

Gender Discrimination against Pre-PhD Students: Evidence from a Field Experiment *

SEBASTIAN BERGER

PHILIPP DOERRENBERG

CHRISTOPH FELDHAUS

RAFAEL GERBER

ANNIKA M. WYSS

November 10, 2023

Abstract

We conduct a field experiment to study gender-based everyday discrimination against pre-PhD level students in the economics profession. Specifically, we investigate if the willingness among professional economists to support a research project depends on the gender of the support requesting student. We invite the members of the professional association of German-speaking economists to participate in a survey and randomly vary (without deception) the gender of the support requester. Our main results are consistent with same-sex discrimination, with pronounced discrimination of female researchers against the female requester. In addition, we examine heterogeneity across tenured and non-tenured economists and observe pronounced same-sex discrimination among tenured researchers. We discuss explanatory mechanisms and conduct a follow-up survey experiment to address concerns relating to confounding factors.

JEL Classification: A11, I23, J16, J44

Keywords: (Everyday) Discrimination, Gender, Academic Research, Economics, Field Experiment

***Berger:** University of Bern. Email: sebastian.berger@iop.unibe.ch. **Doerrenberg:** University of Mannheim. Email: doerrenberg@uni-mannheim.de. **Feldhaus:** University of Bochum. Email: Christoph.Feldhaus@ruhr-uni-bochum.de. **Gerber:** University of Bern. Email: rafael.andreas.gerber@gmail.com. **Wyss:** University of Bern. Email: annika.wyss@psy.unibe.ch. We thank Shelly Lundberg, Andreas Peichl, Jan Schmitz, Sebastian Siegloch and Johannes Voget as well as various seminar/conference participants for helpful comments and suggestions. We are grateful to Alina Pfrang, Adrian Ruchti and Christin Schmidt for excellent research assistance. The present study received ethical approval by the *German Association for Experimental Economic Research (GfeW)* under approval number *uzYAuTy5*. The field experiment is pre-registered on the OSF platform under the following DOI: [10.17605/OSF.IO/GUZKE](https://doi.org/10.17605/OSF.IO/GUZKE). The follow-up survey is pre-registered in the AEA RCT registry under RCT ID: AEARCTR-0011279.

1 Introduction

In a recent overview article, Lundberg and Stearns (2019) report on the state of women in academic economics. They present data showing that the share of women in the profession is very low and that the extent of underrepresentation has remained constant over the last decades. For example, women only represent 13% of all full professors (up from 6% in 1993) and 23% of associate professors (up from 11% in 1993) in PhD-granting economics departments in the US. The share of female assistant professors increased from 20% in 1993 to 29% in 2009, and then decreased over the past decade to 24%. In Germany (the largest country in our study), the representation of women in academic economics does not look more favorably; for example, only about 19% of all tenured university professors in economics are female (Federal Statistical Office 2018).¹

It is clearly important to understand the causal mechanisms behind such aggregate statistics and developments. However, isolating causal relationships is naturally challenging in the context of discrimination and evidence on discrimination in academic research is therefore scarce (references below). A particular dimension of discrimination in academia for which we lack causal evidence is regarding discrimination against young students at the pre-PhD level. This lack of evidence is despite the fact that gender discrimination against pre-PhD students constitutes a potentially important dimension of discrimination: women only represent about 30% of first-year Economics PhD students in the US (Lundberg and Stearns 2019) and about 37% of all Economics PhD students in Germany are female (Federal Statistical Office 2018).²

One potential reason for this underrepresentation lies in everyday discrimination experienced before the decision to start a PhD. It has been shown that (seemingly minor) everyday discrimination potentially has important consequences (along various dimensions) and therefore deserves attention (see Small and Pager 2020 for an overview of everyday discrimination and its consequences). To this end, everyday discrimination experienced in an academic context (in particular through senior researchers) potentially induces young pre-PhD level students to decide against pursuing a PhD. Besides being plainly unfair, the presence of discrimination against young students implies an inefficient situation where some high-potential students choose jobs in industry or public administra-

¹Lundberg and Stearns (2019) conclude (for the US) that, “over the past two decades, women’s progress in academic economics has slowed, with virtually no improvement in the female share of junior faculty or graduate students in decades.” Data from other countries and disciplines also show that women are underrepresented (particularly in quantitative, technical, and science-related fields) and discriminated against in academic research; see e.g., Ceci and Williams (2011), Moss-Racusin et al. (2012), Sheltzer and Smith (2014), Gaule and Piacentini (2018), Bukstein and Gandelman (2019), NSF (2019b) (more related literature further below throughout the Introduction).

²The share of women among doctoral students is similar in other disciplines. For example, women constitute about a third of doctoral recipients in physical sciences and earth sciences and a quarter of those in engineering and in mathematics and computer sciences in the US (NSF 2019a).

tion, rather than a career in academic economics. The pool of young students who select into the economics profession is then inferior relative to a counterfactual world without gender discrimination against pre-PhD level students. That is, (everyday) discrimination against young prospective professional researchers may be an important determinant for the initial decision to enter the economic profession and has implications for the overall quality of the pool of professional economists.³

Motivated by these considerations, the objective of our paper is to shed light on gender-based everyday discrimination against pre-PhD level students by professional academic economists. Our setting focuses on the case of effort provision to support the research project of a Master’s-degree student, and explores if the willingness among professional researchers to support the research project depends on the gender of the support requester. We thus focus on the willingness to pay effort to pre-PhD level students who have not yet decided whether to enter academic economics, while the literature usually focuses on discrimination against economists who have already entered the profession (in situations of evaluations or group-work recognition; e.g., Bagues et al. 2017, Sarsons 2017, Boring 2017, Wu 2018, Mengel et al. 2018, Sarsons et al. 2021, Dupas et al. 2021, Bransch and Kvasnicka 2022, Hengel 2022). Methodologically, we tie to Block et al. (2021) who study *race*-based everyday discrimination (in a non-research setting) using a field experiment in which survey invitations were sent out per email.

We conduct a randomized field experiment (described in Section 2) in which we send emails to all members of the professional association of German-speaking economists (similar to the US counterpart *AEA*). In these emails, the members are invited to participate in a survey and to assess the quality of non-economics journals. The survey invitations and journal assessments are part of an actual research project of two non-fictitious Master’s-degree students at a Swiss university, one of whom is female and one is male. We randomly vary between email recipients if the email is sent and signed by the male or female Master’s student to identify gender-based discrimination.

An important feature of our experimental design is that the entire field experiment is non-deceptive (e.g., non-fictitious students, names and student research project) and the results of the survey were used in a final Master’s thesis. As a result, the effort (journal assessment) that was provided by email recipients does not involve a waste of resources and has a meaningful purpose. This is somewhat in contrast to commonly used correspondence/audit studies (Bertrand and Dufflo 2017; Neumark 2018), where the CVs of fictitious applicants are sent to firms. We believe that a set-up with deceptive

³Bayer and Rouse (2016, page 221), in their overview article, state that the underrepresentation of women (and minority groups) in economics “likely hampers the discipline, constraining the range of issues addressed and limiting our collective ability to understand familiar issues from new and innovative perspectives”. Using a similar argument, Stansbury and Schultz (2023) argue that the lack of socio-economic diversity in the Economics profession implies that the profession is missing out on both talent and perspectives.

elements (fictitious names, meaningless task) would not have been tolerable in our specific case where we contacted a large number of actual economists in a professional association, many of them with high opportunity costs of investing time for the support of an unknown Master’s student.

We find empirical patterns in our experiment that are consistent with same-sex discrimination (see Section 3): male email recipients are more likely to respond to the email of the female student, and female email recipients are more likely to respond to the email of the male student. The differences are quantitatively substantial: the response rate among male recipients is 46% larger if the sender is female (relative to male sender), and the response rate among female recipients is 178% larger if the sender is male (relative to female sender).⁴

A follow-up survey suggests that the name of our male requester is perceived slightly more negatively than other common male names, while the name of our female requester is not differently perceived than other common female names. Although the differences in perceptions are not large and our heterogeneity results indicate that the average effect is indeed the result of gender-based discrimination,⁵ we cannot fully rule out that the unfavorable treatment of the male requester among male recipients is to some extent confounded by name-based discrimination. Importantly, however, our follow-up survey suggests that an unfavorable treatment of the female requester among female recipients cannot be explained with name-based discrimination. Female economists are less supportive of the female requester, *although* the male name is perceived more negatively. Our results thus provide evidence that female economists discriminate against women. We further discuss and empirically explore the issue of potential confounders in Section 4.1.

Considering literature studying discrimination in academia, the finding of same-sex discrimination is for example consistent with the results of Bransch and Kvasnicka (2022) who find that articles published in the top-five economics journals between 1991 and 2010 are less likely to be (co-)authored by females when females are more strongly represented on the editorial board. Studying applications to professorships in Spain, Bagues et al. (2017) find that female evaluators are not significantly more favorable toward female candidates. While not evidence for same-sex discrimination, this is consistent with our finding to the extent that females do *not* discriminate in favor of other females. However, some studies also find same-sex favoritism in science, for example in job promotions (De Paola and Scoppa 2015; Zinovyeva and Bagues 2015), or no effect of shared identity (Milkman et al. 2015). While related, none of these studies provide evidence for same-sex discrimination among female pre-PhD level students in the economics profession.

⁴We find the same pattern when we consider the number of evaluated journals as outcome variable.

⁵For example, we find it difficult to rationalize that the names themselves, rather than the gender of the requesters, have largely opposite effects on male and female recipients.

More generally, our finding of same-sex discrimination is not unprecedented and consistent with previous literature in non-science contexts. For example, a meta study of economic experiments finds significant same-sex discrimination in gender discrimination experiments (Lane 2016). Same-sex discrimination among females is reported by Bagues and Esteve-Volart (2010), who study Spanish public examinations and find that the chances of women to be hired decrease in the share of females in the respective hiring committee. Another example is Bar and Zussman (2019) who find same-sex discrimination in the context of driving test evaluations. We elaborate that our finding of same-sex discrimination is consistent with taste-based discrimination (rather than statistical discrimination) and/or the commonly known notion of intrasexual competition. We further elaborate on the mechanisms explaining our findings in Section 4.2.

