The Emergence of Gender Inequality in a Crowdfunding Market: An Experimental Test of Gender System Theory

J. S. Radford[1]

University of Chicago

June 28, 2016

Acknowledgements
Special thank you to the anonymous reviewers at the *American Sociological Review*, graduate students Jim Murphy and Forest Gregg at the University of Chicago and my colleagues at the Lazer Lab at Northeastern University who have given me extensive feedback on the analysis presented. Thanks also to Professors Elisabeth Clemens, Kristen Schilt, and James Evans at the University of Chicago and David Lazer at Northeastern University who provided guidance at various stages. Finally, this research has benefited from feedback from presentations given at the Social Theory and Evidence Workshop and the Computational Social Science workshop at the University of Chicago, the Economic Sociology Working Group at MIT, Collective Intelligence, the Association for Computational Linguistics, and the American Sociological Association

[1] Corresponding author; 177 Huntington Ave Ste. 1010, Boston, MA 02115; jsradford@uchicago.edu.

Abstract

Research on ascriptive inequality investigates how social groups differ, whether resources are allocated unequally by group differences, and what mechanisms create and sustain this unequal allocation. In the sociology of gender, gender system theory links these three questions into a single theory but has yet to be tested comprehensively. In this study, I perform such a test using data from DonorsChoose, a crowdfunding website for public school teachers in the United States. The data is large and diverse enough to measure the gender differences theorized by gender system theory and allows us to examine whether these gender differences correspond to inequalities in funding. Critically, the data also contain a natural experiment whereby teachers' identity was hidden until 2008. This allows for a direct test of the causes of gender inequality hypothesized by gender system theory. The results show that inequality only emerges after educators' identity was published. And, deanonymization caused inequality to emerge across all types of gender difference. These results provide robust support for gender system theory and contribute to research on the structure and causes of gender inequality.

Research on ascriptive inequality attempts to answer three questions: how are differences between groups defined? To what extent do these differences correspond to unequal allocations of resources and power? What causes this unequal allocation of rewards? Each question has received substantial treatment. Yet these explanations are often disconnected from one another. The present challenge in studying inequality is synthesizing answers to these questions in a coherent and concise theory (Reskin 2000, 2003).

In the sociology of gender, Ridgeway and Correll (2004) and Risman (2004) advanced gender system theory which synthesized several major lines of research on gender. The theory provides a single explanation that links gender differences, inequality, and the causes of inequality. However, due in large part to the theory's high evidentiary demands, there has been little work using it (see Bobbitt-Zeher 2011 for an exception) and no comprehensive test of its central claims. As a result, research on gender inequality continues to be fragmented along the three questions.

In this study, I use a data set from a crowdfunding website for public school educators in the United States called DonorsChoose to test the core arguments of gender system theory. The data from the website is large and varied enough to measure the key gender differences put forward by Ridgeway and Correll and Risman. And, the fact that it is a crowdfunding site allows for a direct test for whether or not resources are unequally distributed according to these gender differences. And, most importantly, the data contain a natural experiment wherein educators were anonymous until 2008. This natural experiment allows for a test of the theory's main causal claim: that gender inequality does not occur until people can see one another as male or female.

Gender System Theory

Gender system theory answers the first two questions of inequality by arguing that gender differences and gender inequality occur along three dimensions: the individual, interactional, and institutional (Ridgeway and Correll, 2004; Risman, 2004). Individual gender refers to someone's sex category and includes male, female, intersex, or transsexual. Interactional gender refers to someone's behavior and is commonly referred to as their gender performance (West and Zimmerman 1987). These include masculine, feminine, transgender, or genderqueer. Institutional gender refers to the gender attached to someone's social position or role and captures how we understand queen, kindergarten teacher, father, or high school gym teacher to be suitable for and typical of men and women. The empirical challenge put forward by gender system theory is in requiring that all three dimensions must be accounted for *simultaneously* when analyzing gender inequality.

Sex, the individual dimension, refers to the ways in which we biologically differentiate people by gender. People label themselves and one another as male or female and reward people differently because of those labels. Historically, individual differences were understood to be phenotypical expressions of biological differences. Only in the past fifty years has the concept of "gender" come to supplant this understanding, defining behavioral differences as social constructs independent of underlying biological features.

Despite the theoretical revolution, the vast majority of studies of gender inequality are based on estimating the effect of a binary variable for male or female. This narrow focus leads many scholars to control away critical factors like occupation, education, and parental status in studies of gender wage inequality (O'Neill 2003, Blau and Kahn 2006, Lips 2013). By focusing only on sex inequality, we fail to account for broader forms of *gender* inequality. Gender system theory requires us to account for differences in gender performance, such as whether there is a pay bump for masculine or feminine behavior, and gendered positions, such as the pay difference between mechanical and biological engineers (Cech 2013a).

Second, research on the interactional dimension examines the way people's behavior is interpreted and rewarded based on their ascribed sex, what is referred to as gender performance. Performance theorists argue that people punish or reward one another's behavior based on whether or not that behavior meets the expectations for their ascribed sex (Butler 1999, 2004; West and Zimmerman 1987, 2009). For example, males are expected to be assertive and analytical and are rewarded for conforming to this. In contrast, women are punished for being assertive or analytical but rewarded for being deferential and caring.

The body of work on gender performance shows that people who behave in ways that violate sex-derived expectations are stigmatized and subordinated (Williams 1992; Connell 1995; Pascoe 2011). However, other work has shown that there are many different ways of being masculine or feminine and the institutional context influences which expectations apply (Wood and Eagly 2002; Cech 2013a). For example, women may be rewarded for assertiveness as mothers, but not as managers. Following gender system theory, when must take all three dimensions into account to determine whether the rewards and punishments of gender performance depend on or transcend the institutional context.

The third dimension, institutional gender, is when we interpret social positions as more appropriate for or desirable to men and women and includes how we come to associate men and women with positions like librarian, homemaker, and chief executive officer. Research has shown how fields become gendered (Light 1999; Goldstein 2014), how differences within fields become gendered (Eagly and Karau 2002; Cech 2013a), and how gender-neutral job descriptions and rules are used to differentiate otherwise equivalent positions as appropriate for men or women (Acker 1990; Bobbitt-Zeher 2011).

Institutional gender and the inequality resulting from it has persisted in the United States despite two generations of gender desegregation (England, Alison, and Wu. 2007). This has led many scholars to treat these differences as unrelated to gender (Lips 2013). However, experimental work shows that individuals value work differently when they believe men or women do that work (Ridgeway 1991; Glick 1991; Glick, Wilk, and Perreault 1995; Alksnis, Desmarais, and Curtis 2008). The implication is that the unequal status we attribute to male-dominated and female-dominated social positions is not due to any real differences in expertise, skill, or productivity, but simply to the perception that men or women do those jobs. In gender system theory calls for research which treats these social positions the same as the binary variable for males and females.

Returning to the three guiding questions of gender inequality, gender system theory answers the first question, how are gender differences constructed, by asserting that they are created along three dimensions: individual (sex), interactional (behavior), and institutional (social position). It answers the second question, how are gender differences unequally rewarded, by asserting that resources are distributed unequally along each of these three dimensions. That is, sex, behavior, and social positional each contribute to gender inequality. The third question, what causes inequality, has a separate answer.

Ridgeway (2009) argues that the gender system only ‘turns on’ when people’s cognitive schema about gender are triggered by sex categorization (Ridgeway 2009). Experiments show that when people identify someone as male or female, they begin interpreting the situation using a gendered lens or frame (Brewer and Lui, 1989; Stangor, Lynch, Duan, and Glas 1992; Conway, Pizzamiglio, and Mount 1996; Wagner and Berger 1997; Ito and Urland, 2003; Ridgeway 2011). This body of research shows that only after we categorize people in a situation as male and female do we then interpret behaviors as masculine or feminine and social positions as female- or male-appropriate.

Critical questions remain for this causal argument. Research on stereotype malleability shows that a variety of information can trigger sex categorization and situational factors can moderate its effect on behavior (Blair 2002; Lenton, Bruder, and Sedikides 2009). For example, research has shown stereotypes can be triggered by describing someone with gender-defining nouns like queen and father or gender-related nouns like ballet, beer-belly, computer, and fashion (Banaji and Hardin 1996; Blair and Banaji 1996). These findings suggest more work is needed to determine whether, in anonymized cases, proxy information might trigger gender frames and thereby create inequality.

The challenge for testing gender system theory is the high evidentiary demands it makes. First, a data set must allow researchers to measure all three dimensions of the gender system. Second, the data must also document a situation where we can determine if these three dimensions are rewarded before and after a variety of potential cognitive triggers are offered. I will argue in the next section that the data from DonorsChoose offers just this kind of analytical power. I then use the data to test gender system theory's answer to each of the three core questions about gender inequality.

In the first stage, I describe how I operationalize the three dimensions of the gender system and show the results of correlational tests which reveal how they combine to form rich patterns of gender difference among educators on DonorsChoose. In the second stage, I account for gender inequality, estimating the extent to which the three dimensions affect whether or not educators receive funding through DonorsChoose. In the final stage, I leverage the natural experiment to test several potential causes of these patterns of gender inequality. The results from each stage provide strong support for gender system theory.

Case: DonorsChoose

Overview

DonorsChoose is a crowdfunding website for educators from public schools in the United States. It was founded in 2000 for educators in New York City to raise money for projects like new books or a class field trip. Since 2000, eligibility has expanded to include educators from all public schools in the United States. Between 2002 and 2010, the period covered by the data made available for this project, more than 90,000 educators from over 40,000 schools posted over 290,000 projects. In that time, donors gave a total of $129 million to educators across the United States.[i]

The experience of seeking funding through DonorsChoose has remained relatively stable since its founding. Let's say a kindergarten teacher wants a set of books for her classroom so students can learn to read. The teacher would create an account on DonorsChoose and fill out basic information about herself and her classroom, including the school, grade level, and subject area. She would then create a project proposal where she writes an essay about why donors should fund her project and provides information about the resources she wants, whether it for essential or enrichment activities, and how many students will be impacted. Once complete, the project information is vetted by DonorsChoose staff and posted on the website.

Once the project is posted, anyone can donate money to the project through the website. Despite large amounts of missing data on donors, the analysis in stage three indicates that most donors are likely to be people the teacher has never met. During the time the project is online, it may also become eligible for matching grants from institutional donors or advertised on a giving page maintained by networks of donors who solicit others to donate to projects through the page. If the project is fully funded, it is closed and the DonorsChoose staff purchases the books and sends them to the teacher. In some cases, the teacher might send thank-you packages to donors. These packages typically include an impact letter from the teacher, thank you letters written by students, and photos of students reading the books. If the project does not meet its goal by the expiration date, the teacher receives nothing.[ii]

Gender on DonorsChoose

In this study, I focus on estimating the extent to which gender affects the likelihood that projects will succeed. This is because all three dimensions of gender can be defined using data provided on project pages. And, the natural experiment discussed below occurs on educators' project pages, making projects the appropriate level of analysis for detecting causation.

Conceptually, I map the three dimensions of the gender system onto information found in educators' projects. I measure the individual dimension, educator's sex, as whether or not educators identify themselves as male or female on the website. I measure the interactional dimension, gender performance, based on whether educators use male-typical or female-typical language when presenting themselves and their projects in their essays. Lastly, I measure the institutional dimension, the gender of social positions, by how male- or female-dominated an educator's job is (e.g. kindergarten teacher versus high school gym teacher).

The Natural Experiment

The natural experiment I leverage is a shift from anonymous to individually-identified project proposals. Prior to 2008, educators' personal information was kept private. Visitors to the website could see project proposals, read essays, and see information about the school and aggregate data on students. However, they could not see any information about who the educators were.

In January 2008, DonorsChoose began publishing educators' identity (e.g. "Mr. Spain") on project pages. Later that year, in December 2008, DonorsChoose updated the design of search results to include educators' identity on the website's search results.[iii] This staged roll-out of information allows for a test of three causal mechanisms. We can assess whether gender affected educators' likelihood of funding before their identity was published (January 2008), when it was published on project pages (January 2008), and when it was published in search results (December 2008).

Stage One: Patterns of Gender Difference

This study rests on simultaneously measuring three separate but related variables representing the three dimensions of gender. I measure individual gender to be educators' binary sex category: "Mr." or "Ms." and "Mrs." Institutional gender is measured as the extent to which an educator's job is more or less female-dominated. Interactional gender is measure by the extent to which educators use male or female-typical language in their essays. The three measures are scaled to be consistent: larger values indicate a male, a more male-typical job role, and more masculine writing.

