sociological science

Supplement to:

Rivera, Lauren A., Katherine Weisshaar, and András Tilcsik. 2024. "Disparate Impact? Career Disruptions and COVID-19 Impact Statements in Tenure Evaluations." Sociological Science 11: 626-648.

ONLINE SUPPLEMENT

Disparate Impact? Career Disruptions and COVID-19 Impact Statements in Tenure Evaluations

Part 1: Additional details on survey design and methods

Name selection

We selected candidates' names (Jennifer and Michael Nelson) to signal gender and to imply that the candidate was likely white and in an age range typical for a tenure candidate (mid-30s). We selected possible last names from the U.S. Census' list of common last names among white individuals (Comenetz 2016) and chose first names from the most common male and female names of children born in 1987 (i.e., the birth year of a 35-year-old candidate) (Social Security Administration 2022). We conducted a pretest of 600 respondents on the Prolific survey platform to inform our selection of the final names. Respondents evaluated name choices in terms of perceived gender, race/ethnicity, social class, education, age, and immigration status. Results from the pretest for the two selected names are provided in Table S10. The results show that both names signaled our intended characteristics, and there were no statistically significant differences in perceptions of the above characteristics between Jennifer and Michael Nelson based on the pretest findings (Gaddis 2017, 2018).

Evaluation Task and Survey Order

The survey questions included four sections: (A) tenure recommendation; (B) evaluation of the candidate's current record; (C) evaluation of the candidate's future record; and (D) social and demographic questions about the respondent. The order of sections A and B were randomized—independent of the experimental treatment—such that a subset of respondents answered questions related to their recommendation for tenure (A) followed by evaluations of the candidate's record (B), and the other subset answered questions about the candidate's record (B) followed by their tenure recommendation (A). Sections C and D always followed sections A and B and were not randomized in order. Randomizing the order of sections A and B avoided order effects, i.e., the possibility that participants respond to the survey stimuli differently when they answered evaluation questions before (or after) the main dependent variable (McFarland 1981).

Aside from the randomly assigned conditions (COVID-19 impact statement and candidate gender) and the order of sections A and B, respondents viewed identical instructions and materials. After consenting to take the survey, respondents were told that they are members of the University Personnel Committee (UPC) at an R1 university, which considers candidates for promotion and tenure. Participants were told that they would receive a summary of evaluations from a subcommittee that reviews candidates' published work, CVs, external letters, and teaching evaluations. They were instructed to review an Assistant Professor candidate for tenure in a laboratory science field. They were told that candidates were given the option of submitting a COVID-19 Impact Statement this year, which would be included as a separate document if the candidate provided a statement.

Survey Population and Sampling
The sample for our survey consisted of tenured full professors in biological and physical sciences (BPS) within the top 100 national universities as ranked by U.S. News and World Report in 2022 (US News & World Report 2022). 35 of the top 100 universities were randomly selected for inclusion in the sample. We focused on a sample of the top 100 national universities ranked by U.S. News and World Report (US News & World Report 2022) because tenure-track faculty in selective, research-intensive universities face "ideal worker norms" conflicting with caregiving, given the focus on publishing and conducting extensive research pre-tenure (Blair-Loy and Cech 2022; Cech and Blair-Loy 2014). These are, therefore, the sites in which research disruptions from the pandemic could have large impacts on tenure decisions.

Survey Invitations and Pilot Study
The faculty in our sample were contacted by email and invited to participate in a survey on faculty evaluations. We did not offer compensation for the survey, given past research indicating the limited effectiveness of incentives on survey response rates among faculty (Dykema et al. 2013; Sauermann and Roach 2013); respondents were told that the survey was confidential and voluntary and that they could opt out of reminders or discontinue the survey at any point.

Before launching the main survey, we randomly selected 75 participants for a pilot survey to refine our questionnaire and experimental design; these respondents were excluded from participation in our main survey. Of the 75 invited to participate in the pilot, 18 responded to the survey in its entirety, yielding a 24% response rate. The pilot survey presented respondents with the main experimental task and evaluation of the tenure candidates and additionally included questions about the perceived realism of the evaluation task. The vast majority of respondents indicated that the evaluation task was "completely realistic" or "realistic."

Sample Characteristics
We launched the survey in February 2023 and, following our pre-registered data collection plan,[1] closed it after six weeks of data collection. Respondents who did not complete the full survey were excluded from the analysis. Of the 2,905 invited survey participants, 602 faculty members completed the survey. This yielded a response rate of about 20.7%, which is comparable to response rates in recent studies with faculty participants (e.g., Moss-Racusin et al. 2012; Williams and Ceci 2015).

Dependent variable coding
All survey evaluation measures of the fictitious tenure candidate followed ordinal scales with higher values indicating higher or more favorable evaluations, as described below. The primary dependent variable indicating the promotion to tenure recommendation was a 1-6 scale ranging from 1 ("Definitely do not promote") to 6 ("Definitely promote"). The extent of advocacy for tenure ranged from 1 ("Not at all") to 7 ("To a very great extent"). Evaluations of the candidate's productivity, research quality, and scholarly impact are on scales from 1 to 5 ("Poor" to "Excellent"). Commitment is measured in terms of how committed the candidate is compared to similar faculty members, on a scale from 1 ("Much less committed") to 7 ("Much more committed"). The estimated hours worked are measured in 10-hour increments from "Under 20 hours per week" to "Over 80 hours per week." Future impact and future productivity questions

[1] https://aspredicted.org/blind.php?x=SDN_6P7

were on a 1-5 scale, with 1 indicating "Much less" and 5 "Much more" productivity/impact compared to similar scholars. Finally, the likelihood of promotion to full professor in the next five years was on a 1-4 scale, from 1 "Highly unlikely" to 4 "Highly likely." Since each faculty respondent viewed one randomly assigned candidate profile in the survey experiment, each evaluation question was measured once per respondent.

Variables were treated as continuous for use in OLS regression models, as described below. For the analysis in Figure 3, the above dependent variables were standardized to a mean of 0 and standard deviation of 1 to allow for comparisons of the experimental treatment in standardized units.

Independent variable coding

The experimental conditions were coded as follows. The COVID-19 impact statement condition was coded as a categorical variable with "No statement" as the reference category (= 0); 1 = Lab impact statement, and 2 = Childcare impact statement. Candidate gender is dichotomous (1 = female candidate; 0 = male candidate). Respondent characteristics such as gender (1 = woman), parental status (1 = parent), and discipline type (1=biological sciences; 0 = physical sciences) were coded dichotomously. Respondents' university ranking was measured as a continuous variable, with lower values indicating more prestigious universities.

Statistical Analysis

We followed our pre-registration plan to conduct statistical analyses.[2] The main statistical analyses involved three OLS regressions. For each dependent variable described above, we conducted separate regressions first with the categorical COVID-19 impact statement experimental treatment as an independent variable (pooling candidate gender), then with candidate gender as an independent variable (pooling impact statements), and then with the COVID-19 impact statement treatment interacted with candidate gender. Table S3 provides the full results of these regressions.