Given the observed patterns of same-sex-discrimination and that we have more men than women in our sample, it is not surprising that the average response rate is higher if the email sender is female. The female sender is 26% more likely to receive a response than the male sender, with this finding being driven by the male email recipients. However, for the reasons discussed above, we cannot fully rule out that this average unfavorable treatment of the male requester is to some extent confounded by name-based discrimination.⁶

Most critical decisions in academic research (including the economics profession) are taken by tenured (senior) professors (e.g., hiring, tenure, promotions, journal submissions). In addition, experienced support by senior researchers is likely to be particularly encouraging or discouraging for young pre-PhD students and tenured professors are important role models. Our next set of analyses therefore explores differences in support between tenured and non-tenured economists. We find no treatment effects among non-tenured economists, while tenured professors are, on average, about 50% more likely to respond to the female student.

In light of the previously reported pattern of same-sex discrimination, we also ex-

⁶While the favorable average treatment of the female requester is seemingly in contrast with the aggregate statistics about the share of women in the profession, this finding is for example consistent with findings in related contexts such as performance evaluation in male-dominated fields of study (Breda and Ly 2015), college-admission decisions (Hanson 2017) or in teachers' evaluations in public schools (Lavy 2008). Our average finding is potentially also consistent with Card et al. (2023) who show that, in the last decades, women are considerably more likely to be selected as members of the National Academy of Science and the American Academy of Arts and Science than men with similar publication and citation records. Milkman et al. (2012, 2015) conduct a study in which US faculty members across various disciplines receive meeting requests from fictitious prospective doctoral students. Milkman et al. (2012) show that minority ethnicities and women were equally likely to get an appointment on the same day, but less likely to get an appointment in the near future. Milkman et al. (2015) show that these effects do not depend on whether sender and recipient share the same identity. In contrast to our study, i) Milkman et al. study discrimination against students already interested in a PhD rather than students who do not express an explicit interest in an academic career, ii) their setting involves a meeting appointment rather than a real-effort task, iii) they use fictitious students, and iv) they cannot make any specific statements about discrimination in the economics sciences.

plore the role of gender among tenured and non-tenured economists. Interestingly, we observe same-sex discrimination among non-tenured females, but not among non-tenured males. Within the group of tenured professors, we find same-sex discrimination among both males and females: tenured men are about 80% more likely to respond to the female student, and tenured women are almost three times more likely to respond to the male student. The differences between tenured males and non-tenured males are not driven by differences in name perceptions among these groups (as we show in the follow-up survey), providing an indication that same-sex discrimination among tenured males is not confounded by name perceptions.

The unfavorable treatment of the female requester by tenured female economists is not confounded by name-based discrimination either (for the reasons discussed before). Generally, this finding is consistent with the empirical observation from other settings that women who have been successful in male-dominated environments tend to behave in ways that block the advancement of junior female colleagues (see Section 4.2 for references and more discussion). Given the important role that female seniors play for young females as role models (Bettinger and Long 2005; Beaman et al. 2012), discrimination against females among tenured women is possibly detrimental for the development of young females in economics. In terms of contribution, we are not aware of any previous papers that explore differences in gender discrimination between tenured and non-tenured economists.

2 The Field Experiment

Setting. We conducted a natural field experiment with members of the association of German-speaking economists in countries Austria, Germany and Switzerland (Name: *Verein fuer Socialpolitik*, henceforth: *VfS*. This is the counterpart to the *AEA*).⁷ During the time of our experiment, the *VfS* had approximately 3,900 members at all career stages (incl. PhD students). We contacted all members who, upon signing up for membership, consented to be contacted for research purposes and therefore appeared on a list with email contacts. This list, which contained the email contacts of about 59% of all members, was used to implement our field experiment and to approach the members of the association. Overall, 2,356 members of *VfS* were part of our field experiment. Summary statistics are below.⁸

⁷See their website for more information: <https://www.socialpolitik.de/En>.

⁸We acknowledge that our sample is not fully representative for the full population of economists in the *VfS* because it only includes those members who agreed that their email contact is made public. However, we consider 59% of members to be a fairly large share which allows some conclusions about the group of professional economists.

Real-Effort Task to Support a Student Research Project. We investigate the willingness to support a student research project of two Master’s-level students at the University of Bern (Switzerland). These students were enrolled in a full-time Master’s program. In Switzerland and other German-speaking countries, it is common to start PhD studies after the Master’s-degree and the students in our study therefore were potential future PhD students.⁹ Since they were not enrolled in a PhD program, the students were not known in the scientific community and, for example, had not attended any academic conferences or had in any other way interacted with members of the community. Being unknown in the community, their names were not directly associated with a particular senior economist. In addition, both students were associated to the same chair at U. Bern.

The research project focuses on the perception of non-economics journals among economists. This focus of the project is of actual relevance: interdisciplinary research is growing and co-author teams from different disciplines are now common. Such co-author teams face the challenge that different journals have different reputations in different fields and the quality of journals in another field is often not known to peer researchers in the own field. This is challenging because researcher careers are dependent on publications in journals, which are evaluated by researchers in the own discipline. Against this background, the research project has the objective to implement a survey among economists in which they are asked to assess the quality of 100 peer-reviewed journals in neighboring disciplines. The ultimate goal is to generate a journal list that assigns quality rankings to non-economics journals and to establish equivalence factors between economics journals and journals in neighboring disciplines. See Appendix D for the list of journals.

To implement the survey, we approached *VfS* members via email and sent them an invitation to support the student research project and to complete the survey. The invitation email contained the context of the survey invitation – support of a student research project – and informed email recipients about the nature of the survey. Email recipients were well aware that they were asked to support the research project of a young pre-PhD level student.¹⁰ The invitation email also informed about the purpose of the

⁹It is very common to obtain a consecutive Master’s degree immediately after finishing undergraduate studies. A Master’s degree is not only pursued by students who are interested in academic research. However, conducting a survey and empirical project during their Master’s studies signals that the two students in our project are potentially interested in pursuing a PhD.

¹⁰The email stated upfront in the first sentence that it is sent by a student and that the survey invitation is part of a student research project. Note that, in contrast to the US, the term “student” (German: “Student”) in the German-speaking university system clearly refers to pre-PhD level students who are enrolled in an undergraduate or Master’s program. PhD students are commonly referred to as “Doktoranden”, “Promovierende” or “Wissenschaftliche Mitarbeiter”; these terms broadly translated with “PhD researcher” or “research employee”. PhD students usually have a paid job as research assistants and are therefore perceived to be part of scientific staff rather than students.

study (evaluating non-economics journals), and explained that recipients could assess up to 100 journals and that they could always stop throughout the survey. The email then included a link to the (Qualtrics) survey.

Randomized Treatment. The experimental variation was embedded in the emails that we sent to *VfS* members and in which they were invited to take the survey. The objective of our field experiment is to investigate if response rates of professional economists depend on the gender of the support-requester. A key part of our field experiment is that one of the two students was female and the other student was male. To address our objective, we randomly varied if the emails were sent and signed by the female student or the male student: one randomly selected half of *VfS* members on the email list received the email from the female student and the other randomly selected half received the email from the male student. The emails were identical except for the name of the sender. The treatment variation appeared at two points in the treatment letter: once in the first sentence where the respective student introduces her/him-self, and once at the end of the letter where the email is signed in the name of the respective student. In addition, the name appeared in the email address of the sender (email addresses had the name of the respective student before the @-symbol, while the part after @-symbol was identical). These three appearances imply that the name, and therefore gender, of the email sender was salient to email recipients. The invitation email is displayed in Appendix B.

The two students in our study both have Swiss sounding names with identical first-name syllables: “Annika Wyss” and “Adrian Ruchti”. Signals about heritage, race and ethnicity are thus held constant. In addition, the emails of both students were sent by the same official email servers of U. Bern. Recall that both students were unknown in the scientific community. Email recipients were not aware that they were part of a randomized intervention; they perceived the email only to be about a survey on non-economics journals. Our experimental approach is similar in spirit to correspondence studies in labor-market contexts where race, ethnicity or gender of a job applicant is varied through the name of the applicant. An important difference to these studies is that we rely on a non-fictitious setting (see below).

In Section 4.1, we discuss and empirically explore (using an additional survey) if the treatment effects could be confounded by differences in the names themselves, rather than the gender, of the two students.

Avoidance of Deception. Our experiment does not include any deceptive elements. In particular, the two students who signed the emails actually exist and they signed the emails with their real names. In addition, the survey invitation was part of an actual research project. The real-effort task (journal assessment) does not involve a waste of resources because it involves a research question of actual value (perceived quality

of non-econ journals). The non-deceptive approach to our research question was a key requirement to receive ethical approval.¹¹ We agree that a non-deceptive approach to experimental research is advantageous for various reasons. In addition, we believe that it would not have been tolerable, both from the view point of *VfS* and the contacted economists, if we had implemented a design where we deceived the members of the association and invited them to work on a meaningless task.¹² A strategy with fictitious names was therefore not an option.

It was then clear that we rely on non-fictitious email senders who sign the survey invitations with their real names. In light of our objective to study gender discrimination against pre-PhD level students, our goal was to find two Master’s-level students of different sex who are interested to support our project and to work on the contextual topic of the survey (perception of non-economics journals). To avoid any confounding factors such as heritage, race and ethnicity (see above), we additionally required two students with similar sounding Swiss-heritage names.

The two students who we identified to meet the criteria were, at the time of the experiment, employed as (part-time) student helpers at the chair where the MA-thesis adviser is employed. Student helpers are very common in German-speaking universities (almost every department/chair employs several student helpers). They are enrolled as undergraduate or Master’s level students and work part-time. Their main tasks include support in administrative tasks and research assistance (e.g., literature reviews, proof-reading). As common with student helpers, the two students had an appearance on the website of the chair where they were employed at the time. This appearance included their name and a picture. It did not feature any other person-specific details (e.g., no resumes or academic achievements). Both students appeared on the exact same website which is comparable to websites appearances of current student helpers (‘Hilfsassistenten’, see here: [LINK](#)). Please note that current web search results of the two names are very different from the search results during the field phase of our experiment.

In Section 4.1, we empirically investigate website click statistics to discuss if website visits potentially confound our treatment effects.

Summary Statistics, Number of Observations and Balance Across Groups.

Panel A of Table 1 presents summary statistics for the overall sample of 2356 email recipients. The overall response rate (measured using a dummy indicating if the email

¹¹University specific IRB boards are not common in German-speaking universities. A common approach among experimental researchers in German-speaking countries is to obtain ethical approval through *German Association for Experimental Economic Research* (GfeW). They generally do not grant approval for experiments involving deception; see the GfeW website.

