	317217 (23.79%)	-0.032	-0.029	-0.019	-0.021	0.014	-1.357	2.920	0.087

Table B4. Gender on Persistence in *Personal/Physical* under 16 Initial Conditions

			Neutral	Female		Male		Ratio	H0: $\lambda_F + \eta_F = \lambda_M + \eta_M$	
Titles	No. Titles (%)	No. Posts (%)	β_1	η_F	$\lambda_F + \eta_F$	η_M	$\lambda_M + \eta_M$	$\frac{\lambda_F+\eta_F}{\lambda_M+\eta_M}$	F stat	p-value
Any title	127029 (100%)	1333515 (100%)	-0.031	-0.034	0.004	-0.025	-0.022	-0.195	16.200	0.0001
Initial: (Gender, 1 if Academic, 1 if Personal)										
(Female,1,0)	2117 (1.67%)	19470 (1.46%)	-0.016	-0.041	-0.033	-0.020	0.005	-6.257	2.120	0.145
(Male,1,0)	15215 (11.98%)	141659 (10.62%)	-0.033	-0.041	-0.016	-0.037	-0.045	0.352	2.100	0.147
(Both,1,0)	1929 (1.52%)	19377 (1.45%)	0.051	-0.089	0.026	-0.098	-0.104	-0.250	25.720	0
(Neither,1,0)	31294 (24.64%)	403977 (30.29%)	-0.044	-0.005	0.012	0.014	0.018	0.686	0.120	0.730
(Female,1,1)	434 (0.34%)	4270 (0.32%)	-0.055	-0.054	-0.0002	-0.008	0.109	-0.002	1.890	0.169
(Male,1,1)	908 (0.71%)	8081 (0.61%)	-0.060	-0.029	0.136	-0.006	-0.024	-5.766	3.020	0.083
(Both,1,1)	446 (0.35%)	4727 (0.35%)	0.030	-0.085	0.093	-0.093	-0.071	-1.312	7.600	0.006
(Neither,1,1)	1815 (1.43%)	22101 (1.66%)	-0.053	-0.039	-0.079	-0.129	-0.051	1.574	0.590	0.443
(Female,0,1)	2394 (1.88%)	23168 (1.74%)	-0.055	0.013	0.012	-0.002	-0.001	-9.110	0.160	0.686
(Male,0,1)	2838 (2.23%)	24970 (1.87%)	-0.026	-0.061	0.055	-0.066	-0.068	-0.800	10.170	0.001
(Both,0,1)	2050 (1.61%)	22024 (1.65%)	-0.015	-0.047	0.019	-0.018	-0.009	-2.021	0.780	0.378
(Neither,0,1)	4065 (3.2%)	45476 (3.41%)	-0.048	0.041	0.043	0.019	0.033	1.288	0.120	0.726
(Female,0,0)	6020 (4.74%)	51126 (3.83%)	-0.060	-0.025	-0.011	0.023	0.042	-0.264	7.880	0.005
(Male,0,0)	22285 (17.54%)	180255 (13.52%)	-0.034	-0.027	-0.001	-0.004	-0.020	0.057	0.590	0.441
(Both,0,0)	5006 (3.94%)	45617 (3.42%)	0.047	-0.103	-0.015	-0.094	-0.059	0.249	6.910	0.009
(Neither,0,0)	28213 (22.21%)	317217 (23.79%)	-0.054	0.007	0.005	0.015	0.011	0.434	0.090	0.770

