

How Gender Determines the Way We Speak About Professionals

Additional details on Materials and Methods.

Study 1. In addition to the review-level variables described in the main paper, for each of 1,674 professors, we recorded the following variables from Rate My Professors: average ratings of helpfulness, clarity, easiness, and interest (all on a 1-5 scale), and hotness (illustrated on Rate My Professors using a picture of a hot pepper; 0 = no pepper to 3 = a red, flaming pepper). Gender was inferred from the professor's name and from the pronouns used in the reviews or on the university's website.

Since these data were collected, Rate My Professors has changed the information included in a review and presented on a professor's page. Some of the variables we have collected are no longer available on the website.

Of the comments that contained a reference to the professor, a small percentage of comments contained more than one reference (1.58%), of which most (69.0%) were the same reference type as the first reference. Thus, in the case of multiple references appearing in the same comment, only the first reference was included in the analyses.

Additional data: Our initial plan was to collect data for all reviews in all departments at Cornell, Columbia, and Brown Universities. We began this process by going through the departments in alphabetical order, starting with accounting. We coded some data in this way before we decided to include a broader range of universities and focus instead on a smaller number of departments (biology, psychology, computer science, history, and economics). Including these initial data leaves all the results reported in this paper essentially unchanged.

Study 2. For programs that broadcast both political and non-political content, only political segments were coded. Clearly scripted speech was not coded. For example, political reports were not coded because their wording was presumably written in advance in accordance with journalism reference conventions (full name is used for the first reference, and surname for all subsequent references, 1). Instead, we focused on interview segments, which are less scripted. *The Rush Limbaugh Show* is unscripted throughout and was thus coded in its entirety.

Study 3. The list of figures in American politics: *Susan Rice, Loretta Lynch, Hillary Clinton, Madeleine Albright, Nancy Pelosi, Sonia Sotomayor, Ruth Bader Ginsburg, Elizabeth Warren, Wendy Davis, Ann Kirkpatrick, Barbara Boxer, Dianne Feinstein, Janet Yellen, Sarah Palin, Joni Ernst, Carly Fiorina, Nikki Haley, Michele Bachmann, Sandra Day O'Connor, Susan Collins, Susana Martinez, Mia Love, Condoleezza Rice, Ashton Carter, Thomas Perez, Mitt Romney, Colin Powell, John Boehner, John Roberts, Antonin Scalia, Ted Cruz, Chris Murphy, Adam Schiff, Harry Reid, Dick (Richard) Durbin, Ben Bernanke, Joe Biden, Ed Markey, Bernie Sanders, Josh Earnest, Paul Ryan, Clarence Thomas, Bill Nelson, Mike Pence, John Edwards, and John Kerry.*

The list of well-known individuals in various other domains: *Maya Angelou, Toni Morrison, Jane Austen, Susan B. Anthony, Danielle Steel, Stephenie Meyer, Florence Nightingale, Emily Dickinson, Sally Ride, Louisa May Alcott, Marie Curie, Helen Keller, Serena Williams, Lindsey Vonn, Virginia Woolf, Amelia Earhart, Margaret Mead, Diane Sawyer,*

Dorothy Day, Margaret Thatcher, Jane Goodall, Angela Merkel, Barbara Walters, Robert Frost, Stephen King, Charles Dickens, Martin Luther King Jr., John Grisham, Neil Gaiman, Francis Galton, Mark Twain, Neil Armstrong, William Faulkner, Robert Oppenheimer, John Steinbeck, David Beckham, Ryan Lochte, Ernest Hemingway, Howard Hughes, Charles Darwin, David Muir, Thomas Merton, Winston Churchill, Carl Sagan, Vladimir Putin, and Peter Jennings).

The list of women was generated first, by brainstorming well-known female figures. The list of men is made up of well-known male figures that are roughly matched, on average across the lists, on age, years active, and position or profession.

Studies 5a-b. Participants in this study were asked to imagine themselves as an employer at a funding agency:

Imagine that you work for a funding agency. Your job is to evaluate research proposals by academics who wish to have their projects funded.

Submission is open to researchers at all levels; the funding agency receives proposals from academics who are well-established in their field, as well as from researchers with little experience, who are still learning. Thus, the research proposals vary in their quality and in how much they deserve funding.

They then read four 1-paragraph summaries of research proposals. The following is an example proposal:

Hastings is interested in exploring the causes and consequences of biodiversity. Neighboring populations of animals may be similar, and yet differ in important ways. Several hypotheses have been suggested regarding the potential benefits of biodiversity. Hastings plans to use neighboring salmon populations to shed light on how such diversity emerges, and why it is maintained.

In Study 5b, Prior to answering questions, participants read: “You may feel like you do not know enough to make these judgments, but we would like you to base your answers on what you have read and give us your best guess about the questions.”

Supplementary results.