No study has examined how all three dimensions vary simultaneously. As such, this study is able to describe unmeasured gender differences. Yet, previous studies do provide us with some expectations. We expect gender to be consistent – men should be behave in more masculine ways and work in male-typical roles. And, consistent with studies of tokenization, the degree of segregation should affect the prevalence of gendered behavior and identity.

Hypotheses

In the first stage of my analysis, I investigate the extent to which gender differences among educators on DonorsChoose covary. Because this is the first study to quantitatively define all three dimensions simultaneously, the first hypothesis to test is whether or not the three measures are in fact statistically distinct. By using statistical tests of reliability, we can determine whether these measures capture statistically distinct phenomena.

> ***Hypothesis 1A*: Gender Separability:** *Measures of the three dimensions should be statistically distinguishable.*

Though the measures cannot be too correlated, they should have some relationship. We know men and women segregate and, as with research on tokenization mentioned next, feminine behavior should be prevalent in female-dominated positions. *Gender is consistent*. However, we know little about how the three dimensions correlate as a whole. For example, we do not know whether men in female-dominated occupations are more masculine than men in male-dominated occupations.[iv]

> ***Hypothesis 1B*: Gender Consistency.** *The three measures should be positively correlated with one another.*[v]

Even though we expect gender to be consistent, we have evidence to suggest that it is not linear. Work on tokenization (Kanter 1977) and status emergence (Ridgeway 1990) has shown that segregation causes perceived differences in sex to arise only after reaching a tipping point of between 70 and 85 percent. [vi] People do not begin to associate a particular job as male- or female-typical until men or women make up three quarters of the workforce. More recent work has shown that segregation also causes behavioral expectations to become salient at this tipping point (Turco 2010, Wingfield and Miles 2014). By estimating all three dimensions simultaneously, gender system theory allows us to examine whether segregation corresponds to non-linear changes in behavior or sex or both.

***Hypothesis 1C*: Gender Non-Linearity**. *The relationship between the three dimensions should be non-linear with a tipping point for when women make up about 70 percent or more of the population.*[vii]

Methods

1. Measuring Individual Gender as Sex

Individual gender is the classification of individuals into sex categories: male, female, intersex, and transsexual. On DonorsChoose, educators classify themselves. Despite the fact that educators' names were not published until 2008, they still provided their titles "Mr.," "Mrs.," and "Ms." For this analysis, the 'male' variable is coded as '1' for educators who reported their title as "Mr." and '0' for those who reported "Ms." or "Mrs." Only 1,848 educators (.6 percent) did not provide a gendered title and were excluded from the analysis. Eighty percent of these reported "Dr." as their title and the remaining twenty percent entered no title.

Figure 1 here

Figure 1 shows the representation of male educators on DonorsChoose throughout the period of interest. Male educators account for between eight and sixteen percent of the projects posted on DonorsChoose. The frequency of male and female educators becomes distinctly seasonal from 2007 onward. The dashed line for the raw number of male educators indicates that this is likely due to both an increase in females and decrease in males during the summer months. Thus, I add fixed effects for the month a project was posted to control for seasonality in funding.

2. Measuring Institutional Gender with Occupational Segregation

Institutional gender is the gendered meaning attached to social positions. The classic estimate for this is gender segregation (Reskin, McBrier, and Kmec 1999; Reskin and Bielby 2005; Dinovitzer and Hagan 2014). The more segregated a social position, the more gendered it is taken to be. Here I measure segregation as the proportion of male educators in a given job role.

Specifically, I create a generalized linear model using occupational characteristics like grade level and type of school to predict whether a project was posted by a male educator. I refer to this as a project's "occupational score."

To build this model, I selected all occupational variables which significantly predicted the sex of educators (Table 1). I defined a variable as occupational if educators could not change it at the time they created the project. These included the grade-level and subjects they teach and the kind of school in which they work. But it did not include the type of resource they requested, the cost of the project, or the number of students impacted. After evaluating model fit, the variables I included were grade-level, subject area, whether the school was traditional or alternative (i.e. a charter, magnet, or year-round school), and whether the school was in an urban, suburban, or rural area.

Table 1 here

Figure 2 here

I use a multivariate generalized linear model with the four gendered occupational characteristics to predict the probability that a project is posted by a male educator. This predicted probability is my occupational score. Table 1 reports the results. The most female typical job was teaching language in a traditional, rural, elementary school. The most male-typical job role was teaching social science or sports in alternative, urban, high schools.

Figure 2 is a plot of the relationship between the predicted and observed probabilities with a line of perfect fit. The figure shows how closely the model estimates gender segregation in general, but also shows heteroscedasticity in that the estimates fit less well as the estimated proportion of male educators increases. This provides early evidence for a nonlinear relationship and will be discussed further below. Lastly, the distribution of occupational scores is non-normal. I take the log of the score to produce unbiased measures for stages two and three.[viii]

3. Measuring Interactional Gender with Text

I measure interactional gender by the extent to which educators use male-typical or female-typical language in their essays. Specifically, I measure the masculinity or femininity of an essay as the likelihood that the educator who wrote the essay is male given the words they use. I use standard supervised learning techniques to estimate this likelihood (Ciot, Sonderegger, and Ruths 2013; Evans and Aceves 2016).

Inferring the sex of an author through supervised learning is a common task in natural language processing that typically attains an accuracy of around eighty percent (Ciot, Sonderegger, and Ruths 2013). It generally involves 1) labelling the sex of authors in a data set (having labels is what makes it "supervised"), 2) selecting a set of features (typically words) to use to predict author sex 3) training an algorithm like a naïve Bayes classifier on a small subsample of the data, and 4) applying the trained model to predict the sex of authors in the unanalyzed, test data (Manning, Raghavan, and Schutz 2008: Chapters 13-15).

Given that DonorsChoose provides the sex of all educators, I already have labeled data. The next step is feature selection. Previous work has shown that few words consistently predict author sex across domains and those that do are weakly predictive (Argamon, Koppel, Fine, and Shimoni 2003; Mehl, Gossling, and Pennebaker 2006; Newman, Groom, Handelman, and Pennebaker 2008). Instead, the most distinguishing features tend to be domain-specific words like "children" and "poverty" (Argamon, Koppel, Pennebaker, and Schler 2007; Schwartz et al 2013).

Thus, I chose to use all words in educators' essays, what's called an open vocabulary, regardless of their syntax or phrase ordering, called a "bag-of-words" approach (Schwartz et al 2013).[ix] Using these words, I transform essays into weighted counts of words using term-frequency inverse document frequency (TFIDF). This produces an E x W matrix where each row represents an essay, each column is a word, and the value in each cell is the TFIDF-weighted estimate of the frequency word *j* occurred in essay *i*. [x]

Once the features have been selected, quantified, and vectorized; the next step is to select a model. The most commonly used models in supervised learning are naïve Bayes (NB) and support vector machines (SVM) (Manning, Raghavan, and Schutz 2008). I use NB because its behavior is well understood (Zhang 2008) and it produces estimates for which features are most discriminative (Table 2 below), allowing for a validation that the model is detecting gender differences. I also performed the analysis with SVM which cannot directly validated but tends to be the best performing algorithm. Given the convergent results between NB and SVM, I only discuss the former.

I use cross-validation for the final two steps: fitting the NB model on random subsamples of essays, the training set, and producing estimates for the remaining essays, the so-called test set. The reason for using different samples to train and test a model is to prevent over-fitting which occurs when random differences are taken to be significant. By splitting the data repeatedly, the many random differences wash out.

I assessed the efficiency of the model finding that a training sample size of 2,000 essays minimizes the size of the training sample while maximizing its marginal accuracy (see Figure 3A). Thus, for cross-validation, I train the NB classifier on two thousand randomly selected essays and then predict author sex for the remaining essays.[xi] I repeated this train/test split thirty times and compute the average log likelihood that the author of an essay is male.[xii]

There are two additional challenges that I address. First, in a highly segregated occupation like education, the words that best predict an author's sex are occupational (e.g. "kindergarten," "twelfth grade"). Since I'm measuring occupation using other data, I need to find the words men and women use independent of their occupational role. Second, the rarity of male educators produces an imbalanced sample which biases estimates (Rennie, Shih, Teevan, and Karger 2003; Frank and Bouckaert 2006).

To solve both of these problems, I balance my sample by sex and occupation. I generate a sample of 228,455 essays from the whole sample of over 290,000 essays exactly matched on their occupational variables. I then randomly selected an equal number of males and females from within these occupational groups. The result is a sample of 56,980 projects half written by male educators and half by female educators and balanced on occupational characteristics.

Figure 3 here

Figure 3A shows the accuracy of the classifier for different training sample sizes. As the size of the training sample increases, the accuracy of the classifier increases. The accuracy of the final classification was 64 percent for the NB classifier.[xiii] Figure 3B shows that the accuracy of both classifiers increases as the threshold for making a guess increases. Specifically, the closer to 50:50 the estimate is (zero in the figure), the more inaccurate the classifier. The more extreme the likelihood estimates, the more accurate the classifiers become. This supports the argument that larger estimates correspond with extreme gender differentiation.

Finally, Table 2 lists the top 24 most informative words for males and females from the NB classifier. Without a linguistic theory of gender, interpreting why certain words distinguish male and female texts should be treated with caution. As in other studies, the words the NB classifier reported as most informative are domain-specific; indicating that the differences the classifier detects are substantive, rather than random noise or irrelevant verbiage. Female educators tend to refer to students as “children,” “learners,” or “readers”; use verbs like “help,” “want,” “need,” “feel,” and “love” and ask for “material,” “supplies,” and “books.” Male educators tend to refer to “kids” or “people;” use few distinct verbs, ask for “computers,” “equipment,” “technology,” and “video” and refer to “college,” “course,” and “instruction.”

Table 2 here

Descriptive Results

Hypotheses 1A and 1B act as boundary conditions on how these three measures should relate to one another. According to 1A, the variables cannot be so similar that they measure the same underlying phenomenon. To test this, I calculated the Cronbach’s alpha among the three measures. The standardized alpha was 0.62 which is well below the level for statistical redundancy and supports Hypothesis 1A that the variables capture different phenomena.

Hypothesis 1B sets a lower bound on the relationship among the three variables, stating that they should all be consistent: males should use masculine language and occupy more male-typical roles and the reverse for females. I find that male and female educators do indeed differ in the language they use and their occupational roles. The mean gender language score for men was .26 with a standard deviation of .75 while, for women, the mean was -.26 with a standard deviation of .70 ($t = -85.5$, $p < .001$). This means that men and women’s essays differ but still show significant overlap. In fact, 25 percent of male educators wrote essays using more feminine language than the average female educator.

Individual and institutional gender were also positively correlated. Females had an average occupational score of 11.1 while males had an average score of 21.2 ($t = -139.5$, $p < .001$). Because occupational score is just the predicted number of men in a job, we can interpret this difference as saying that, for the average female, only one in ten colleagues is male while, for the average male, two in ten colleagues are male.

The correlation between language and occupational score was positive and significant ($r = .301$, $t = 75.3$, $p < .001$). Educators in female-typed job roles wrote more feminine essays and vice versa. Important to note here is that this correlation between language and occupation is not an artifact of measurement. As detailed above, the language score is estimated after controlling for segregation and occupation. Thus, even after controlling for the fact that kindergarten teachers are overwhelmingly female, we still find that male and female kindergarten teachers use more feminine language.

Figure 4 here

Finally, to test the gender non-linearity hypothesis (Hypothesis 1C), I examined the curvilinearity of the relationship between the three measures. First, visual inspection indicates a linear relationship for sex and occupational score (Figure 2). However, this relationship is heteroskedastic, becoming less predictive as jobs become more gender-balanced. Figure 4A shows a very slight s-shaped curve in the relationship for sex and language score which is substantiated by linear modelling.[xiv] In the middle of the distribution, the relationship is linear, meaning there are no special subfields where men or women unexpectedly predominate. At the extremes, the relationship becomes near zero, meaning occupational characteristics stop predicting further segregation – a point I return to in the discussion.

Figure 4B shows large curvilinearity between language and occupational scores. As the occupation score for a project decreases, that is, as the expected proportion of men in a job role decreases, not only does the language become more feminine, but the language becomes increasingly more feminine. Explicitly, for projects with occupation scores less than .25, reflecting job roles in which males are predicted to create less than twenty-five percent of the projects, the correlation between language and occupation scores is .28 ($p < .001$). For projects with occupational scores greater than .25, this correlation is only .07 ($p < .001$). This means that there are few gender differences in the essays of middle school and high school teachers. But, there are comparably large gender differences in essays for middle school and elementary school teachers. This shift in behavior differences corresponds to the emergence of salient behavioral expectations in segregated occupations observed by Turco (2010).