Table B5. Examples of Adjacent Posts

Title	$Gender_{t,p-1}$	post $p-1$	$Gender_{t,p}$	post p
Initial: (Neither, 1, 0)				
"Queen's University job market candidates are up. 2017-2018"	1(female)	"Her quantity is pretty outstanding. 3 publications, 4 working papers, and 3 works in progress which seem like they are actually real things on the go. A nice mix of solo and single authored papers. There won't be very many fresh PhD's on the market that can match that. I didn't look closely enough to comment on quality. So I don't know if her publications are typical of what she is able to churn out, or if some of her working papers are top field or better caliber. Either way, it looks like she should generate some interest. If not from top schools, from any school that counts top 100 publications as their signs of success from faculty. Because it looks like she can produce those at an impressive rate."	2 (neutral)	"Stop with the self-promotion you little shiets"
"Is two leading field journals enough to get tenure at a top 30 department?"	1	I am sure there is something that we don't know. Otherwise, this is a very weak record especially given the fact that she took 9 years to get tenure. In fact, I can't think of any decent phd granting econ department which would grant tenure to this file.	1	"Collegial externalities - she looks nice, great gender."
Initial: (Both, 1, 0)				
"Importance of Looks in Academic Job market"	0	"All that matters for men is what shows in a dress shirt. So abs only matter so long as they are flat. Definition won't do anything. Shoulders show better than any other muscle group. But hygiene shows best."	2	"It's really all about the JMP."

C. Trend Analysis

The main pages of the forum record a rough time stamp of the latest post of each thread, such as "1 day ago", "1 month ago", "6 months ago", '1 year ago", "2 years ago" etc. From these time stamps, I obtain the month of the latest update for threads initiated or updated within one year[29], from November 2016 to late October 2017, and the number of years relative to Oct 2017 for threads labeled "1 year ago" or earlier. At the mean time, I record the time stamps of the *First* post of each thread, i.e. the time when a thread started. The time stamps are in the same format as those of the latest updates. I integrate the current dataset with my scraping as of the end of Sept 2016. As a result, I identify the month of the *First* post for threads initiated in the past two years, from October 2015 to October 2017, or no. years relative to Oct 2017 for earlier threads.

I construct four time series of the mean number of *Academic/Professionl* and *Personal/Physical* terms respectively, as follows:

- by Month of the First Post (Start Month): available from Oct 2015 to Oct 2017
- by Month of the Latest Update: available from Oct 2016 to Oct 2017
- by Year of the First Post (Start Year): -6 to 0, relative to Oct 2017
- by Year of the Latest Update: -3 to 0, relative to Oct 2017

Main Findings

- *Female* posts show smaller month-to-month fluctuations in *Academic/Professional* than *Male* posts, with an exception in August 2017, the same month when the NY Times article[30] reveals this study. Figure C1 shows that discussions about women are more academically oriented for the first time in Aug 2017, but there is a slow decline going back to the pre-trend after that.
- *Male* posts show smaller month-to-month fluctuations in *Personal/Physical* than *Female* post. The "intervention" in Aug 2017 does not make a notable difference in this topic. (Figure C5; Figure C6)

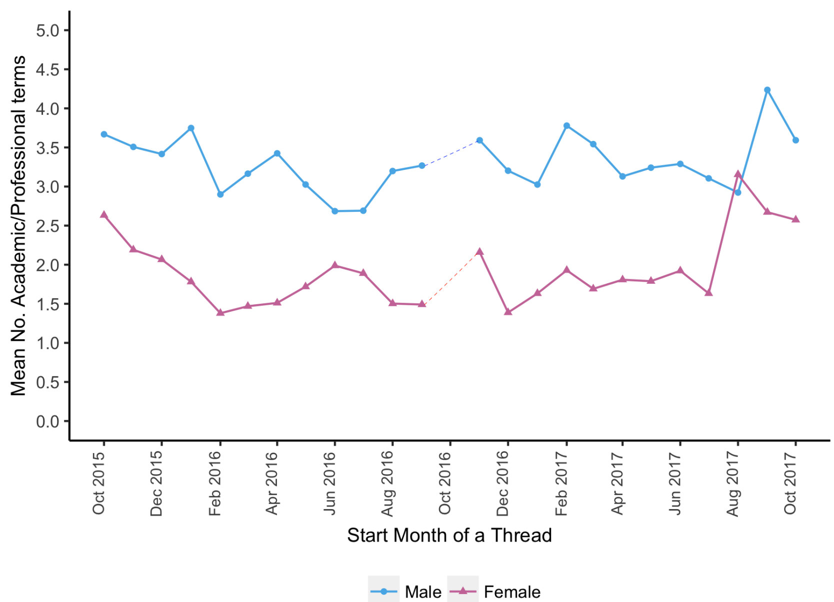
[29] Relative to 10/28/2017.
[30] Wolfers, Justin. 2017."Evidence of a Toxic Environment in Economics". New York Times. 18 August.