¹²The non-deceptive approach of our study is somewhat in contrast to commonly used correspondence/audit studies (Bertrand and Duflo 2017; Neumark 2012; Neumark 2018), where CVs of fictitious applicants are sent to firms.

recipient opened the survey link) was 11%. 5.31 journals were evaluated on average. Among those who opened the survey link, about 50 journals were evaluated. To put our response rate into perspective, note that Block et al. (2021), who sent out survey invitations to study race-based discrimination, have response rates of 2% for a sample of the overall population and 5.7% for a sample of elected officials.

Using the list of *VfS* members, we derive information about gender and academic title/rank (no academic title, PhD and/or professor) for our sample of email recipients. We use the academic title/rank information to create a dummy variable indicating if an email recipient is tenured. In the German-speaking university system, non-tenured post-doctoral researchers and assistant professors are usually not referred to as professors. We therefore classify an email recipient as tenured if she/he is listed as a professor in the name list.¹³ Reflecting the gender composition in academic economics, 18% of our sample is female. A little bit less than half of the sample is tenured.

Panel B of Table 1 shows the number of observations for different combinations of pre-determined characteristics. 1,943 email recipients were male, 967 of those non-tenured and 976 tenured. 413 email recipients were female, 280 of them non-tenured and 133 tenured. We will analyze treatment effects within each of the four cells. Note that about half of the respondents received the male-sender letter.

Panel C of Table 1 shows balance in pre-defined characteristics across experimental groups (this includes all subgroups – full cross between tenure and gender – which we consider in our heterogeneity analyses). Displayed p-values are based on non-parametric ranksum tests for differences across treatment groups (t-tests exhibit almost identical p-values). As shown in the Table, all characteristics are balanced across experimental groups. We also regress a treatment group dummy on a set of covariates (again including dummies indicating subgroups) and find that neither single variable nor the overall model explain treatment group status (p-value F-statistic: 0.333).

3 Results

This section presents the field-experimental results. Our main outcome variable is a dummy indicating whether the email recipient opened the survey link. Mean differences in this variable across groups measure differences in response rates and therefore the general willingness to support the email sender. This outcome variable is in spirit an extensive-margin decision and compares to the outcome variable used in labor-market correspondence studies (call-back yes/no). Additional analyses explore treatment effects

¹³A few non-tenured assistant professors might be listed as 'professors' in the list. While our tenure variable may thus include a few misclassifications regarding tenure status, it certainly measures if someone has the rank of a professor or not.

on the number of assessed journals, both across all recipients and conditional on assessing at least one journal.

Our exhibition focuses on simple bar graphs (along with 95% confidence bands) and statistical inference is based on non-parametric pairwise ranksum (Wilcoxon-Mann-Whitney) tests for differences across experimental groups. Regression results, presented in table form in Appendix A, mirror differences in averages across groups and we therefore do not discuss the regression results in what follows. The Figures in the main body of the text focus on our main outcome variable, the response rate. While the following write-up also lays out the results for the number of assessed journals, for reasons of brevity, we do not present the corresponding results in Figures and instead leave them to Appendix A (in regression table form; see Table A.2 therein).

We first study average treatment effects across all email recipients (Section 3.1), and subsequently examine heterogeneity with respect to gender (3.2) and tenure status (3.3) of email recipients.

3.1 Average Differences

Figure 1 shows the mean response rate across experimental groups. The overall response rate is 9.35% if the email sender is male and 11.78% if the email sender is female. This difference is statistically significant with a p-value of 0.056 and quite sizable: female senders are 25.99% more likely to receive a response than male senders $((11.78-9.35)/9.35)$.

The differences in response rates between female and male senders are reflected in the (unconditional) total number of journals that are evaluated by email recipients (not reported in a figure). Email recipients on average evaluate 4.75 journals if the email sender is male and 5.86 if the email sender is female (p-value for this difference: 0.054).

We do not find any treatment effects on the number of journals conditional on assessing at least one journal. Among email recipients who assessed at least one journal, 50.82 journals were assessed if the email sender is male and 49.77 if the email sender is female (p-value: 0.961). We also do not see any significant differences in treatment effects on the intensive margin when we consider heterogeneity with respect to gender and tenure status. We therefore do not report intensive-margin results in subsequent analyses.

3.2 Male vs. Female Recipients

In a next step, we study heterogeneity in treatment effects with respect to the gender of the email recipient. Figure 2 presents average response rates across experimental groups separately for female and male email recipients. We see that male email recipients are more likely to respond to the female sender, while female email recipients are more likely

to respond to the male sender.

In particular, male email recipients responded to 9.2% of emails coming from the male student and to 13.39% of emails by the female student. This difference across experimental treatments is statistically significant with a p-value of 0.004. In relative terms, men are 46% more likely to respond to the female sender than to the male sender $((13.39-9.2)/9.2)$.

This pattern is even more pronounced among female email recipients. Women respond to 10.05% of all emails by the male student and to 3.61% of all emails by the female student. This difference is statistically significant with a p-value of 0.011. The response rate among women is thus almost three times as high if the mail is sent by the male student, relative to female sender (relative difference: $178\% = (10.05-3.61)/3.61$).

These patterns of same-sex discrimination translate to the (unconditional) total number of evaluated journals. Male economists evaluate 4.47 journals for the male sender, and 6.73 journals for the female sender (p-value: 0.003). Female economists transcribe 5.98 journals for the male sender and only 1.44 journals for the female sender (p-value: 0.005).

3.3 Tenured vs. Non-Tenured Recipients

Given that tenured senior professors have a particularly relevant role in the profession, we further explore if the treatment has differential effects among tenured and non-tenured email recipients. The results of this heterogeneity analysis are summarized in Figure 3, where we depict average response rates by treatment and tenure-status of email recipients.

We find no effects among non-tenured recipients. In this group, the response rate is 9.82% if the email comes from the male student and 10.22% if it comes from the female student (not statistically significant; p-value: 0.814). The result picture is different among tenured email recipients: they respond to 8.86% of requests by the male sender and to 13.6% of requests by the female sender. This difference is statistically significant with a p-value of 0.012. We thus see a favorable treatment of female students among this group, which is quite sizable: tenured economists are 53.5% more likely to respond to the woman than to the man $((13.6-8.86)/8.86)$.

These differences in response rates are reflected in the (unconditional) number of evaluated journals. We see no difference in the number of evaluated journals between male and female email senders among the group of non-tenured economists (means 5.13 vs 4.77, with p-value 0.918). Tenured professors on average discriminate against the male sender: 4.35 evaluated journals if sender is male and 7.15 evaluated journals if sender is female (p-value: 0.008).

Motivated by the observed patterns of same-sex discrimination, we also study the role of gender among tenured and non-tenured economists. The main results are presented

in Figure 4 which shows average response rates by treatment, tenure-status and gender. We first consider the group of non-tenured economists (left part of Figure 4). There is no evidence for a differential treatment of the male and female sender among non-tenured male email recipients (among this group, response rate is 10.37% if sender is male and 12.10% if sender is female; p-value: 0.394). We do see strong patterns of same-sex discrimination among non-tenured female recipients: they respond to 8.11% of all emails sent by the male student and to 3.03% of all emails by the female student (statistically significant with p-value 0.068). That is, non-tenured female economists are 167.66% more likely to respond to the male sender than to the female sender $((8.11-3.03)/3.03)$.

Among tenured professors (right part of Figure 4), we observe patterns consistent with same-sex discrimination among both male and female email recipients. Tenured male recipients respond to 8.1% of emails sent by the male student and to 14.73% of the female student (statistically significant with p-value 0.001). Tenured male recipients are thus 81.9% more likely to respond to the female sender, relative to male sender $((14.73-8.1)/8.1)$. The pattern is even more pronounced among tenured female professors: they respond to 14.08% of the male student email and to 4.84% of the female student email. Although the number of observations is quite small in the group of tenured female professors, the difference is statistically significant with a p-value of 0.074. Measured in relative terms, tenured female economists are 190.9% more likely to respond to the male sender than to the female sender $((14.08-4.84)/4.84)$.

Differences in response rates again translate to differences in the (unconditional) total number of evaluated journals. Non-tenured male economists do not discriminate on this margin (5.31 vs 5.69, p-value: 0.463), whereas non-tenured women discriminate against the female sender: 4.56 evaluated journals if sender is male and only 1.25 evaluated journals if sender is female (p-value: 0.066). Among tenured professors, we find patterns consistent with same-sex discrimination in the number of journals among both male and female economists. Tenured male economists evaluate 3.69 for male senders and 7.83 for female senders (p-value: 0.000). Tenured female economists transcribe almost 5 times as many journals for the male sender than for the female sender: 8.93 vs 1.86 with p-value of 0.031.

4 Discussion of Results

In this Section, we first address concerns about potential confounding factors (4.1) and then discuss potential explanations for our findings (4.2).

4.1 Potential Confounders

There are two natural concerns relating to confounding factors. First, there might be differences across the two names in our study (aside from gender differences) that drive the treatment effect. As a result, gender-based discrimination cannot be isolated from name-based discrimination. Second, email recipients searched online for the sender of the survey invitation. The website appearance could then add to the decision whether to start the survey or not. Eventually, these potential confounders constitute our cost of having a non-deceptive setting with non-fictitious email senders. In the following, we present conceptual and empirical arguments (including an additional survey experiment and an analysis of website clicks) to address these concerns.

4.1.1 Name-based Discrimination

We discuss the confounding role of name-based discrimination along three points.

Consistency with Empirical Findings. First, the empirical results of our experiment are plausibly inconsistent with name-based discrimination. For example, we find evidence of opposite treatment effects for female and male email recipients, and for tenured and non-tenured economists. These findings are inconsistent with the notion that perceptions of names drive the results: we consider it very implausible that the names have largely opposite effects on male vs female recipients and tenured vs non-tenured recipients. In addition, we find no treatment effects among non-tenured men, while we find an effect among tenured men. It is not plausible that non-tenured and tenured men (i.e., both of same sex) perceive the names considerably differently.

Names of Students in Experiments. Second, as mentioned previously, the two students in the two treatment conditions (recall: “Annika Wyss” and “Adrian Ruchti”) both have Swiss sounding names with identical first-name syllables. In contrast to e.g. “Emily” and “Greg” vs. “Lakisha” and “Jamal” in the seminal study by Bertrand and Mullainathan (2004), the two names in our setting are thus not differently related to factors such as heritage, race and ethnicity. We also see that the two names are equally common; *Google* searches produce a very similar number of search hits for both names. These aspects suggests that survey respondents were not differently familiar with the two names.

Follow-up Survey Experiment on Name Perceptions. Third, to further investigate the potential role of name-based discrimination, we conduct a survey experiment which we describe in the following.