To compare the statistical significance of the effects of experimental treatments and variation in the effects, we rely on a threshold of $p<.05$ for statistical significance. In addition to examining statistical significance in OLS regression models, we used the "margins" command in Stata (Williams 2012) to compare the average marginal effects of independent variables from the regression results (for example, comparing the marginal effects of the COVID-19 impact statement within candidate gender, displayed in Figure 3 and Table S6). All graphs displaying estimates include error bars, which, as indicated in the figure captions, present either the standard errors of estimates or the 95% confidence intervals derived from the marginal effects and regression models.

In addition to the main statistical analysis, our pre-registration plan described analyses that explored respondent variation in treatment effects. To conduct these analyses, we again used OLS regressions, and this time, we interacted the COVID-19 impact statement experimental treatment with respondent characteristics (e.g., gender, parental status, university prestige). Table S7 displays these regression results, and Figures S2-S4 present the results graphically. Our exploratory analyses of interaction effects included variation across respondents' field,

[2] https://aspredicted.org/blind.php?x=SDN_6P7

public/private university, caregiving status, personal COVID-19 risk, and more, but we limit the presentation of results in the Supplementary Materials to university prestige and respondents' gender and parental status (Figures S2-S4 and Table S7), given the general lack of statistical significance of other interactions. Marginal effects estimated across values of respondents' university ranking are displayed graphically in Figure S2, derived from Model 2 in Table S7 and presented in Table S8, with 95% confidence intervals estimated from the "margins" command in Stata.

We conducted similar OLS regressions for the LIWC output measures of interest, examining the main effect of the COVID-19 impact statements on language use in the open-ended responses (see Table S9 and Figure S1). While we examined a range of LIWC measures, such as certitude and tentative language, most measures did not reveal statistically significant variation in effects across experimental treatments. We present two measures that were statistically significantly different across COVID-19 impact statement treatments ($p<0.05$): "time" and "focus on past." The regression results for these measures are presented in Table S9 and displayed graphically in Figure S1.

Part 2: Survey Materials

The first Summary document below is for “Jennifer Nelson” and the second is for “Michael Nelson.” Directly following the Summary document, respondents who were randomly assigned one of the COVID-19 impact statement conditions viewed one of the following impact statements. Respondents assigned to “no statement” did not view any statement. The first statement displayed below is the childcare impact statement, followed by the lab impact statement.

SUMMARY OF EVALUATION OF TENURE CASE
CANDIDATE: DR. JENNIFER NELSON
REPORT FOR THE UNIVERSITY PERSONNEL COMMITTEE

1. SUMMARY

Dr. Jennifer Nelson is an Assistant Professor at the university and is being considered for promotion to Associate Professor with tenure. The subcommittee's overall summary of the case's strengths and weaknesses is below.

2. RESEARCH

Strengths
Overall, the subcommittee viewed as strengths of the case the systematic and programmatic nature of Dr. Nelson's research agenda. We were especially impressed by the originality and rigor of certain research articles. The external letter writers also praised the quality of some of Dr. Nelson's papers and noted that she has a positive reputation in the field. Her teaching evaluations and CV indicate that she is a competent teacher and an active departmental citizen.

Weaknesses
However, there were also weaknesses of the case. While Dr. Nelson has successfully published in well-regarded journals and has procured external funding, there were some concerns about productivity, especially the quantity of published work. The subcommittee believed that, while the number of publications and ongoing projects was around the ballpark of what is typically expected for tenure, the case is close to the threshold and is not clear-cut.

Summary of External Research Evaluations
The external reviewers differ in their overall assessment of the research record. Seven external reviewers were asked to evaluate the quality and impact of the candidate's scholarly achievements. The letters echoed similar strengths and weaknesses as the university subcommittee. Several external reviewers praised Dr. Nelson's reputation in her subfield as well as the quality of her research, which were both viewed as very good. Paper 5 was described as a particularly important contribution. However, letter writers differed in their assessments of the contribution of other articles (e.g., Paper 3). In addition, several letter writers expressed concerns about Dr. Nelson's volume of research, with some questioning whether the quantity of published work was sufficient for attaining tenure. Selected quotations are included below:

> Reviewer (A): "Dr. Jennifer Nelson is the whole package – a strong researcher, an accomplished teacher, and a good departmental citizen. I look forward to following her career."
>
> Reviewer (B): "While Dr. Nelson produces solid research, my main concern is whether she has been sufficiently productive to date to warrant tenure."
>
> Reviewer (C): "You asked me to assess Dr. Nelson's research productivity. I typically measure productivity by both the quality and quantity of scholarship. While many of her papers are strong, I find the record to be light."
>
> Reviewer (D): "Dr. Nelson has an eye for interesting problems and much of her work demonstrates theoretical and methodological sophistication. Paper 5 is a clear demonstration of these strengths."

Reviewer (E): “Some of the papers (e.g., Paper 5) are excellent, but I was less impressed with some of the other articles (e.g., Paper 3).”

Reviewer (F): “A cohort analysis shows that Dr. Nelson lags behind other researchers who graduated in the same year in terms of the number of publications and citations.”

Reviewer (G): “I am enthusiastic about the contributions Dr. Nelson’s research makes to her subfield.”

3. TEACHING

Dr. Nelson has been a valuable contributor to teaching at the undergraduate and graduate levels. She has taught a mix of introductory and elective courses. These courses were generally well-received. Teaching evaluations indicate an effective lecture style and solid management of the overall classroom environment. She has also advised a number of senior thesis students and served on several dissertation committees.

4. SERVICE

Dr. Nelson has been an active participant in departmental activities and has shown a willingness to serve the broader university community as well. Our summary assessment is that she has made service contributions at the expected level.

SUMMARY OF EVALUATION OF TENURE CASE
CANDIDATE: DR. MICHAEL NELSON
REPORT FOR THE UNIVERSITY PERSONNEL COMMITTEE

1. SUMMARY

Dr. Michael Nelson is an Assistant Professor at the university and is being considered for promotion to Associate Professor with tenure. The subcommittee's overall summary of the case's strengths and weaknesses is below.

2. RESEARCH

Strengths
Overall, the subcommittee viewed as strengths of the case the systematic and programmatic nature of Dr. Nelson's research agenda. We were especially impressed by the originality and rigor of certain research articles. The external letter writers also praised the quality of some of Dr. Nelson's papers and noted that he has a positive reputation in the field. His teaching evaluations and CV indicate that he is a competent teacher and an active departmental citizen.