- There is no clear seasonal pattern due to job market.

- By start year, threads **initiated** 5 or 6 years ago contain more *Academic/Professional* terms on average but decline since then (Figure C3). There is also a notable increase in *Personal/Physical* from 5 years ago to 4 years ago (Figure C7).

- There are small year-to-year variations in *Academic/Professional* (Figure C4) and *Personal/Physical* (Figure C8) for threads **initiated or updated** in the last four years.

Figure C1. Mean #*Academic/Professional* by Start Month of Threads



Notes: Threads initiated in Oct 2016 are not identified. The latest dataset indicates the month a thread started from Nov 2016 to Oct 2017. The time stamps for threads started in Oct 2015 to Sept 2016 are preserved from an earlier round of scraping in Sept 2016.

Figure C2.Mean $\#$*Academic/Professional* by Month of the Latest Update

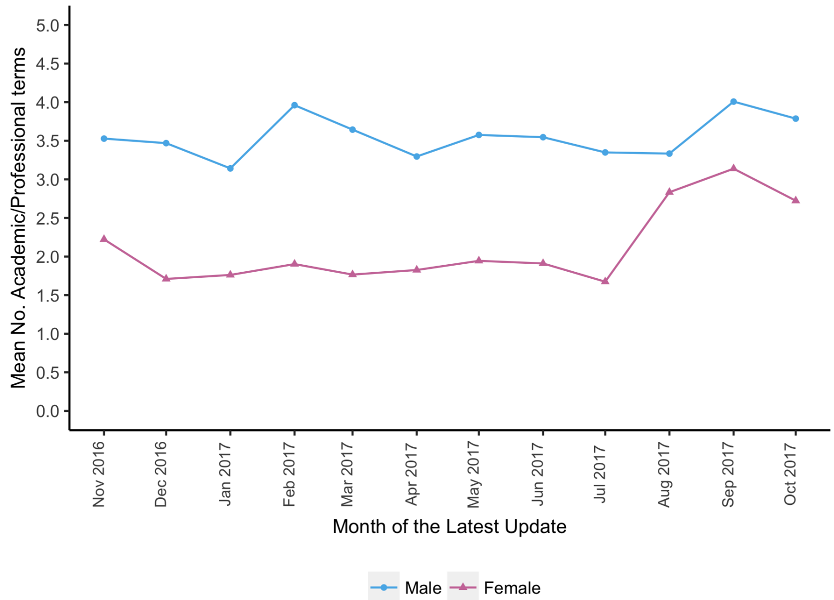


Figure C3. Mean $\#$*Academic/Professional* by Start Year

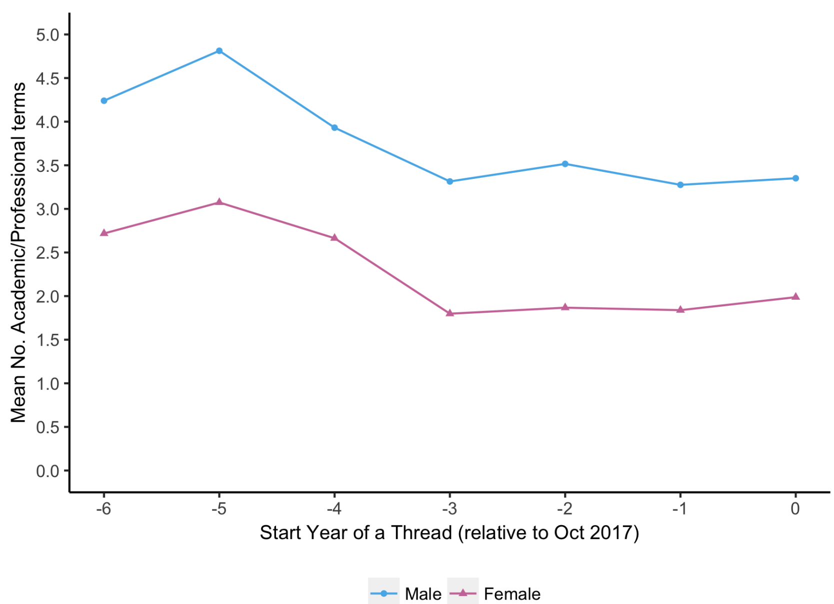


Figure C4. Mean #*Academic/Professional* by Year of the Latest Update

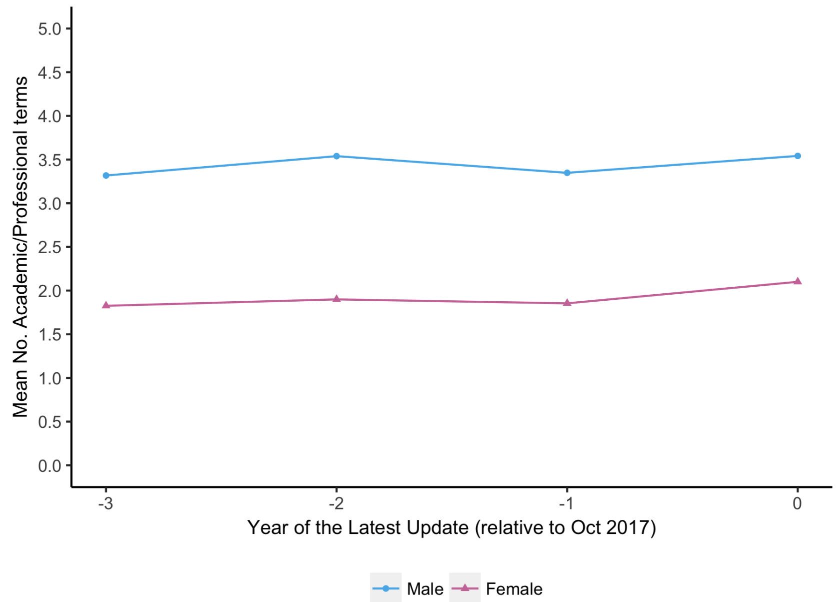
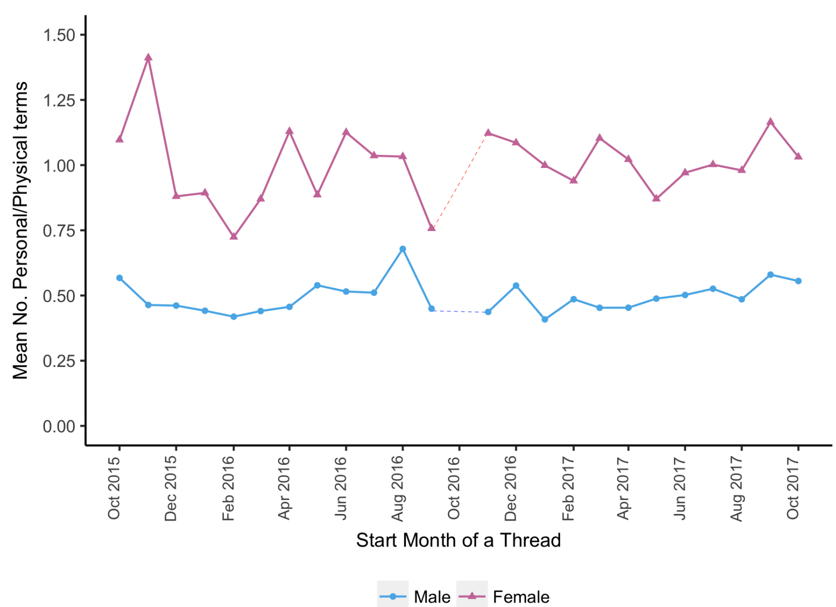


Figure C5. Mean #*Personal/Physical* by Month of the Latest Update



Notes: Threads initiated in Oct 2016 are not identified. The latest dataset indicates the month a thread started from Nov 2016 to Oct 2017. The time stamps for threads started in Oct 2015 to Sept 2016 are preserved from an earlier round of scraping in Sept 2016.

