

Supplementary Appendices for “Role Congruity and Expectation Disconfirmation: The Role of Gender in Performance Evaluations”

Table of Contents

<i>APPENDIX A: SUPPORTING INFORMATION FOR STUDY 1 (SCHOOLS CONTEXT)</i>	3
Figure A1. Example of Introductory Screen for Study 1 (Female Principal Condition)	3
Figure A2. Example Scorecard for Study 1 (Female Principal, High Performance Condition)	4
Table A1. Descriptive Statistics and Frequencies for Study 1	4
Table A2. Treatment Group Balance Across Key Demographic Variables for Study 1	6
Table A3. Regression Coefficients for Study 1 (Education Context)	7
Robustness Checks for Study 1 (Education Context): Including Control Variables	7
Figure A3. Average marginal effects of gender treatment for study 1 with controls	8
Table A4. Regression Coefficients for Study 1 with Controls	8
Robustness Checks for Study 1 (Education Context): Heterogeneous Treatment Effects	10
Figure A4. Average marginal effects of gender treatment for study 1 by respondent gender	11
Table A5. Regression Coefficients for Study 1 with Interactions by Respondent Gender	12
Figure A5. Average marginal effects of gender treatment for study 1 by respondent age	13
Table A6. Regression Coefficients for Study 1 with Interactions by Respondent Age	14
Figure A6. Average marginal effects of gender treatment for study 1 by partisanship	15
Table A7. Regression Coefficients for Study 1 with Interactions by Respondent Partisanship	16
Figure A7. Average marginal effects of gender treatment for study 1 by respondent ideology	17
Table A8. Regression Coefficients for Study 1 with Interactions by Respondent Ideology	18
<i>APPENDIX B: SUPPORTING INFORMATION FOR STUDY 2 (POLICE CONTEXT)</i>	19
Figure B1. Example of Introductory Screen for Study 2 (Male Police Captain Condition)	19
Figure B2. Example Scorecard for Study 2 (Male Captain, High Performance Condition)	20
Table B1. Descriptive Statistics and Frequencies for Study 2	20
Table B2. Treatment Group Balance Across Key Demographic Variables for Study 2	22
Table B3. Regression Coefficients for Study 2 (Police Context)	23
Robustness Checks for Study 2 (Police Context): Including Control Variables	23
Figure B3. Average marginal effects of gender treatment for study 2 with controls	24
Table B4. Regression Coefficients for Study 2 with Controls	24
Robustness Checks for Study 2 (Police Context): Heterogeneous Treatment Effects	26
Figure B4. Average marginal effects of gender treatment for study 2 by respondent gender	27

Table B5. Regression Coefficients for Study 2 with Interactions by Respondent Gender	28
Figure B5. Average marginal effects of gender treatment for study 2 by respondent age	29
Table B6. Regression Coefficients for Study 2 with Interactions by Respondent Age	30
Figure B6. Average marginal effects of gender treatment for study 2 by partisanship	31
Table B7. Regression Coefficients for Study 2 with Interactions by Respondent Partisanship	32
Figure B7. Average marginal effects of gender treatment for study 2 by respondent ideology	33
Table B8. Regression Coefficients for Study 2 with Interactions by Respondent Ideology	34

APPENDIX A: SUPPORTING INFORMATION FOR STUDY 1 (SCHOOLS CONTEXT)

Figures A1 and A2 present examples of the treatments that were shown to survey respondents for study 1. Figure A1 presents the female principal treatment condition. The male principal condition used the name “Scott” instead of “Amy” and masculine gender pronouns instead of feminine ones.

Figure A2 presents the scorecard for the female principal, high performance condition. The high performance condition showed three green upward-pointing arrows and one red downward-pointing arrow. The moderate performance condition showed two green and two red arrows. The low performance condition showed one green and three red arrows. We randomly assigned the location of the arrows within each performance condition to offset the possibility that respondents view particular performance metrics as more important than others. For example, each respondent allocated to the high performance condition had an equal probability of viewing the red arrow in either the graduation rate box, college admission box, reading pass rates box, or the math pass rates box. Only the location of the arrows changed. The relative location of the performance criteria on the scorecards remained fixed across all conditions.

Table A1 presents descriptive statistics and frequencies for the sample used in study 1. Only participants who passed all of the screening checks in MTurk are included in the sample. Table A2 presents balance checks for study 1 across key respondent demographics, including gender, age, political ideology, and partisanship. Table A3 presents the regression coefficients for the main models for study 1 presented in the text. The models were estimated using ordinary least squares with interaction terms included for the treatment groups.

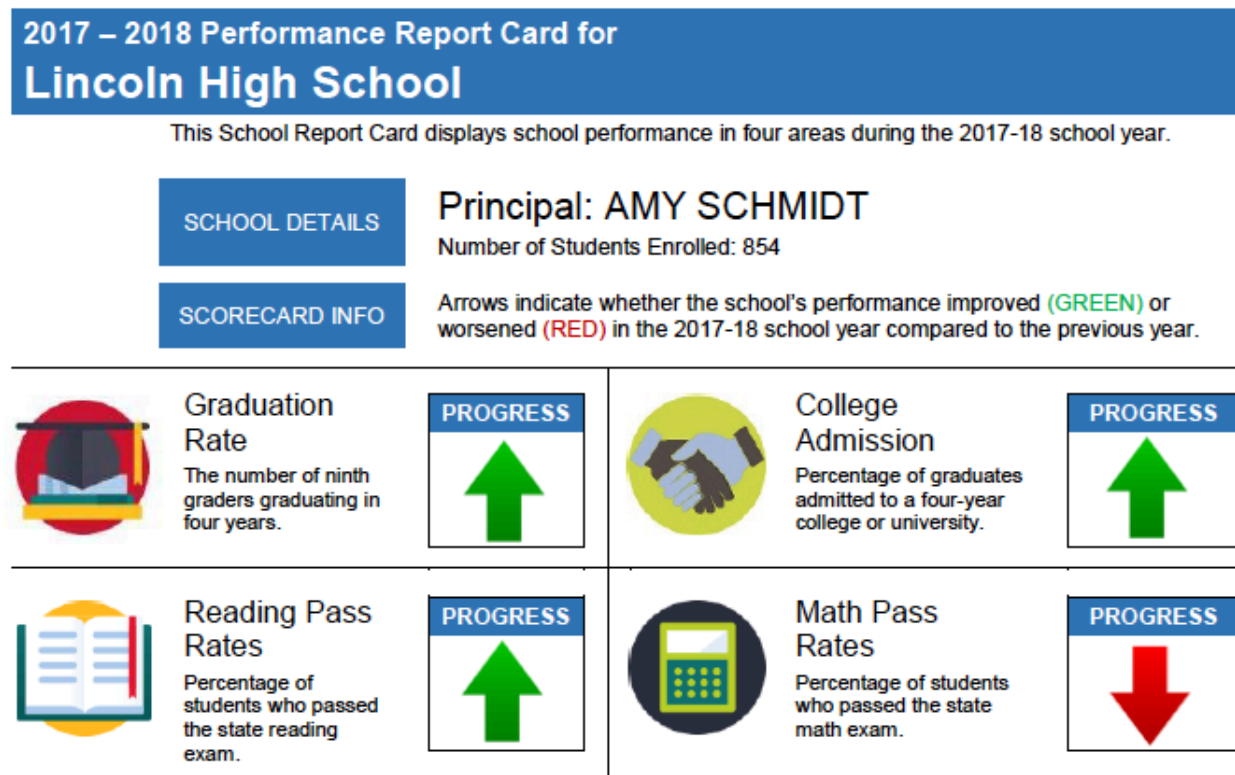
Figure A1. Example of Introductory Screen for Study 1 (Female Principal Condition)