Weaknesses
However, there were also weaknesses of the case. While Dr. Nelson has successfully published in well-regarded journals and has procured external funding, there were some concerns about productivity, especially the quantity of published work. The subcommittee believed that, while the number of publications and ongoing projects was around the ballpark of what is typically expected for tenure, the case is close to the threshold and is not clear-cut.

Summary of External Research Evaluations
The external reviewers differ in their overall assessment of the research record. Seven external reviewers were asked to evaluate the quality and impact of the candidate's scholarly achievements. The letters echoed similar strengths and weaknesses as the university subcommittee. Several external reviewers praised Dr. Nelson's reputation in his subfield as well as the quality of his research, which were both viewed as very good. Paper 5 was described as a particularly important contribution. However, letter writers differed in their assessments of the contribution of other articles (e.g., Paper 3). In addition, several letter writers expressed concerns about Dr. Nelson's volume of research, with some questioning whether the quantity of published work was sufficient for attaining tenure. Selected quotations are included below:

> Reviewer (A): "Dr. Michael Nelson is the whole package – a talented researcher, an accomplished teacher, and a good departmental citizen. I look forward to following his career."
>
> Reviewer (B): "While Dr. Nelson produces solid research, my main concern is whether he has been sufficiently productive to date to warrant tenure."
>
> Reviewer (C): "You asked me to assess Dr. Nelson's research productivity. I typically measure productivity by both the quality and quantity of scholarship. While many of his papers are strong, I find the record to be light."
>
> Reviewer (D): "Dr. Nelson has an eye for interesting problems and much of his work demonstrates theoretical and methodological sophistication. Paper 5 is a clear demonstration of these strengths."

Reviewer (E): “Some of the papers (e.g., Paper 5) are excellent, but I was less impressed with some of the other articles (e.g., Paper 3).”

Reviewer (F): “A cohort analysis shows that Dr. Nelson lags behind other researchers who graduated in the same year in terms of the number of publications and citations.”

Reviewer (G): “I am enthusiastic about the contributions Dr. Nelson’s research makes to his subfield.”

3. TEACHING

Dr. Nelson has been a valuable contributor to teaching at the undergraduate and graduate levels. He has taught a mix of introductory and elective courses. These courses were generally well-received. Teaching evaluations indicate an effective lecture style and solid management of the overall classroom environment. He has also advised a number of senior thesis students and served on several dissertation committees.

4. SERVICE

Dr. Nelson has been an active participant in departmental activities and has shown a willingness to serve the broader university community as well. Our summary assessment is that he has made service contributions at the expected level.

OPTIONAL COVID-19 IMPACT STATEMENT FROM THE CANDIDATE

I have two young children, both of whom require constant supervision because of their young ages. Due to local protocols, my children's daycare was closed completely from March 2020 until March 2021. I was unable to use a babysitter during this time because of COVID risks. During this first year of the pandemic, I estimate this closure decreased the time I was able to devote to research by approximately 65 percent. My children both returned to daycare in March 2021, but from March 2021 through the summer of 2022, childcare access was unpredictable because of recurrent exposures and quarantines that occurred with no advance notice. I estimate these disruptions decreased the time I was able to devote to research by approximately 35 percent during this second year of the pandemic.

OPTIONAL COVID-19 IMPACT STATEMENT FROM THE CANDIDATE

I lead a research laboratory on campus, where I conduct the bulk of my research. Due to university protocols, my lab was closed completely from March 2020 until March 2021. I was unable to use alternative space for research during this time because of COVID risks. During this first year of the pandemic, I estimate this closure decreased the time I was able to devote to research by approximately 65 percent. My lab reopened in March 2021, but from March 2021 through the summer of 2022, lab access was unpredictable because of recurrent exposures and quarantines among members of my research team that occurred with no advance notice. I estimate these disruptions decreased the time I was able to devote to research by approximately 35 percent during this second year of the pandemic.

Part 3: Supplementary Tables and Figures

Table S1. Descriptive Statistics of Sample and Dependent Variables**.**

	Mean	Standard Deviation	N
Respondent Characteristics			
Social and Demographic Characteristics			
Gender (woman = 1)	0.210	0.407	596
Race/Ethnicity			
Asian	0.062	0.242	577
Hispanic/Latino	0.005	0.072	577
White	0.932	0.251	577
Parental status (parent = 1)	0.572	0.495	596
Youngest child's age	20.036	11.741	335
Caretaking status (caretaker = 1)	0.305	0.461	591
COVID-19 risk (high risk = 1)	1.394	0.489	596
Career Measures			
University ranking (1-100)	48.617	27.56	480
Biological sciences field	0.392	0.489	515
Number of tenure cases	32.43	23.573	602
Years as full prof.			
0-4 years	0.093	0.290	378
5-9 years	0.169	0.376	378
10-14 years	0.201	0.401	378
15-19 years	0.230	0.421	378
Over 20 years	0.307	0.462	378
Agreement that "science is a calling" (1-5)	3.983	0.929	580
Own pandemic productivity (1-7)	3.367	0.980	578
Dependent Variable Distributions			
Recommendation for tenure (1-6)	4.711	0.886	602
Extent of tenure advocacy (1-7)	4.596	1.280	602
Productivity (1-5)	3.272	0.966	602
Research quality (1-5)	4.176	0.738	602
Research impact (1-5)	3.887	0.681	602
Commitment (1-7)	4.390	1.015	602
Hours worked per week (1-7)	2.977	0.953	602
Future productivity (1-5)	3.352	0.859	602
Future impact (1-5)	3.520	0.810	602
Likelihood of full prof. in 5 years (1-4)	3.005	0.615	602

Note: N displays the non-missing number of cases across each measure. Across all dependent variables, higher values indicate more favorable evaluations. University ranking measure is from the institution's U.S. News and World Report 2022 ranking; lower values indicate more prestigious institutions. Lower values of own pandemic productivity indicate lower perceived productivity.

Table S2. Comparison of survey respondents' characteristics to all invited participants.

	Mean (SD) among survey respondents	Mean (SD) among invited survey participants
Region		
Northeast	0.30	0.31
South	0.31	0.30
Midwest	0.18	0.19
West	0.22	0.21
Public university (=1)	0.67	0.66
University ranking (1-99)	48.62 (27.56)	46.57 (27.78)
Female professor (=1)	0.21	0.20

Note: Invited participants include both faculty who completed the survey and survey non-respondents. N = 602 survey participants and N = 2980 invited participants.

Table S3. OLS Regression Models Predicting Tenure Promotion Recommendation**.**

	Model 1: Promotion	Model 2: Promotion
COVID-19 Impact Statement Condition (ref = No Statement)		
Lab Impact Statement	0.189*	0.178
	(0.087)	(0.125)
Childcare Impact Statement	0.333***	0.370**
	(0.088)	(0.123)
Female Candidate		0.143
		(0.124)
Impact Statement × Candidate Gender		
Lab Statement × Female Candidate		0.020
		(0.175)
Childcare Statement × Female Candidate		-0.071
		(0.175)
Constant	4.537***	4.465***
	(0.062)	(0.088)
Observations	602	602
R-squared	0.024	0.029

Note: *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Model 1 presents regression model displayed graphically in Fig. 1; Model 2 presents regression model displayed in Fig. 2. Standard errors are displayed in parentheses.