Figure C6. Mean #*Personal/Physical* by Month of the Latest Update

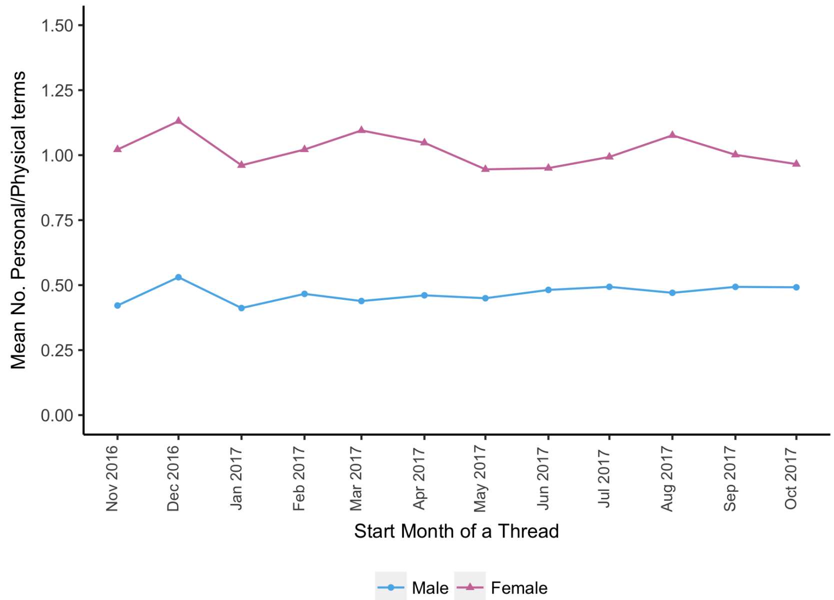


Figure C7. Mean #*Personal/Physical* by Start Year

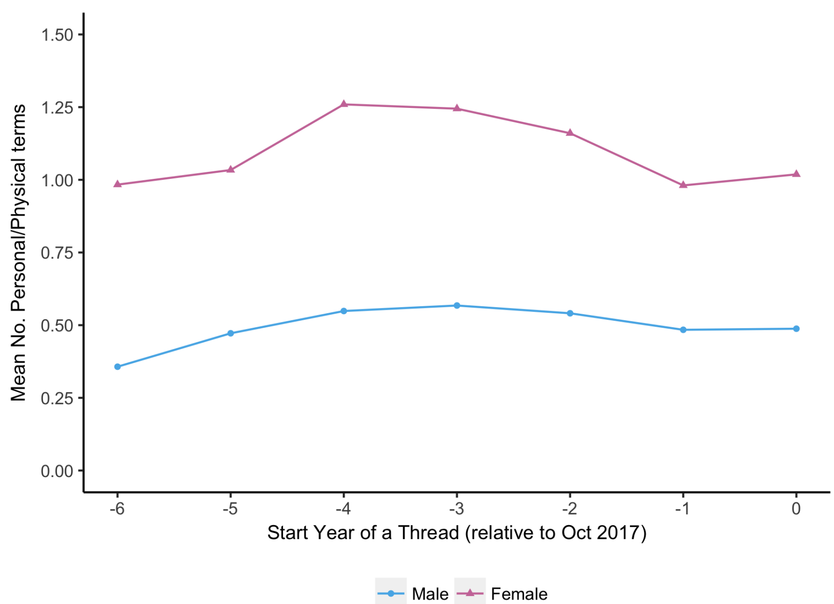


Figure C8. Mean $\#Personal/Physical$ by Year of the Latest Update

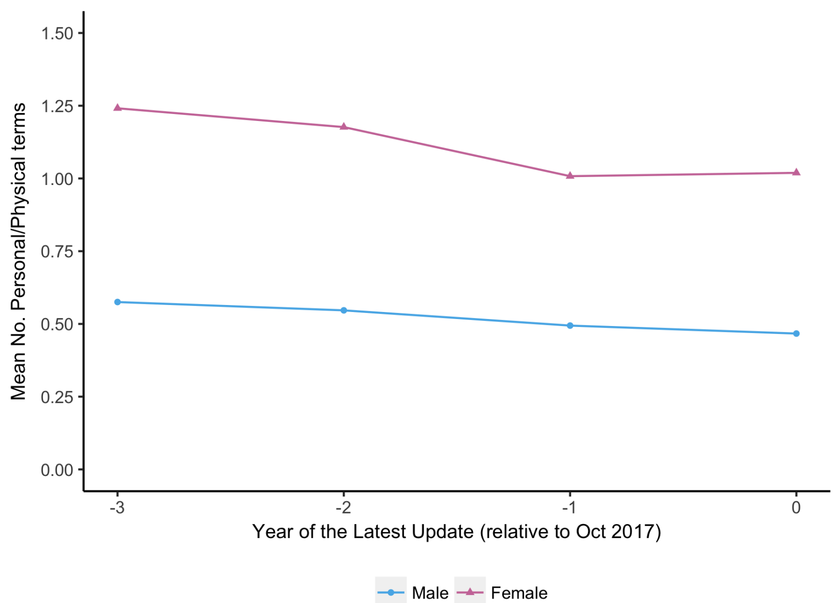


Table C1: *Academic/Professional* - Prior vs. Post August 2017

	Number of *Academic/Professional* Words							
	Level 1		Level 2		Level 3		Level 4	
	Prior	Post	Prior	Post	Prior	Post	Prior	Post
$Female_i$	-1.374	-0.819	-1.573	-0.788	-1.702	-1.063	-1.556	-0.606
	(0.020)	(0.106)	(0.025)	(0.130)	(0.029)	(0.153)	(0.032)	(0.182)
Constant	2.992	3.168	3.362	3.492	3.520	3.654	3.433	3.461
	(0.015)	(0.066)	(0.019)	(0.080)	(0.020)	(0.088)	(0.022)	(0.100)
R^2	0.009	0.003	0.010	0.002	0.009	0.003	0.011	0.001
F Stat.	4679.138	60.159	3831.112	36.659	3363.783	48.348	2312.590	11.068
N	415,168	20,449	302,501	15,788	263,149	13,161	185,644	8,939

Notes: "Prior" restricts the sample to threads *initiated* before August 2017, while "Post" look at threads created from August to October 2017. Standard errors in parentheses are clustered at the thread level. Each post contains at least 3 and at most 300 words. "Level 1" to "Level 4" refer to increasingly restrictive levels of gender classifiers to identify gender-related posts (see Figure 1).