Procedure and Sample. We commissioned the commercial survey provider *Norstat* to conduct the survey. The survey sample includes 495 respondents from Germany (the country where most email recipients come from) and is intended to be roughly comparable to our field-experimental sample of economics academics. Specifically, the survey provider invites respondents with a tertiary education degree (university) who are of working age (i.e., between 26 and 65 years, with proportional distribution among four age groups).¹⁴ Compared to our field experimental sample of economists, we deliberately oversample women in the survey (to meet their share of 50% in the general population). The survey experiment was pre-registered (see title page) and went to the field in April 2023. The median response time to complete the survey was a little less than four minutes.

Survey Design. The objective of the survey is to investigate if the names of the two students in our field experiment are associated with any specific traits, characteristics or attributes. To this end, we investigate if the names in our experiment are perceived and rated differently than the 10 most common names in Switzerland (which is the country of our field experiment's students). We compose a list of full names containing random combinations of the 10 most common Swiss first names and surnames.¹⁵ Each survey respondent is asked to rate three female names and three male names. The names to be rated always include the actual names of the two students in our field experiment (one female, one male). The remaining two male and two female names are randomly selected from the list of common Swiss names. The order of all six names is randomized. Obviously, survey respondents were not able to know which of the presented names are in the focus of the survey.

After providing some basic demographics, respondents are asked the following three questions for each of the six names.

- i) I can imagine naming my (potential) child *FIRST-NAME* (7-point scale from 'not at all true' to 'very true').
- ii.) I rate a person with the full name *FIRST-NAME SURNAME* as [...] (where this Q is asked for each of the following attributes: trustworthy, competent, selfish, attractive, sympathetic, reliable. 7-point scale from 'not at all true' to 'very true' for each of the attributes).
- iii) Based on my perception of the name *FIRST-NAME SURNAME*, I would like to work together with this person (7-point scale from 'not at all true' to 'very true').

¹⁴We acknowledge that, obviously, our field experimental sample includes many professors and people with a PhD degree. It is very difficult, and not our intention, to gather a survey sample that is exactly comparable. We focus on survey respondents with tertiary education to exclude respondents who are very different in terms of education than our field experimental sample.

¹⁵Sources for the names: FIRST NAMES and LAST NAMES.

The attributes in question ii) are chosen to reflect different aspects, such as character traits, outer appearance and competence. We code all variables such that higher values imply more positive perceptions.

Survey Results. For each of the survey questions, we compare the name of the student in our experiment to other names of same gender to abstract from gender differences across names (that is, we compare our female (male) student to other female (male) names). A direct comparison of the male and female name is confounded by gender differences and we thus focus on the comparison to names of the same gender. We build a simple additive linear index for survey question ii) in which the six attributes elicited in this question are summarized. Aside from presenting comparisons among the full sample of all survey respondents, we also show separate comparisons for the samples of male and female survey respondents, as these directly speak to our observed patterns of same-sex discrimination. We present the results in balance tables which include, for each survey question, the means for the name of interest and other names of same gender as well as the difference between the two groups. To pinpoint the statistical significance of this difference, we run OLS regressions of the respective outcome variable on a dummy indicating the name in our field experimental study (with standard errors that take into account that one survey respondent answers several questions for several names). The results are presented in Table 2, in which Panel A reports the findings for the female name and Panel B for the male name.

The overall finding in Panel A is that our female student's name, *Annika Wyss*, is not differently perceived than other common female names. The differences are economically small and mostly statistically insignificant. The patterns are very similar across all attributes elicited in question ii) and which we summarize in an index (results for the single attributes not reported). The only difference we detect is regarding the question about naming ones own child. Though the magnitude of the difference is not large (0.3 where the maximum difference could be six), we see that people are more likely to name their child *Annika*, the name in our field experiment, than other common first names. That is, the only difference that we see indicates a more positive perception of our female student's name. All described patterns are very similar in the populations of male and female respondents.

Panel B reveals that our male student's name, *Adrian Ruchti*, is perceived more negatively than other common male names. In the sample of all respondents, we see negative and statistically significant differences for survey questions ii) and iii). The absolute economic magnitude of these differences is in the range of 0.04 and 0.4 (where the maximum possible difference is six). The findings are similar for the single attributes of question ii). Evaluated in relative terms (with the mean of the comparison group of other male names as the reference point), the differences are between 1% and 8%. While

differences clearly exist, we do not view these differences as very large economically. Looking at the sub-samples of female and male survey respondents, we observe the more negative perceptions in both groups, though the differences are more pronounced among male respondents.

One finding in our main experiment is that tenured male recipients are less likely to respond to the male email sender, whereas non-tenured male recipients have similar response rates for the male and female sender. If name-based discrimination was the driver behind this finding, we should see that name perceptions are different between tenured and non-tenured males. Because tenured academics tend to be older than younger ones, we shed light on this by comparing name perceptions of older and younger males in the follow-up survey. We run regressions of the name perception outcomes on an indicator for the male sender's name (which is zero for other male names) along with an interaction between this indicator and an indicator for older survey respondents (older than 45 years old; note that we only have four age groups in our data). Panel A of Table 3 shows that there are no differences in perceptions of the male name between younger and older males, suggesting that older and younger males do not perceive the name differently and indicating that our main experimental results are not driven by name perceptions. The remaining panels of the table show similar analyses for the female name in our experiment (Panel B) and among the sample of female survey respondents (C and D).

To summarize, the female name in our study is not perceived differently than other females, whereas our male name is perceived somewhat more negatively than other male names. What do these findings imply for the interpretation of our main field experiment? Although the magnitudes are not very large, we cannot fully rule out that any unfavorable treatment of our male requester is to some extent confounded by name perceptions. One finding that speaks against this interpretation is that we find differences between younger (non-tenured) and older (tenured) males in our experiment, but do not see such differences between younger and older males in name perceptions.

Importantly, our survey results do not provide any support that an unfavorable treatment of the female student is biased due to differences in name perceptions. For example, we find discrimination against the female student by female economists, although the male name is perceived more negatively (also by females in our survey). That is, if name perceptions played a role at all, this confounder would work for us and the true gender effect in the main experiment might even be larger.

4.1.2 Website Searches

As described in Section 2, the two students in our study were employed as student helpers during the field phase of our experiment and, in this role, appeared on the website of the University chair that employed them. This web appearance included their name and a

picture, but it did not feature any other person-specific details. Another concern is that email recipients searched online for the email sender and that the website appearances then confound the gender effect.

We identify the following main indications that website searches are not a plausible driver of our treatment effects.

Visual Appearance of Student in Experiments. First, the invitation email clearly states that the survey invitation is sent by a pre-PhD level student at U. Bern. It is not likely that many professional economists will bother to invest time into the web search of unknown Master’s level students at a small university (this is supported by click stats — see below). In addition, both students are listed on the same website and it becomes clear from the online appearance that they are associated with the same chair. Furthermore, both students are of Swiss heritage and their visual appearance does not differ in terms of heritage, ethnicity or background.

Consistency with Empirical Findings. Second, our empirical findings of heterogeneous treatment effects also suggest that web searches do not play an important role. Treatment heterogeneity with respect to gender and tenure-status is plausibly inconsistent with web searches driving the results: Why would the web appearances have opposite effects on male vs female recipients and tenured vs non-tenured recipients? Furthermore, it is not plausible to us that non-tenured and tenured men (i.e., both of same sex) perceive the website appearances considerably differently.

Empirical Analysis of Website Clicks. Third, the number of website clicks suggest ex-post that web appearances do not confound the gender effect: The website of the chair where the students were employed (LINK HERE, including appendant sub-websites) did not have any unusual high traffic during the field phase of the experiment. The experiment was in the field for 35 days between March 06 and April 09, 2018. Table 4 presents an overview of the average number of daily website visitors during the 35 days of the field phase and the 35 days before and after the experiment. To check for year trends, the table additionally includes the average number of daily visitors during the same time periods in the year before (2017) and after (2019) the experiment. As shown in the table, the number of visitors during the field phase of the experiment in 2018 is not at all unusual. This finding clearly suggests that recipients of the survey did not widely search online for the two email senders.¹⁶

¹⁶We also used *Google Trends* to investigate the number of web searches for the two students. However, *Google* does not provide any results because this trends analysis does not contain enough data. *Google* searches for the names of current student helpers at the chair show that the university website always appears on top of the results list. This suggests that general web searchers for our two students should be reflected in the university-website visits which we analyze.

4.2 Mechanisms

What are the explanatory mechanisms that can rationalize our field experimental findings? While our experiment was not designed to disentangle the mechanisms behind its findings, we rely on economic reasoning and previous literature to discuss potential mechanisms. We follow the literature and discuss our results along the distinction between statistical and taste-based discrimination. The observed pattern of same-sex discrimination stands out and, in the case of same-sex discrimination among females, is likely not confounded by name-based discrimination. We thus focus on same-sex discrimination in our discussion of mechanisms.

Statistical discrimination (Arrow 1972; Phelps 1972) in our context implies that the gender of the email sender serves as a signal for the ability of the email sender to use the survey results. Statistical discrimination would then explain our results if email recipients believed that email senders of opposite gender are not well qualified and/or able to produce a good research study based on the survey. As a result, it would then not be worth the effort to support the survey project of the opposite-gender email sender. We argue that this is not a likely explanation for our results; we find it implausible that female and male email recipients in our study hold different beliefs about which gender is more or less worth the effort of being supported.

Taste-based discrimination (going back to Becker 1957¹⁷) is the more plausible explanation for our findings in our view. Bar and Zussman (2019) extend the Becker-type, utility-based model of taste-based discrimination to include gender preferences. The intuition behind their extension is that people reward members of groups whose company they enjoy. Such a model explains same-sex discrimination because (most) people enjoy the company of people of opposite gender. This model can thus explain our finding of same-sex discrimination being prevalent among both men and women: email recipients support the survey project of opposite-gender request senders simply because they have a taste for people of opposite gender and have a non-pecuniary joy of supporting their project. An additional facet of taste-based discrimination is related to intrasexual competition (e.g., Hunt et al. 2009; Rosvall 2011), according to which people are steadily in competition with people of own gender. This evolutionary channel is originally rooted in competition over mates of opposite gender and also consistent with same-sex discrimination among both male and female email recipients.

As referenced in the Introduction, the meta analysis by Lane (2016) finds significant same-sex discrimination in economic experiments. Rationalizing this result and explaining why gender is the only identity category which is associated with bias against the in-group rather than the out-group, the paper argues that there is an 'obvious intuition

¹⁷Becker focuses on racial relations and argues that discrimination against black people is explained by white people having a non-pecuniary cost from interacting with non-whites.

why gender is different from the other identity categories': sexual attraction towards the out-group (i.e., the opposite gender) and 'chivalry' (following Eckel and Grossman 2001). Our results are consistent with the findings and explanations in Lane (2016).

