

LETTER

Accessibility and Equity in the Research Process: Gender Bias in Elite Interview Recruitment

Margaret A. T. Kenney and John Salchak

Travers Department of Political Science, University of California, Berkeley, CA, USA

Corresponding author: Margaret A. T. Kenney; E-mail: margaret_kenney@berkeley.edu

(Received 8 October 2024; revised 12 March 2025; accepted 15 March 2025)

Abstract

Researchers' racial and gender identities influence their outcomes in academia and the field of political science. This letter interrogates how researcher identity affects the research process: specifically elite interview recruitment. Within an ongoing research project we embed a pre-registered audit experiment randomizing the gender of the researcher conducting outreach to estimate whether there are differences in interviews scheduled holding all other confounders constant. We find that when outreach is conducted by a woman, elites are more likely to schedule an interview. This letter contributes to our understanding of bias and inequality during the research process. In addition, our study offers a new approach to audit experiments that limits deception and wasted time for elites.

Keywords: audit experiment; gender; qualitative interviews

Edited by: Daniel J. Hopkins and Brandon M. Stewart

1. Introduction

The field of political science remains dominated by white scholars (APSA Task Force 2011), with scholars of color experiencing harassment and exclusion in the IR security field (for example) (Zvobgo and Loken 2020). Gender disparities are also prevalent, affecting women's academic success, citations, and research agendas (e.g., Hardt *et al.* 2019 and Dion, Sumner, and Mitchell 2018). In this letter, we focus on one facet of researcher identity—gender—to understand how inequities may manifest not only in outcomes for scholars, but also in the research process itself. In particular, we seek to understand how identity affects qualitative interview recruitment. Despite extensive research on how identity affects interview procedures and answers, no study has probed whether interviewer identity results in differential access to interviews (West and Blom 2017). There is anecdotal evidence that gender bias may negatively affect women's recruitment of elite interviews (Mosley 2013). However, potential bias in interview recruitment has been difficult to measure, given confounders that prevent the causal effect from being isolated.

In this letter, we embed an audit experiment within an existing research project to estimate how a researchers' gender affects recruitment of elite interviews.¹ Audit studies enable researchers to evaluate whether participants engage in discriminatory behavior in a real-world setting. In this study, each subject receives an identical email from a UC Berkeley student inviting the recipient to participate in an elite, semi-structured interview. To preserve internal validity, we use email aliases so that the students are matched on all other background characteristics, including race and academic institution.² We chose names that had been pre-tested by Elder and Hayes (2023) and were strongly associated with a particular

¹ IRB protocol 2023-05-16402 approved, Section A.6 of the Supplementary Material.

² Preregistration [here](#).

gender and a similar set of characteristics. Therefore, any non-response effects can be explained solely by gender identity.

We find that elites are more likely to schedule an interview when receiving the outreach from a woman (relative to a man). 16% of those contacted by “Mary Williams” scheduled an interview, while only 8% scheduled an interview when contacted by “Jake Miller.” These differential response rates are driven by female elites’ unwillingness to reply under the male treatment condition. Female elites had a 15% response rate to the female treatment, but just a 2% response rate to the male treatment.

This letter offers two primary contributions. First, this letter speaks to the expanding literature that seeks to understand and ameliorate bias and inequity in the field of political science. We add to this important work by interrogating how inequities may arise in the process of the research itself and the extent to which methods are accessible. Bias based on identity can hinder researchers from recruiting a large enough sample to be representative and generalizable. The differential responses by male and female elites also suggest that research may be systematically omitting certain voices because of gender bias.

In addition, this letter suggests a new and more ethical approach to audit experiments. Audit experiments have produced important results to better understand political phenomena. However, this method has been critiqued for wasting the time of respondents. We design an audit experiment nested in another research study to avoid wasted time for elites. This study also engages only in limited identity deception, rather than traditional audit experiments which often combine identity, activity, motivation, and misinformation deception. We argue that it is not necessary to abandon audit experiments, but instead we encourage scholars to think creatively about how to design audit experiments in an optimally ethical manner. While there are trade-offs in terms of power, we believe that protecting the time of subjects outweighs the potential costs of this approach. Therefore, we offer an avenue forward for the use of audit experiments in the discipline to test research questions about inequality and bias while avoiding the usual concerns with this method.

