

Supplementary Information for  
“Accessibility and Equity in the Research Process”

Margaret Kenney and John Salchak

University of California, Berkeley

April 24, 2025

# Contents

<b>A Appendix</b>	<b>2</b>
A.1 Outreach Email . . . . .	2
A.2 Coding Guide . . . . .	3
A.3 Discussion of Name Choice . . . . .	4
A.4 Coefficient Plot Regression Tables . . . . .	4
A.5 Robustness Checks . . . . .	8
A.6 Ethical and Human Subjects Principles . . . . .	8
A.7 Power Analysis . . . . .	10
A.8 Exploratory Tests . . . . .	11
A.9 Substantive Nature of Interviews and Study Population . . . . .	15
A.10 Anonymized Pre-Analysis Plan . . . . .	18
A.11 Process from Outreach to Interview . . . . .	22

# A Appendix

## A.1 Outreach Email

Dear [NAME],

I hope this email finds you well. My name is Mary Williams / Jake Miller, and I'm a research assistant on a University of California, Berkeley research project about public commenting at international organizations. The research team would like to feature your invaluable perspective in the study.

Might you be willing to discuss your participation with the public commenting process in the context of a Zoom interview sometime during the next 3 weeks? If you are willing to interview, please reply and we will share a Calendly by which we can coordinate an agreeable time.

If these times don't work for you, please reach out and I will accommodate to find a time at which our schedules sync. I would be more than happy to provide copies of my questions in advance, and/or additional information about the study. If there are others in your network that would be willing to participate, I kindly request that you share it with them.

Best wishes,

Mary Williams / Jake Miller

## A.2 Coding Guide

In this section, we discuss the coding scheme to better understand the replication data and code.

**id:** anonymous identification for the elites to preserve confidentiality

**treat:** indicates treatment assignment - coded as 1 if receiving Jake Miller treatment, 0 if receiving Mary Williams treatment.

**gender:** gender of elite; 1 if male, 0 if female.

**gender\_match:** if the treatment assignment and elite gender are the same, coded as 1; 0 otherwise.

**delivered:** if the email was delivered, coded as 1; 0 if the email bounced.

**respond:** if the elite responded to the email, coded as 1; 0 otherwise.

**schedule:** the primary outcome variable; coded as 1 if the elite scheduled and attended the interview; 0 otherwise.

**time\_reminder:** coded as 1 if the elite required a reminder before responding, 0 otherwise.

**time\_respond:** the number of days from when the email was sent until the elite responded

**text:** text of the response email for sentiment analysis

**west:** whether the elite is from a Western country

### A.3 Discussion of Name Choice

We chose the names for the email aliases specifically to avoid administering a bundled treatment. [Elder & Hayes \(2023\)](#) pretest a range of first and last names for race, gender, and other traits (such as whether the individual is competent, hardworking, intelligent, professional, warm, working class, etc.) The authors find that there is a great deal of heterogeneity in the characteristics associated with names, even between two names that signal similar race and gender ([Elder & Hayes 2023](#)). Therefore, we leverage their experimental results to identify names that (a) clearly signal gender and (b) are most similar across all other traits.

Both the names Jake and Mary were significantly more likely to be considered white; in addition, Jake was overwhelmingly identified as a male name and Mary was overwhelmingly identified as a female name. Additionally, given that the first names are relatively common in both the U.S. and international context, the gender is easily recognizable. The last names Williams and Miller also cued similar attributes. Tables [3](#) and [4](#) illustrate how these names vary across the various traits which [Elder & Hayes \(2023\)](#) test.

Among firm representatives, we used the name Mary Miller to increase internal validity. However, we had concerns about OECD bureaucrats recognizing the audit study if they had the same last name, so this was changed before moving to the larger sample to Mary Williams.