And what about a rationalization for our finding of pronounced same-sex discrimination among tenured females? These findings are potentially consistent with the empirically observed phenomenon (sometimes referred to as the 'queen bee effect') that women who have been successful in male-dominated environments tend to behave in ways that block rather than help the advancement of junior female colleagues (Staines et al. 1974; Mathison 1986; Ellemers et al. 2004; Ellemers et al. 2012).

5 Conclusion

We conduct a field experiment to study gender-based everyday discrimination against pre-PhD level students. We invite members of a professional association of academic Economists to participate in a survey and randomly vary if the emailed survey invitations are sent and signed by a female or male student. We find that female economists discriminate against the female email sender, while male economists discriminate against the male email sender. We also find heterogeneous effects for tenured and non-tenured economists.

Our empirical setting does not allow any conclusion regarding the types of strategies that should be pursued to mitigate discrimination in academic economics (such as mentoring programs, see Blau et al. 2010). We interpret our paper as a documentation of everyday discrimination against young students in economics and we leave it for future research to investigate how such discrimination should be combated.¹⁸

¹⁸Looking at settings outside academia, Flory et al. (2021) is an example of a study investigating *how* diversity (within firms) can be increased.

References

- Arrow, K. J. (1972). Some mathematical models of race in the labor market. In A. Pascal (Ed.), *Racial Discrimination in Economic Life*. Lexington, MA: Lexington Books.
- Bagues, M., M. Sylos-Labini, and N. Zinovyeva (2017). Does the gender composition of scientific committees matter? *American Economic Review* 107(4), 1207–38.
- Bagues, M. F. and B. Esteve-Volart (2010). Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment. *The Review of Economic Studies* 77(4), 1301–1328.
- Bar, R. and A. Zussman (2019). Identity and Bias: Insights from Driving Tests. *The Economic Journal* 130(625), 1–23.
- Bayer, A. and C. E. Rouse (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives* 30(4), 221–42.
- Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *Science* 335(6068), 582–586.
- Becker, G. S. (1957). *The Economics of Discrimination*. Chicago, IL: The University of Chicago Press.
- Bertrand, M. and E. Duflo (2017). Field Experiments on Discrimination. *Handbook of Economic Field Experiments* 1, 309–393.
- Bertrand, M. and S. Mullainathan (2004, aug). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* 94(4), 991–1013.
- Bettinger, E. P. and B. T. Long (2005). Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review Papers and Proceedings* 95(2), 152–157.
- Blau, F. D., J. M. Currie, R. T. A. Croson, and D. K. Ginther (2010). Can mentoring help female assistant professors? interim results from a randomized trial. *American Economic Review P&P* 100(2), 348–52.
- Block, R., C. Crabtree, J. B. Holbein, and J. Q. Monson (2021). Are americans less likely to reply to emails from black people relative to white people? *Proceedings of the National Academy of Sciences* 118(52).
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics* 145, 27–41.

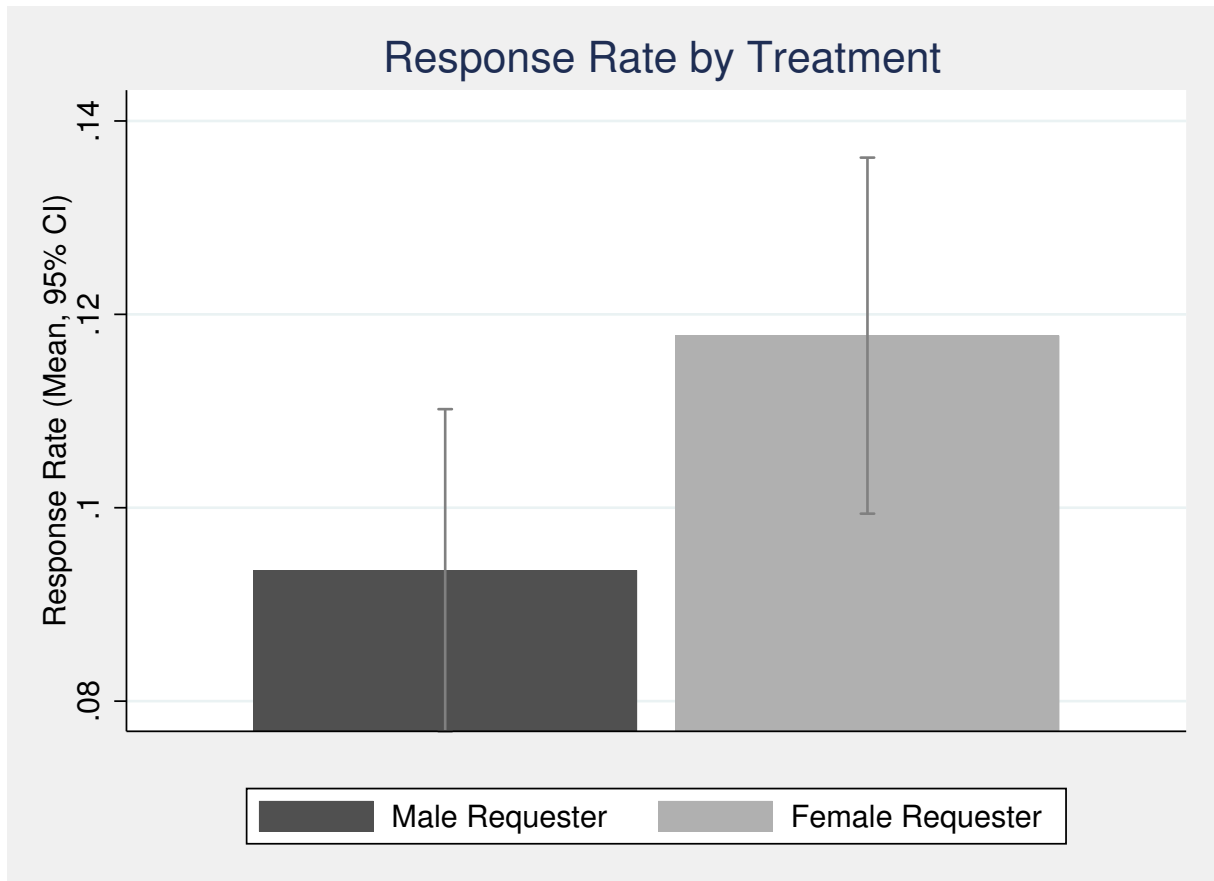
- Bransch, F. and M. Kvasnicka (2022). Male gatekeepers: Gender bias in the publishing process? *Journal of Economic Behavior & Organization* 202, 714–732.
- Breda, T. and S. T. Ly (2015). Professors in core science fields are not always biased against women: Evidence from france. *American Economic Journal: Applied Economics* 7(4), 53–75.
- Bukstein, D. and N. Gandelman (2019). Glass ceilings in research: Evidence from a national program in uruguay. *Research Policy* 48(6), 1550 – 1563.
- Card, D., S. DellaVigna, P. Funk, and N. Iriberry (2023). Gender gaps at the academies. *Proceedings of the National Academy of Sciences* 120(4), e2212421120.
- Ceci, S. J. and W. M. Williams (2011). Understanding current causes of women’s underrepresentation in science. *Proceedings of the National Academy of Sciences* 108(8), 3157–3162.
- De Paola, M. and V. Scoppa (2015). Gender discrimination and evaluators’ gender: Evidence from italian academia. *Economica* 82(325), 162–188.
- Dupas, P., A. S. Modestino, M. Niederle, J. Wolfers, and T. S. D. Collective (2021). Gender and the dynamics of economics seminars. NBER Working Paper 28494.
- Eckel, C. and P. Grossman (2001). Chivalry and solidarity in ultimatum games. *Economic Inquiry* 39(2), 171–188.
- Ellemers, N., F. Rink, B. Derks, and M. K. Ryan (2012). Women in high places: When and why promoting women into top positions can harm them individually or as a group (and how to prevent this). *Research in Organizational Behavior* 32, 163–187.
- Ellemers, N., H. Van den Heuvel, D. de Gilder, A. Maass, and A. Bonvini (2004). The underrepresentation of women in science: Differential commitment or the queen bee syndrome? *British Journal of Social Psychology* 43(3), 315–338.
- Federal Statistical Office (2018). Personal an hochschulen. Fachserie 11, reihe 4.4, 2018. personal an hochschulen, available online: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/personal-hochschulen-2110440187004.html>.
- Flory, J. A., A. Leibbrandt, C. Rott, and O. Stoddard (2021). Increasing Workplace Diversity: Evidence from a Recruiting Experiment at a Fortune 500 Company. *Journal of Human Resources* 56(1), 73–92.
- Gaule, P. and M. Piacentini (2018). An advisor like me? advisor gender and post-graduate careers in science. *Research Policy* 47(4), 805 – 813.

- Hanson, A. (2017). Do college admissions counselors discriminate? evidence from a correspondence-based field experiment. *Economics of Education Review* 60, 86 – 96.
- Hengel, E. (2022). Publishing While Female: are Women Held to Higher Standards? Evidence from Peer Review. *The Economic Journal* 132(648), 2951–2991.
- Hunt, J., C. J. Breuker, J. A. Sadowski, and A. J. Moore (2009). Male-male competition, female mate choice and their interaction: determining total sexual selection. *Journal of Evolutionary Biology* 22(1), 13–26.
- Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review* 90, 375–402.
- Lavy, V. (2008). Do gender stereotypes reduce girls’ or boys’ human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics* 92(10-11), 2083–2105.
- Lundberg, S. and J. Stearns (2019). Women in economics: Stalled progress. *Journal of Economic Perspectives* 33(1), 3–22.
- Mathison, D. L. (1986). Sex differences in the perception of assertiveness among female managers. *The Journal of Social Psychology* 126(5), 599–606.
- Mengel, F., J. Sauermann, and U. Zoelitz (2018). Gender Bias in Teaching Evaluations. *Journal of the European Economic Association* 17(2), 535–566.
- Milkman, K. L., M. Akinola, and D. Chugh (2012). Temporal distance and discrimination: An audit study in academia. *Psychological Science* 23(7), 710–717.
- Milkman, K. L., M. Akinola, and D. Chugh (2015). What happens before? a field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology* 100(6), 1678–1712.
- Moss-Racusin, C. A., J. F. Dovidio, V. L. Brescoll, M. J. Graham, and J. Handelsman (2012). Science faculty’s subtle gender biases favor male students. *Proceedings of the National Academy of Sciences* 109(41), 16474–16479.
- Neumark, D. (2012). Detecting discrimination in audit and correspondence studies. *Journal of Human Resources* 47(4), 1128–1157.
- Neumark, D. (2018). Experimental Research on Labor Market Discrimination. *Journal of Economic Literature* 56(3), 799–866.
- NSF (2019a). Doctorate recipients from US universities: 2019. National center for science and engineering statistics, national science foundation. NSF 21-308. alexandria, va. available at <https://nces.gov/pubs/nsf21308/>.