This letter proceeds in three parts. First, we outline our pre-registered expectations for how gender affects elites’ willingness to schedule an interview. Next, we describe the audit experiment design and consider the unique ethical implications. Finally, we discuss the results of the audit experiment and offer suggestions for future research.

2. Theoretical Expectations

Gender discrimination remains prevalent and well-documented across fields, despite some progress. Stereotypes about women negatively affect their standing in society and the workplace (see, for example, Catalyst 2021; Hegewisch 2013; Helmer *et al.* 2017; UNESCO 2021). In male-dominated fields in particular, there is anecdotal evidence that women experience gendered dynamics while conducting interviews and recruiting participants, potentially limiting response rates for women researchers (McDowell 1997; Mosley 2013). Therefore, we expect that elites will be more likely to respond to male-presenting names in interview outreach. Alternatively, women are perceived as “kind, sensitive, and relationship-oriented” and men as “dominant, ambitious, and achievement-oriented” which could motivate the opposite results (Heilman and Caleo 2018). Gender discrimination is not the only marginalized identity; however, this study provides a first cut with gender.

H1: Elites will be more likely to schedule an interview when receiving outreach from a male-presenting name.

We also consider which elites are most likely to respond to outreach and participate in research. Women are more likely to take on non-promotable tasks (Babcock *et al.* 2022). Therefore, we expect that female elites would be more likely to schedule an interview. This hypothesis considers whether qualitative interviewing overburdens specific groups.

H2: Female elites will be more likely to schedule an interview.

Alternatively, this relationship could be explained by women's propensity to support other young women interested in a male-dominated field. While research on women's support for other women in the workplace is somewhat mixed, Carr and Kelan (2016) illustrate that women do support and align themselves with each other in the workplace.

H3: Elites will be more likely to schedule an interview with a matched gendered-name in outreach.

3. Research Design

We conduct an audit experiment, inviting elites to participate in a 15–30 minute semi-structured interview with the research team. The researchers initiate this outreach via email with identical written text to allow for control between conditions. Audit studies are critiqued for involving significant deception, harm to future researchers in the discipline, and waste of public or private resources (Desposato 2022). This audit study did not result in any wasted time for elites. The interviews were subsequently carried out for a separate research project under an approved IRB protocol. Nesting the audit experiment in another project allows the researchers to limit deception, be sensitive to elites' time, and avoid negative effects for future researchers, while still gaining causal leverage.

We manipulate the source of the outreach: (1) male-presenting name (Jake Miller) and (2) female-presenting name (Mary Williams). We created email aliases under these names through our academic institution, and the individuals are presented as research assistants. The profiles of these researchers are matched on all other covariates that could affect willingness to respond and engage in the interview. The outreach email is identical other than its source (Section A.1 and A.11 of the Supplementary Material). The names also cue similar attributes in regard to gender, class, education and thus are not a bundled treatment (Elder and Hayes 2023).³

Finally, we block on elite gender. Given that we believe that elite gender will affect response rates, this approach ensures equal treatment assignment. In addition, this covariate adjustment increases the precision of the average treatment effect (ATE) estimate (Hartman and Huang 2024).

3.1. Sampling and Analysis

We sample participants in two ways. The interviews focused on how firm professionals and IO bureaucrats engage on economic issues in international organizations (Kenney 2025).⁴ We contacted tax professionals that engaged with IO bureaucrats on international taxation rules, compiling a list of those who recently submitted public comments. OECD bureaucrats were sampled via the OECD website and publications. We targeted bureaucrats that engage with non-state actors and organize public consultations. In total, we emailed 229 elites. As pre-registered, we excluded subjects if the emails did not deliver, resulting in a sample size of 173 elites.

We test the hypotheses with difference in means tests to find the ATE of researcher gender on likelihood of scheduling an interview. The independent variable is gender treatment and the dependent variable is whether an interview was scheduled and subsequently carried out. We use *p*-values to interpret all tests, with the criterion for claiming statistical significance being a *p*-value of 0.1 (as pre-registered).