### A.4 Coefficient Plot Regression Tables

I include the full regression tables for the coefficient plots in the main text - Figures 1 and 2. Figure 1 shows the average treatment effect of outreach gender (“Mary Williams” / “Jake Miller”) on interview scheduling. The full OLS regression results in the firm elite

Table 3: First Name Balance

	Jake	Mary
African American	2.499771	2.831652
Asian	2.328687	2.405568
Hispanic	2.362136	2.406222
White	3.945939	3.678110
Democrat	3.041157	3.046837
Man	4.374605	2.135272
Republican	3.33567	3.13897
Woman	2.014949	3.953641
Aggressive	2.900788	2.695599
Assertive	3.110825	3.086516
Compassionate	3.082252	3.400587
Competent	3.330506	3.319707
Foreign	2.007969	2.399853
Hardworking	3.493659	3.359041
Honest	3.319555	3.299730
Intelligent	3.243345	3.303757
Likable	3.341856	3.366658
Middle Class	3.140405	3.112615
Professional	3.040392	3.319479
Strong	3.332094	3.168000
Traditional	3.210653	3.511391
Upper Class	3.208183	3.078484
Violent	2.610680	2.359294
Warm	3.173973	3.202834
Working Class	3.294605	3.246691
Athletic	3.448394	2.931095

sample, the OECD bureaucrat sample, and the pooled samples are displayed in Table 5.

Figure 2 shows how the gender of the elite affects their likelihood to respond to the email outreach and schedule an interview. I use probit models to test the relationship between elite gender and the two outcome variables (response / scheduling interview), controlling for treatment assignment. The full results are presented in Table 6.

Table 4: Last Name Balance

	Miller	Williams
Democrat	3.134889	3.121611
Man	2.991123	2.804673
Republican	3.130969	3.080332
Woman	3.049383	3.115233
African American	2.875238	3.076493
Aggressive	2.747843	2.802694
Asian	2.268507	2.286086
Assertive	3.093104	3.067163
Athletic	3.073214	3.191895
Competent	3.250346	3.324387
Foreign	2.398375	2.470902
Hardworking	3.275995	3.318650
Hispanic	2.226398	2.313378
Honest	3.192190	3.259648
Intelligent	3.237538	3.330921
Likable	3.223820	3.285796
Middle Class	3.173692	3.158950
Professional	3.261700	3.323377
Strong	3.110046	3.150186
Traditional	3.164427	3.167270
Warm	3.076227	3.207768
White	3.324333	3.129970
Working Class	3.210162	3.313702
Compassionate	3.199746	3.242694
Upper Class	3.152962	3.183218
Violent	2.543100	2.557482

Table 5: Average Treatment Effect: Interview Scheduled

	<i>Sample</i>		
	Pooled	Firm	OECD Bureaucrat
Jake Miller	−0.082* (0.050)	−0.143 (0.102)	−0.044 (0.053)
Num.Obs.	173	57	116
R2	0.016	0.032	0.006

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Note:* This table shows the OLS regression results for the average treatment effect of outreach gender on whether the elite scheduled an interview. As pre-registered, the independent variable is the outreach gender – 1 if “Jake Miller” and 0 if “Mary Williams.” The dependent variable is whether the elite scheduled and attended an interview, taking the value of 1 if yes, 0 if not. I test these results in the pooled sample of firm representatives and OECD bureaucrats (Model 1), amongst only the sample of firm representatives (Model 2), and amongst only the sample of OECD bureaucrats (Model 3).

Table 6: Elite Interviewee Gender and Response Rate

	<i>Dependent Variable</i>	
	Interview Scheduled	Email Response
Elite Gender	0.376 (0.265)	0.160 (0.202)
Treatment Assignment	−0.414 (0.257)	−0.290 (0.199)
Num.Obs.	173	173

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Note:* This table shows the probit regression results to understand how elite gender affected response rate. The independent variable is the elite gender – 1 if male and 0 if female. The first model uses the dependent variable of whether the elite scheduled and attended an interview, taking the value of 1 if yes, 0 if not. The second model uses the dependent variable of whether the elite responded to the email, taking the value of 1 if yes and 0 if not.



## A.5 Robustness Checks

In the pre-analysis plan, we discussed running probit models as a robustness check of the results of Hypothesis 1. Table 7 presents the results of this robustness check. The results are robust to this specification: elites are more likely to schedule an interview when contacted by Mary Williams, rather than Jake Miller ( $p = 0.0986$ ). The results for the email response outcome measure are also in the same direction but continue to fail to reach standard levels of significance ( $p = 0.131$ ).