However, Figure 4C shows that this curvilinearity differs by sex. Across the entire distribution, male educators' language is more masculine than the language of female educators in the same job role. Male educators' language also gets more masculine as the occupational score increases. However, female educators' language levels off once segregation falls below roughly 75 percent female (an occupational score greater than .25). For males with relatively high occupational scores, the correlation between language and occupation remains significant and positive ($r = .12$, $p < .001$) while, for females, the correlation becomes insignificant and near zero ($r = .013$). [xv] Finally, women in less segregated positions write gender-neutral essays (meaning their language score is roughly zero). Men with lower occupational scores write gender neutral essays.

Stage One Discussion

These results provide support for all three hypotheses. Testing the gender separability hypothesis (1A), I find that each of the three dimensions are measured in a statistically distinguishable way. Second, tests of the gender consistency hypothesis (1B) confirmed that male educators write more masculine essays and were more prevalent in male-typical roles. And, all educators wrote more feminine essays in female-type job roles.

Finally, tests of the non-linearity hypothesis showed that though there were strong linear components to almost all of the relationships, there were several significant nonlinearities that are consistent with research on tokenization. Occupation and language were highly curvilinear. And, consistent with prior work, educators' language became increasingly feminine only when occupational score fell below 25 percent (i.e. in jobs where women make up over 75 percent of the workforce).

However, contrary to existing work, when I separately analyze these changes for males and females (Figure 4C), I find that two different changes in behavior are happening simultaneously. For women, language is gender neutral except when women make up the 75 percent super-majority, at which point their language becomes more feminine. For men, their language also becomes more feminine when women make up a super-majority. Men however do not use gender neutral language in less segregated roles. Instead, men's language continues to become more masculine in less segregated roles. These two unexpected findings suggest that future work on tokenization should separately analyze the effect of segregation on minority and majority members and that behavioral changes will differ based on who is in the majority and minority (Wingfield and Miles 2014).

In addition, the heteroskedasticity in the relationship between sex and occupational score might be explained in two ways. First, it may mean that the forces of segregation are weaker in the less segregated subfields. Men and women still segregate in urban, charter schools; but to a lesser extent than they do in rural, elementary schools. Alternatively, it could indicate a shift in the forces of segregation where the coefficients that predict male- and female-typical roles change size or direction across the distribution. For example, women might prefer traditional schools when teaching elementary students, but prefer alternative schools when teaching high school students. Re-running the model in Table 1 on subsets of the data showed no qualitative change in which variables predict segregation or in what direction. Instead, the same variables simply varied in strength, supporting the first interpretation that the forces of segregation become weaker in less segregated subfields. [xvi]

Stage Two: Estimating Gender Inequality

Having validated my three measures of gender, I move to the second stage of this study which addresses the question of the extent to which these differences correspond to unequal allocations of resources and power. I do this by investigating whether men, people in male-typical job roles, or people using masculine language are more or less likely to receive funding on DonorsChoose.

Existing research on the gender pay gap suggests that women should make about 23 percent less than men. But, this gap should decline or disappear completely once we take into account occupational variables and compare “apples to apples” (O’Neill 2003, Blau and Kahn 2006). The problem with this apples to apples comparison, called the “decompositional” approach, is that it defines gender inequality narrowly as inequality between males and females controlling away broader gender differences (Lips 2013) and ignoring differences in behavior (Goldin 2014; Card, Cardoso, and Kline 2015).

In this second stage, I present a gender system theory approach to estimating gender inequality and compare it to estimates from decomposition-based methods. The results indicate that gender system theory produces a better fitting model and that measures of inequality based on decomposition *grossly underestimate* gender inequality. I also look at the size and direction of the correlations to test the core, but as yet untested assertions of gender system theory. In the third stage, I test whether these patterns of inequality occur as a result of the publication of educator's sex or regardless of it.

Hypotheses: Patterns of Gender Inequality

Using gender system theory to measure inequality on DonorsChoose means estimating the direct effect of all three dimensions and the effect of their interactions on the likelihood of educators receiving funding. As a proof of concept, I test whether the three dimensions add unique information for estimating gender inequality. If gender system theory holds, using all three variables and their interaction terms should produce a model that better fits funding than models just using sex.

> ***Hypothesis 2A*: Best Fitting Model**. *Using all three gender variables and their interactions should produce a better fitting model of the likelihood of funding.*

Second, I expect a universal male advantage in the likelihood of funding. Males, people using masculine language, and people in male-typed jobs should all be more likely to receive funding. Previous research indicates this should remain true even though education is considered a feminine occupation (Williams, 1995; Gorman, 2005; Dinovitzer and Hagan, 2014).

> ***Hypothesis 2B*: Universal Male Advantage.** *Males, people using more masculine language (higher language scores), and people in more male-typical job roles (higher occupational scores) should be more likely to be funded.*

Third, as mentioned earlier, performance theorists assert a gender conformity hypothesis: people are rewarded for behaving in ways consistent with their sex category (West and Zimmermann 2009). This will be tested directly here: men using masculine language and women using feminine language should be more likely to receive funding than men using feminine language or women using masculine language. As an extension, psychologist have found that gender conformity differs by institutional location (Wood and Eagly 2002). Therefore, the three-way interaction between sex, language, and job role should also be significant.

> ***Hypothesis 2C*: Gender Conformity**. *The interaction effect for sex and language should be positively and significantly related to the likelihood of funding.*
>
> ***Hypothesis 2D*: Institutional Gender Conformity**. *The three-way interaction effect for sex, language, and occupation should be positive and significantly related to the likelihood of funding*.

Finally, research using the decompositional approach suggests that occupational variables should explain most of the gender inequality (O'Neill 2003, Blau and Kahn 2006). While I approach this in a fundamentally different way, I still expect the male- and female-typicality of job roles, the institutional dimension, to have the largest effect size.[xvii]

> ***Hypothesis 2E*: Dominance of Institutional Inequality**. *The occupational score should be the largest predictor of funding*.

Modelling Inequality

In this study, I estimate gender inequality as the extent to which my variables of interest predict a binary variable for whether educators were successfully funded.[xviii] The controls in the model were chosen from among the remaining characteristics educators provide to donors as well as monthly fixed effects (see Table 3). Control variables were: 1) the (logged) price of the project request, 2) a three-tiered variable for the poverty level of the school, calculated by DonorsChoose and based on the percentage of students with government subsidized lunch; 3) A five category typology for the resource an educator requests; and 4) the number of students impacted or “reached” Control variables for macroeconomic factors and educators’ social networks are addressed in the third stage of this analysis.

Table 3 here

There are several variables I do not use in the model. First, educators reported whether or not the project was to fund resources for essential or enrichment activities and whether the resources would be used by future students. However, neither of these variables significantly affected a project’s chances of success and did not affect the estimated coefficients of other variables.[xix] I also exclude whether projects were eligible for matching grants and whether they received donations from a giving page, both of which strongly predict funding. These variables are causal mediators for the relationship between gender and funding and should not be estimated as controls. Path analysis using structural equation modelling (not shown) showed the results for both stages two and three remain the same when controlling for these.

Inequality Results

Table 4 here

The first hypothesis to test is the core tenet of gender system theory – that models using all three dimensions of gender to predict inequality have a better fit than those just using sex (Hypothesis 2A). In Table 4, I report the results of a series of nested models starting with only the male/female binary and concluding with all controls and all three dimensions, their interaction terms, and controls. First, the decreasing AIC values indicate that adding these variables improves the fit of the model. A direct comparison of the model fit statistics using

ANOVA revealed that the full model (Model 1D) provided a better fit than the sex only model (1C) and a direct effects only model (not shown) model at a $p < .001$ level. This confirms Hypothesis 2A that including the three dimensions and their interactions produces a better fitting model of funding.

Comparing the models, we can examine how estimates of gender inequality get distorted when we focus only on the male/female binary. Model 1C represents a classic approach to gender inequality: estimating sex inequality with standard controls. Calculating the estimated probability of a male receiving funding, Model 1C shows that male educators are 27 percent more likely to receive funding than female educators.[xx] Thus, if sex was our only measure of gender inequality, we would say that a man's project is 27 percent more likely to be successful on DonorsChoose than a woman's project.

When I estimate the effect of the other gender variables on the likelihood of funding in Model 1D, the coefficient for sex decreases thirty percent from .097 to .069. A coefficient of .069 translates into a 20 percent higher likelihood of funding for male educators. In a decompositional lens, we would interpret this reduction to mean that language use and job role lead to a raw inequality of seven percentage points difference (from 27 percent to 20 percent) in funding for males and females.

However, in a gender system theory framework, we add the coefficients together instead of looking at the change in coefficient for sex. From Model 1D, we still conclude that male educators have a twenty percent higher likelihood of funding than the average educator. But, we also conclude that male educators using masculine language in a male-type job role, have an *84 percent higher likelihood* of funding as compared to the average educator.[xxi]

Thus, in contrast to sex-only models, when we account for all three dimensions, gender inequality triples from 27 percent in Model 1C to 84 percent in Model 1D (a 750 percent increase). And, in contrast to the decompositional approach, the estimated effect of language and occupation and the interaction effects on the likelihood of funding is 64 percent rather than seven percent. In stage three, I show that sex, occupation, and language only significantly affect educators' likelihood of funding after educators' identity is published and therefore this additional 64 percent increase is driven by gender, adding causal evidence that these variables should not be controlled away.

In Hypothesis 2B, I posited that gender inequality should show a universal male advantage: males, people using more masculine language, and people in male-typical job roles should all be more likely to receive funding. Model 1D indicates this is true among the direct effects.[xxii]

Among the interaction effects, monotonicity tended to hold true, but several key effects were insignificant. The one negative coefficient was the interaction between sex and occupation, indicating that men in more feminine positions (a male kindergarten teacher) and women in more masculine positions (a female high school gym teacher) were more likely to receive funding.

Based on the gender conformity hypothesis (Hypothesis 2C), I expected a positive coefficient for the interaction of sex and language score on funding. Women using feminine language and men using masculine language should be more likely to receive funding. However, this effect was insignificant and near zero. In Hypothesis 2D, I expected the three-way interaction to be significant because prior research suggests that gender conformity varies by the social position. This too was insignificant.

Instead, the interaction effect for occupation and language was positive and significant, indicating a higher likelihood of success when behavior fits the job role *irrespective of sex*. For example, elementary school English teachers are more likely to receive funding when they use feminine language, no matter whether they are male or female, than elementary English teachers who use masculine language.

Finally, I assess Hypothesis 2E, that the gender of the job role would be the largest predictor of funding, by comparing the size of the coefficients. In Model 1D, the coefficients for the variables of interest are standardized and thus the coefficients can be directly compared. As predicted, the occupation score was the strongest predictor, almost four times larger than any other coefficient. A one standard deviation increase in the occupational score equates to a 44 percent increase in likelihood of funding. The next largest coefficients were for the other two direct effects and for the interaction between occupation score and language score. These differences ranged in size from a 7 to 9 percent change in likelihood per standard deviation. Lastly, the smallest significant coefficient was for the interaction between male and occupation, which equated to a four percent change in the likelihood of funding.

Discussion of Gender Inequality

The results here demonstrate the importance of taking all three dimensions of gender into account when measuring gender inequality. The best model of inequality was the one that included all three dimensions with their interaction effects. And, when estimating the total amount of gender inequality, the gender system model revealed three times as much gender inequality as compared to a model only using sex and increased the estimate of language and occupation-based inequality from seven percent to 64 percent.

These results demonstrate the importance of theory for interpreting models. Gender system theory argues we should include occupational and linguistic differences while most research omits these and decompositional research controls them away. While the argument that occupation and language should be accounted for has thus far been theoretical, in the next section, I provide empirical evidence for it, showing that inequality along these differences only emerges after teachers' identity is made public.

The structure of gender inequality was largely as expected in size and direction. Occupational gender produced the most inequality. Males, people using masculine language, and people in male-typical roles were all more likely to receive funding. This amounts to direct support for the glass escalator hypothesis (Williams 1992), that males still receive premiums even when working in feminine occupations. The exception to monotonicity, the negative coefficient for sex and occupation, may be evidence of a *glass escalator bonus*. Not only do men still receive a patriarchal dividend in female-dominated roles, but there may be an extra bump for being a man in a female-dominated role.