Amy Schmidt is the principal of Lincoln High School. She has several support staff and oversees a faculty of approximately 50 teachers and teacher aids. Her school has 854 students and the campus consists of three buildings, including two academic buildings and a gymnasium.

On the next screen you will be shown an annual school-level report card with information on Lincoln High School’s performance on a number of aspects related to student academic achievement.

For the purpose of assessment, the card reports upward pointing green arrows for areas where the school has improved (↑) and downward pointing red arrows for areas where the school has worsened (↓).

Figure A2. Example Scorecard for Study 1 (Female Principal, High Performance Condition)



Note: The high performance condition showed three green and one red arrow. The moderate performance condition showed two green and two red arrows. The low performance condition showed one green and three red arrows. We randomly assigned the location of the arrows within each performance condition.

Table A1. Descriptive Statistics and Frequencies for Study 1

Variable	Operationalization	N	%	Mi n	Ma x	Mea n	SD
<i>Dependent Variables</i>							
Principal Performance	How would you rate Principal {Amy/Scott} Schmidt's performance?	2,447		0	10	5.58	2.20
Principal Adequacy	How would you rate Principal {Amy/Scott} Schmidt's adequacy?	2,446		0	10	5.72	2.22
Principal Competence	How would you rate Principal {Amy/Scott} Schmidt's competence?	2,447		0	10	5.98	2.20
School Performance	How would you rate Lincoln High School's overall performance?	2,446		0	10	5.46	2.34
<i>Attention Checks</i>							

Gender Check	Indicator. Correct recall of the principal's gender (Correct = 1)	2,447	0	1	.93	.25
Performance Check	Indicator. Correct recall of number of improved areas (Correct = 1)	2,447	0	1	.93	.26
<u>Demographics</u>						
Gender	Respondent's gender (dichotomized)	2,423	100.00			
	"Male"	1,122	46.31			
	"Female"	1,301	53.69			
Age	Respondents' age (derived from birth year)	2,446	19	83	38.29	12.56
Education	Highest level of education	2,446	1	7	4.22	1.26
	"Less than high school"	6	.25			
	"High school graduate"	231	9.44			
	"Some college"	622	25.43			
	"2-year college"	312	12.76			
	"4-year college"	939	38.39			
	"Professional degree"	306	12.51			
	"Doctorate"	30	1.23			
Race/Ethnicity	Respondent's race or ethnicity (categories are not mutually exclusive)					
	"African American or Black"	256	10.46			
	"Asian American or Asian"	195	7.97			
	"Hispanic or Latino/Latina/Latinx"	195	7.97			
	"Middle Eastern or Arab"	16	.65			
	"Multiracial"	58	2.37			
	"Pacific Islander"	13	.53			
	"White or Caucasian"	1,815	74.17			
	"American Indian/Native American"	40	1.63			
	"Other/Self-Description"	16	.65			
Employment	Employment Status	2,446	100.00			
	"Full-time"	1,295	52.94			
	"Part-time"	299	12.22			
	"Unemployed, looking for work"	143	5.85			
	"Unemployed, not looking for work"	107	4.37			
	"Self-employed"	294	12.02			
	"Retired"	102	4.17			
	"Student"	157	6.42			
	"Currently unable to work"	49	2.00			
Income	Yearly household income in \$10,000 brackets	2,446	1	13	5.97	3.40
Marital Status	Indicator for single, never married (Yes = 1)	2,447	0	1	.43	.5
Liberal-Conservative	<i>Where would you place yourself among the following categories?</i>	2,383	1	7	3.55	1.73

	"Extremely liberal"	264	11.08				
	"Liberal"	583	24.46				
	"Slightly liberal"	343	14.39				
	"Moderate; middle of the road"	510	21.40				
	"Slightly conservative"	240	10.07				
	"Conservative"	341	14.31				
	"Extremely conservative"	102	4.28				
Democrat-Republican	<i>In politics today, how do you consider yourself?</i>	2,355		1	5	3.44	1.51
	"Republican"	414	17.58				
	"Independent, but lean Republican"	273	11.59				
	"Independent, with no party preference"	395	16.77				
	"Independent, but lean Democrat"	410	17.41				
	"Democrat"	863	36.65				
Work in K-12	Indicator: Ever worked in K-12 education (Yes = 1, No = 0)	2,446		0	1	.16	.37
Child in K-12	Indicator: Child enrolled in K-12 education (Yes = 1, No = 0)	2,446		0	1	.27	.44

Table A2. Treatment Group Balance Across Key Demographic Variables for Study 1

<i>DV</i>	<i>Gender Condition</i>	<i>Performance Condition</i>			<i>Test for Mean Differences</i>		
		Low (3)	Moderate (4)	High (5)	3 vs. 4	3 vs. 5	4 vs. 5
<i>Rater Gender</i>	Male (1)	0.52	0.56	0.55	NS	NS	NS
	Female (2)	0.51	0.53	0.53	NS	NS	NS
<i>Test for Mean Differences</i>	1 vs. 2	NS	NS	NS			
<i>Rater Age</i>	Male (1)	37.18	37.97	38.44	NS	NS	NS
	Female (2)	38.52	39.15	38.44	NS	NS	NS
<i>Test for Mean Differences</i>	1 vs. 2	NS	NS	NS			
<i>Liberal-Conservative Orientation</i>	Male (1)	3.37	3.52	3.68	NS	-.31*	NS
	Female (2)	3.63	3.66	3.45	NS	NS	.21+
<i>Test for Mean Differences</i>	1 vs. 2	.26*	.23+	NS			
<i>Republican-Democrat Orientation</i>	Male (1)	3.59	3.45	3.41	NS	.18+	NS

	Female (2)	3.34	3.41	3.44	NS	NS	NS
<i>Test for Mean Differences</i>	1 vs. 2	.26*	NS	NS			

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. NS = Not statistically significant. Estimates are from two-tailed *t* tests for mean group differences.