Table S4. OLS Regression Models Predicting All Evaluation Measures (Unstandardized).

	Model 1		Model 2		Model 3	
	Tenure Promotion Recommendation					
Impact Statement (ref = No Statement)						
Lab Statement	0.189*	(0.087)			0.178	(0.125)
Childcare Statement	0.333***	(0.088)			0.370**	(0.123)
Female Candidate			0.122+	(0.072)	0.143	(0.124)
Impact Statement × *Candidate Gender*						
Lab Statement × Female					0.020	(0.175)
Childcare Statement × Female					-0.071	(0.175)
Constant	4.537***	(0.062)	4.650***	(0.051)	4.465***	(0.088)
R-squared	0.024		0.005		0.029	
	Extent of Advocacy					
Impact Statement (ref = No Statement)						
Lab Statement	0.259*	(0.125)			0.430*	(0.179)
Childcare Statement	0.607***	(0.126)			0.636***	(0.176)
Female Candidate			0.212*	(0.104)	0.349*	(0.177)
Impact Statement × *Candidate Gender*						
Lab Statement × Female					-0.337	(0.250)
Childcare Statement × Female					-0.044	(0.251)
Constant	4.308***	(0.089)	4.490***	(0.074)	4.131***	(0.126)
R-squared	0.038		0.007		0.049	
	Productivity					
Impact Statement (ref = No Statement)						
Lab Statement	0.239*	(0.096)			0.316*	(0.136)
Childcare Statement	0.340***	(0.096)			0.488***	(0.134)
Female Candidate			0.144+	(0.079)	0.296*	(0.135)
Impact Statement × *Candidate Gender*						
Lab Statement × Female					-0.153	(0.191)
Childcare Statement × Female					-0.291	(0.191)
Constant	3.080***	(0.068)	3.200***	(0.056)	2.929***	(0.096)
R-squared	0.022		0.006		0.032	
	Research Quality					
Impact Statement (ref = No Statement)						
Lab Statement	0.035	(0.074)			0.072	(0.105)
Childcare Statement	0.106	(0.074)			0.094	(0.104)
Female Candidate			0.059	(0.060)	0.076	(0.104)
Impact Statement × *Candidate Gender*						
Lab Statement × Female					-0.074	(0.147)

Table S4. OLS Regression Models Predicting All Evaluation Measures (Unstandardized).

	Model 1		Model 2		Model 3	
Childcare Statement × Female					0.028	(0.148)
Constant	4.129***	(0.052)	4.147***	(0.043)	4.091***	(0.074)
R-squared	0.004		0.002		0.006	
	Research Impact					
Impact Statement (ref = No Statement)						
Lab Statement	0.030	(0.068)			0.059	(0.097)
Childcare Statement	0.109	(0.068)			0.114	(0.096)
Female Candidate			0.041	(0.055)	0.064	(0.096)
Impact Statement × *Candidate Gender*						
Lab Statement × Female					-0.058	(0.136)
Childcare Statement × Female					-0.007	(0.136)
Constant	3.841***	(0.048)	3.867***	(0.039)	3.808***	(0.068)
R-squared	0.005		0.001		0.006	
	Commitment					
Impact Statement (ref = No Statement)						
Lab Statement	0.299**	(0.100)			0.286*	(0.142)
Childcare Statement	0.336***	(0.100)			0.279*	(0.140)
Female Candidate			0.280***	(0.082)	0.234+	(0.141)
Impact Statement × *Candidate Gender*						
Lab Statement × Female					0.021	(0.199)
Childcare Statement × Female					0.128	(0.199)
Constant	4.179***	(0.071)	4.250***	(0.058)	4.061***	(0.100)
R-squared	0.022		0.019		0.042	
	Hours Worked Per Week					
Impact Statement (ref = No Statement)						
Lab Statement	0.318***	(0.094)			0.365**	(0.135)
Childcare Statement	0.224*	(0.094)			0.213	(0.133)
Female Candidate			0.033	(0.078)	0.056	(0.134)
Impact Statement × *Candidate Gender*						
Lab Statement × Female					-0.091	(0.189)
Childcare Statement × Female					0.025	(0.189)
Constant	2.796***	(0.067)	2.960***	(0.055)	2.768***	(0.095)
R-squared	0.020		0.000		0.021	
	Future Productivity					
Impact Statement (ref = No Statement)						
Lab Statement	0.274**	(0.083)			0.276*	(0.119)
Childcare Statement	0.485***	(0.084)			0.514***	(0.118)
Female Candidate			0.150*	(0.070)	0.176	(0.118)

Table S4. OLS Regression Models Predicting All Evaluation Measures (Unstandardized).

	Model 1		Model 2		Model 3	
Impact Statement × *Candidate Gender*						
Lab Statement × Female					-0.006	(0.167)
Childcare Statement × Female					-0.051	(0.167)
Constant	3.100***	(0.059)	3.277***	(0.049)	3.010***	(0.084)
R-squared	0.054		0.008		0.062	
	Future Research Impact					
Impact Statement (ref = No Statement)						
Lab Statement	0.114	(0.081)			0.045	(0.115)
Childcare Statement	0.192*	(0.081)			0.031	(0.113)
Female Candidate			0.080	(0.066)	-0.072	(0.114)
Impact Statement × *Candidate Gender*						
Lab Statement × Female					0.135	(0.161)
Childcare Statement × Female					0.329*	(0.161)
Constant	3.418***	(0.057)	3.480***	(0.047)	3.455***	(0.081)
R-squared	0.009		0.002		0.019	
	Likelihood of Promotion to Full in 5 Years					
Impact Statement (ref = No Statement)						
Lab Statement	0.114+	(0.061)			0.151+	(0.087)
Childcare Statement	0.229***	(0.061)			0.269**	(0.086)
Female Candidate			0.070	(0.050)	0.123	(0.086)
Impact Statement × *Candidate Gender*						
Lab Statement × Female					-0.073	(0.121)
Childcare Statement × Female					-0.075	(0.122)
Constant	2.891***	(0.043)	2.970***	(0.035)	2.828***	(0.061)
R-squared	0.023		0.003		0.028	

Note: *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Standard errors of coefficients are in parentheses. Model 1 estimates the effect of the COVID-19 impact statement condition, pooling candidate gender. Model 2 estimates the effect of candidate gender, pooling impact statement conditions. Model 3 includes an interaction of COVID-19 impact statement condition and candidate gender. N for all models is 602.

Table S5. OLS Regression Models Predicting Evaluation Models (Standardized).