I tested this enhanced glass escalator hypothesis by splitting the sample into low and high occupational score subgroups and re-estimating Model 1D (analysis not shown). If the hypothesis is true, then the negative coefficient should be largest in the most female-typical roles (i.e. subgroups with the lowest occupational scores). The results indicate that the coefficient exists regardless of whether or not the educators are in jobs with high or low occupational scores. This suggests not only a bonus for men in female-typical roles, but also for females in male-typical roles: a token bonus. It should be noted however that this token bonus is only a marginal bonus. Women in male-typical roles are still less likely to receive funding than men in the same role because men in general are more likely to receive funding.

The other unexpected finding was that the interaction for sex and language was near zero and insignificant. The data failed to support the gender conformity hypothesis. One explanation is that only men (or women) are punished for gender bending (Kane 2006). To test this, I separate the data for males and females and re-estimate the effect of gendered language on the likelihood of funding (not shown). I find that both men and women are rewarded for using masculine language and punished for feminine language all other things being equal.

The evidence suggests that, instead of punishing (rewarding) educators whose behavior deviates from (conforms to) their sex, educators are punished when their behavior deviates from their job role. For example, the results in stage one show that female authors tend to ask for "material" especially "books" for students who they describe as "children" or "readers" who "need," "feel," or would "love" them. Male educators ask for "computers" or "technology" for students they describe as "students" for "instruction" in a "course" for "success" in "college."

The results here indicate that educators who talk about "children who love to read" in a female-typical job like a rural elementary school are more likely to get funded than the same teacher talking about "giving students computers to prepare them for success in college." And the opposite is true for teachers in urban charter high schools (more male-typical locations).

It is important to restate the fact that language and occupation are measured independently. The language score is estimated on a sample of essays balanced on sex and occupation. It is not capturing a hidden difference in grade-level or school type, but the ways in which the presentation of self and one's classroom tends to differ between male and female educators.'

Stage Three: the Emergence of Gender Inequality in Framing

The third and final stage of my analysis investigates what caused gender inequality to occur on DonorsChoose. Currently, the prevailing explanation of gender inequality is that it arises from everyday acts of discrimination which steer men and women into different kinds of work. In this "gender pipeline" model (Alper 1993), wages, status, and prestige are already distributed to different social positions. This inequality becomes gendered because men and women segregate into these positions. In this section, I provide direct evidence that much of this structural inequality is caused by gender framing by showing that there was little or no inequality on any of the three dimensions of gender until educators' identity was published.

The primary model for explaining how gender inequality is produced is the gender pipeline model wherein men and women are pushed and pulled into different domains from high school to college to their first job (Correll 2001, Cech 2013b; Morgan, Gelbgiser, and Weeden 2013). Once in a career, men and women continue to segregate into subspecialties or, for women, out of the career pipeline altogether (Acker 1990; Williams 2000; Wolfinger, Mason, and Gouldin 2009; Cech 2013a). In this view, gender inequality is the result of men and women experiencing different, often discriminatory, pushing and pulling forces at different points in the life course.

The pipeline model focuses on why men and women make the choices they do to explain gender inequality but fails to address the question of why different choices are rewarded differently. For example, the pipeline model explains how women are pushed out of technology jobs and transition to other industries, but fails to explain why technology jobs are better paid than those in other industries. The result is a large collection of mechanisms which explain the variety of reasons men and women tend to segregate into unequal pipelines (Morgan, Gelbgiser, and Weeden 2013: 990). But the model does not explain how women consistently end up in less prestigious, lower paying roles at nearly every juncture across the life course in nearly all fields.

In this stage, I use gender system theory to show how such macro patterns of inequality emerge on DonorsChoose from the micro-level mechanism of gender framing. Before January 2008, no information about educators' identity was published by the website. After January 2008, DonorsChoose began publishing educators' name: "Mr. Spain" and "Ms. Kinsley." The results I show here indicate that publishing these names caused the patterns of inequality described in stage two to emerge. This provides direct support for gender system theory and for the argument that occupational inequality, long believed to be exogenous to gender, is the result of gender discrimination (England et al 2007).

Given my reliance on the timing of changes to the website, I examine two other contemporaneous changes which could account for the emergence of gender inequality. First, I examine the potential role of the financial crisis and show that the crisis had relatively little effect on the market. Second, I investigate whether the nationalization of the website affected the patterns of inequality that emerged. I show that educators using the website before it nationalized were insulated from the discriminatory effects of the publication of their identities by their social networks. However, following the publication of their identities, networks between educators and donors eventually became sources of gender inequality

Hypotheses

Gender system theory asserts that inequality is the result of a discriminatory process whereby sex categorization triggers individuals' to think evaluate situations in gendered ways. We can test this directly on DonorsChoose because of the natural experiment.

> ***Hypothesis 3A*: Sex Triggers Framing.** *If publishing educators' identity triggered donors' gender frames, gender inequality should emerge starting in January 2008.*

A profound corollary of this framing mechanism is that, in the absence of publishing educators' identities, there would be no gender inequality. This is particularly difficult to believe on DonorsChoose because donors should be able to easily infer educators' sex. First, education is one of the most segregated occupations in the United States. Kindergarten teachers everywhere are highly likely to be women (97 percent of teachers below fourth grade are female in my sample). Second, since education is compulsory, everyone in the U.S. should have extensive exposure to this field. In the same way that queen, father, fashion, and beer-belly are commonly understood signals about gender, so too should "kindergarten teacher." This leads to a competing hypothesis.

> ***Hypothesis 3B*: Proxies Trigger Framing.** *If proxy information like an educators' job role or the language they use can triggering gender schema, gender inequality should occur prior to 2008.*[xxiii]

Finally, not only can we test what kinds of information might elicit gender frames, but whether more information elicits stronger reactions (Blair 2002). On DonorsChoose, the fact that educators' identity was published in search results at the end of 2008 means we can test whether increasing the prevalence of identifying information led to more inequality after this change.

> ***Hypothesis 3C*: Trigger Intensity**. *The amount of gender inequality should increase after December 2008.*

The Impact of Publishing Sex

To observe the impact of publishing sex on the likelihood of funding, I used a regression discontinuity analysis, estimating the reduced (Model 1B) and full model (Model 1D) from stage two on annual slices of data from 2007, 2008, and 2009.[xxiv] Table 5 shows the results which depict the emergence of inequality from 2007 to 2009. Hypothesis 3A predicts that discrimination only arises once educators' identity is known in 2008, while hypothesis 3B predicts discrimination could happen earlier. Before sex is published in 2008 (Model 2B), only language and occupation scores were significantly different from zero but only at the $p < .05$ level. In contrast, the model for 2008 (Model 3B) shows the pattern of inequality found in stage two. Each effect found in stage two becomes significant at the $p < .01$ or better level and the coefficients increase in size from 2007 by a factor of between two and ten. The only exception is the direct effect of language which decreased in size and became statistically indistinguishable from zero.

Finally the hypothesis that more triggers cause more discrimination (Hypothesis 3C) can be adjudicated by examining how these estimates change from 2008 (Model 3B) to 2009 (Model 4B). The results in Table 5 indicate the estimates found in 2008 remain in 2009, though the direct effect of the language score quadruples in size and is statistically significant.

Table 5 here

If we compare the estimated probability of funding for a male using masculine language in a male-type role, the cumulative log odds in 2007 are .432, equivalent to a roughly 54 percent higher likelihood of funding than the average educator. In 2008, this educator has a cumulative log odds of .711, a 103 percent higher likelihood of funding. For projects posted in 2009, this inequality increases further with a cumulative log odds of .862, equivalent to 137 percent higher likelihood of funding. [xxv]

The evidence for Hypothesis 3A, that proxies trigger inequality is weak and ambiguous. The weak effects among projects posted in 2007 remain weak when adding data for projects posted in 2006. [xxvi] However, the direct effect for occupational score becomes significant at the $p<.001$ level. This suggests that some occupational inequality existed prior to the publication of sex in 2008. This could indicate two phenomena. First, gender schema might have been weakly triggered prior to 2008. Or, some occupational inequality may be attributable to causes beyond gender. These questions are unanswerable here.

Hypothesis 3B is strongly supported. Not only do the expected variables become significant with the exception of one, but they increase in size substantially from 2007 to 2008. This indicates that publishing teachers identity led to inequality along all three dimensions, not just sex. Finally, Hypothesis 3C was weakly supported. Between 2008 and 2009, some coefficients increase in size as predicted but not uniformly so. Thus, we cannot say with certainty that publishing educators' identity in search results led to an increase in gender inequality.

Alternative 1: The Financial Crisis

One significant moderator for gender inequality could be macro-economic trends. In late 2008, the financial sector of the U.S. economy collapsed. The Dow Jones Industrial Average reached a peak of almost 14,000 in October 2007 and, by October 2008, closed below 10,000 for the first time in six years. The index eventually reached its nadir of 6,500 in March 2009.

These changes correspond roughly to changes in gender discrimination on DonorsChoose and offer an alternative mechanism for explaining the emergence of gender discrimination: increased competition. Thébaud and Sharkey (2016) found that, during the recession, women-led firms were more likely to be denied small business loans as compared to male-led firms. Specifically, the financial crisis may have caused a decline in the amount of donations resulting in economic scarcity which could then have caused inequality to emerge. However, scarcity appears not to have occurred (Figure 5).

Figure 5 here

During the period of interest, the number of donations per project actually increased from two to five, the average amount of money requested declined from $900 to $600, and the likelihood of receiving at least one donation increased from 65 to 85 percent. As Figure 5 shows, both the probability of success and the probability of receiving a large donation remain within a relatively stable, seasonal cycle throughout the period of interest with the exception of reaching a brief minimum during the summer of 2008. In essence, the increase in the number of projects was actually met with more donations spread across more projects. Finally, no change was observed when Models 2-4 from Table 5 were re-run with a control for the value of the Dow Jones Industrial Average at the beginning of the month when a project was posted (not shown).

These results actually provide no evidence regarding whether or not increased competition can affect discrimination. In this case, there was very little increase in scarcity. One possible reason for this is the growth of funding from institutions through matching grants. Another possibility is that both demand and supply adjusted simultaneously. While there is no indication that the overall number of donors or aggregate amount of money requested decreased during the period, they may not have grown as much as they would have otherwise.

Alternate 2: Nationalization and Educators' Social Networks

Another potential confounding effect is the expansion of eligibility to the whole United States beginning in the fall of 2007. Prior to that, educators in only 11 states including New York, South Carolina, and California were eligible to seek funding through DonorsChoose. The nationalization of the website was met with a quick influx of new educators and donors from vastly different parts of the United States. These demographic changes could affect patterns of inequality by bringing new donors with different preferences into the market or by changing the kinds of projects available for funding. In this section I compare the patterns of inequality for educators on the website before nationalization to those coming after. I find that new educators faced more discrimination than existing ones. But when I add controls for educators' social networks, I find inequality re-emerges. I also find that gender framing likely affected the formation of new networks after 2008.

First, I test whether or not gender inequality may have been caused by unseen changes in the population. I flagged educators who posted at least one project before September 2007 as having been pre-existing educators and everyone else was flagged as a new educator.[xxvii]

Table 6 here

Table 6 shows the results of the inequality model for new and old educators before and after sex is published. The results indicate that educators who were already on the website were not discriminated against except according to their occupational score (Models 5A and 6A). But, the patterns of inequality do replicate for educators who were not on the website prior to the Fall of 2007 (Model 5B and 6B). The only exception is the weak degree of statistical significance for gendered language and the strong, small, and negative interaction between male and probability of an educator being male.

These results indicate that publishing educators' identity did not affect educators who were already on the website. One explanation is that these educators had built up networks of donors willing to fund them regardless of changes to the website. Unfortunately, directly measuring the social networks of educators and donors is impossible given a lack of data. Instead, I derive two network indicators.

The first is whether or not a project received a donation from donors in the same zip code as the educator. This is an indicator for whether educators mobilized local, offline networks to generate online donations. DonorsChoose provides little data on donors. Donors are not required to provide personal information when they make donations. And, the information they do provide is relatively minimal and truncated for anonymization. I use zip code because it is commonly provided for donations: 48 percent of the 1.1 million donations have the donor's zip code. However, the zip code is truncated to the first three digits. This means I can only match educators and donors who share in common the first three digits of their zip code. [xxviii]

The second network indicator is whether or not a project received a donation from donors who have already given to the educator. This approximates the extent to which educators develop ties with their donors. For this measure of repeat donations, I counted any donation to an educator from a donor account which had given a donation to that educator. This included repeat donations to the same project. The vast majority of donors (71 percent) are recorded as donating only once. Thus, most donors are not repeat givers. However, some of these are likely to be repeat donors who donated without logging into a donor account.