4. Experimental Results

The results for Hypothesis 1 are the opposite of our expectations: interviewees were more likely to schedule an interview if the outreach was conducted by Mary Williams, relative to Jake Miller. In substantive terms, the response rate for Mary Williams (16%) was double that in the Jake Miller

³Section A.3 of the Supplementary Material.

⁴Section A.9 of the Supplementary Material.

Table 1. Overall effect size: Interview scheduling rate.

Treatment	Full sample	Male interviewees	Female interviewees
Mary Williams	16.3% (N = 86)	17.3% (N = 52)	14.7% (N = 34)
Jake Miller	8.0% (N = 87)	13.0% (N = 46)	2.4% (N = 41)
Gender differential	8.2% ($p = 0.099$)	4.3% ($p = 0.56$)	12.3% ($p = 0.071$)
90% CI	[0%, 16%]	[−8%, 16%]	[1%, 23%]

Note: Difference in means test. DV: interview scheduled and attended (1 if yes; 0 if no). IV: gender outreach treatment (1 if Jake; 0 if Mary). N is the number of observations in each subgroup.

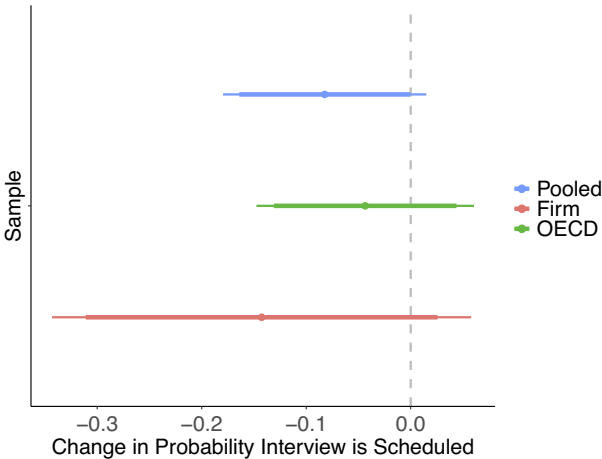


Figure 1. Effect of male vs female interview request.
Note: OLS regression results, robust SEs. IV: outreach gender treatment (1 if male; 0 if female). DV: interview scheduled and attended (1 if yes; 0 if no). Stacked 95% and 90% CIs. Full regression table in Section A.4 of the Supplementary Material.

treatment (8%). These results are significant in the full sample (see Table 1 and Figure 1) and robust to pre-registered robustness checks (Section A.5 of the Supplementary Material). The 90% CI ranges from 0% to 16%. When breaking down the results into the two substantive groups – OECD bureaucrats and firm representatives—the coefficient signs remain the same, but do not reach standard levels of significance ($p = 0.16$ and $p = .41$ respectively). These results can be explained by the limited power in the individual samples, which is a fundamental barrier in running an audit experiment that does not waste time for elites.

We offer two potential explanations for these results. First, the results seem to be driven by female elites’ unwillingness to schedule interviews when contacted by a male interviewer. The difference is pronounced, with a 14.7% interview schedule rate for Mary Williams, as opposed to 2.4% for Jake Miller, among female elites. Female elites’ unwillingness to respond to email outreach from a male contact has important implications, given that it may result in systematically omitting certain voices from qualitative research. We acknowledge, however, that we have limited power to detect heterogeneous effects.

We therefore offer a second potential explanation. We suggest that the results may be driven by women’s perception by others as non-threatening, relationship-oriented, and kind (Heilman and Caleo 2018). Therefore, elites may feel more comfortable scheduling an interview with a woman.

Contrary to our expectations in Hypothesis 2, male elites are more likely to schedule an interview (relative to female elites). Figure 2 displays these results. However, the results do not reach standard levels of significance ($p = 0.156$).