Table 7: Average Treatment Effect: Gender Outreach

	<i>Dependent Variable</i>	
	Interview Scheduled	Email Response
Jake Miller	-0.419* (0.254)	-0.299 (0.199)
Num.Obs.	173	173

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

*Note:* This table shows the pre-registered probit regression results to show robustness. The independent variable is the outreach gender – 1 if “Jake Miller” and 0 if “Mary Williams.” The first model uses the dependent variable of whether the elite scheduled and attended an interview, taking the value of 1 if yes, 0 if not. The second model uses the dependent variable of whether the elite responded to the email, taking the value of 1 if yes and 0 if not.

In addition, we use a difference in proportions test to bolster our results.<sup>6</sup> We reject the null hypothesis, at the  $p = 0.0973$  level, that the proportions of scheduled interviews is the same across the two treatment groups.

## A.6 Ethical and Human Subjects Principles

In this section, we consider the ethical implications of the project and its compliance with APSA Principles and Guidance for Human Subjects Research. Importantly, because these interviews were carried out for another research project, there was no wasted time for

---

<sup>6</sup>We thank an anonymous reviewer for this suggestion.

Table 8: Difference in Proportions Test

Mary Williams		Jake Miller	
Interviews Scheduled	14	7	
Sample Size	86	87	
X-squared	P-Value	Confidence Interval (Low)	Confidence Interval (High)
2.7487	0.0973	-0.1791	0.0144

elites. We believe that preventing wasted time offers additional benefits beyond avoiding using public or private resources to conduct the study. This characteristic also preserves the realism of the treatment. If there were not interviews subsequently scheduled, subjects might suspect that they are part of an audit study. Furthermore, being conscious of elites' time also avoids alienating elites from research projects.

This study did involve limited deception. Following previous audit studies, the participants will not know that they will be a part of this specific research project, so as to simulate a realistic response to the outreach. However, the email outreach that they will receive is typical of emails that they receive on a regular basis with the only manipulation being the name of the researcher (female / male name). Therefore, there are not significant risks to deceiving the subjects and deception is necessary to understand the extent of gender discrimination. The subjects will not be debriefed. A debrief is not necessary given that the outreach is similar to what they receive on a regular basis, and will not result in any adverse effects for the subject. In addition, with the careful attention given to confidentiality and with only aggregate results shared in future publications, there are very limited risks to the participants themselves. Finally, given that we hope to continue our relationships with these elites in our future research, debriefing may result in lower levels of trust in future work with them which we hope to avoid.

Given that respondents did not know that they were being studied, and subsequently could not consent to participation, we obtained a consent waiver from IRB. This consent

waiver was justified given that there was no more than minimal risk of harm to the subjects. The experimental treatment is in line with normal stimuli that elites experience in their daily lives. In addition, the audit experimental set up necessitates a consent waiver in order to observe how the subjects will react in a realistic setting (without knowing that the results are part of the research). The experimental outreach is the same as what these professionals receive regularly, with the only experimental stimuli being the name change of the person contacting them. Therefore, not receiving a debrief will not adversely affect their rights / welfare, especially given that the interviews are in fact being carried out and therefore do not waste time for the subjects.

## A.7 Power Analysis

We hoped to obtain 0.95 power to detect a small effect size at the standard 0.05 alpha error probability. We based our pre-registration power calculations off an alpha error probability of 0.05, rather than the 0.1 threshold, as we wanted to ensure that we had an adequate sample size for the study without wasting the time of more elite interviewees than needed. Using the more conservative alpha error probability in the power calculations allowed us to justify recruitment of a larger set of participants. A larger sample size in this case was important given that we did not have a clear sense of what the effect size would be in the entire sample –we had an effect size amongst firm representatives that was used in the power calculation, but it was unclear whether that effect size would generalize. Therefore, in an *ex ante* power analysis, we find that the required sample size is a minimum of 98 (49 / group) for each set of elites, using the effect size from the pretest data (made up of firm representatives). Before pre-registration of the OECD bureaucrat population, we conducted a power analysis based on the effect size among firm representatives. Based on these results, it appeared possible to conduct this experiment and have adequate power without pooling the two groups. However, the effect size was significantly lower amongst OECD bureaucrats