The proportion of donations from local or repeat donors per project ranged widely from between one third to two thirds. Educators are therefore not succeeding based solely on attracting new funders or mobilizing existing ones. And, the two variables overlapped extensively. Educators with more donations from within their zip code also had more repeat donations (Pearson $r = .40$, $p <.001$) and projects that received a donation from someone within the same zip code were twice as likely to receive repeat donations ($t = 208.8$, $p < .001$). Among projects posted in 2008, educators who had been on the website before nationalization were almost twice as likely as new teachers to receive a repeat donation ($t = -.11.6$, $p < .001$). However, in 2008, educators who had been on DonorsChoose prior to nationalization were not more likely to receive donations from people in the same three-digit zip-code.

Table 7 here

Table 7 shows models for the likelihood of successful funding with these two network measures are added as controls. The results show that gender discrimination re-emerges for pre-nationalization educators. Entering the variables one at a time (not shown) indicated that the suppression effect is due almost solely to receiving donations from repeat donors, not from local donors in the same zip code.[xxix] This means that relationships with previous donors insulated educators from the effects of having their identity published.

Table 8 here

If networks of repeat donors insulated educators from inequality, it stands to reason that, after educators' identity was published, new ties between educators and their repeat donors may reflect the new discriminatory preferences. Specifically, to the extent that first time donations are influenced by gender, the subsequent tie formation process between educators and donors should also be affected by gender.

To test this, I estimated the likelihood that projects would receive a repeat donation in 2007, 2008, and 2009 conditional on the three dimensions of gender (see Table 8). There was very little gender inequality in repeat donations in 2007 and 2008 (Models 9 and 10). However, the patterns of inequality emerge in 2009, all in the same direction as before (Model 11A). This gendered shift in repeat giving was especially strong among new educators (Model 11B).

This emergence of gender inequality in tie formation follows after the emergence of gender inequality in the system as whole (3A). In essence, as new educators came onto the website and the website began publishing their identities, donors began discriminating by gender and the new ties that were then formed began to reflect the same patterns of discrimination.

Discussion: The Causes of Inequality

The analysis in stage three provides direct support for the argument that gender inequality emerged as a result of DonorsChoose having published educator's identity. What is most surprising is that the biggest effect was not discrimination by sex. Rather, it was discrimination by gendered occupational characteristics. That is to say that, once donors began seeing educators' names, not only did it make them slightly more likely to donate to educators named "Mr" but it made them much more likely to donate to educators in male-typical job-roles like a high school history teacher in an urban charter school. Following gender system theory, we can infer that seeing educators' sex triggered donors' gender frames which impacted how donors then interpreted educators' language, occupation, and sex.

More research is needed on the other two mechanisms tested here. That educators' language and job role might trigger gender stereotypes and therefore gender discrimination (Hypothesis 3B) is supported by prior research, but not this study. In this case, the effect may have been too weak or noisy to detect. Similarly, the hypothesis that publishing educators' identity in more places would increase gender discrimination (Hypothesis 3C) received weak support. There was an increase in gender inequality, but it was not large and not consistent across each variable tested.

Finally, testing the two alternative hypotheses revealed that social networks mediated the emergence of inequality. Educators created online relationships with some of their donors. These networks of donors insulated educators from the effects of DonorsChoose publishing their identities. However, the formation of ties with new donors was affected by gender preferences and, eventually, educators identifying as male, using male language, and in more male typical job roles were also more likely to receive repeat donations. In this way, social networks transitioned from suppressing gender inequality to embedding it in the social structure of the community.

Conclusions

The results here provide strong support for the basic claims of gender system theory. Gender differences are constructed along three dimensions: sex category (individual), gender performance or behavior (interactional), and gendered social positions (institutional). Social rewards are then allocated unequally across these three dimensions. And, finally, the data here provides strong evidence that gender inequality only emerges when people know one another's sex.

There are several shortcomings to this study however. First, the generalizability of the patterns of inequality found here is unclear. I found gender to be consistent, each dimension was positively correlated to the others, and some of their joint distributions were curvilinear. However, Acker (2006) and Reskin (2003) argue that patterns of gender difference should differ by organization. Future research should examine the relationship between the three dimensions of the gender system on other websites, organizations, and occupations.

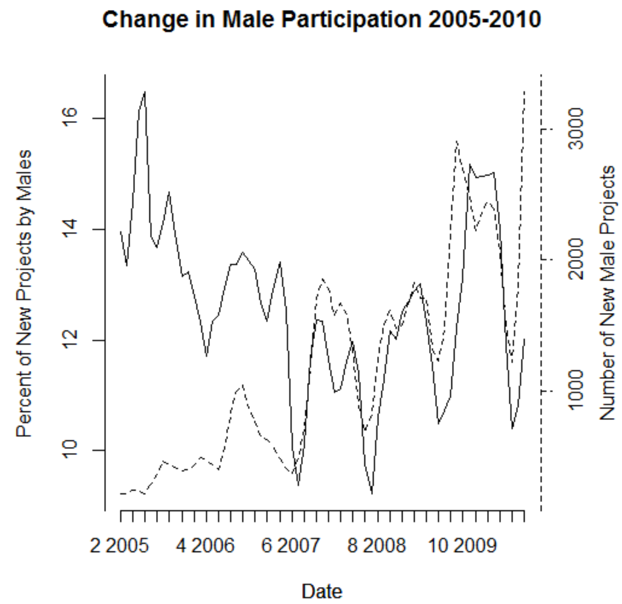
The same is true for the structure of inequality. Is gender inequality always monotonically biased towards males, masculine language, and male-typical roles? Is gender conformity rewarded in other domains, even if it's not on DonorsChoose? Following Reskin and Acker, we again expect that the patterns found here may not necessarily replicate in other contexts. Furthermore, work on intersectionality indicates that these patterns are different for African American males and females (Wingfield and Miles 2014). The ultimate question is whether or not the patterns of gender difference and inequality are stable across contexts and, if not, what explains the differences?

Finally, further research is needed on the mechanisms driving inequality. The results here supported the hypothesis that publishing sex led to discrimination. But better controlled experiments are needed to more cleanly test the other two hypotheses that publishing more information in more places enhances inequality along the three dimensions of gender and that institutional or behavioral cues can trigger the unequal distribution of rewards.

As has been the argument throughout, what would separate these new studies from earlier studies is gender system theory. We need to expand our existing methods of studying gender inequality to include all three dimensions of gender simultaneously. Without accounting for the three dimensions put forward by gender system theory, gender scholarship will continue to suffer from substantial omitted variable biases.

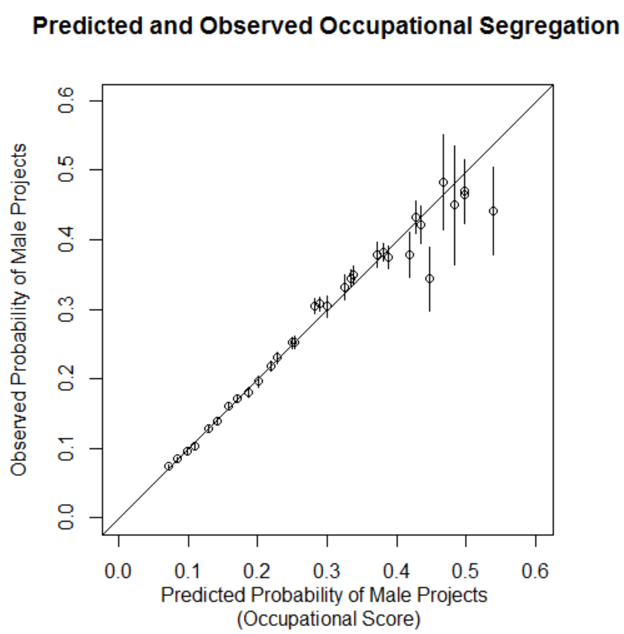
Figures and Images

Figure 1: Change in male participation (per five months)



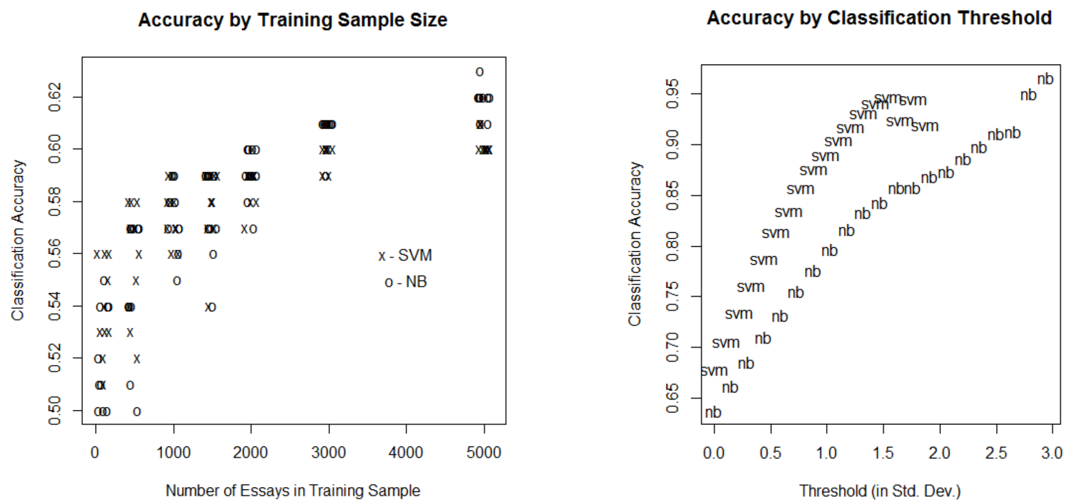
Note: The lines represent the projects posted by male teachers during a five month window. For example, the fourteen percent at February 2005 indicates that males posted fourteen percent of the projects between January and March 2005.

Figure 2: Predicted and Observed Occupational Segregation



Note: Predicted probability of male projects is produced by the model of occupational segregation (Table 1). These values are binned to produce the observed number of males in each bin.

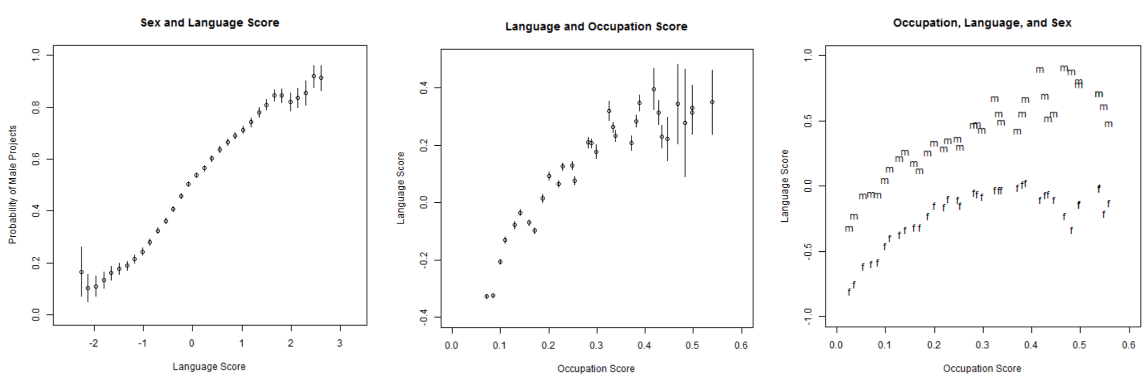
Figure 3: Classification Accuracy by the Number of Training Essays and Classification Threshold



(a): Model Efficiency (b) Classification Threshold

Note: The accuracies reported in (a) are slightly lower than reported in the text as those reported here are not standardized. The classification threshold in (b) corresponds to the point at which we include the classifier's guess and is a measure of tolerance. At zero standard deviations away from the mean, we accept all guesses as the log probability approaches a coin toss. At one standard deviation away, we take only those estimates that are one standard deviation or more away from the mean log probability. Thus, as the classifier becomes more extreme, we expect it to be more accurate.