Table A3. Regression Coefficients for Study 1 (Education Context)

	(1) Performanc e	(2) Adequacy	(3) Competenc e	(4) Sch. Perf.
Female principal	-0.117 (0.106)	-0.150 (0.112)	-0.150 (0.114)	-0.147 (0.111)
Moderate performance	1.965*** (0.108)	1.834*** (0.113)	1.797*** (0.115)	2.238*** (0.113)
High performance	3.851*** (0.108)	3.646*** (0.114)	3.525*** (0.116)	4.129*** (0.113)
Female principal x Moderate perf.	0.099 (0.151)	0.168 (0.159)	0.155 (0.162)	0.069 (0.158)
Female principal x High perf.	0.071 (0.150)	0.209 (0.159)	0.163 (0.161)	0.126 (0.157)
Constant	3.682*** (0.076)	3.909*** (0.081)	4.230*** (0.082)	3.384*** (0.080)
R-squared	0.524	0.479	0.452	0.539
Observations	2447	2446	2447	2446

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses. Male principal and low performance are the omitted reference categories.

Robustness Checks for Study 1 (Education Context): Including Control Variables

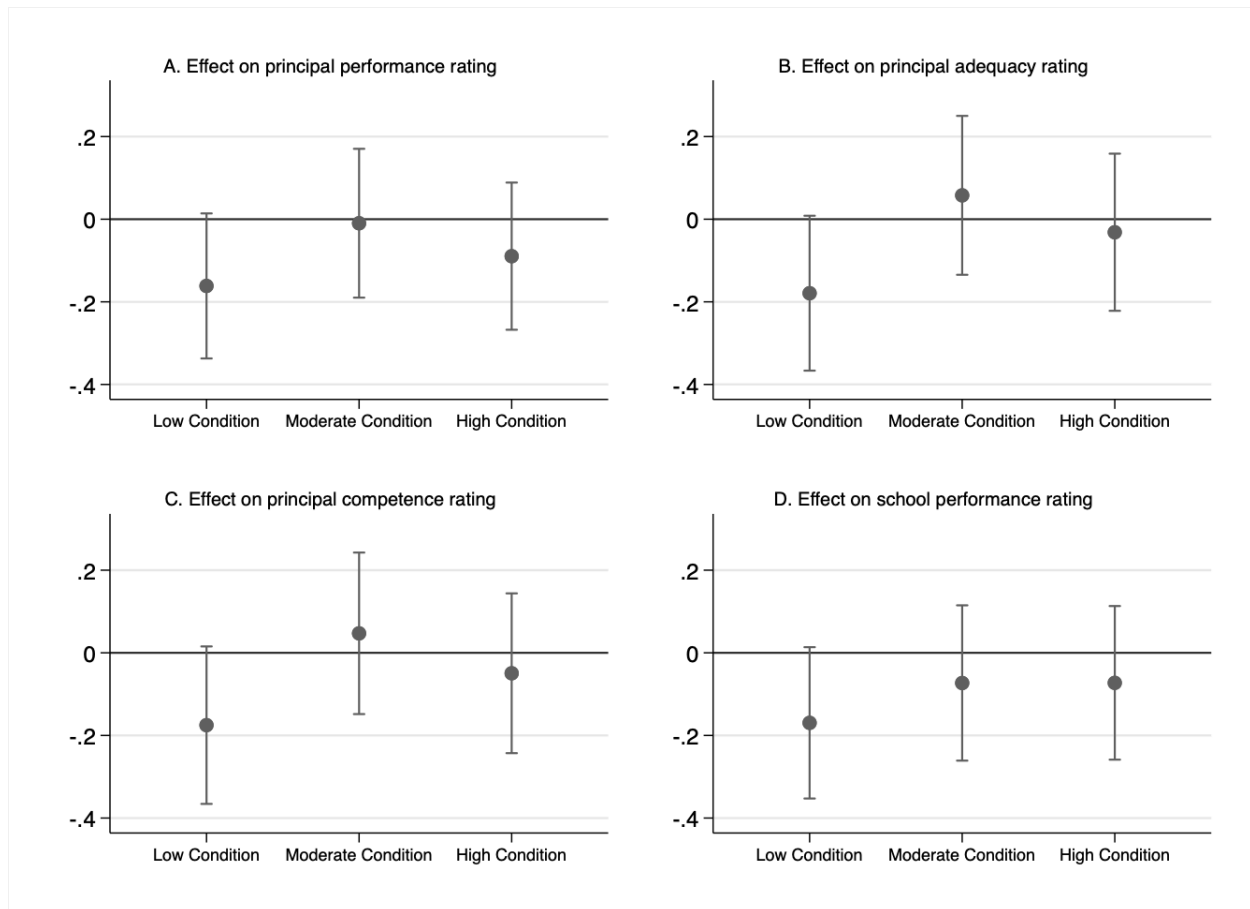
Figure A3 presents the average marginal effects of the gender treatment across each performance treatment condition when control variables are included in the regression models. The main results are robust to the inclusion of multiple variables controlling for respondent demographics and comprehension checks. Several of the marginal effects approach, but do not meet, conventional levels of statistical significance when control variables are included in the models.

Specifically, the marginal effect of the principal's gender is close to statistical significance in the low performance condition in each of the four models, as shown in panels

A-D in Figure A3. The p-value for the average marginal effect of gender in the low performance condition is 0.130 in the principal performance model, 0.116 in the principal adequacy model, 0.131 in the principal competence model, and 0.128 in the school performance model.

Table A4 presents the regression coefficients for the models including control variables.