	Tenure Rec.	Advocacy	Productivity	Research Quality	Research Impact	Commitment	Hours Worked	Future Productivity	F In
Impact Statement (ref = No Statement)									
Lab Statement	0.201	0.336*	0.327*	0.098	0.087	0.282*	0.383**	0.321*	0
	(0.141)	(0.140)	(0.141)	(0.143)	(0.143)	(0.140)	(0.142)	(0.139)	(
Childcare Statement	0.418**	0.497***	0.505***	0.127	0.168	0.275*	0.223	0.599***	0
	(0.139)	(0.138)	(0.139)	(0.141)	(0.141)	(0.138)	(0.140)	(0.137)	(
Female Candidate	0.162	0.273*	0.306*	0.103	0.095	0.230+	0.059	0.205	-(
	(0.140)	(0.138)	(0.139)	(0.141)	(0.141)	(0.139)	(0.140)	(0.137)	(
Impact Statement × Candidate Gender									
Lab Statement × Female	0.022	-0.264	-0.158	-0.100	-0.085	0.021	-0.096	-0.006	0
	(0.197)	(0.195)	(0.197)	(0.200)	(0.200)	(0.196)	(0.198)	(0.194)	(
Childcare Statement × Female	-0.080	-0.034	-0.301	0.039	-0.011	0.126	0.027	-0.059	0
	(0.198)	(0.196)	(0.197)	(0.200)	(0.200)	(0.196)	(0.199)	(0.194)	(
Constant	-0.278**	-0.363***	-0.355***	-0.115	-0.116	-0.325**	-0.219*	-0.398***	-(
	(0.099)	(0.098)	(0.099)	(0.101)	(0.101)	(0.099)	(0.100)	(0.098)	(
Observations	602	602	602	602	602	602	602	602	6
R-squared	0.029	0.049	0.032	0.006	0.006	0.042	0.021	0.062	0

Note: *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Table displays regressions models underlying Fig. 3. Standar
in parentheses.

Table S6. Marginal Effects of Impact Statements Within Candidate Gender (Standardized).

Marginal Effect of Impact Statement Compared to No Statement	Tenure Rec.	Advocacy	Productivity	Research Quality	Research Impact	Commitment	Hours Worked	Future Productivity	F I
Male Candidates									
Lab Statement	0.201	0.336*	0.327*	0.098	0.087	0.282*	0.383**	0.321*	0
	(0.141)	(0.140)	(0.141)	(0.143)	(0.143)	(0.140)	(0.142)	(0.139)	(
Childcare Statement	0.418**	0.497***	0.505***	0.127	0.168	0.275*	0.223	0.599***	0
	(0.139)	(0.138)	(0.139)	(0.141)	(0.141)	(0.138)	(0.140)	(0.137)	(
Female Candidates									
Lab Statement	0.223	0.072	0.169	-0.002	0.002	0.303*	0.287*	0.314*	0
	(0.138)	(0.137)	(0.138)	(0.140)	(0.140)	(0.137)	(0.139)	(0.136)	(
Childcare Statement	0.338*	0.462***	0.204	0.165	0.157	0.401**	0.250+	0.539***	0
	(0.140)	(0.139)	(0.140)	(0.142)	(0.142)	(0.139)	(0.141)	(0.138)	(

Note: *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Marginal effects are displayed graphically in Figure 3. S
are indicated in parentheses. Significance indicators are two-tailed t-tests relative to the within-gender, no st
estimate.

Table S7. OLS Regression Models Predicting Tenure Recommendation with Respondent Interactions.

	Interaction with Univ. Ranking	Interaction with Resp. Gender	Interaction with Resp. Parental Status
Lab Impact Statement	-0.050	0.127	0.297*
	(0.147)	(0.099)	(0.135)
Childcare Impact Statement	-0.036	0.281**	0.324*
	(0.173)	(0.099)	(0.132)
Respondent Gender (Female Candidate = 1)		-0.220	
		(0.146)	
Lab Impact Statement × Female Respondent		0.311	
		(0.213)	
Childcare Impact Statement × Female Respondent		0.226	
		(0.215)	
University Ranking	-0.003		
	(0.002)		
Lab Impact Statement × University Ranking	0.004		
	(0.003)		
Childcare Impact Statement × University Ranking	0.005+		
	(0.003)		
Respondent Parental Status (Parent = 1)			0.017
			(0.125)
Lab Impact Statement × Parent			-0.171
			(0.179)
Childcare Impact Statement × Parent			0.025
			(0.178)
Constant	4.722***	4.582***	4.521***
	(0.127)	(0.071)	(0.091)
Observations	480	596	596
R-squared	0.016	0.028	0.027

Note: *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Table displays regressions models underlying Figures S2, S3, and S4. Standard errors are displayed in parentheses.

Table S8. Marginal Effects of Impact Statements Across Respondent University Ranking.

Marginal Effect of Impact Statement Compared to No Statement, Estimated Across Respondent University Ranking	Lab Impact Statement	Childcare Impact Statement
Respondent's University Ranking (Lower is more prestigious)		
10	-0.011	0.018
	(0.123)	(0.147)
20	0.029	0.072
	(0.102)	(0.123)
30	0.068	0.126
	(0.088)	(0.104)
40	0.108	0.181+
	(0.083)	(0.091)
50	0.147	0.235*
	(0.089)	(0.087)
60	0.187+	0.289**
	(0.105)	(0.094)
70	0.226+	0.343**
	(0.127)	(0.110)
80	0.266+	0.397**
	(0.151)	(0.131)
90	0.305+	0.452**
	(0.178)	(0.155)
Observations	480	480

Note: *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Significance tests are two-tailed t-tests relative to the no statement control. Standard errors are displayed in parentheses. Marginal effects are displayed graphically in Figure S2. Ranking is from U.S. News and World Report (2022).

Table S9. OLS Regression Models Predicting LIWC Focus on Past and Time Dimensions of Open-Ended Tenure Rationale Responses**.**

	LIWC Focus on Past Dimension	LIWC Time Dimension
COVID-19 Impact Statement Condition (ref = No Statement)		
Lab Impact Statement	1.251**	0.712*
	(0.384)	(0.299)
Childcare Impact Statement	1.011**	1.343***
	(0.385)	(0.300)
Constant	1.343***	0.672**
	(0.272)	(0.212)
Observations	602	602
R-squared	0.020	0.032

Note: *** p<0.001, ** p<0.01, * p<0.05, + p<0.10. Standard errors are displayed in parentheses. Table displays regressions models underlying Fig. S1. Measures are the percent of words within the open-ended text responses that contain words in the "Focus on Past" and "Time" temporality dimensions.

Table S10. Pretest results of name perceptions.