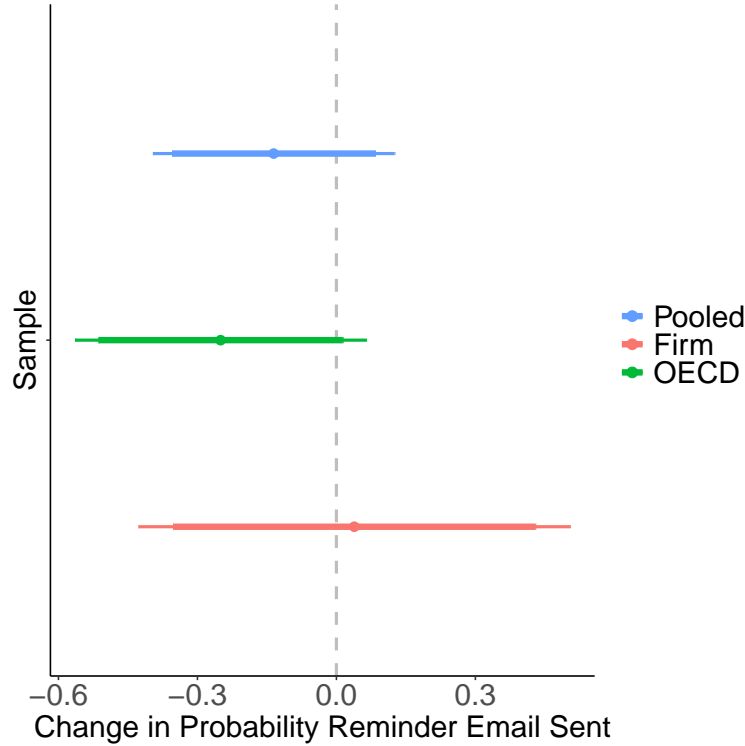
(due to a lower response rate) and thus we ultimately chose to pool the two groups for our main results in order to have adequate power. The response rates varied between the data used for the power analysis (firm representatives) – 20% of those contacted scheduled an interview – and the OECD bureaucrats – 10% of those contacted scheduled an interview. Therefore, to achieve adequate power, we had to pool the data from OECD bureaucrats and firm representatives.

## A.8 Exploratory Tests

To leverage all of the data compiled in this audit experiment, we also conduct additional exploratory tests that were not pre-registered. First, we consider whether the gender of the person conducting outreach affected response times. We use two measures: (1) whether the person needed a reminder one week after to respond and (2) the time until response measured in days. These results are again not statistically significant (see Figure 3 and 4 respectively). The coefficients for the time to response seem to suggest that the response times are longer among those who received outreach from Mary Williams (though again, this is insignificant).

Next, we consider whether there were substantive differences in the email responses received across treatment groups. Qualitatively, no one explicitly questioned the credentials or knowledge of either research assistant. However, it is possible that the email responses differed in more subtle ways. We consider two tests of the content of email responses across treatment groups. We caution, however, that there is very limited statistical power in these analyses, given that there were only 56 responses, but we include these analyses given the potential interest of readers. First, we conduct a sentiment analysis to test whether the gender of the outreach contact affected the tone of email responses. Table 9 shows the results. The results are insignificant and the coefficients are substantively small. Therefore, there do not seem to be any major differences in the tone of responses based on gender.

Figure 3: Effect of Male vs Female Interview Request



*Note:* OLS regression results, robust SEs. IV: outreach gender treatment (1 if male; 0 if female). DV: reminder email required (1 if yes; 0 if no). Stacked 95% and 90% confidence intervals.

Next, we also consider whether more respondents asked questions to Mary, relative to Jake, in their email responses. We use a binary variable coded as 1 if the respondent asked a question before agreeing to schedule an interview, and 0 otherwise. Table 10 shows the results. There is no statistically significant difference in whether respondents asked questions between treatment groups.

Finally, given that there are some emails that were unsuccessfully delivered as not all email addresses collected were still active, we wanted to ensure that these delivery rates were not significantly related to the gender of the elite or their treatment assignment.<sup>7</sup> There are not any statistically significant relationships between the likelihood of delivery and gender

---

<sup>7</sup>As described in the pre-analysis plan, we dropped all respondents whose emails did not deliver given that they did not receive the treatment and could not respond to the request.