Figure A3. Average marginal effects of gender treatment for study 1 with controls



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table A4.

Table A4. Regression Coefficients for Study 1 with Controls

	(1)	(2)	(3)	(4)
	Performance	Adequacy	Competence	Sch. Perf.

Female principal	-0.161 (0.107)	-0.179 (0.114)	-0.175 (0.116)	-0.170 (0.111)
Moderate performance	2.001*** (0.108)	1.854*** (0.115)	1.826*** (0.117)	2.306*** (0.113)
High performance	3.903*** (0.108)	3.729*** (0.115)	3.596*** (0.117)	4.228*** (0.113)
Female principal x Moderate perf.	0.152 (0.153)	0.237 (0.163)	0.222 (0.166)	0.096 (0.160)
Female principal x High perf.	0.072 (0.152)	0.148 (0.162)	0.126 (0.165)	0.097 (0.159)
Principal gender correct	0.123 (0.132)	0.280* (0.141)	0.198 (0.143)	0.106 (0.137)
Performance check correct	-1.233*** (0.126)	-0.954*** (0.135)	-0.862*** (0.137)	-1.304*** (0.132)
Female respondent	-0.056 (0.066)	-0.066 (0.071)	-0.057 (0.072)	-0.191** (0.069)
Age	-0.006+ (0.003)	-0.007* (0.003)	-0.007* (0.003)	-0.006+ (0.003)
Education	-0.003 (0.027)	0.004 (0.029)	-0.019 (0.030)	-0.013 (0.028)
African American or Black	-0.149 (0.166)	-0.226 (0.177)	-0.104 (0.180)	-0.127 (0.173)
Asian American or Asian	-0.237 (0.175)	-0.328+ (0.187)	-0.392* (0.190)	-0.211 (0.183)
Hispanic or Latino/Latina/Latinx	-0.080 (0.150)	-0.153 (0.161)	-0.009 (0.163)	0.088 (0.157)
Middle Eastern or Arab	-0.310 (0.394)	-0.442 (0.421)	-0.706+ (0.428)	0.131 (0.411)
Multiracial	-0.029 (0.211)	-0.213 (0.225)	0.039 (0.229)	-0.083 (0.220)
Pacific Islander	0.127 (0.502)	0.368 (0.536)	0.417 (0.545)	0.169 (0.524)
White or Caucasian	0.043 (0.145)	0.002 (0.154)	0.095 (0.157)	0.039 (0.151)
Self-identify or Other	0.025 (0.494)	0.393 (0.527)	0.413 (0.536)	-0.388 (0.515)
American Indian or Native American	0.074 (0.251)	-0.145 (0.268)	-0.075 (0.272)	0.144 (0.262)
Employed part-time in an organization	0.042 (0.102)	-0.006 (0.109)	0.034 (0.110)	-0.024 (0.106)
Unemployed - looking for work	-0.276* (0.140)	-0.346* (0.150)	-0.260+ (0.152)	-0.151 (0.147)
Unemployed - not looking for work	-0.076 (0.160)	-0.204 (0.170)	0.082 (0.173)	-0.082 (0.166)

Retired	0.058 (0.175)	-0.036 (0.187)	0.027 (0.190)	0.124 (0.183)
Student	-0.301* (0.138)	-0.302* (0.148)	-0.243 (0.150)	-0.243+ (0.144)
Currently unable to work	-0.182 (0.241)	-0.256 (0.257)	-0.225 (0.262)	-0.178 (0.251)
Self-employed	0.066 (0.105)	0.070 (0.112)	0.082 (0.114)	0.131 (0.110)
Income category	-0.020+ (0.010)	-0.014 (0.011)	-0.023* (0.011)	-0.016 (0.011)
Single or unmarried	-0.157* (0.077)	-0.138+ (0.082)	-0.121 (0.083)	-0.216** (0.080)
Political ideology	0.023 (0.030)	0.019 (0.032)	0.018 (0.032)	-0.008 (0.031)
Political partisanship	0.031 (0.034)	0.047 (0.036)	0.030 (0.037)	0.011 (0.035)
Worked in K-12	0.280** (0.088)	0.171+ (0.094)	0.238* (0.096)	0.220* (0.092)
Child in K-12	-0.022 (0.077)	-0.075 (0.083)	-0.005 (0.084)	0.015 (0.081)
Constant	4.978*** (0.345)	4.825*** (0.368)	5.201*** (0.374)	4.991*** (0.360)
R-squared	0.556	0.503	0.477	0.575
Observations	2294	2294	2294	2294

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.

Robustness Checks for Study 1 (Education Context): Heterogeneous Treatment Effects

Next, we explore potential heterogeneous treatment effects across respondent gender, age, partisanship, and political ideology. Some individuals, such as males, older respondents, Republicans, and ideologically conservative respondents, might be more easily swayed by gender stereotypes, and thus more likely to perceive incongruities between women and leadership roles. To examine whether the treatment conditions have different effects on different groups of respondents (i.e., heterogeneous treatment effects), we include three-way interaction terms in each of the regression models that interact the two treatment groups with each of these respondent demographics.