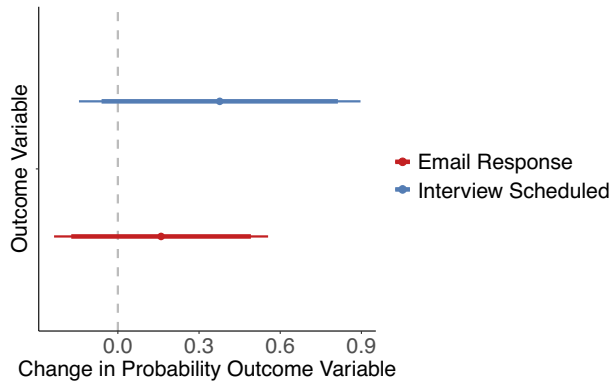


Figure 2. Effect of elite gender on response rate.

Note: Probit regression results. IV: elite gender (1 if male; 0 if female). DV: outcome variable (1 if yes; 0 if no). Stacked 95% and 90% CIs. Regression tables in Section A.4 of the Supplementary Material.

Table 2. Interaction effect of matched gender and interviewee gender.

	<i>Dependent variable</i>	
	Interview scheduled	Email response
Matched gender	0.123* (0.066)	0.163 (0.107)
Interviewee gender	0.149** (0.058)	0.165* (0.094)
Interaction effect	−0.165* (0.099)	−0.221 (0.145)
Num.Obs.	173	173
R2	0.030	0.020
* $p < 0.1$, ** $p < 0.05$		

Note: OLS regression models, robust SEs. Matched gender: 1 if elite and treatment are same gender; 0 otherwise. Interviewee gender: 1 if male; 0 if female. Interaction effect: interaction of matched gender and interviewee gender. DV: outcome variable (1 if yes; 0 if no).

Finally, we test Hypothesis 3, which predicts that when the gender is the same between the interviewer and the interviewee, there will be a greater likelihood of scheduling an interview.⁵ We find that this expectation holds: Table 2 shows that there is a positive and statistically significant relationship between matched gender and likelihood of scheduling an interview. Importantly, these results are moderated by the interviewee's gender. The interaction effect is negative and statistically significant. Therefore, the positive effect of matched gender on the likelihood of scheduling an interview is enhanced when the interviewee is female.

⁵Interaction effect not pre-registered.

5. Discussion and Conclusion

In this letter, we conduct an audit experiment to test the effect of researcher gender on elite interview recruitment. We find that elites are more likely to schedule an interview if they receive outreach from the female alias. While these results are counter to our initial expectations, we believe that it is important to consider how these biases affect the accessibility of qualitative interviewing.

Qualitative research is costly, both in time and material resources. However, interviews are relatively “costless” forms of qualitative research given the increased use of Zoom interviews. This letter aimed to understand how even the most “costless” form of qualitative research may provoke equity concerns based on researcher identity. The results suggest that qualitative interviews, even those conducted via representative sampling, may be limited by gender biases. Therefore, it is important to consider how to mitigate this bias for those who are still interested in pursuing qualitative research. We offer two recommendations.

First, collaborating with scholars of varying identities can enhance interview sample size. When representativeness is the goal, collaborating with research assistants or colleagues of other genders could increase sample size. Our results indicate that male authors in particular would benefit from collaborating with female co-authors. As co-authored work has increased in recent years, there has been considerable debate about how to value said work in a scholar’s broader research agenda. The results here imply that co-authored work may often be necessary to increase qualitative interview sample size or representativeness.

Second, researchers that anticipate limited response rates given these results should consider alternative strategies to enhance their sample size. For example, oversampling and sending additional follow-up emails could increase their sample size. These strategies may be particularly necessary when male researchers are reaching out to female interviewees.

There are several avenues for future research in this space. This experiment tests the effects of gender; however, there are many other characteristics which could cause non-response bias. Experimental manipulation of these factors could expand knowledge about the barriers faced in qualitative work. Future work should also test the mechanisms that drive these gendered effects and potential interventions to ameliorate non-response differences.

Acknowledgments. We thank Ryan Brutger, Susan Hyde, Michaela Mattes, and Jason Wittenberg for helpful feedback and advice on this research. We also thank participants at the UC Berkeley IR/CP Workshop and the Spring 2023 Major Themes in Comparative Analysis: Research Design course. Lastly, we thank the editorial team at Political Analysis and the three anonymous reviewers.