Study 1. Reviews of female professors (vs. those of male professors) were marginally less likely to include an open-ended comment, $M_s = .94$ ($SE = .003$) vs. $.95$ ($SE = .005$), $\chi^2 = 3.09$, $p = .079$, and were less likely to contain a reference, $M_s = .24$ ($SE = .009$) vs. $.27$ ($SE = .006$), $\chi^2 = 6.79$, $p < .01$.

In addition to the effects of target gender on references described in the main text, female professors were marginally more likely to be referred to as Dr. surname ($M = .27$, $SE = .021$) than were men ($M = .23$, $SE = .013$), $\chi^2 = 3.75$, $p = .053$. None of the other types of reference differed significantly by professor gender except as described in the main text, $p > .19$.

Female professors were rated as less helpful, $M_s = 3.61$ ($SE = .04$) vs. 3.76 ($SE = .03$), $t(1548) = 2.85$, $p < .005$, and clear, $M_s = 3.56$ ($SE = .04$) vs. 3.72 ($SE = .03$), $t(1548) = 2.85$, $p < .005$, than male professors. However, helpfulness and clarity, as well as interest in the course, negatively related to use of surname, $\chi^2_{\text{helpfulness}} = 70.39$, $p < .001$, $\chi^2_{\text{clarity}} = 48.26$, $p < .001$, $\chi^2_{\text{interest}} = 31.39$, $p < .001$. Thus, target gender appears to have independent effects on use of surname and on ratings of helpfulness and clarity. Ratings of how easy the course was were not significantly related to gender and to use of surname, $p = .79$.

Study 2. Because the first reference to a given target within a segment may be more likely to include a full name as a way to introduce the target, we also analyzed the data without these first references; as predicted, speakers referred to men more than to women by surname $M_s = 38.25\%$ vs. 17.66% , $z = -4.55$, $p < .001$, $OR = .15$ [0.06, 0.33].

The effect of target gender on surname reference was qualified by an interaction with speaker gender, $z = 3.82$, $p < .001$; only male speakers were significantly more likely to refer to men than women by surname, $M_s = 32.76\%$ vs. 12.65% , $z = -4.53$, $p < .001$, $OR = .17$ [0.08, 0.36], whereas female speakers were not, $M_s = 18.26\%$ vs. 15.62% , $p = .11$. It is worth noting, however, that there were many more observations for male speakers than female speakers ($N_s = 7,849$ vs. $1,723$), making the analyses with only male speakers better powered; it would be important to analyze a dataset with a greater number of female speakers to determine whether they also show the gender effect. The effect of gender on surname reference did not significantly interact with target or speaker political affiliation, $ps > .07$. Pundits were also more likely to refer to women (29.51%) than to men (11.30%) by first name, $z = 4.35$, $p < .001$, $OR = 10.65$ [3.67, 30.95], and were more likely to refer to men (37.99%) than to women (17.78%) by their title along with their surname (e.g., Senator Schumer), $z = -4.72$, $p < .001$, $OR = .26$ [0.15, 0.46]. Pundits did not differ significantly in their likelihood of referring to men and women by full name or by title along with full name, $ps > .08$.

Study 3. The gender bias in surname use emerged for both political, $\chi^2 = 32.13$, $p < .001$, $M_{men} = .38$ (.02) vs. $M_{women} = .31$ (.02), and non-political figures, $\chi^2 = 213.55$, $p < .001$, $M_{men} = .45$ (.02) vs. $M_{women} = .18$ (.02), but was larger for non-political figures, $\chi_{interaction}^2 = 120.93$, $p < .001$, perhaps because the latter were better known overall. The results were slightly different when we used a mixed linear model (see Methods section for details on why we used two different methods of analysis); the bias was only statistically significant for non-political figures, $F(1, 43.96) = 18.73$, $p < .001$, and not statistically significant for political figures, $p = .37$.

The effect of gender on surname was qualified by an interaction with party affiliation of target, $\chi^2 = 22.68$, $p < .001$: participants were significantly more likely to use surname when referring to Republican men vs. women, $\chi^2 = 49.88$, $p < .001$, but not when referring to Democratic politicians, $p = .53$. As reported in the main text, participants were more likely to report using full name when referring to women than to men, $\chi^2 = 208.84$, $p < .001$. They were also more likely to use first name only when referring to women than to men, $\chi^2 = 10.69$, $p < .005$. However, this latter result was not significant using a mixed linear model, $p = .24$, nor when Hillary Clinton was excluded from the analysis, $p = .80$, suggesting this result may have been driven by a single politician rather than a broader trend. As reported in the main text, targets whose first names were perceived to be more common were more likely to be referred to by surname. However, this result was not significant when using a mixed linear model, $p = .55$.

Study 4. Participants most commonly referred to the scientist in their essays by full name (60.27% of references), followed by surname only (17.47%) and first name only (16.78%). A small percentage of references referred to the scientist by professional title (Dr. Surname/Full Name; 3.43%) or common title (Mr./Mrs. Surname/Full Name; 2.05%).