- NSF (2019b). Women, minorities, and persons with disabilities in science and engineering: 2019. National science foundation, national center for science and engineering statistics. special report NSF 19-304. alexandria, va. available at <https://www.nsf.gov/statistics/wmpd>.
- Phelps, E. (1972). The statistical theory of racism and sexism. *American Economic Review* 62(4), 659–61.
- Rosvall, K. A. (2011). Intrasexual competition in females: evidence for sexual selection? *Behavioral Ecology* 22(6), 1131–1140.
- Sarsons, H. (2017). Recognition for group work: Gender differences in academia. *American Economic Review: Papers and Proceedings* 107(5), 141–45.
- Sarsons, H., K. Gërkhani, E. Reuben, and A. Schram (2021). Gender differences in recognition for group work. *Journal of Political Economy* 129(1), 101–147.
- Sheltzer, J. M. and J. C. Smith (2014). Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences* 111(28), 10107–10112.
- Small, M. L. and D. Pager (2020). Sociological perspectives on racial discrimination. *Journal of Economic Perspectives* 34(2), 49–67.
- Staines, G., C. Tavriss, and T. E. Jayaratne (1974). Queen bee syndrome. *Psychology Today* 7(55).
- Stansbury, A. and R. Schultz (2023). The economics profession’s socioeconomic diversity problem. *Journal of Economic Perspectives* 37(4), 207–30.
- Wu, A. H. (2018). Gendered language on the economics job market rumors forum. *AEA Papers and Proceedings* 108.
- Zinovyeva, N. and M. Bagues (2015). The role of connections in academic promotions. *American Economic Journal: Applied Economics* 7(2), 264–92.

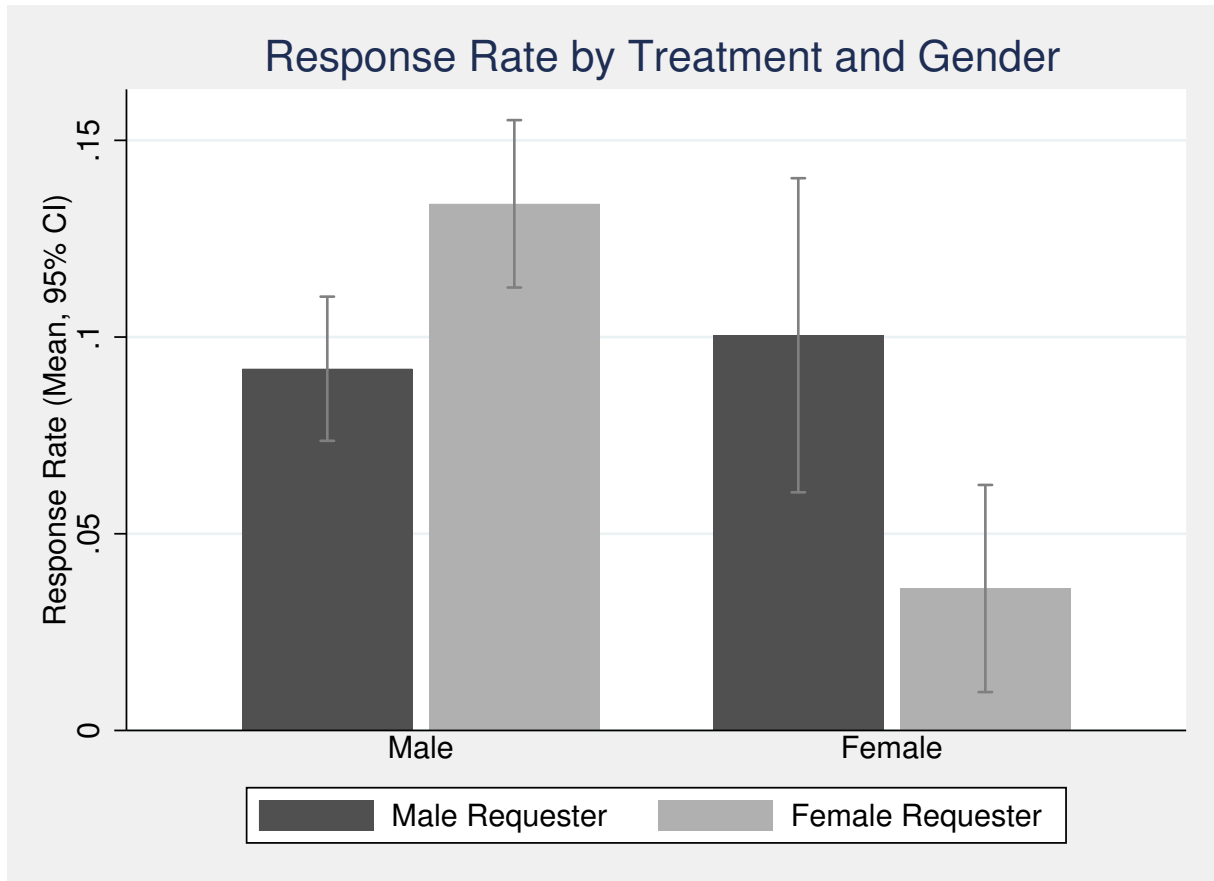
Main Figures and Tables

Figure 1: Response Rate by Experimental Group



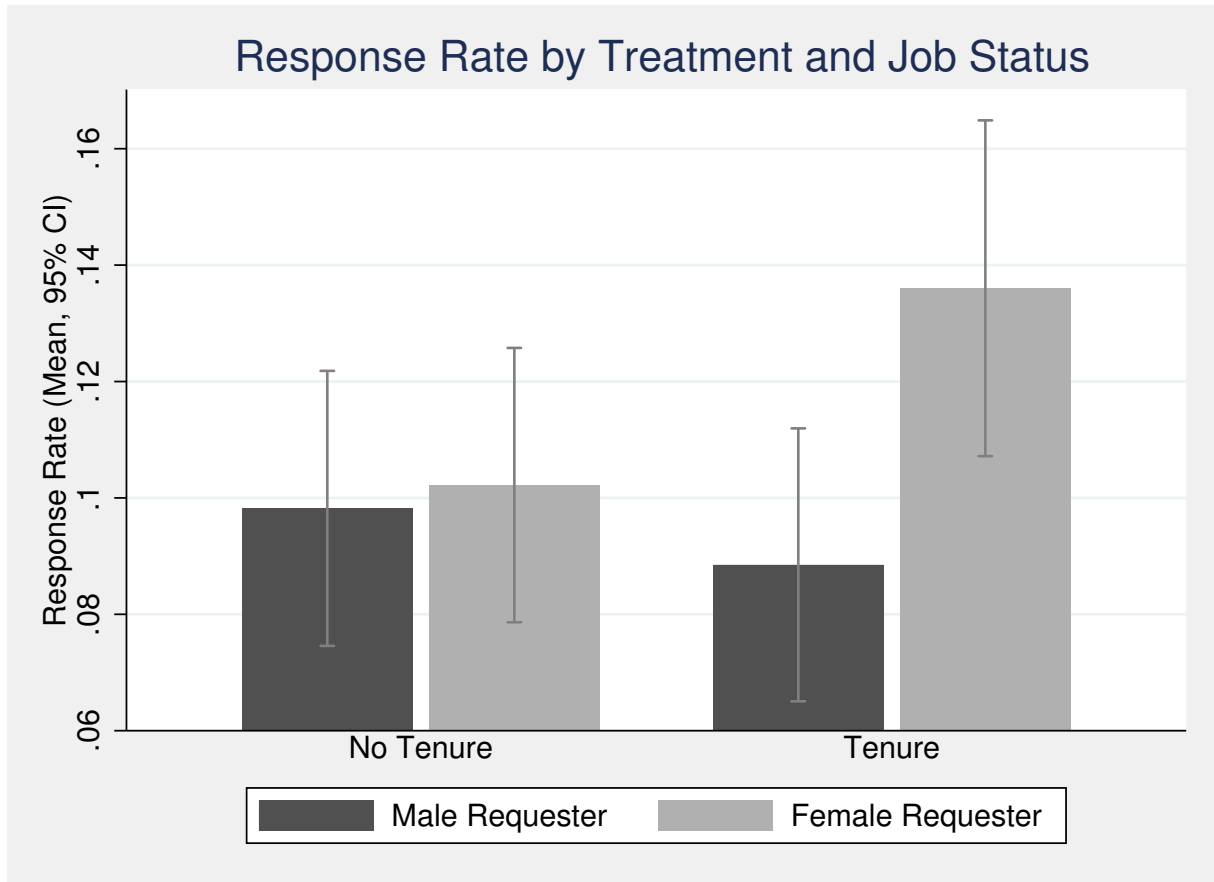
Notes: Average response rate by experimental group, along with 95% confidence bars. The outcome variable measures the share of survey-invitation recipients who clicked on the survey link (as described in Section 2). Respondents in experimental group “Male Requester” received a survey-invitation mail that was signed by a male Master’s degree student. Respondents in experimental group “Female Requester” received a survey-invitation mail that was signed by a female Master’s degree student. Total number of observations is 2356 with proportional distribution across experimental groups. The p-value of a pair-wise ranksum test for differences across experimental groups is: 0.056. Data are obtained from a field experiment as described in Section 2.