Table 9: Sentiment Analysis

	(1)
Jake Miller	0.013 (0.044)
Num.Obs.	56
R2	0.002
* $p < 0.1$ , ** $p < 0.05$ , *** $p < 0.01$	

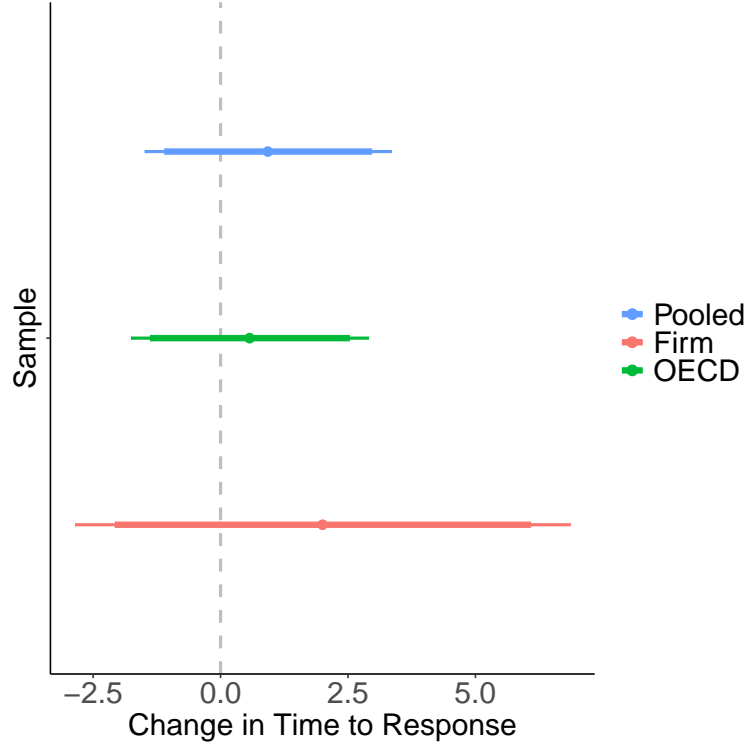
*Note:* This table illustrates whether outreach gender affected the sentiment of email responses. We use OLS regression to test whether there is a relationship. The independent variable is the treatment assigned – 1 if Jake Miller and 0 if Mary Williams. The dependent variable is the sentiment of the email response; We use the Harvard-IV dictionary to classify positive and negative words in the SentimentAnalysis package. Higher values indicate more positive email responses, and lower values indicate more negative email responses.

Table 10: Number of Questions

	(1)
Jake Miller	-0.042 (0.115)
Num.Obs.	56
R2	0.002
* $p < 0.1$ , ** $p < 0.05$ , *** $p < 0.01$	

*Note:* This table illustrates whether outreach gender affected the number of questions asked in email responses. We use OLS regression to test whether there is a relationship. The independent variable is the treatment assigned – 1 if Jake Miller and 0 if Mary Williams. The dependent variable is a binary variable, taking the value of 1 if the respondent asked a question before agreeing to interview and 0 otherwise.

Figure 4: Effect of Male vs Female Interview Request



*Note:* OLS regression results, robust SEs. IV: outreach gender treatment (1 if male; 0 if female). DV: time to response, measured in days. Stacked 95% and 90% confidence intervals.

or treatment. In addition, the rates of dropped emails are very similar across gender and treatment group. See Tables 11 and 12.

Table 11: Balance Test: Email Delivery

Treatment	Elite Gender	Count
0	0	19
0	1	13
1	0	10
1	1	14

Table 12: Email Delivery Balance

	Model 1	Model 2
Elite Gender	0.063 (0.057)	
Jake Miller		0.055 (0.057)
Num.Obs.	229	229
R2	0.005	0.004
* p < 0.1, ** p < 0.05, *** p < 0.01		

## A.9 Substantive Nature of Interviews and Study Population

The substantive focus of the interviews was how and why firms choose to engage with bureaucrats (and vice versa) at international organizations on issues of technical economic policy. We chose this substantive area to test differences in interview recruitment across researcher identity given that the topic was neutral, with no explicit connection to gendered or sensitive issues. If anything, the substantive topic biased against finding that women researchers were more effective. The male dominated nature of technical economic policy motivated our predictions that women researchers would have more difficulty in recruiting interview participants. Therefore, the substantive nature of the interviews should not bias the results toward favorable recruitment for women researchers.

We acknowledge that the use of a neutral substantive topic to test these expectations could circumscribe the preferences toward interviewing with a woman researcher – as compared to gendered or sensitive issues. These findings, therefore, may perhaps be an underestimate of preferences toward women researchers on more sensitive or gendered issues. We do not think that the present experiment can speak to the effect sizes in these specific cases, and suggest future research to better understand these unique situations in which interviews focus on a gendered substantive topic.