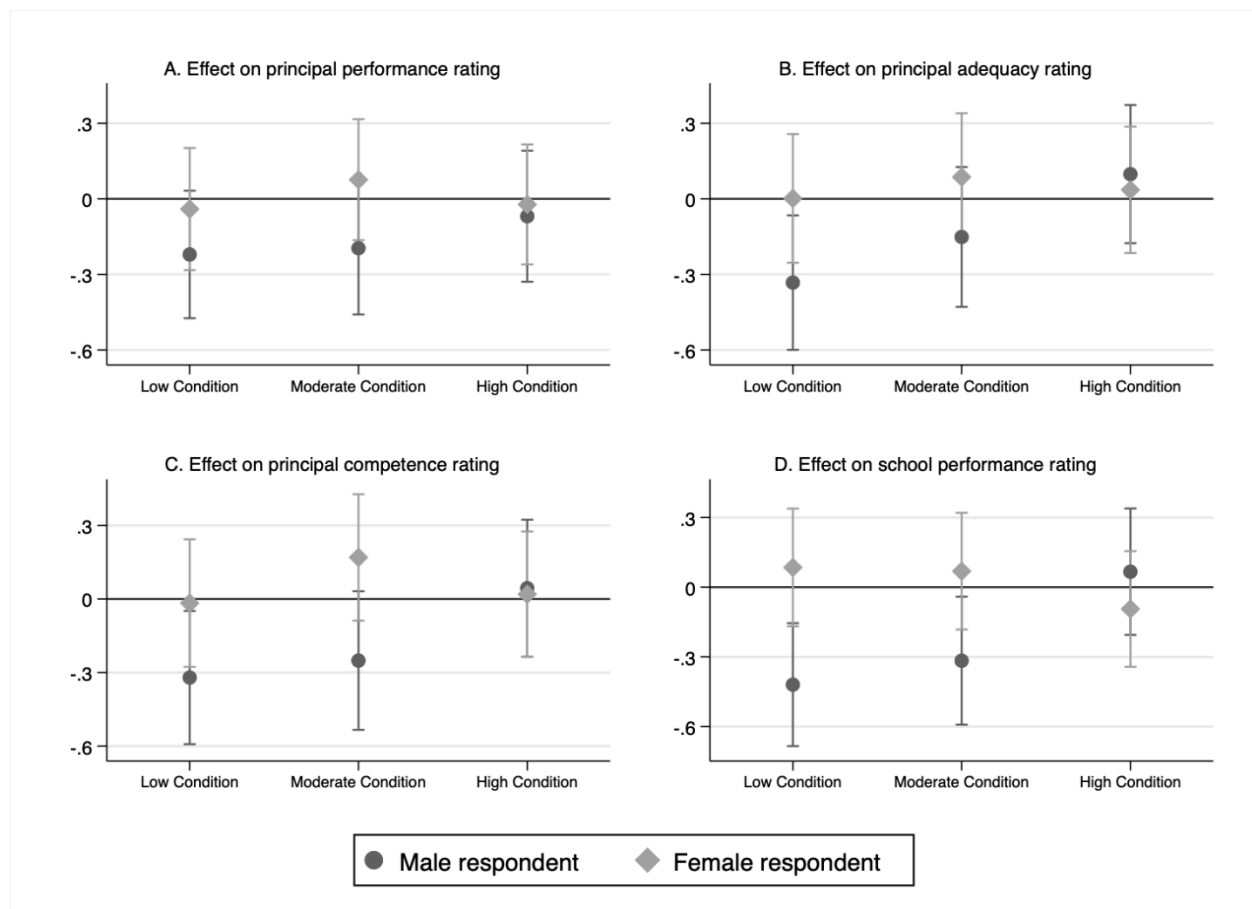
We first examine whether male and female respondents react differently to male and female public managers. Figure A4 shows the average marginal effects of the principal's gender across each performance condition for both male and female respondents. The figure shows that, in general, female respondents and male respondents evaluate male and female principals similarly. The effect of the principal's gender remains statistically indistinguishable from zero, except in a few cases. Specifically, the female principal treatment condition has a statistically

significant, negative effect for male respondents allocated to the low performance condition in three of the four models. Compared to male respondents in the male principal, low performance condition, male respondents in the female principal, low performance condition provide lower evaluations in terms of principal adequacy (p-value= 0.040), principal competence (p-value= 0.052), and school performance (p-value= 0.009), as shown in panels B, C, and D in Figure A4. Male respondents in the moderate performance condition also rate female principals lower on school performance (p-value= 0.059), as shown in panel D of Figure A4.

Meanwhile, the principal's gender does not have a statistically significant effect on female respondents across all of the performance conditions and dependent variables. These findings provide some limited evidence that male respondents evaluate female principals less favorably than male principals under conditions of low performance.

Table A5 presents regression coefficients for the models that include interactions by respondent gender.

Figure A4. Average marginal effects of gender treatment for study 1 by respondent gender



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table A5.

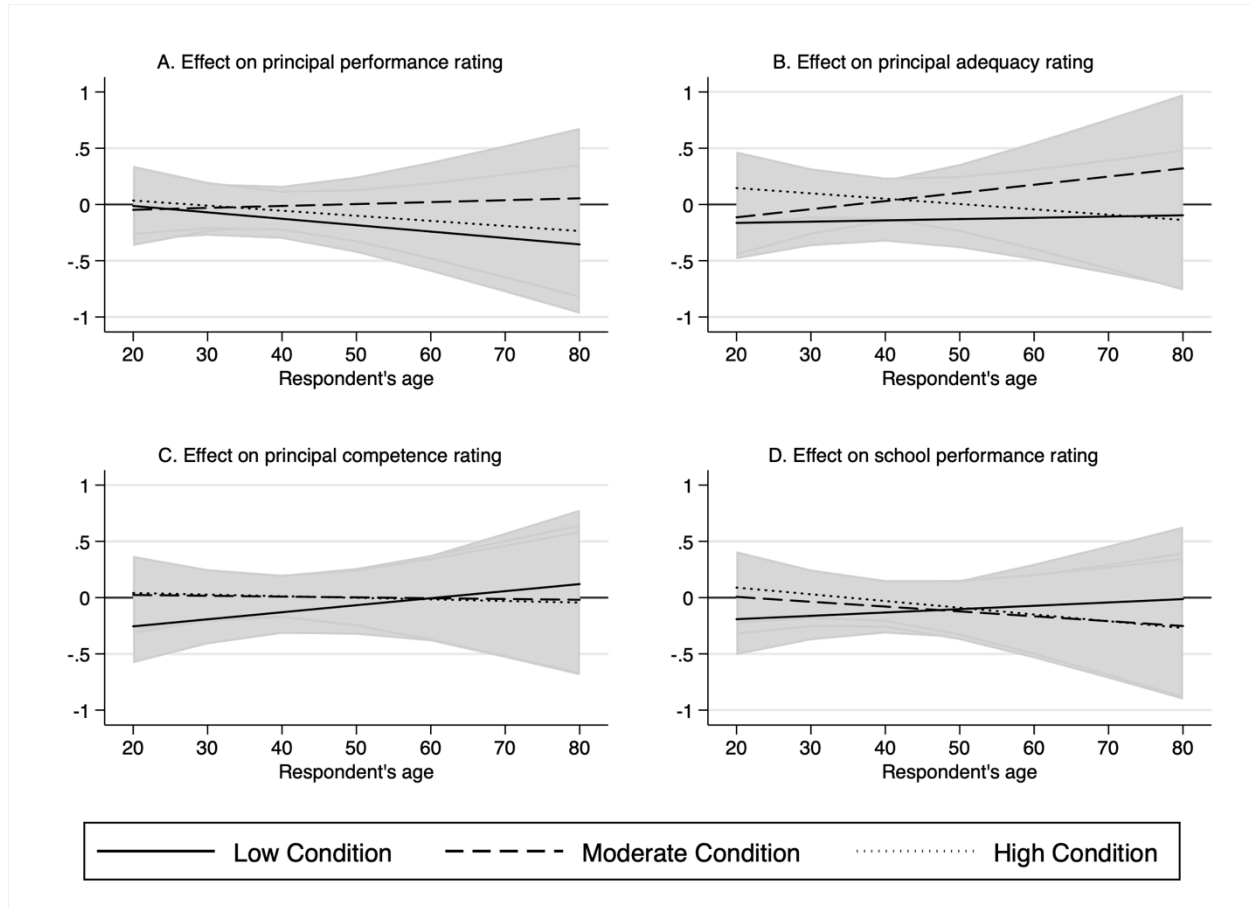
Table A5. Regression Coefficients for Study 1 with Interactions by Respondent Gender