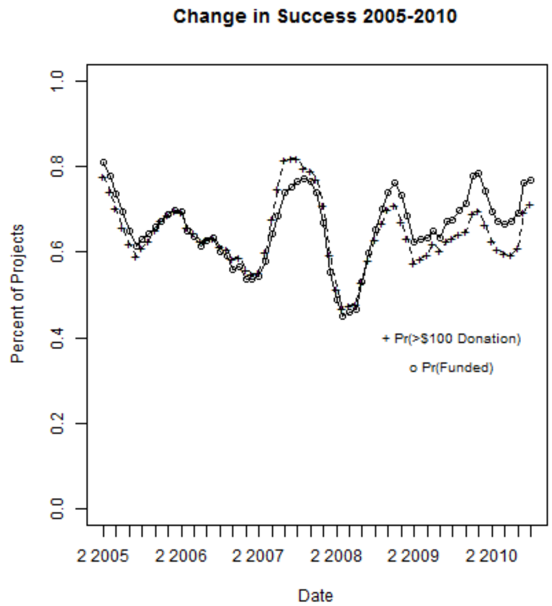
Figure 4: Joint Distributions between Sex, Language, and Occupation



(a) Sex and Language (b) Language and Occupation (c) Sex, Language, and Occupation

Note: The range of the Y-axis in (b) and (c) differs from the X-axis in (a) as the X-axis in (a) is a raw score and the Y-axis values in (b) and (c) are mean language scores.

Figure 5: The Changing Rate of Success and Large Donor Funding 2005-2010



Note: Large donor funding is defined as the probability of receiving more than one donation of over $100. This cut-off of $100 is provided by DonorsChoose as one form of protecting donors' anonymity.

Table 1. Generalized Linear Model Estimates for Occupational Segregation

Variable	Estimate (Standard Error)
Intercept	-0.68*** (0.02)
Grade PreK-2	-2.26*** (0.02)
Grade 4-5	-1.21*** (0.02)
Grade 6-8[a]	-0.43*** (0.02)
Special Ed	-0.53*** (0.03)
Arts	0.20 *** (0.02)
Language	-0.56*** (0.01)
Social Science	0.25*** (0.02)
Sports	0.67*** (0.03)
Applied[b]	-0.17*** (0.02)
Non-traditional	0.16*** (0.02)
Rural	-0.46*** (0.02)
Suburban[c]	-0.21*** (0.02)
N	228,655

Note: Dependent variable is whether or not a project is created by a male or female educator. All categories are as defined by DonorsChoose.org
*** $p < .001$ (two-tailed z-tests)
[a]Null category is Grades 9-12
[b]Null category is STEM
[c]Null category is urban

Table 2: Most Informative Words by Sex

Male Words		Female Words	
kids	college	children	love
people	instruction	learners	want
student	course	readers	need
equipment	lack	supplies	feel
technology	quality	books	help
computers	funding	materials	fun
video	schools	read	families
instruments	based	skills	like
projector	american	learn	order
laptop	ability	paper	activities
physics	program	grade	able
music	material	make	possible

Note: Words are based on extracting the top 200 most informative words per iteration of the NB classifier and using the Shannon Information score to measure how consistently the word indicates male or female authorship across all iterations.

Table 3: Descriptive Statistics of Variables Used to Estimate the Probability of Successful Funding

Variable	Mean	Std	Min	Max	N
Funded	0.70	0.46	0	1	56838
Male	0.50	0.50	0	1	56838
Occupation Score	0.21	0.12	0.01	0.60	56838
Language Score	0.00	0.78	-2.95	3.42	56838
Ask for Books	0.20	0.40	0	1	56838
Ask for Other	0.07	0.25	0	1	56838
Ask for Supplies	0.35	0.48	0	1	56838
Ask for Technology	0.36	0.48	0	1	56838
Ask for Trips	0.02	0.13	0	1	56838
Ask for Visitor	0.00	0.05	0	1	56838
Students Reached	116	170	1	4000	56838
Project Price	718.26	1260.03	10.98	46323.17	56838
School Poverty	1.84	0.41	0	2	56838
Used by Future Students[a]	0.90	0.30	0	1	56838
Word Count[a]	281.12	104.97	53.00	1373.00	56838
LIX Readability[a]	43.17	6.51	19.10	141.00	56838

[a] Variables not reported in final models as discussed in text

Table 4: Probability of Successful Funding by Gender

Variables	Model 1A	Model 1B	Model 1C	Model 1D
(Intercept)	0.859*** (0.009)	0.615*** (0.021)	6.137*** (0.116)	6.714*** (0.122)
Male[S]	0.019* (0.009)	0.062*** (0.010)	0.097*** (0.010)	0.069*** (0.011)
Occupation Score[S,L]		0.231*** (0.018)		0.365*** (0.020)
Language Score[S]		-.201*** (0.019)		0.088*** (0.021)
Male[S] x Occupation Score[S,L]		-0.033*** (0.009)		-0.042*** (0.009)
Male[S] x Language Score[S]		0.001 (.010)		0.002 (0.010)
Occupation Score[S,L] x Language Score[S]		0.068*** (0.016)		0.079*** (0.017)
Male[S] x Occupation Score[S,L] x Language Score[S]		0.017* (0.008)		0.002 (0.009)
Ask for Visitors[a]			-0.104 (0.191)	-0.233 (0.195)
Ask for Trips[a]			0.346*** (0.083)	0.220** (0.084)
Ask for Books[a]			0.206*** (0.030)	0.269*** (0.030)
Ask for Technology[a]			-0.338*** (0.024)	-0.424*** (0.025)
Ask for Other[a]			-0.177*** (0.041)	-0.163*** (0.042)
Students Reached[L]			-0.032*** (0.009)	-0.099*** (0.010)
Project Price[L]			-0.987*** (0.016)	-1.081*** (0.017)
School Poverty			0.590*** (0.023)	0.576*** (0.023)
AIC	69204	68879	61927	61013
N	56838	56838	56838	56838
df	2	27	8	27

*** p < .001, ** p < .01, * p < .05 (two-tailed z-tests)

Note: The dependent variable is the likelihood of funding. Monthly fixed effects are included but not reported for space.

[S] Variable is standardized

[L] Log of the variable is used

[a] Null category is supplies

Table 5: Probability of Successful Funding Before (2007), During (2008), and After (2009) Sex Identification

Variables	2007		2008		2009	
	Model 2A	Model 2B	Model 3A	Model 3B	Model 4A	Model 4B
(Intercept)	0.808*** (0.059)	8.516*** (0.383)	0.233*** (0.045)	9.330*** (0.312)	0.663*** (0.041)	5.336*** (0.239)
Male^{S}	0.023 (0.027)	0.049 (0.031)	0.050* (0.020)	0.080*** (0.024)	0.087*** (0.019)	0.072*** (0.021)
Occupation Score^{S,L}	0.014 (0.049)	0.139* (0.058)	0.302*** (0.038)	0.496*** (0.046)	0.430*** (0.037)	0.579*** (0.040)
Language Score^{S}	-0.281*** (0.054)	0.151* (0.062)	-0.242*** (0.041)	0.037 (0.049)	-0.108** (0.036)	0.139*** (0.039)
Male^{S} x Occupation Score^{S}	0.005 (0.023)	-0.007 (0.027)	-0.049** (0.018)	-0.066** (0.021)	-0.036* (0.018)	-0.047* (0.019)
Male^{S} x Language Score^{S}	0.001 (0.027)	0.003 (0.030)	0.003 (0.020)	-0.009 (0.024)	0.013 (0.019)	0.007 (0.020)
Occupation Score^{S,L} x Language Score^{S}	-0.021 (0.045)	-0.009 (0.050)	0.098** (0.034)	0.138*** (0.040)	0.112*** (0.032)	0.123*** (0.034)
Male^{S} x Occupation Score^{S,L} x Language Score^{S}	0.038 (0.022)	0.018 (0.025)	0.015 (0.017)	0.010 (0.02)	-0.010 (0.017)	-0.024 (0.018)
Ask for Visitors		-0.130 (0.217)		1.486*** (0.341)		1.645*** (0.357)
Ask for Trips		-0.127 (0.518)		0.110 (0.674)		0.285 (0.624)
Ask for Books		0.107 (0.089)		0.270*** (0.067)		0.412*** (0.059)
Ask for Technology		-0.633*** (0.078)		-0.423*** (0.059)		-0.397*** (0.049)
Ask for Other		0.004 (0.130)		-0.334*** (0.096)		-0.189* (0.085)
# of Students Reached^{L}		-0.095** (0.029)		-0.105*** (0.023)		-0.171*** (0.020)
Price of Project^{L}		-1.466*** (0.053)		-1.685*** (0.047)		-0.818*** (0.033)
School Poverty		0.641*** (0.070)		0.770*** (0.055)		0.681*** (0.044)
AIC	8875	7102	14382	11184	17990	16635
N	7352	7352	10955	10955	16035	16035
df	8	27	8	27	8	27

*** $p < .001$, ** $p < .01$, * $p < .05$ (two-tailed z-tests)

Note: The dependent variable is the likelihood of funding. Monthly fixed effects are included but not reported for space.

^{S} Variable is standardized

^{L} Log of the variable is used

Table 6. Probability of Successful Funding Before and After Sex Identification for New and Pre-existing Teachers on DonorsChoose

Variables	2007		2008-2009[a]	
	Same Teachers	New Teachers	Same Teachers	New Teachers
	Model 5A	Model 5B	Model 6A	Model 6B
(Intercept)	8.095*** (0.587)	9.009*** (0.527)	6.195*** (0.516)	6.466*** (0.199)
Male^{S}	0.024 (0.050)	0.062 (0.039)	0.038 (0.043)	0.077*** (0.017)
Occupation Score^{S,L}	0.167 (0.094)	0.121 (0.073)	0.477*** (0.085)	0.565*** (0.032)
Language Score^{S}	0.259* (0.106)	0.103 (0.077)	0.101 (0.091)	0.084** (0.032)
Male^{S} x Occupation Score^{S}	0.016 (0.042)	-0.015 (0.035)	-0.006 (0.038)	-0.065*** (0.015)
Male^{S} x Language Score^{S}	-0.025 (0.049)	0.021 (0.038)	0.037 (0.047)	0.002 (0.016)
Occupation Score^{S,L} x Language Score^{S}	-0.008 (0.082)	-0.009 (0.064)	-0.024 (0.076)	0.164*** (0.027)
Male^{S} x Occupation Score^{S,L} x Language Score^{S}	0.011 (0.039)	0.018 (0.033)	0.047 (0.036)	-0.018 (0.014)
Ask for Visitors	-0.031 (0.732)	-0.146 (0.748)	0.035 (0.564)	0.683 (0.836)
Ask for Trips	0.177 (0.283)	-0.718 (0.375)	1.097*** (0.302)	1.812*** (0.356)
Ask for Books	0.170 (0.152)	0.090 (0.110)	0.407** (0.129)	0.348*** (0.046)
Ask for Technology	-0.596*** (0.138)	-0.635*** (0.095)	-0.536*** (0.116)	-0.382*** (0.039)
Ask for Other	0.015 (0.208)	0.032 (0.168)	0.012 (0.186)	-0.303*** (0.066)
# of Students Reached^{L}	-0.074 (0.050)	-0.109** (0.036)	-0.181*** (0.044)	-0.133*** (0.015)
Price of Project^{L}	-1.353*** (0.079)	-1.568*** (0.074)	-1.050*** (0.069)	-1.111*** (0.029)
School Poverty	0.485*** (0.144)	0.690*** (0.081)	0.540*** (0.138)	0.735*** (0.035)
AIC	2406	4712	3061	25479
N	2449	4922	2851	24175
df	27	27	27	27

*** p < .001, ** p < .01, * p < .05 (two-tailed z-tests)
Note: The dependent variable is the likelihood of funding. Monthly fixed effects are included but not reported for space.
[s] Standardized Variable
[l] Logged Variable
[a] 2008 and 2009 are combined because there are few teachers who persisted from 2007 to 2009.