A logistic regression was performed to test the effect of researcher gender on the likelihood that participants used a surname reference in their essay. The model explained 12.5% (Nagelkerke R^2) of the variance in use of surname in the essay and correctly classified 85.2% of

cases. Similar logistic regressions were run for full name and first name references. Researcher gender did not significantly predict use of these references, $ps > .88$.

Four participants in the female condition referred to the scientist by male pronouns. Excluding these participants leaves the results essentially unchanged.

Studies 5a and 5b. The effect of surname on judgments of fame was qualified by a marginal interaction with participant gender, $\chi^2 = 3.68, p = .055$; follow up analyses revealed that female participants were more likely to select the researcher referred to by surname (vs. full name) as better known, ($M = .65 (SE = .04)$ vs. $M = .41 (SE = .04)$), $\chi^2 = 17.77, p < .001$; male participants were only marginally more likely to do so, ($M = .55 (SE = .04)$ vs. $M = .46 (SE = .04)$), $\chi^2 = 3.36, p = .067$. There was no evidence that the effect of surname on judgments of being distinguished or eminent differed between male and female participants; the effect on judgments of being distinguished or eminent did not significantly interact with participant gender, $ps = .52$ and $.25$, respectively.

Study 6. We found no evidence that the effect of reference type on judgments of fame differed between male and female participants, $p = .19$. To determine whether the difference between the conditions was the result of demand effects, we included a hypothesis probe at the end of the study, asking participants what they thought “this study was about or what it was trying to test”. Five hundred and thirteen participants responded to the hypothesis probe; of those, only 3.9% guessed that the study had anything to do with names, and none mentioned a difference in reference type. Given that none of the participants who responded guessed our hypothesis or identified the independent variable, it does not appear that the difference between the conditions is the result of demand effects.

We also ran two pre-registered replications of this study (preregistrations details available at <http://aspredicted.org/blind.php?x=ue2d2m> and <http://aspredicted.org/blind.php?x=ts45du>). These studies were identical to the original study, except that we modified our attention check in the second replication study to make it slightly more difficult (almost all participants passed the attention check in the previous two studies, so we reasoned that it may have been too easy) and increased our sample size by 15% to make up for the potential increase in excluded participants. We found that in both studies, participants rated researchers referred to by surname as better known than researchers referred to by full name, and this difference was statistically significant in one of the studies, $Ms = 4.85 (SD = 1.71)$ vs. $4.53 (SD = 1.71)$, $t(1078) = .001$ (one-tailed), and was directional but did not reach statistical significance in the other study, $Ms = 4.89 (SD = 1.67)$ vs. $4.72 (SD = 1.75)$, $t(932) = .075$ (one-tailed).

Study 7. We found no evidence that the effect of reference type on judgments of fame, status, or likelihood of winning a prize differed between male and female participants, $ps > .69$.

Study 8. We found no evidence that the effect of reference type on eminence, choice of the NSF winner, or the allocation of the NSF cash award differed between male and female participants, $ps > .53$

Additional studies.

We conducted several additional studies in which our main goal was to test different questions than the ones addressed in the paper; however, the design sometimes also allowed us to re-test the questions that were the focus of this paper. We report these results here. We ran five studies using a similar design to Studies 5-8 (participants read research proposals about scientists who were referred to by full name or surname). The results were all consistent with our reported results and conclusions: in all five studies, participants judged the researcher referred to by surname to be better known than the researcher referred to by full name, and this difference was statistically significant (in four studies) or directional (in one study). One additional study was similar but employed a between-subjects design; its sample size was small compared to our other studies ($N = 209$), and though the result did not reach statistical significance ($p = .21$, two-tailed), it was in the predicted direction, i.e., participants rated researchers referred to by surname as better known than researchers referred to by full name, $M_s = 4.73$ ($SD = 1.61$) vs. 4.44 ($SD = 1.72$). We used the data from this latter study to compute the sample size for Study 6 and its replications. Consistent with the results of Study 7, in another study we found that reference type did not significantly affect judgments of deservingness of a career award when fame was not first brought to mind, which motivated us to run Study 7. We are also in the process of designing a paradigm for a follow-up project on the potential behavioral consequences of the gender reference bias in a new domain. To that end, we have conducted three exploratory studies in a different domain; in these studies (two using a within-subjects and one a between-subjects design), participants read reviews of popular science books by authors referred to by surname or full name. We found no statistically significant differences between conditions in how well known participants judged the authors to be. Note, however, that each study had some methodological issue (in one study the full name appeared once in the surname condition, in another there was only one reference, and it did not appear until halfway through the text). There were also considerable differences between these studies and the studies reported in this paper; for example, they were in a different domain where fame may play a different role, and the descriptions were longer and may not have engaged the participants. We are still improving the design of this new paradigm. These studies were not run with the purpose of being included in this paper, but we are reporting them here for the sake of transparency.

References

1. Associated Press (2016) *The 2016 Associated Press Stylebook and Briefing on Media Law* (Basic Books, New York, NY).