	(1) Performance	(2) Adequacy	(3) Competence	(4) Sch. Perf.
Female principal	-0.221 (0.154)	-0.333* (0.162)	-0.320+ (0.165)	-0.420** (0.161)
Moderate performance	1.833*** (0.160)	1.701*** (0.169)	1.692*** (0.172)	1.991*** (0.167)
High performance	3.487*** (0.160)	3.212*** (0.168)	3.136*** (0.171)	3.594*** (0.167)
Female principal x Moderate performance	0.025 (0.222)	0.181 (0.234)	0.069 (0.238)	0.103 (0.232)
Female principal x High performance	0.152 (0.221)	0.431+ (0.233)	0.364 (0.237)	0.486* (0.231)
Female respondent	-0.413** (0.154)	-0.477** (0.162)	-0.408* (0.165)	-0.766*** (0.161)
Female principal x Female respondent	0.180 (0.213)	0.334 (0.225)	0.304 (0.229)	0.504* (0.223)
Moderate performance x Female respondent	0.290 (0.217)	0.303 (0.229)	0.223 (0.233)	0.509* (0.227)
High performance x Female respondent	0.662** (0.217)	0.793*** (0.229)	0.692** (0.233)	0.998*** (0.227)
Female prin. x Moderate perf. x Female	0.092 (0.304)	-0.096 (0.320)	0.117 (0.326)	-0.119 (0.318)
Female prin. x High perf. x Female	-0.133 (0.302)	-0.397 (0.318)	-0.328 (0.324)	-0.665* (0.316)
Constant	3.909*** (0.111)	4.167*** (0.117)	4.457*** (0.120)	3.790*** (0.116)
R-squared	0.526	0.483	0.454	0.544
Observations	2423	2423	2423	2423

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.

Next, we examine whether the gender treatment has different effects across respondent age. Figure A5 presents the average marginal effects of the principal's gender for each performance condition across respondent age. The gender treatment remains statistically insignificant across respondent age at all performance levels. Table A6 presents regression coefficients for the models that include interactions by respondent age.

Figure A5. Average marginal effects of gender treatment for study 1 by respondent age



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table A6.

Finally, we test for heterogeneous treatment effects across respondent partisanship and political ideology. Figure A6 presents the average marginal effects of the principal's gender for each performance condition across respondent partisanship. We measure partisanship using a 5-point scale, where 1 = Republican, 2 = Independent but lean Republican, 3 = Independent with no party preference, 4 = Independent but lean Democrat, and 5 = Democrat. We find evidence that Republican and Republican-leaning respondents in the low performance condition tend to evaluate female principals less favorably than male principals across all four dependent variables, as shown in panels A-D of Figure A6. The marginal effect of the gender treatment is negative and statistically significant in the low performance condition for Republicans, independents who lean Republican, and independents with no party preference. These respondents evaluate female principals lower than male principals in terms of principal performance (p-value for Republican respondents = 0.006, p-value for independents who lean Republican = 0.010, p-value for independents = 0.064), principal adequacy (p-values = 0.046, 0.039, 0.065), principal competence (p-values = 0.063, 0.050, 0.069), and school performance

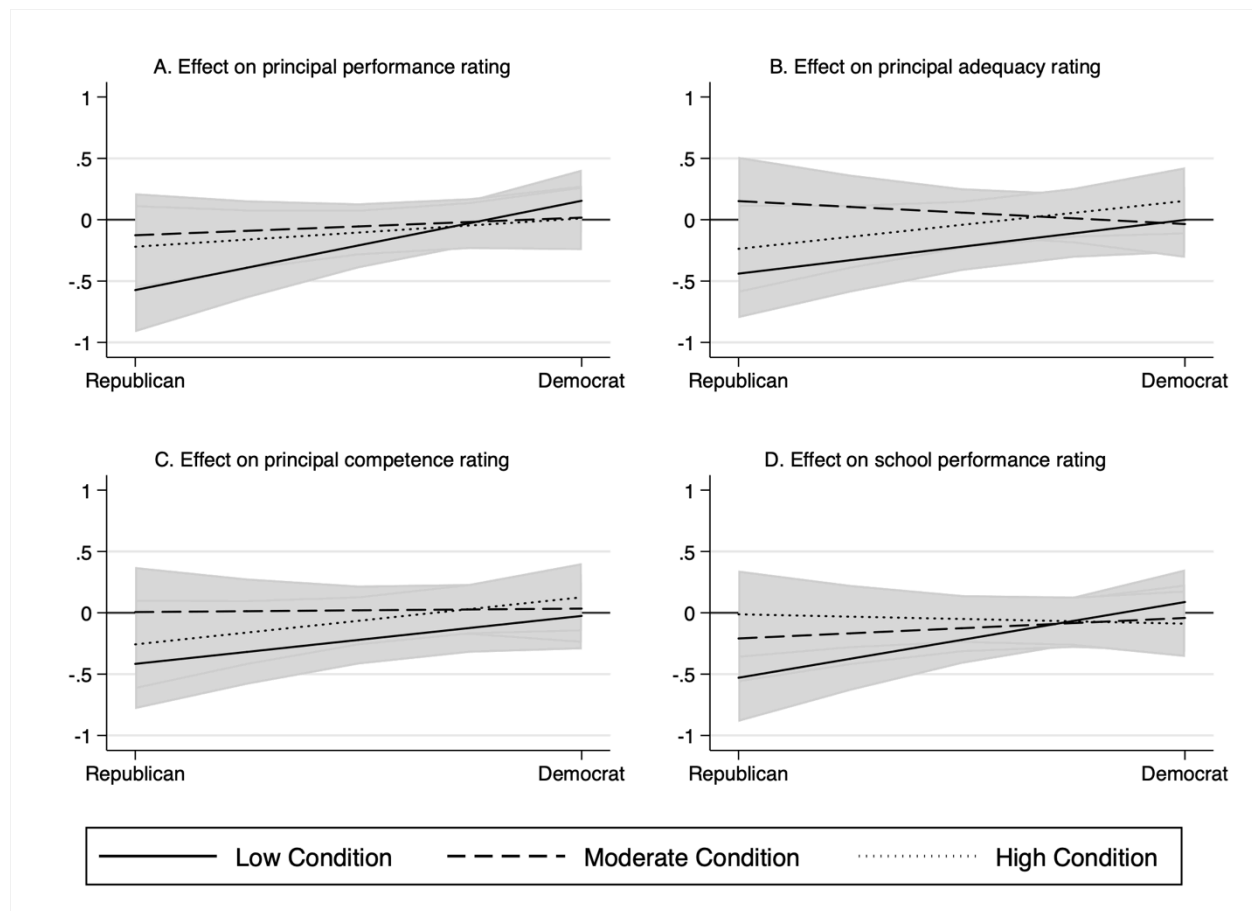
(p-values = 0.015, 0.018, 0.062). Table A7 presents regression coefficients for the models that include interactions by respondent partisanship.