Table 7: Network Effects on the Probability of Successful Funding for New and Pre-existing Teachers on DonorsChoose

Variables	2007		2008-2009	
	Same Teachers	New Teachers	Same Teachers	New Teachers
	Model 7A	Model 7B	Model 8A	Model 8B
(Intercept)	8.766*** (0.633)	10.018*** (0.572)	7.699*** (0.65)	7.839*** (0.221)
Male^{S}	0.007 (0.054)	0.038 (0.042)	0.174*** (0.052)	0.062*** (0.018)
Occupation Score^{S,L}	0.109 (0.102)	0.081 (0.079)	0.540*** (0.105)	0.531*** (0.035)
Language Score^{S}	0.224 (0.116)	0.138 (0.082)	0.045 (0.107)	0.092** (0.035)
Male^{S} x Occupation Score^{S}	0.017 (0.046)	-0.005 (0.038)	-0.136** (0.046)	-0.045** (0.017)
Male^{S} x Language Score^{S}	0.017 (0.053)	0.027 (0.041)	0.071 (0.055)	-0.007 (0.017)
Occupation Score^{S,L} x Language Score^{S}	0.012 (0.088)	-0.019 (0.069)	0.023 (0.090)	0.146*** (0.030)
Male^{S} x Occupation Score^{S,L} x Language Score^{S}	0.042 (-0.120)	0.011 (0.035)	-0.008 (0.043)	-0.012 (0.015)
# of Donations in Zip Code^{L}	1.367*** (0.166)	1.973*** (0.123)	1.915*** (0.180)	1.711*** (0.061)
# of Repeat Donations^{L}	2.576*** (0.253)	1.845*** (0.217)	3.627*** (0.215)	2.672*** (0.074)
Ask for Visitors	-0.383 (0.802)	-0.357 (0.858)	0.345 (0.698)	-0.356 (0.851)
Ask for Trips	0.18 (0.297)	-0.828* (0.386)	1.298*** (0.353)	1.922*** (0.425)
Ask for Books	0.157 (0.161)	0.106 (0.116)	0.397** (0.148)	0.311*** (0.049)
Ask for Technology	-0.581*** (0.149)	-0.601*** (0.101)	-0.603*** (0.140)	-0.332*** (0.043)
Ask for Other	0.108 (0.217)	0.009 (0.178)	0.060 (0.221)	-0.279*** (0.071)
# of Students Reached^{L}	-0.060 (0.052)	-0.103** (0.038)	-0.115* (0.052)	-0.119*** (0.017)
Price of Project^{L}	-1.551*** (0.086)	-1.803*** (0.081)	-1.580*** (0.091)	-1.421*** (0.033)
School Poverty	0.533*** (0.153)	0.767*** (0.088)	0.782*** (0.167)	0.724*** (0.038)
AIC	2125	4193	2249	21793
N	2449	4903	2851	24139
df	29	29	29	29

*** p < .001, ** p < .01, * p < .05 (two-tailed z-tests)
Note: Dependent variable is the probability the project receives full funding. The zip code and repeat donations measures are not standardized.
^{s} Standardized Variable
^{l} Logged Variable

Table 8: Predicting Repeat Donations Before, During, and After Sex is Published

Variables	2007	2008	2009	
	All Teachers	All Teachers	All Teachers	New Teachers
	Model 9	Model 10	Model 11A	Model 11B
(Intercept)	-2.181*** (0.303)	-3.42*** (0.238)	-3.343*** (0.115)	-3.215*** (0.127)
Male^{S}	0.038 (0.028)	-0.006 (0.021)	0.046*** (0.013)	0.059*** (0.015)
Occupation Score^{S,L}	0.056 (0.054)	0.041 (0.039)	0.175*** (0.025)	0.207*** (0.027)
Language score^{S}	0.011 (0.059)	0.028 (0.043)	0.032 (0.026)	0.001 (0.028)
Male^{S} x Occupation score^{S}	-0.017 (0.025)	-0.024 (0.018)	-0.026* (0.012)	-0.045*** (0.013)
Male^{S} x Language Score^{S}	0.008 (0.028)	-0.018 (0.021)	0.089*** (0.013)	0.110*** (0.014)
Occupation Score^{S,L} x Language Score^{S}	-0.006 (0.048)	0.037 (0.034)	-0.035 (0.022)	-0.016 (0.023)
Male^{S} x Occupation Score^{S,L} x Language Score^{S}	0.013 (0.022)	0.004 (0.016)	-0.019 (0.001)	-0.038*** (0.011)
Ask for Visitors	0.089 (0.452)	0.011 (0.503)	0.892*** (0.133)	0.035 (0.449)
Ask for Trips	0.298 (0.161)	0.568*** (0.169)	0.316*** (0.086)	0.349** (0.114)
Ask for Books	0.098 (0.069)	0.045 (0.055)	-0.018 (0.030)	-0.002 (0.032)
Ask for Technology	-0.285*** (0.071)	-0.125* (0.051)	-0.126*** (0.026)	-0.095*** (0.028)
Ask for Other	-0.207 (0.116)	-0.013 (0.084)	-0.225*** (0.050)	-0.341*** (0.058)
# of Students Reached^{L}	-0.046 (0.025)	-0.032 (0.019)	-0.033*** (0.010)	-0.032** (0.011)
Price of Project^{L}	0.127*** (0.037)	0.259*** (0.032)	0.426*** (0.015)	0.390*** (0.016)
School Poverty	-0.222*** (0.060)	0.042 (0.049)	-0.071** (0.025)	-0.087** (0.026)
AIC	8458	13904	39103	34579
N	7352	10955	16035	14751
df	27	27	27	27

*** p < .001, ** p < .01, * p < .05 (two-tailed z-tests)
Note: The dependent variable is the number of donations from donors who've donated to the teacher previously. The model is a generalized linear model with a Poisson link function.
^{s} Standardized Variable
^{l} Logged Variable

References

Acker, Joan. 1990. “Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations.” *Gender & Society* 4(2):139–58.

Acker, Joan. 2006. “Inequality Regimes: Gender, Class, and Race in Organizations.” *Gender & Society* 20(4):441–64.

Alksnis, Christine, Serge Desmarais, and James Curtis. 2008. “Workforce Segregation and the Gender Wage Gap: Is ‘Women’s’ Work Valued as Highly as ‘Men’s’?” *Journal of Applied Social Psychology* 38(6):1416–41.

Alper, Joe. 1993. “The Pipeline Is Leaking Women All the Way Along.” *Science* 260(5106):409–11.

Argamon, Shlomo, Moshe Koppel, Jonathan Fine, and Anat Rachel Shimoni. 2003. “Gender, Genre, and Writing Style in Formal Written Texts.” *Text* 23(3):321–46.

Argamon, Shlomo, Moshe Koppel, James W. Pennebaker, and Jonathan Schler. 2007. “Mining the Blogosphere: Age, Gender and the Varieties of Self-Expression.” *First Monday* 12(9).

Banaji, Mahzarin R. and Curtis D. Hardin. 1996. “Automatic Stereotyping.” *Psychological Science* 7(3):136–41.

Blair, Irene V. 2002. “The Malleability of Automatic Stereotypes and Prejudice.” *Personality and Social Psychology Review* 6(3):242–61.

Blair, Irene V. and Mahzarin R. Banaji. 1996. “Automatic and Controlled Processes in Stereotype Priming.” *Journal of Personality and Social Psychology* 70(6):1142-1163.

Blau, Francine D. and Lawrence M. Kahn. 2006. “The US Gender Pay Gap in the 1990s: Slowing Convergence.” *Industrial & Labor Relations Review* 60(1):45–66.

Bobbitt-Zeher, Donna. 2011. “Gender Discrimination at Work: Connecting Gender Stereotypes, Institutional Policies, and Gender Composition of Workplace.” *Gender & Society* 25(6):764–86.

Brewer, Marilynn B. and Layton N. Lui. 1989. “The Primacy of Age and Sex in the Structure of Person Categories.” *Social Cognition* 7(3):262–74.

Butler, Judith. 1999. *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

Butler, Judith. 2004. *Undoing Gender*. New York: Routledge.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2015. “Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women.” *National Bureau of Economic Research*. Retrieved September 14, 2015 (http://www.nber.org/papers/w21403).

Cech, Erin A. 2013a. “Ideological Wage Inequalities? The Technical/Social Dualism and the Gender Wage Gap in Engineering.” *Social Forces* 91(4):1147–82.

Cech, Erin A. 2013b. "The Self-Expressive Edge of Occupational Sex Segregation." *American Journal of Sociology* 119(3):747–89.

Ciot, Morgane, Morgan Sonderegger, and Derek Ruths. 2013. "Gender Inference of Twitter Users in Non-English Contexts" Pp. 1136–45 in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*.

Cole, Stephen R. and Miguel A. Hernán. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology* 168(6):656–64.

Connell, R. W. 1995. *Masculinities*. Berkeley: University of California Press.

Conway, Michael, M. Teresa Pizzamiglio, and Lauren Mount. 1996. "Status, Communality, and Agency: Implications for Stereotypes of Gender and Other Groups." *Journal of Personality and Social Psychology* 71(1):25.

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106(6):1691–1730.

Dinovitzer, Ronit and John Hagan. 2014. "Hierarchical Structure and Gender Dissimilarity in American Legal Labor Markets." *Social Forces* 92(3):929–55.

Eagly, Alice H. and Steven J. Karau. 2002. "Role Congruity Theory of Prejudice toward Female Leaders." *Psychological Review* 109(3):573–98.

England, Paula, Paul Allison, and Yuxiao Wu. 2007. "Does Bad Pay Cause Occupations to Feminize, Does Feminization Reduce Pay, and How Can We Tell with Longitudinal Data?" *Social Science Research* 36(3):1237–56.

Evans, James A. and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory" *Annual Review of Sociology*. 42(10).

Frank, Eibe and Remco R. Bouckaert. 2006. "Naive Bayes for Text Classification with Unbalanced Classes." Pp. 503–10 in *Proceedings of the 10th European Conference on Principle and Practice of Knowledge Discovery in Databases*, *PKDD'06*.

Glick, Peter. 1991. "Trait-Based and Sex-Based Discrimination in Occupational Prestige, Occupational Salary, and Hiring." *Sex Roles* 25(5-6):351–78.

Glick, Peter, Korin Wilk, and Michele Perreault. 1995. "Images of Occupations: Components of Gender and Status in Occupational Stereotypes." *Sex Roles* 32(9-10):565–82.

Goldin, Claudia and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of 'blind' Auditions on Female Musicians." *The American Economic Review*. 90(4): 715-741.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104(4):1091–1119.

Goldstein, Dana. 2014. *The Teacher Wars: A History of America's Most Embattled Profession*. New York: Random House.

Gorman, Elizabeth H. 2005. "Gender Stereotypes, Same-Gender Preferences, and Organizational Variation in the Hiring of Women: Evidence from Law Firms." *American Sociological Review* 70(4):702–28.

Ito, Tiffany A. and Geoffrey R. Urland. 2003. "Race and Gender on the Brain: Electrocortical Measures of Attention to the Race and Gender of Multiply Categorizable Individuals." *Journal of Personality and Social Psychology* 85(4):616–26.

Kane, Emily W. 2006. "'No Way My Boys Are Going to Be Like That!': Parents' Responses to Children's Gender Nonconformity." *Gender & Society* 20(2):149–76.

Kanter, Rosabeth Moss. 1977. "Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women." *American Journal of Sociology* 82(5):965–90.

Kulik, C. T. and M. Olekalns. 2012. "Negotiating the Gender Divide: Lessons From the Negotiation and Organizational Behavior Literatures." *Journal of Management* 38(4):1387–1415.

Lenton, Alison P., Martin Bruder, and Constantine Sedikides. 2009. "A Meta-Analysis on the Malleability of Automatic Gender Stereotypes." *Psychology of Women Quarterly* 33(2):183–96.

Light, Jennifer S. 1999. "When Computers Were Women." *Technology and Culture* 40(3):455–83.

Lips, Hilary M. 2013. "The Gender Pay Gap: Challenging the Rationalizations. Perceived Equity, Discrimination, and the Limits of Human Capital Models." *Sex Roles* 68(3-4):169–85.

Manning, Christopher D., Raghavan, Prabhakar, and Schütze, Hinrich. 2008. *Introduction to Information Retrieval*, Oxford: Cambridge University Press.

Mehl, Matthias R., Samuel D. Gosling, and James W. Pennebaker. 2006. "Personality in Its Natural Habitat: Manifestations and Implicit Folk Theories of Personality in Daily Life." *Journal of Personality and Social Psychology* 90(5):862–77.

Morgan, Stephen L., Dafna Gelbgiser, and Kim A. Weeden. 2013. "Feeding the Pipeline: Gender, Occupational Plans, and College Major Selection." *Social Science Research* 42(4):989–1005.

Newman, Matthew L., Carla J. Groom, Lori D. Handelman, and James W. Pennebaker. 2008. "Gender Differences in Language Use: An Analysis of 14,000 Text Samples." *Discourse Processes* 45(3):211–36.

O'Neill, June. 2003. "The Gender Gap in Wages, circa 2000." *The American Economic Review* 93(2):309–14.

Ott, E. Marlies. 1989. "Effects of the Male-Female Ratio at Work: Policewomen and Male Nurses." *Psychology of Women Quarterly* 13(1):41–57.

Pascoe, C. J. 2011. *Dude, You're a Fag: Masculinity and Sexuality in High School*. Berkeley: University of California Press.

Pedregosa, F. et al. 2011. “Scikit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research* 12:2825–30.

Pennebaker, James W. and Laura A. King. 1999. “Linguistic Styles: Language Use as an Individual Difference.” *Journal of personality and social psychology* 77(6):1296.