Figure 2: Response Rate by Experimental Group and Gender



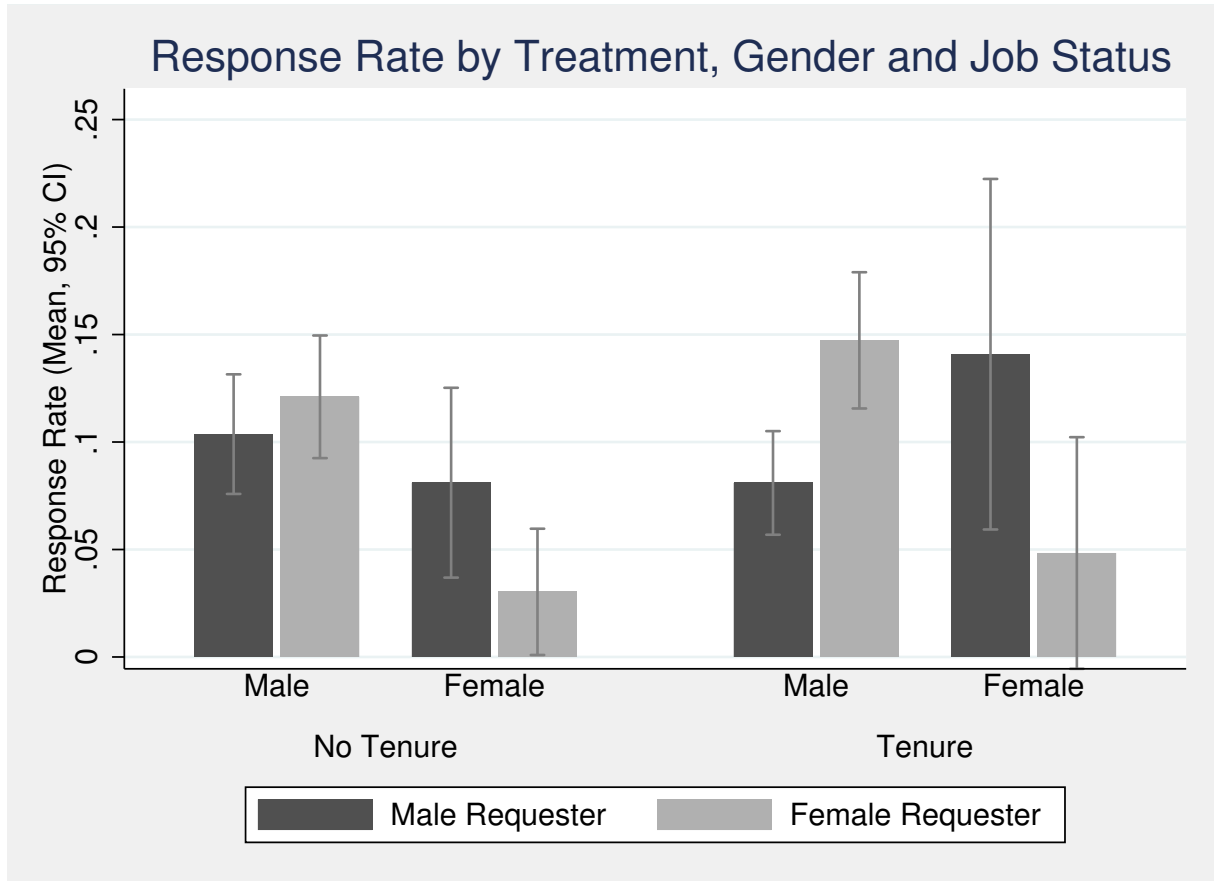
Notes: Average response rate by experimental group and gender, along with 95% confidence bars. The outcome variable measures the share of survey-invitation recipients who clicked on the survey link (as described in Section 2). Respondents in experimental group “Male Requester” received a survey-invitation mail that was signed by a male Master’s degree student. Respondents in experimental group “Female Requester” received a survey-invitation mail that was signed by a female Master’s degree student. Total number of observations is 2356 with proportional distribution across experimental groups. The overall sample consists of 1943 men and 413 women. P-values from pair-wise ranksum tests for differences across experimental groups are as follows: 0.004 among all men; 0.011 among all women. Data are obtained from a field experiment as described in Section 2.

Figure 3: Response Rate by Experimental Group and Job Status



Notes: Average response rate by experimental group and job status (tenured/non-tenured), along with 95% confidence bars. The outcome variable measures the share of survey-invitation recipients who who clicked on the survey link (as described in Section 2). Respondents in experimental group “Male Requester” received a survey-invitation mail that was signed by a male Master’s degree student. Respondents in experimental group “Female Requester” received a survey-invitation mail that was signed by a female Master’s degree student. Total number of observations is 2356 with proportional distribution across experimental groups. The overall sample consists of 1109 tenured professors and 1247 non-tenured economists. P-values from pair-wise ranksum tests for differences across experimental groups are as follows: 0.814 among all non-tenured economists; 0.012 among all tenured professors. Data are obtained from a field experiment as described in Section 2.

Figure 4: Response Rate by Experimental Group, Gender and Job Status



Notes: Average response rate by experimental group, gender and job status (tenured/non-tenured), along with 95% confidence bars. The outcome variable measures the share of survey-invitation recipients who who clicked on the survey link (as described in Section 2). Respondents in experimental group “Male Requester” received a survey-invitation mail that was signed by a male Master’s degree student. Respondents in experimental group “Female Requester” received a survey-invitation mail that was signed by a female Master’s degree student. Total number of observations is 2356 with proportional distribution across experimental groups. The overall sample consists of 967 non-tenured male economists, 280 non-tenured female economists, 976 tenured male professors, 133 tenured female professors. P-values from pair-wise ranksum tests for differences across experimental groups are as follows: 0.394 among all non-tenured male economists, 0.068 among all non-tenured female economists, 0.001 among all tenured male professors, 0.074 among all tenured female professors. Data are obtained from a field experiment as described in Section 2.

Table 1: Summary Statistics, Number of Observations, and Balance

Panel A: Summary Statistics			
Variable	N	mean	sd
Response Rate	2356	0.11	0.31
No Evaluated Journals	2356	5.31	20.06
No Evaluated Journals (cond. on response)	249	50.23	39.42
Gender (1 = female)	2356	0.18	0.38
Job Status (1 = tenured)	2356	0.47	0.50
Panel B: Number of Observations (by Group)			
	No Tenure	Tenure	Total
Male	967	976	1,943
Female	280	133	413
Total	1,247	1,109	2,356
Panel C: Balance Across Treatment Groups			
	Male Sender	Female Sender	p-value Diff
Gender (1 = female)	0.186	0.164	0.164
Job Status (1 = tenured)	0.480	0.461	0.345
Female/No Tenure	0.126	0.112	0.294
Female/Tenured	0.060	0.053	0.410
Male/No Tenure	0.394	0.427	0.099
Male/Tenured	0.420	0.409	0.568

Notes: Variable *Response Rate* measures the share of survey-invitation recipients who clicked on the survey link. Variable *No Evaluated Journals* measures the number of journals that were evaluated by the survey-invitation recipient. Variables *Gender* and *Job Status* are dummy variables indicating if the survey-invitation recipient is female and tenured, respectively. Further variables in Panel C describe balance within subgroups that we consider in our heterogeneity analyses (e.g., *Female/No Tenure* is a dummy variable indicating females without tenure). All variables are described in more detail in Section 2. Data are obtained from a field experiment as described in Section 2.

Table 2: Follow-up Survey: Perceptions of Female and Male Names

(1)	(2)	(3)	(4)
Panel A: Perceptions of Female Name			
Variable	Other Female Names	Annika Wyss	Difference
Sample: All Survey Respondents			
name child	3.497	3.829	0.332***
index attributes	4.519	4.546	0.027
work together	4.672	4.629	-0.043
Sample: Female Survey Respondents			
name child	3.085	3.473	0.388***
index attributes	4.547	4.620	0.073
work together	4.639	4.636	-0.003
Sample: Male Survey Respondents			
name child	3.931	4.197	0.265**
index attributes	4.498	4.478	-0.020
work together	4.710	4.621	-0.088
Panel B: Perceptions of Male Name			
Variable	Other Male Names	Adrian Ruchti	Difference
Sample: All Survey Respondents			
name child	3.278	3.240	-0.038
index attributes	4.418	4.164	-0.254***
work together	4.562	4.162	-0.400***
Sample: Female Survey Respondents			
name child	3.058	3.237	0.179
index attributes	4.486	4.366	-0.120**
work together	4.606	4.319	-0.287***
Sample: Male Survey Respondents			
name child	3.522	3.233	-0.289**
index attributes	4.356	3.973	-0.383***
work together	4.524	4.013	-0.511***

Notes: The table reports means in name perceptions for the names in our field experimental study (column 3) and other common Swiss names of the same gender (column 2). Panel A is regarding the female name in our field experiment (*Annika Wyss*) and Panel B is regarding the male name (*Adrian Ruchti*). Perceptions are elicited using three questions (column 1), each on a 7-point scale. Variable *index attributes* is a linear index summarizing six attributes that were elicited in survey question two (see Section 4.1 for details). Column (4) reports the difference between columns (3) and (4), along with stars indicating the level of statistical significance of this difference: * significant at 10%; ** significant at 5%; *** significant at 1%. Statistical significance is obtained from OLS regressions of the respective outcome variable of column (1) on a dummy indicating the name in our field experimental study (with standard errors that take into account that one survey respondent answers several questions for several names). The sample of all survey respondents in Panel A includes 495 unique individuals. The samples in Panels B (females) and C (males) include 248 and 244 unique individuals, respectively (3 individuals indicated a diverse sex).

Table 3: Follow-up Survey: Differences in Name Perceptions betw. Young and Old

	(1)	(2)	(3)
	name child	index attributes	work together
Panel A: Male Survey Respondents			
Adrian Indicator	-0.221 (0.162)	-0.437*** (0.101)	-0.492*** (0.144)
Old Indicator	0.173 (0.225)	-0.299** (0.118)	-0.348** (0.157)
Adrian \times Old	-0.134 (0.231)	0.110 (0.135)	-0.038 (0.190)
constant	3.435*** (0.172)	4.502*** (0.089)	4.701*** (0.117)
N	238	217	229
Panel B: Female Survey Respondents			
Adrian Indicator	0.524*** (0.191)	-0.002 (0.077)	-0.123 (0.119)
Old Indicator	0.974*** (0.216)	-0.067 (0.111)	-0.219 (0.161)
Adrian \times Old	-0.662** (0.260)	-0.241** (0.117)	-0.328* (0.170)
constant	2.548*** (0.151)	4.519*** (0.081)	4.717*** (0.114)
N	242	219	226
Adrian Indicator	0.524*** (0.175)	0.095 (0.071)	0.032 (0.110)
Old Indicator	0.636*** (0.226)	-0.121 (0.110)	-0.225 (0.159)
Adrian \times Old	-0.257 (0.234)	-0.046 (0.109)	-0.070 (0.163)
constant	2.750*** (0.160)	4.608*** (0.081)	4.753*** (0.110)
N	241	220	224

Notes: The table reports results from regressions (using OLS) of the respective perception variable on i) an indicator for the name in our field experimental study (either *Adrian Ruchti* or *Annika Wyss*), ii) an indicator for old survey respondents (older than 45 years), iii) and the interaction between variables i) and ii). The indicator for the male (female) name is zero for other male (female) names, implying that its coefficient provides the difference in perception between the male (female) name in our study and other male (female) names. In Panel A, the sample in all regressions is restricted to male respondents. In Panel B, the sample in all regressions is restricted to female respondents. The three specifications are different with respect to the perception outcome variable. Stars indicate the level of statistical significance of the coefficients: * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Average Number of Daily Website Visitors During Different Time Periods

	30 Jan—05 March	06 March—09 April	10 April—14 May
2017	77.94	75.97	53.89
2018	71.17	60.26	60.40
2019	83.51	70.40	69.46

Notes: Reported are the average numbers of daily website visitors during different time periods in the year of the field experiment (2018), the year before the experiment (2017) and the year after the experiment (2019). The field experiment was in the field for 35 days between 06 March and 09 April 2018. The time period between 30 January and 05 March represents the 35 days before the experimental field phase. The time period between 10 April and 14 May represents the 35 days after the experimental field phase. The statistics are for visitors of the following website and the appendant sub-websites: <http://www.iop.unibe.ch/>. The daily averages are calculated based on daily data that were obtained from the communication/marketing department at University of Bern (<http://www.kommunikation.unibe.ch>).