Table A6. Regression Coefficients for Study 1 with Interactions by Respondent Age

	(1) Performanc e	(2) Adequacy	(3) Competenc e	(4) Sch. Perf.
Female principal	0.101 (0.338)	-0.187 (0.357)	-0.380 (0.363)	-0.251 (0.354)
Moderate performance	2.121*** (0.347)	1.857*** (0.366)	1.644*** (0.372)	2.031*** (0.363)
High performance	3.817*** (0.340)	3.450*** (0.358)	3.271*** (0.365)	3.579*** (0.356)
Female principal x Moderate performance	-0.183 (0.490)	-0.073 (0.517)	0.418 (0.526)	0.344 (0.513)
Female principal x High performance	0.025 (0.477)	0.427 (0.503)	0.449 (0.512)	0.461 (0.499)
Age	0.001 (0.006)	-0.006 (0.006)	-0.008 (0.007)	-0.008 (0.006)
Female principal x Respondent Age	-0.006 (0.008)	0.001 (0.009)	0.006 (0.009)	0.003 (0.009)
Moderate performance x Respondent Age	-0.004 (0.009)	-0.000 (0.009)	0.004 (0.009)	0.006 (0.009)
High performance x Respondent Age	0.001 (0.009)	0.005 (0.009)	0.007 (0.009)	0.015 (0.009)
Female prin. x Moderate perf. x Age	0.007 (0.012)	0.006 (0.013)	-0.007 (0.013)	-0.007 (0.013)
Female prin. x High perf. x Age	0.001 (0.012)	-0.006 (0.013)	-0.008 (0.013)	-0.009 (0.012)
Constant	3.655*** (0.239)	4.119*** (0.252)	4.528*** (0.256)	3.677*** (0.250)
R-squared	0.524	0.480	0.452	0.540
Observations	2446	2446	2446	2446

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.

Figure A6. Average marginal effects of gender treatment for study 1 by partisanship



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table A7.

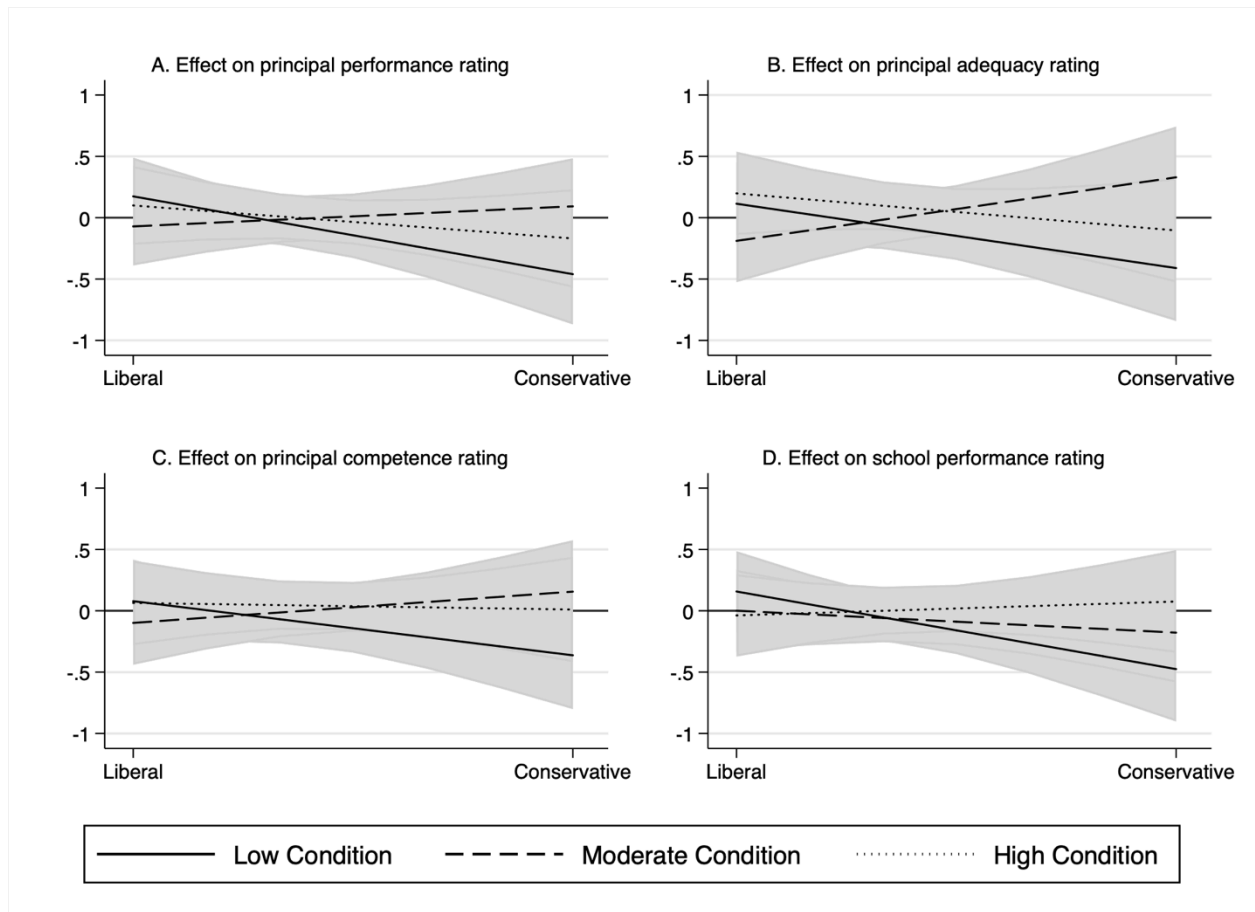
Figure A7 presents the average marginal effects of the principal's gender for each performance condition across respondents' political ideology. We measure political ideology using a 7-point scale where 1 = extremely liberal, 2 = liberal, 3 = slightly liberal, 4 = moderate or middle of the road, 5 = slightly conservative, 6 = conservative, and 7 = extremely conservative. We find that conservative respondents in the low performance condition evaluate female principals less favorably than male principals on two of the four dependent variables. Respondents who are slightly conservative, conservative, and extremely conservative rate female principals lower in terms of principal performance (p-values, respectively, are 0.085, 0.066, 0.064), as shown in panel A of Figure A7, and in terms of school performance (p-values = 0.079, 0.066, 0.067), as shown in panel D of Figure A7, compared to male principals under conditions of low performance. Table A8 presents regression coefficients for the models that include interactions by respondent political ideology.