Next, we describe the populations under study: firm representatives and OECD bureaucrats. To ensure anonymity of subjects, we are unable to offer specific details about the individuals that were included in this study. We provide broad descriptions that offer as much information as possible about their roles, substantive expertise, and geographic locations.

We interviewed firm and industry representatives to understand why firms choose to engage (or not) with OECD bureaucrats on international tax regulations, in addition to learning about their specific commenting and lobbying strategy. The public comments submitted to the OECD are signed by the individual who prepared them (for example, the Vice President of Global Tax at Amazon). [M. A. T. Kenney \(2025\)](#) therefore compiled a list of the individuals at firms who participated in the ten most recent public comment sessions, in the hopes of acquiring a sample that is the most up to date. These individuals were tax experts – in many cases, attorneys or accountants – that either coordinated internal tax policy for an individual firm or represented the industry or business association on tax issues. This population included firm representatives from a range of geographic locations: Africa, Asia, North America, and Europe. In addition, the population consisted of representatives of large multinational corporations, small firms, and industry / trade associations in high, middle, and low income countries.

Next, we conducted semi-structured interviews of OECD bureaucrats to understand their perspective in coordinating stakeholder engagement and reviewing public comments. To conduct outreach, [M. A. T. Kenney \(2025\)](#) collected the names and emails of OECD bureaucrats that (1) had their information publicly available and (2) appeared to engage actively in stakeholder engagement. This population included bureaucrats across a broad range of directorates – all of which focused broadly on technical, economic policy.

We consider these populations to be elites, as contrasted by the mass public, given that they occupy a privileged role in the policymaking process. Firm representatives that work on

international taxation are part of an elite epistemic community and transnational network ([Christensen 2021](#)). Therefore, their niche subject knowledge offers them an opportunity to have a direct effect on international tax policy. OECD bureaucrats have a more direct hand in creating international regulatory policy, reflecting a broader trend of international organization rulemaking on issues that were previously solved in the domestic context. OECD bureaucrats write policy documents and international regulations, informed by their engagement with non-state actors. Both populations of elites work on technical economic policy, and their roles require advanced degrees.

## A.10 Anonymized Pre-Analysis Plan

### Hypotheses

1: Elites will be more likely to schedule an interview with a male-presenting name in outreach (relative to a female-presenting name).

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

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

### Design Plan

#### *Study type*

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

#### *Blinding*

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

*Is there any additional blinding in this study?*

No response

#### *Study design*

We will conduct an audit study, inviting elites to participate in a 15-30 minute semi-structured interview with a member of the research team. The researchers initiate this outreach via email with identical written text to allow for control between conditions. In addition, 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.

This process allows the researchers to limit deception and negative effects for future researchers, while still gaining casual leverage over the situation at hand. Audit studies are critiqued for involving significant deception, harm to future researchers in the discipline, and waste of public or private resources (Desposato 2022). Therefore, the design reflects the desire to limit the detrimental effects of audit studies while maintaining experimental control.

We will 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 helping the study team.

The profiles of these researchers are matched on all other covariates that could affect willingness to respond and engage in the interview (e.g. age, institutional affiliation, race). Additionally, the outreach email is identical other than its source. These names are consistent with Elder & Hayes (2023)’s recommendations the names cue similar attributes in regard to gender, class, education and thus do not provide a bundled treatment. Both the names Jake and Mary were significantly more likely to be considered white; in addition, Jake was overwhelmingly identified as a male name and Mary was overwhelmingly identified as a female name. Additionally, given that the first names are relatively common in both the U.S. and international context, the gender is easily recognizable.

It is possible that names cue gender differently across international contexts. After receiving the results, we will explore whether the results are different based on whether interviewees are located in the West (or not). However, the majority of participants in this process are from Western Europe or the United States. We expect that these names will be familiar enough to cue gender correctly. Next, we will block on gender of the interviewee; given that we believe that we believe that interviewee gender will affect the way that they respond to male / female names in outreach, this approach allows us to have equal numbers of treatment

/ control assignment between the two groups.

### *Randomization*

We use simple randomization, with pre-treatment blocking on interviewee gender to ensure approximately equal assignment to treatment and control across male and female interviewees.