Names Pretest Measure	Jennifer Nelson	Michael Nelson	P-value of difference between means or proportions
Class	2.84	2.90	0.65
Education	2.15	2.15	0.95
Immigrant	4.38	4.16	0.20
Age Group	3.60	3.61	0.95
Gender			
Woman	0.96	0.02	0.00
Man	0.04	0.98	0.00
I don't know	0.00	0.00	--
Race/Ethnicity			
Asian or Asian American	0.00	0.02	0.29
Black or African American	0.04	0.07	0.56
Hispanic or Latino	0.00	0.00	--
White or Caucasian	0.90	0.87	0.61
I don't know	0.06	0.04	0.74

Note: Averages are displayed for class, education, immigrant, and age group variables. Measures are coded as follows. Class: 1-4; 2 = "upper middle class"; 3 = "middle class." Education: 1-5; 2 = "Bachelor's degree"; 3 = "Some college". Immigrant: 1-5; 4 = "Somewhat unlikely"; 5 = "Extremely unlikely." Age: 1-7; 3 = 25-34; 4 = 35-44. Proportions are displayed for gender and race/ethnicity variables. P-values are reported from two-tailed t-tests. N = 50 for Jennifer Nelson; N = 45 for Michael Nelson.

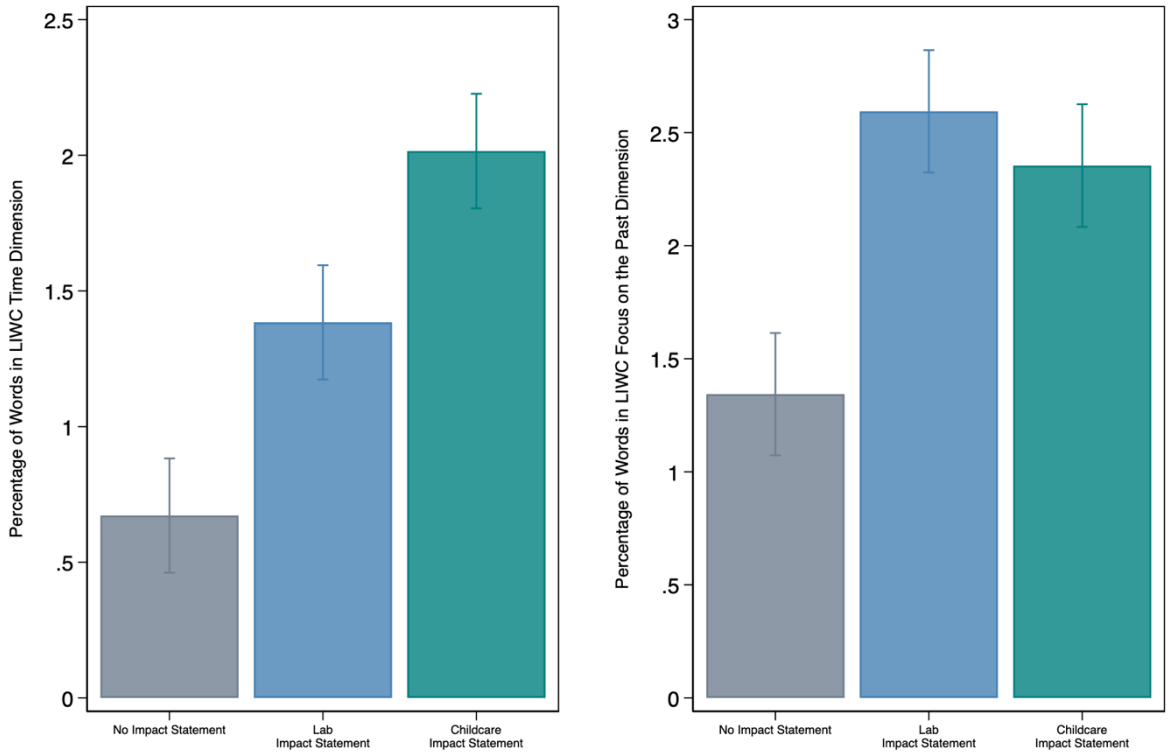

Fig. S1. Variation in LIWC Temporal Dimensions Within Tenure Rationale Open-Ended Responses. Estimates from OLS regression models (see Table S9) using LIWC analysis of open-ended text responses. Measures indicate the percentage of words in open-ended responses belonging to LIWC "Time" and "Focus on Past" dimensions. Error bars display standard errors.

Figure S1. Figure S1 depicts the results from LIWC analyses of the open-ended response, displaying the average percentage of words in the "Time" and "Focus on Past" dimensions. The regression results are displayed in Table S9. We find that the open-ended text rationale for both the lab statement and childcare statement have greater percentages of words in these dimensions: for example, the lab impact statement is associated with an increased mention of the "time" dimension by about 0.711 percentage points (p=0.017), and the childcare impact statement by about 0.134 percentage points (p<0.000008). The effect is higher on the "time" measure for the childcare statement compared to the lab statement (p=0.036), but is not statistically different across the lab and childcare statements for the "focus on past" measure (p=0.534). To interpret the substantive effects of the LIWC outcomes, we find that about 23% of respondents include one or more words in these temporal LIWC dimensions in open-ended responses for the "no statement" condition. In the lab condition, this rate increases to 44%, and to 53% in the childcare impact statement condition (both statistically significant compared to no statement at p<0.001). These results, therefore, suggest that respondents viewing either impact statement consider the temporal nature of the pandemic disruption as a consideration in their tenure rationale.