Table C2: *Personal/Physical* - Prior vs. Post August 2017

	Number of *Personal/Physical* Words							
	Level 1		Level 2		Level 3		Level 4	
	Prior	Post	Prior	Post	Prior	Post	Prior	Post
$Female_i$	0.711	0.689	0.691	0.645	0.607	0.525	0.596	0.507
	(0.009)	(0.038)	(0.011)	(0.042)	(0.012)	(0.051)	(0.013)	(0.054)
Constant	0.407	0.416	0.452	0.464	0.441	0.459	0.520	0.531
	(0.002)	(0.011)	(0.003)	(0.013)	(0.003)	(0.014)	(0.005)	(0.019)
R^2	0.041	0.039	0.033	0.031	0.024	0.019	0.023	0.019
F Stat.	6006.779	327.805	3911.239	232.295	2387.962	105.261	1953.126	87.879
N	415,168	20,449	302,501	15,788	263,149	13,161	185,644	8,939

Notes: "Prior" restricts the sample to threads *initiated* before August 2017, while "Post" look at threads created from August to October 2017. Standard errors in parentheses are clustered at the thread level. Each post contains at least 3 and at most 300 words. "Level 1" to "Level 4" refer to increasingly restrictive levels of gender classifiers to identify gender-related posts (see Figure 1).