## **Sampling Plan**

### *Existing Data*

Registration prior to creation of data

### *Explanation of existing data*

No response

### *Data collection procedures*

Given that the interviews focus on the bureaucrat's experience with the public commenting process, we collected all publicly available emails for OECD bureaucrats that appeared to work on issues that had subsequent public consultations.

### *Sample size*

The sample size will be 166 individuals.

### *Sample size rationale*

The power analysis indicated that we need a minimum sample size of 93 to detect the effect size at 0.05 significance level. Therefore, we use the maximum number of publicly available emails (166) given the potential for emails to bounce, etc.

## **Stopping rule**

No response

## **Variables**

### *Manipulated variables*

We randomize the gender of the outreach source – the outreach email will either come from a female-presenting name (Mary Williams) or a male-presenting name (Jake Miller).

### *Measured variables*

The outcome measure / dependent variable is whether the elite schedules an interview. We will also measure whether the elites responded to the email.

## **Analysis Plan**

### *Statistical models*

We use OLS regression. The independent variable (treatment) and dependent variable (interview scheduled) will be included. There will be no control variables in the confirmatory tests. We will use p-values to interpret all of the tests described above, with the criterion for claiming statistical significance being a p-value of 0.1.

### *Transformations*

No response

### *Inference criteria*

P values with the criterion for claiming statistical significance being a p value of 0.1

### *Data exclusion*

We will exclude interviewees whose emails bounce back, as this implies that the individual no longer works in the role and thus did not receive the treatment.

### *Missing data*

No response

### *Exploratory analysis*

It is possible that names cue gender differently across international contexts. After receiving the results, we will explore whether the results are different based on whether interviewees

are located in the West (or not). However, the majority of participants in this process are from Western Europe or the United States. We expect that these names will be familiar enough to cue gender correctly.

## **A.11 Process from Outreach to Interview**

In this section, we provide additional information about engagement with respondents after the initial outreach email. The email aliases of Jake Miller and Mary Williams engaged with respondents up until the interview took place. Mary Williams / Jake Miller sent one follow up email to those who did not respond after a week. These follow up emails were standardized across treatment groups. The email aliases also were used to answer interviewee questions and to provide respondents with the scheduling link and IRB consent form. The answers to questions were standardized across the Jake Miller / Mary Williams treatment conditions. To provide even greater control across treatment conditions, one author answered all respondent questions through the email aliases. The most common question from respondents was to have more information about the study and to have copies of the interview questions (90%). The other subset of questions was how the research team had identified their name (10%).

The majority of email responses that did not convert to an interview consisted of individuals that emailed to politely decline the interview request (52%). There was also a subset of respondents that did not reply after responding with an initial willingness to participate (34%). These respondents did not ask additional questions that could have influenced their participation. The remaining individuals responded with a question, and ultimately did not schedule an interview (8%).

We consider whether there is differential attrition between treatment groups. First, we test whether there are differences across treatment groups in those who responded with a

question, but did not ultimately schedule an interview. The number of respondents who followed up with a question, and ultimately chose not to schedule, was similar across treatment groups. There were 8 people who followed up with a question to Mary Williams – and 5 people for Jake Miller – but did not ultimately choose to schedule an interview. There is not a significant difference between these proportions. In addition, the attrition biases against the main results in which Mary Williams receives more interviews scheduled than Jake Miller.

Table 13: Attrition after Email with Question

Treatment	Number
Jake Miller	5
Mary Williams	8

When the interviewee scheduled a meeting time via the Calendly link, interview subjects received a confirmation email with a Zoom link that stated that they would be meeting with a member of the UC Berkeley research team, rather than stating the PIs’ names. We chose to frame the aliases as research assistants in order to allow their exit from the research process to happen seamlessly and without raising suspicions that participants were engaged in an audit study. It is common practice for research assistants to work in a scheduling capacity / provide additional information about a research study. This design therefore allowed for the aliases to be engaged with research subjects up until the interview itself without raising suspicions about being under study. None of our participants asked about the research assistants’ role during the interview or afterward.

## Author Biographies

Margaret Kenney is a PhD Candidate in Political Science at the University of California, Berkeley.



John Salchak is a PhD Student in Political Science at the University of California, Berkeley.