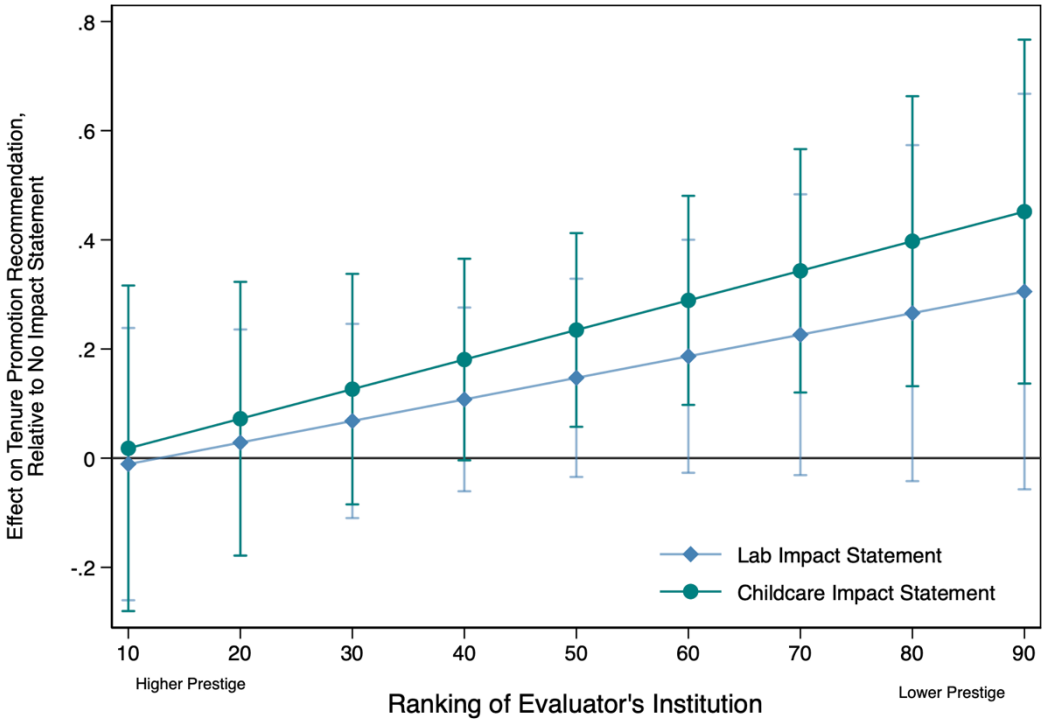

Figure S2. Variation in the effect of COVID-19 impact statements on tenure recommendation across evaluator's university prestige. Estimates are from OLS regressions predicting the tenure recommendation, with an interaction of impact statement × evaluator's university ranking. Estimates display the marginal effect compared to no statement across values of the evaluator's university—lower values indicating higher-prestige universities—with 95% confidence intervals. The full regression model is provided in Table S7 and marginal effects in Table S8. Positive values on the *y*-axis indicate a more favorable evaluation compared to no statement (y=0); statistically significant estimates at p<0.05 do not cross the *y*-axis.

Figure S2 displays the marginal effects of the lab and childcare impact statement on recommendation for tenure, estimated across values of respondents' university ranking, with 95% confidence intervals. When the confidence intervals cross 0 on the *y*-axis, this indicates that an effect is not statistically significant at p<0.05. Lower values on the university ranking indicate more prestigious institutions. The marginal effect estimates are displayed in Table S8 and derived from Model 2 in Table S7. While the interaction terms in the regression model are not statistically significant at p<0.05 (p=0.087 for the interaction with the childcare statement), we find evidence of variation by respondent institution in the marginal effects estimates. Specifically, the effects of impact statements are close to 0 and not statistically significant at the most prestigious institutions, but effect sizes are larger in less prestigious universities, becoming statistically significant for the childcare statement and marginally significant (p<0.10) for the lab statement. For example, the estimated marginal effect of the childcare statement is 0.018 and not statistically significant for respondents in a university ranked at about 10—e.g., Caltech, Northwestern, Duke—and increases to a statistically significant effect of 0.235 (p=0.011) at universities ranked about 50 (e.g., The Ohio State University, Purdue University, Villanova). At universities ranked about 90 (e.g., University at Buffalo—SUNY, University of Delaware), the effect is estimated to be about 0.452, nearly half a point on the scale, and is statistically

significant (p=0.006). It is important to note that for this analysis, since university was self-reported and some respondents opted not to answer this question, results should be interpreted with caution.

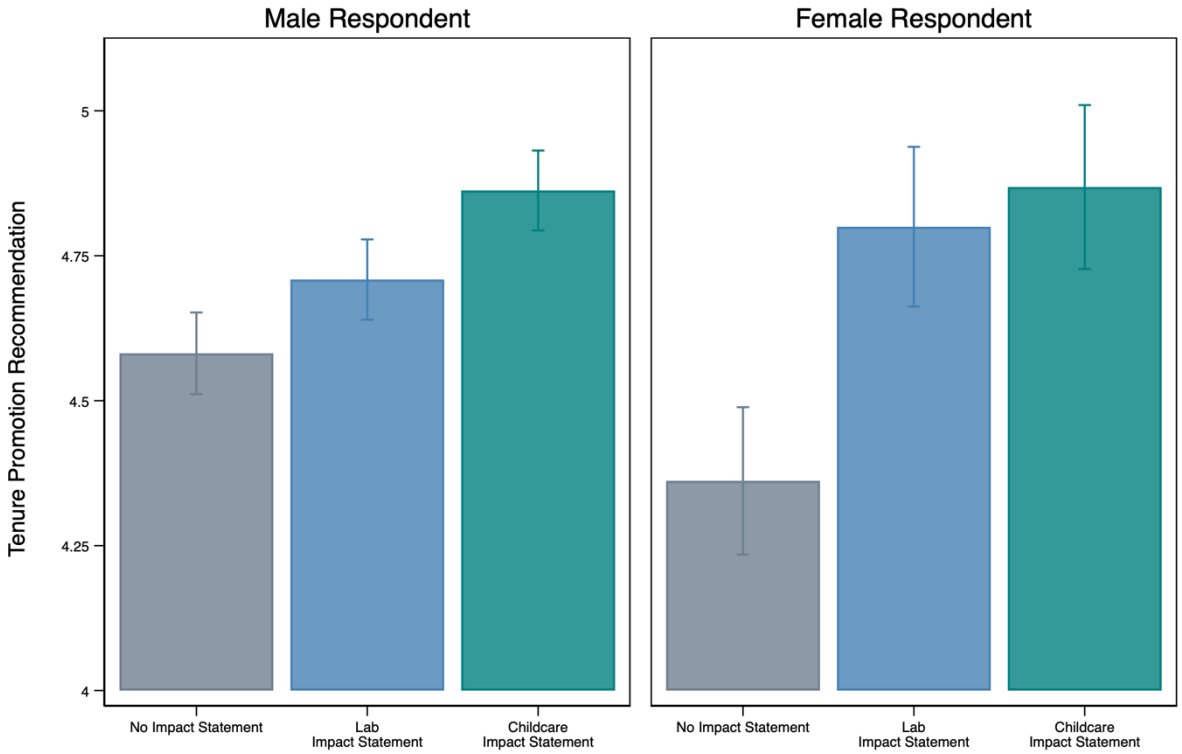

Fig. S3. Recommendation of promotion with tenure across COVID-19 impact statement experimental conditions and respondent gender**.** Estimates from OLS regression models with impact statement × candidate gender interaction term (see Table S7). Error bars display standard errors. Tenure recommendation measure is a 1-6 scale with 6 indicating a stronger recommendation for tenure.

Figure S3 illustrates variation in the tenure recommendation across the COVID-19 impact statement and respondent gender. The regression model for these results is displayed in Table S7. While there is no statistically significant interaction effect of the impact statement with respondent gender, there is some evidence of variation in responsiveness to the treatments across respondent gender. Among both male and female respondents, the childcare statement has a statistically significant and positive effect relative to no statement (p<0.01 for both). The lab impact statement does not have a statistically significant effect compared to no statement among male respondents (p=0.21), but does have a statistically significant positive effect among female respondents (p<0.05). Put differently, these findings suggest that male respondents may be less responsive to the lab disruption statement than female respondents, although the gender difference in effect sizes is not statistically significant (p=0.145).