Appendix

Table of Contents

A	Field Experimental Results: Regression Tables	1
B	Treatment Letter	3
C	Questionnaire of Follow-Up Survey Experiment on Name Perceptions	4
D	List of Journals to Be Evaluated	6

A Field Experimental Results: Regression Tables

Table A.1: Regression Results: Response Rate

	(I)	(II)	(III)	(IV)	(V)
Treat (Female Requester)	0.024* (0.013)	0.024* (0.013)	0.042*** (0.014)	0.004 (0.017)	0.003 (0.017)
Female		-0.041*** (0.015)	0.009 (0.022)		-0.030* (0.017)
Tenure		0.008 (0.013)		-0.010 (0.017)	-0.013 (0.017)
Treat x Female			-0.106*** (0.028)		
Treat x Tenure				0.044* (0.025)	0.052** (0.026)
Treat x Tenure x Female					-0.069* (0.036)
constant	0.094*** (0.008)	0.097*** (0.011)	0.092*** (0.009)	0.098*** (0.012)	0.105*** (0.013)
N	2356	2356	2356	2356	2356

Notes: The table presents the results of OLS regressions (linear probability model). The outcome variable, *Response Rate*, is a dummy variable indicating if a survey-invitation recipient clicked on the survey link. Respondents in treatment group *Female Requester (Treat)* received a survey-invitation mail that was signed by a female Master's degree student. The coefficient for this group is relative to the experimental group that received a survey-invitation mail that was signed by a male Master's degree student. Variables *Female* and *Tenure* are dummy variables indicating if the survey-invitation recipient is female and tenured, respectively. All variables are described in more detail in Section 2. Specifications (I)-(V) differ with respect to the included right-hand-side variables. Significance stars indicate precision obtained from robust standard errors (presented in parentheses): * significant at 10%; ** significant at 5%; *** significant at 1%. The data are obtained from a field experiment as described in Section 2.

Table A.2: Regression Results: Number of Evaluated Journals

	(I)	(II)	(III)	(IV)	(V)
Treat (Female Requester)	1.109 (0.826)	1.086 (0.823)	2.259** (0.925)	-0.364 (1.084)	-0.391 (1.079)
Female		-1.617 (1.021)	1.504 (1.623)		-0.798 (1.158)
Tenure		0.631 (0.853)		-0.779 (1.121)	-0.872 (1.151)
Treat x Female			-6.793*** (1.933)		
Treat x Tenure				3.162* (1.669)	3.770** (1.735)
Treat x Tenure x Female					-5.175** (2.280)
constant	4.753*** (0.561)	4.751*** (0.733)	4.473*** (0.597)	5.128*** (0.796)	5.321*** (0.848)
N	2356	2356	2356	2356	2356

Notes: The table presents the results of OLS regressions. The outcome variable, *Number of Evaluated Journals*, measures the number of journals that were evaluated by the survey-invitation recipient in the survey (including zeros). Respondents in treatment group *Female Requester (Treat)* received a survey-invitation mail that was signed by a female Master's degree student. The coefficient for this group is relative to the experimental group that received a survey-invitation mail that was signed by a male Master's degree student. Variables *Female* and *Tenure* are dummy variables indicating if the survey-invitation recipient is female and tenured, respectively. All variables are described in more detail in Section 2. Specifications (I)-(V) differ with respect to the included right-hand-side variables. Significance stars indicate precision obtained from robust standard errors (presented in parentheses): * significant at 10%; ** significant at 5%; *** significant at 1%. The data are obtained from a field experiment as described in Section 2.

B Treatment Letter

Dear member of Verein fuer Socialpolitik,

I am Adrian Ruchti [*Treatment Variation: Annika Wyss*], a student from the University of Bern (Switzerland), and I am contacting you in order to invite you to participate in a survey as part of a student research project (link below). We are contacting all members of the Verein fuer Socialpolitik in an effort to gather information about how economists in Germany, Austria, and Switzerland rate journals in the boundary disciplines of economics.

Our project goal is to create a journal list (similar to rankings such as from the Handelsblatt or the VHB) that assigns quality ratings to journals that economists may publish in, but which are not listed as an “economics journal”. It will help us establish equivalence factors.

For example, is the journal “Nature” a journal similar to a very top journals (e.g., American Economic Review) or not. By participating in our survey, you will help us provide subjective ratings to a list of journals. This list is very long (i.e., more than 100 journals). On each page of the survey, you will have the option “Stop rating journals”, which will immediately bring you to the end of the survey. Thus, your time investment is voluntary. Importantly, even rating a few journals will help us in our project.

Should you have any question, please feel free to reply to this email. Thanking you in advance for your participation.

Sincerely,

Adrian Ruchti [*Treatment Variation: Annika Wyss*]

Follow this link to the Survey:

Link to Survey

Or copy and paste the URL below into your internet browser:

Survey URL

C Questionnaire of Follow-Up Survey Experiment on Name Perceptions

Each respondents answers questions 5, 6, and 7 for each of six different names (three female, three male). The set of names always includes the names of the names in our experiment (one male, one female), and the remaining two female names and two male names are randomly pulled from a list of the most common Swiss names. The order of questions is randomized. Screenshots of the original survey screens (in German) are available upon request.

Opening Screen

Thank you for participating in our survey.

In the following, you are asked to rate 3 male and 3 female first and last names according to your perception.

Your cooperation in this survey is extremely important for the validity of our study. The survey can be interrupted at any time and continued at the same point by clicking on your personal link again.

Question 1

What is your highest academic qualification?

Question 2

How old are you?

Question 3

Please indicate your gender.

Question 4

In which federal state is your main residence?

Question 5

Please rate the following question on a scale of -3 (not at all true) to 3 (very true):

I can imagine naming my (potential) child FIRST-NAME.

Question 6

Please rate the following question on a scale of -3 (not at all true) to 3 (very true):

I rate a person with the full name FIRST-NAME SURNAME as [...].

- Trustworthy
- Competent
- Selfish
- Attractive
- Sympathetic
- Reliable

Question 7

Please rate the following question on a scale of -3 (not at all true) to 3 (very true):

Based on my perception of the name FIRST-NAME SURNAME, I would like to work with this person.

D List of Journals to Be Evaluated

- Nature Sustainability
- Nature Energy
- Nature Communications
- Scientific Reports
- Proceedings of the National Academy of Science (PNAS)
- PLOS One
- Science Advances
- New England Journal of Medicine
- Journal of the American Medical Association (JAMA)
- Cell
- The Lancet
- Behavioral and Brain Science
- Trends in Cognitive Science (TiCS)
- Journal of Personality and Social Psychology
- Journal of Experimental Social Psychology
- Food Quality and Preferences
- Journal of the American Mathematics Society
- American Journal of Mathematics
- Journal of the American Statistical Society
- Proceedings of the Royal Society (A, B)
- Philosophical Transactions of the Royal Society
- Journal of Theoretical Biology
- Psychological Review
- Psychological Science
- Psychoneuroendocrinology
- Evolution and Human Behavior
- Neuron
- Nature Human Behavior
- Nature Reviews Neuroscience
- Columbia Law Review
- Stanford Law Review
- Journal of Legal Studies
- Behavioral Science and Policy
- Behavioural Public Policy
- Physical Review Letters

- Journal of the Economic Science Association
- Agricultural Systems
- American Criminal Law
- American Journal of Preventive Medicine
- American Journal of Public Health
- Annual Review of Sociology
- Appetite
- Behavioral Medicine
- Berkeley Technology Law Journal
- Bioinformatics
- Biological Psychiatry
- BMC Public Health
- Boston University Law Review
- Cerebral Cortex
- Cognition, Brain, Behavior
- Computers in Human Behavior
- Current Opinion in Neurobiology
- Developmental Neuroscience
- Education, Citizenship and Social Justice
- Ethics & Behavior
- European Law Review
- Experimental Psychology
- Food Policy
- Frontiers in Behavioral Neuroscience
- Group Processes & Intergroup Relations (GRIP)
- Handbook of Health Economics
- Health Affairs
- Human Brain Mapping
- Human Nature
- Il Nuovo Cimento
- Industrial Health
- Journal of Applied Psychology
- Journal of Applied Social Psychology
- Journal of Cross-Cultural Psychology
- Journal of European Competition Law & Practice
- Journal of Experimental Social Psychology
- Journal of Genetic Psychology

- Journal of Health Economics
- Journal of Language and Social Psychology
- Journal of Moral Education
- Journal of Occupational Health Psychology
- Journal of Social Issues
- Journal of Social Justice
- Law and Human Behavior
- Legal and Criminological Psychology
- Nature Climate Change
- Nature Neuroscience
- NeuroImage
- Neurology
- Personality and Social Psychology Review
- Philosophical Psychology
- Physiology & Behavior
- PLOS Computational Biology
- Population Health
- Psychological Sciences and Social Sciences
- Psychology, Crime & Law
- Psychology, Health, & Medicine
- Public Choice
- Science
- Sex Roles
- Social Cognition
- Social Cognitive and Affective Neuroscience (SCAN)
- Social Justice Research Stress & Health
- The American Journal of Clinical Nutrition (AJCN)
- The Proceedings of the Nutrition Society