Rennie, Jason D., Lawrence Shih, Jaime Teevan, and David R. Karger. 2003. “Tackling the Poor Assumptions of Naive Bayes Text Classifiers.” Pp. 616–23 in *International Conference on Machine Learning*, Vol. 3.

Reskin, Barbara F. 2000. “The Proximate Causes of Employment Discrimination.” *Contemporary Sociology* 29(2):319–28.

Reskin, Barbara F. 2003. “Including Mechanisms in Our Models of Ascriptive Inequality.” *American Sociological Review* 68(1):1-21.

Reskin, Barbara F. and Denise D. Bielby. 2005. “A Sociological Perspective on Gender and Career Outcomes.” *Journal of Economic Perspectives* 71–86.

Reskin, Barbara F., Debra B. McBrier, and Julie A. Kmec. 1999. “The Determinants and Consequences of Workplace Sex and Race Composition.” *Annual Review of Sociology* 25:335–61.

Ridgeway, Cecilia L. 1991. “The Social Construction of Status Value: Gender and Other Nominal Characteristics.” *Social Forces* 70(2):367–86.

Ridgeway, Cecilia L. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. New York: Oxford University Press.

Ridgeway, Cecilia L. and Shelley J. Correll. 2004. “Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations.” *Gender & Society* 18(4):510–31.

Risman, Barbara J. 2004. “Gender as a Social Structure: Theory Wrestling with Activism.” *Gender & Society* 18(4):429–50.

Robins, James M., Miguel Angel Hernán, and Babette Brumback. 2000. “Marginal Structural Models and Causal Inference in Epidemiology.” *Epidemiology* 11(5):pp. 550–60.

Roth, Louise Marie. 2004. “The Social Psychology of Tokenism: Status and Homophily Processes on Wall Street.” *Sociological Perspectives* 47(2):189–214.

Schwartz, H. Andrew et al. 2013. “Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach” edited by T. Preis. *PLoS ONE* 8(9):e73791.

Stangor, Charles, Laure Lynch, Changming Duan, and Beth Glas. 1992. “Categorization of Individuals on the Basis of Multiple Social Features.” *Journal of Personality and Social Psychology* 62(2):207–18.

Thebaud, Sarah and Amanda Sharkey. 2016. “Unequal Hard Times: The Influence of the Great Recession on Gender Bias in Entrepreneurial Financing.” *Sociological Science* 3:1–31.

Turco, Catherine J. 2010. “Cultural Foundations of Tokenism: Evidence from the Leveraged Buyout Industry.” *American Sociological Review* 75(6):894–913.

West, Candace and Don H. Zimmerman. 1987. “Doing Gender.” *Gender & Society* 1(2):125–51.

West, Candace and Don H. Zimmerman. 2009. “Accounting for Doing Gender.” *Gender & Society* 23(1):112–22.

Williams, Christine L. 1992. “The Glass Escalator: Hidden Advantages for Men in the ‘Female’ Professions.” *Social Problems* 39(3):pp. 253–67.

Williams, Joan. 1999. *Unbending Gender: Why Family and Work Conflict and What to Do About It*. Oxford: Oxford University Press.

Wingfield, Adia Harvey and Ranell L. Myles. 2014. “Still a Man’s World? Revisiting Men Who Do Women’s Work: Still a Man’s World?” *Sociology Compass* 8(10):1206–15.

Wolfinger, Nicholas H., Mary Ann Mason, and Marc Goulden. 2009. “Stay in the Game: Gender, Family Formation and Alternative Trajectories in the Academic Life Course.” *Social Forces* 87(3):1591–1621.

Wood, Wendy and Alice H. Eagly. 2002. “A Cross-Cultural Analysis of the Behavior of Women and Men: Implications for the Origins of Sex Differences.” *Psychological Bulletin* 128(5):699–727.

Yoder, Janice D. 1991. “Rethinking Tokenism: Looking beyond Numbers.” *Gender & Society* 5(2):178–92.

Zhang, Harry. 2004. “The Optimality of Naive Bayes.” *AA* 1(2):3.

---

[i] The data for this study come from a publically accessible repository of data provided by DonorsChoose. The data was originally provided in 2011 for a data analysis competition by the website. Since this time, DonorsChoose has made this data publically available on an ongoing basis via its website: http://data.donorschoose.org/.

[ii] The policy for what happens to the money donated to unsuccessful projects has changed over time. Recently, the policy has changed to allow donors to send their donation to educators in the form of gift cards.

[iii] These dates and changes are derived from the data provided by DonorsChoose and snapshots of the website's content and layout collected by the Internet Archive.

[iv] In some cases, Williams (1995) notes that men in female-dominated occupations tend not to talk about their work, indicating they may use the same masculine language as men in other jobs. On the other hand, she also found that men in female-dominated occupations sometimes try to compensate for gender stigma, indicating men might actually use more masculine language in female-typical roles compared to men in more male-typical roles.

[v] Based on the way I measure gender, women should use more feminine words by definition and women should be in more feminine-typed job roles. However, I use a sex- and occupation-balanced sample when estimating gendered language. A positive association between language and sex and between language and occupation is not guaranteed.

[vi] Kanter defined this threshold at 15 percent, though later research added an intermediate stage of tokenization between 15 and 35 percent (Ott 1989). See Yoder (1991) for a summary.

[vii] By my operationalization, there have to be more women in female-typical job roles. However, this is not necessarily true across all job roles. For example, men tend to teach social science courses and courses in higher grade levels. However, there are proportionately fewer men teaching courses that are both high school and social science than would be expected. Similarly, by definition women should use more feminine language. But if an essay is estimated to be 95% female, that does not mean there is a 95 percent probability that the author of that essay is female. This divergence is what defines non-linearity.

[viii] Because of the degree of segregation within education, the occupational score is skewed. To normalize the distribution, I tested two corrections, first, taking a logarithm of the estimated probability and, second, inverting the probability to create inverse probability weights (Robins, Hernn, and Brumback 2000, Cole and Hernn 2008). Both methods eliminated heteroskedasticity and yielded the same results in the analysis. For clarity and consistency, I report the results based on taking the logarithm because it maintains the same sign as the other variables of interest – larger values are masculine and smaller values feminine.

[ix] The exception to this is that I exclude the most common words such as "the" "that" and "for" (what are called stopwords). While prior research has shown that men and women use some of these stopwords differently (Mehl, Gossling, and Pennebaker 2006; Schwartz et al 2013), there is no theoretical explanation for this difference. I ran models including and excluded stopwords and found no difference in model accuracy or in tests of gender inequality discussed in sections two and three. I also ran models using the most frequent 5,000 words and found no difference. Finally, I added in features derived from LIWC which are the basis for the analysis in Newman et al. (2008). These did not substantially alter the individual predictions nor the overall behavior of the classifier.

[x] Various specifications for TFIDF weighting were assessed including different vocabulary sizes, normalization, and smoothing but these did not alter the model. Thus, I do not discuss these details here. All text processing and supervised learning were done with the scikit-learn package in python (Pedregosa et al 2011).

[xi] This type of cross validation is called a hold-p-out cross validation (p is 54,980 in this case) and I use it because of the large amount of text that must be processed and the little amount of data needed to maximize the model's accuracy. There are several other systematic ways to split a corpus into training and testing samples. The most common is K-fold cross validation in which a sample is broken up into k subsamples and all texts except for the kth-sample are used for training and the kth for testing. I performed a 10-fold cross validation and the results presented in stage two and three remain unchanged.

[xii] The final estimate is the average, standardized log-likelihood that an educator is male. Specifically, the NB classifier gives each essay two predictions, one that it is written by a male and one that it is written by a female. Since these are my only two categories, these probabilities sum to one. I combine these two probabilities into a single likelihood measure by dividing the probability an essay is written by a male author by the probability it is written by a female author and then computing the log of this (making it a log likelihood). I then standardize all of the log likelihoods in the iteration by taking the z-score so that each iteration contributes the same amount of variance to the final estimate.

[xiii] The accuracy of the SVM was 68 percent. This 60-65 percent accuracy is substantially lower than the 80 percent accuracy I find when not controlling for occupational characteristics. This is a testament to how important institutional gender may be to current approaches to gender inference in text classification.

[xiv] The curvilinearity is so small I considered excluding its mention here. However, the results using data from the SVM classifier revealed a similar and stronger curvilinearity. So, I include it here.

[xv] Model not shown

[xvi] These models are not shown here but are provided in the replication material. As an example, one comparison involved re-running the model for rural and urban schools separately and including all of the same predictor variables except the dummy variables for rural, suburban, and urban.

[xvii] Note that, according to gender system theory, all of these hypotheses depend on donors' gender frames being triggered according to gender system theory. If frames are not triggered, these inequalities would not occur. I bracket this question now and return to it in the third stage where I test whether these hypotheses hold before sex is published when educators' identities are made public.

[xviii] There are two alternative approaches I could have chosen. First, I could have used a hierarchical model common in educational research. I decided against this because the nested social structure of observations within teachers within schools were the exception rather than the rule. Most educators only posted one project and among educators posting multiple projects, many change grades, subject areas, and schools. A hierarchical model would limit the number of analyzable cases and shrinks the scope of the analysis to some of the most atypical behavior on the website. Second, I measure inequality by successful funding, but there are other measures I could have used such as the time to first donation or the total amount of money donated. However, the former is subject to unobservable

features like the DonorsChoose search algorithm. The latter is censured and educators receive nothing unless their project is fully funded.
[xix] These variables are no longer published on DonorsChoose.
[xx] Note, the coefficients are log odds ratios because the model is a logistic regression. And, since the male variable is z-score standardized, it is no longer a {0,1}. When standardized, males have a value of 2.65 and females have a value of -0.377. Thus, the odds for male funding success is calculated as exp(2.65*.097), which equals an odds ratio of 1.27 and for women it is exp(-0.377*.097) which equals .724.
[xxi] Using the standardized values for male (2.65), language score (1.0), and occupation score (1,0), the odds calculation is: exp((.069*2.65) + (.365*1) + (.088*1) + (-.042*2.65) + (.002*2.65) + (.079*1) + (.002*2.65)) = exp(0.61) = 1.84
[xxii] Also notable is that the coefficient for language score changes sign once control variables are added. This is due to the correlation between language and the type of resources educators ask for. Once the type of resource is controlled for in the model, the language score becomes positive.
[xxiii] It is important to note that if Hypothesis 3B is supported, then all data from all periods would show patterns of gender inequality. Such a result could be due to framing caused by proxy triggers or by a hidden variable unaffected by whether or not educators' identity is published. Thus, support for Hypothesis 3B would mean that the basic causal argument for gender system theory would be untestable by this natural experiment. However, Hypothesis 3B is not supported by the evidence while Hypothesis 3A is.
[xxiv] The second estimation issue was the window during which projects were available for funding. As a policy, projects were given eight months to attract funding in the fall of 2007 while, in the fall of 2008, they were given only five months. In addition, projects that received funding before their expiration were closed immediately. Because of this, it is not feasible to include projects based on when they were live due to censoring and changing exposure periods. Instead, I use the month and year a project *was posted* to segregate data. Thus, there is some contamination between samples.
[xxv] I calculated the cumulative log odds in the same way as footnote xxi. As an example, assuming standardized values of one for male, language score, and occupation score, in 2007 (Model 2B) the calculation is (.049*2.65) + (.139*1) + (.151*1) + (-.007*2.65) + (-.003*2.65) + (-.009*1) + (.018*2.65) = 0.432.
[xxvi] If you include all the data prior to 2007, that is from 2003 onward, the coefficients remain the same, but the significance of the occupational score is p<.001. The reason I do not include data from before 2007 in the main text is that the further from 2008 the data, the less comparable the data are.
[xxvii] Results indicated that there was no difference between educators in eligible and ineligible states who did not post before nationalization. In the following, I combine them into the same category of new educators. Comparing new and existing educators introduces a range of unobserved differences. The most important are selection effects for whether or not an educator posts more than one project. Those who post multiple projects over multiple years should not be assumed to be the same as those who post only once. In addition, if educators are more likely to succeed in their second project than their first, that is, if they get better at winning funding over time, existing educators should have an additional advantage. While I have tested the differences I can observe, such as multiple posting, unobserved differences limit the inferences that can be drawn from this analysis.
[xxviii] I also created a variable for whether or not educators received a donation from within their state because donors provide their state in 71 percent of donations. The same results hold.
[xxix] I also controlled for the number of repeat donations conditional on the number of total donations a project received. This revealed the suppression effect was not an artifact of indirectly controlling for the number of donations.