Conflicts of Interests. The authors do not have any conflicts of interest regarding this paper.

Data Availability Statement. Replication code for this article is available at Kenney and Salchak (2025). A preservation copy of the same code and data can also be accessed via Dataverse at <https://doi.org/10.7910/DVN/8APIHS>.

Supplementary Material. For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2025.10004>.

References

- APSA Task Force. 2011. “Political Science in the 21st Century.”
- Babcock, L., B. Peyser, L. Vesterlund, and L. Weingart. 2022. *The No Club: Putting a Stop to Women’s Dead-End Work*. New York: Simon and Schuster.
- Carr, M., and E. Kelan. 2016. “Femininities at Work: How Women Support Other Women in the Workplace.” In *British Academy of Management (BAM2016) Conference: Thriving in Turbulent Times*, Newcastle: British Academy of Management.
- Catalyst. 2021. “Women in the Workforce: Global” (Quick Take).
- Desposato, S. 2022. “Public Impacts From Elite Audit Experiments: Aggregate and Response Delay Harms.” *Political Studies Review* 20 (2): 217–227. <https://doi.org/10.1177/14789299211059657>
- Dion, M. L., J. L. Sumner, and S. M. Mitchell. 2018. “Gendered Citation Patterns across Political Science and Social Science Methodology Fields.” *Political Analysis* 26 (3): 312–327. <https://doi.org/10.1017/pan.2018.12>

- Elder, E. M., and M. Hayes. 2023. "Signaling Race, Ethnicity, and Gender with Names: Challenges and Recommendations." *The Journal of Politics* 85 (2): 764–770. <https://doi.org/10.1086/723820>
- Hardt, H., A. E. Smith, H. J. Kim, and P. Meister. 2019. "The Gender Readings Gap in Political Science Graduate Training." *The Journal of Politics* 81 (4): 1528–1532. <https://doi.org/10.1086/704784>
- Hartman, E., and M. Huang. 2024. "Improving Precision Through Design and Analysis in Experiments with Noncompliance." *Political Science Research and Methods* 12 (3): 557–572. <https://doi.org/10.1017/psrm.2023.38>
- Hegewisch, A. 2013. "The Gender Wage Gap: 2013; Differences by Race and Ethnicity, No Growth in Real Wages for Women." Institute for Women's Policy Research.
- Heilman, M. E., and S. Caleo. 2018. "Gender Discrimination in the Workplace." In *The Oxford Handbook of Workplace Discrimination*, edited by A. J. Colella and E. B. King, Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199363643.013.7>
- Helmer, M., M. Schottdorf, A. Neef, and D. Battaglia. 2017. "Gender Bias in Scholarly Peer Review." *eLife* 6: e21718. <https://doi.org/10.7554/eLife.21718>
- Kenney, M., and J. Salchak. 2025. "Replication Data for: Accessibility and Equity in the Research Process: Gender Bias in Elite Interview Recruitment." Harvard Dataverse. <https://doi.org/10.7910/DVN/JB0A7R>
- Kenney, M. A. T. 2025. "Firm Influence in International Organizations: The Case of OECD Tax Evasion Regulations." Working Paper.
- McDowell, L. 1997. *Capital Culture: Gender at Work in the City*. Oxford: John Wiley & Sons.
- Mosley, L. 2013. *Interview Research in Political Science*. Ithaca: Cornell University Press. <https://doi.org/10.7591/9780801467974>
- UNESCO. 2021. "UNESCO Research Shows Women Career Scientists Still Face Gender Bias." <https://www.unesco.org/en/articles/unesco-research-shows-women-career-scientists-still-face-gender-bias>.
- West, B. T., and A. G. Blom. 2017. "Explaining Interviewer Effects: A Research Synthesis." *Journal of Survey Statistics and Methodology* 5 (2): 175–211. <https://doi.org/10.1093/jssam/smw024>
- Zvobgo, K., and M. Loken. 2020. "Why Race Matters in International Relations" *Foreign Policy*.