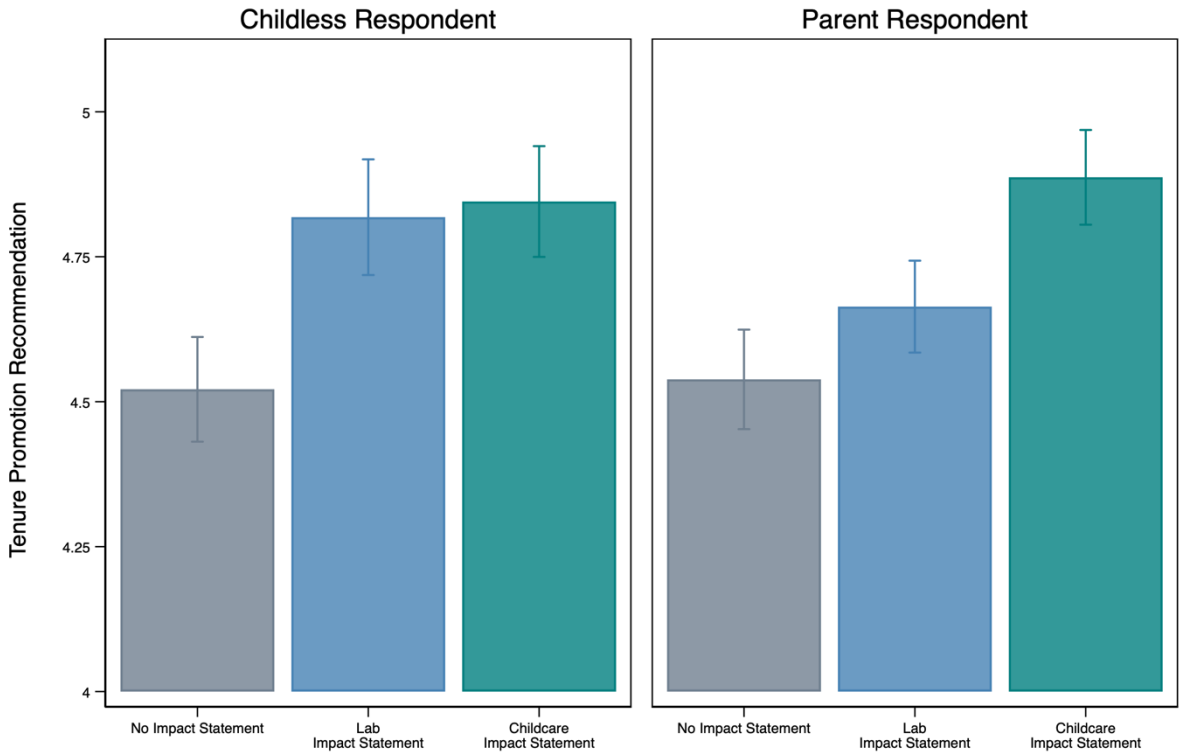

Fig. S4. Recommendation of promotion with tenure across COVID-19 impact statement experimental conditions and respondent parental status. Estimates from OLS regression models with impact statement × candidate parental status interaction term (see Table S7). Error bars display standard errors. Tenure recommendation measure is a 1-6 scale with 6 indicating a stronger recommendation for tenure.

Figure S4 depicts the tenure recommendation levels across the COVID-19 impact statement conditions and respondents' parental status, with the regression model estimates displayed in Table S7. There is no statistically significant interaction effect of parental status on tenure recommendation, but we find some differences in evaluations across parents and non-parents. Childless respondents have significantly higher tenure recommendation levels when viewing either the lab or childcare statement, compared to no statement (p<0.05), and both statements increase recommendations to a similar extent (p=0.85). Respondents who are parents do not recommend candidates with the lab impact statement significantly higher than candidates with no statement (p=0.286), but do recommend candidates with childcare impact statements significantly higher than those with no statement (p<0.01). Parents recommend candidates with the childcare statement at a marginally significantly higher level than candidates with a lab statement (p=0.052). Overall, these results suggest that childless respondents' evaluations are positively impacted by either type of COVID-19 impact statement, whereas respondents who are parents are most responsive to the childcare impact statement.

References

Blair-Loy, Mary, and Erin A. Cech. 2022. *Misconceiving Merit: Paradoxes of Excellence and Devotion in Academic Science and Engineering*. University of Chicago Press.

Cech, Erin A., and Mary Blair-Loy. 2014. "Consequences of Flexibility Stigma Among Academic Scientists and Engineers." *Work and Occupations* 41(1):86–110. doi: 10.1177/0730888413515497.

Comenetz, Joshua. 2016. "Frequently Occurring Surnames in the 2010 Census." Retrieved (https://www2.census.gov/topics/genealogy/2010surnames/surnames.pdf).

Dykema, Jennifer, John Stevenson, Lisa Klein, Yujin Kim, and Brendan Day. 2013. "Effects of E-Mailed versus Mailed Invitations and Incentives on Response Rates, Data Quality, and Costs in a Web Survey of University Faculty." *Social Science Computer Review* 31(3):359–70. doi: 10.1177/0894439312465254.

Gaddis, S. 2017. "How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies." *Sociological Science* 4:469–89. doi: 10.15195/v4.a19.

Gaddis, S. Michael. 2018. "An Introduction to Audit Studies in the Social Sciences." Pp. 3–44 in *Audit studies: Behind the scenes with theory, method, and nuance*, edited by S. M. Gaddis. Cham: Springer International Publishing.

McFarland, Sam G. 1981. "Effects of Question Order on Survey Responses." *Public Opinion Quarterly* 45(2):208. doi: 10.1086/268651.

Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." *Proceedings of the National Academy of Sciences* 109(41):16474–79. doi: 10.1073/pnas.1211286109.

Sauermann, Henry, and Michael Roach. 2013. "Increasing Web Survey Response Rates in Innovation Research: An Experimental Study of Static and Dynamic Contact Design Features." *Research Policy* 42(1):273–86. doi: 10.1016/j.respol.2012.05.003.

Social Security Administration. 2022. "Popular Baby Names in 1987." (https://www.ssa.gov/cgi-bin/popularnames.cgi).

US News & World Report. 2022. "The Best National Universities in America." *US News & World Report*. (https://www.usnews.com/best-colleges/rankings/national-universities).

Williams, Richard. 2012. “Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects.” *The Stata Journal* 12(2):308–31. doi: 10.1177/1536867X1201200209.

Williams, Wendy M., and Stephen J. Ceci. 2015. “National Hiring Experiments Reveal 2:1 Faculty Preference for Women on STEM Tenure Track.” *Proceedings of the National Academy of Sciences* 112(17):5360–65. doi: 10.1073/pnas.1418878112.