Table A7. Regression Coefficients for Study 1 with Interactions by Respondent Partisanship

	(1) Performance	(2) Adequacy	(3) Competence	(4) Sch. Perf.
Female principal	-0.756** (0.272)	-0.549+ (0.288)	-0.514+ (0.292)	-0.684* (0.285)
Moderate performance	1.522*** (0.283)	1.448*** (0.298)	1.606*** (0.303)	1.888*** (0.296)
High performance	3.676*** (0.280)	3.521*** (0.295)	3.685*** (0.300)	4.050*** (0.293)
Female principal x Moderate performance	0.593 (0.385)	0.747+ (0.406)	0.513 (0.413)	0.433 (0.403)
Female principal x High performance	0.478 (0.383)	0.214 (0.405)	0.161 (0.412)	0.691+ (0.401)
Political partisanship	-0.109* (0.053)	-0.056 (0.056)	-0.036 (0.057)	-0.080 (0.056)
Female principal x Respondent Partisanship	0.182* (0.072)	0.109 (0.076)	0.098 (0.077)	0.154* (0.075)
Moderate performance x Respondent Partisanship	0.131+ (0.074)	0.111 (0.078)	0.059 (0.079)	0.108 (0.077)
High performance x Respondent Partisanship	0.052 (0.073)	0.043 (0.078)	-0.042 (0.079)	0.030 (0.077)
Female prin. x Moderate perf. x Partisanship	-0.146 (0.102)	-0.156 (0.108)	-0.091 (0.110)	-0.113 (0.107)
Female prin. x High perf. x Partisanship	-0.124 (0.102)	-0.012 (0.108)	-0.002 (0.109)	-0.173 (0.107)
Constant	4.077*** (0.206)	4.115*** (0.217)	4.365*** (0.221)	3.664*** (0.215)
R-squared	0.528	0.483	0.457	0.545
Observations	2355	2355	2355	2355

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models were estimated using ordinary least squares. Standard errors are in parentheses.

Figure A7. Average marginal effects of gender treatment for study 1 by respondent ideology



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table A8.

Table A8. Regression Coefficients for Study 1 with Interactions by Respondent Ideology

	(1) Performanc e	(2) Adequacy	(3) Competenc e	(4) Sch. Perf.
Female principal	0.280 (0.246)	0.201 (0.260)	0.152 (0.264)	0.263 (0.257)
Moderate performance	2.125*** (0.248)	1.998*** (0.262)	1.848*** (0.266)	2.253*** (0.259)
High performance	3.834*** (0.254)	3.643*** (0.268)	3.497*** (0.272)	4.055*** (0.265)
Female principal x Moderate performance	-0.378 (0.349)	-0.477 (0.368)	-0.293 (0.373)	-0.232 (0.364)
Female principal x High performance	-0.135 (0.351)	0.047 (0.371)	-0.079 (0.376)	-0.319 (0.366)
Political ideology	0.040 (0.048)	0.008 (0.051)	0.004 (0.051)	0.009 (0.050)
Female principal x Respondent Ideology	-0.106+ (0.064)	-0.087 (0.067)	-0.074 (0.068)	-0.105 (0.066)
Moderate performance x Respondent Ideology	-0.049 (0.065)	-0.047 (0.069)	-0.011 (0.070)	-0.002 (0.068)
High performance x Respondent Ideology	0.001 (0.065)	0.004 (0.069)	0.011 (0.070)	0.024 (0.068)
Female prin. x Moderate perf. x Ideology	0.133 (0.089)	0.174+ (0.094)	0.116 (0.095)	0.076 (0.092)
Female prin. x High perf. x Ideology	0.061 (0.090)	0.037 (0.095)	0.065 (0.096)	0.124 (0.094)
Constant	3.539*** (0.179)	3.868*** (0.189)	4.201*** (0.191)	3.334*** (0.186)
R-squared	0.525	0.481	0.455	0.544
Observations	2383	2383	2383	2383

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.

APPENDIX B: SUPPORTING INFORMATION FOR STUDY 2 (POLICE CONTEXT)

Figures B1 and B2 present examples of the treatments that were shown to survey respondents for study 2. Figure A1 presents the male police captain treatment condition. The female treatment condition used the name “Amy” instead of “Scott” and feminine gender pronouns instead of masculine ones.

Figure B2 presents the scorecard for the male captain, high performance condition. The high performance condition showed three green checkmarks and one red X. The moderate performance condition showed two green checkmarks and two red X’s. The low performance condition showed one green checkmark and three red X’s. We randomly assigned the location of the checkmarks and X’s within each performance condition to offset the possibility that respondents view particular performance metrics as more important than others. For example, each respondent allocated to the high performance condition had an equal probability of viewing the red X in either the clearance rate box, non-violent crimes box, violent crimes box, or the response times box. Only the location of the checkmarks and X’s changed. The relative location of the performance criteria on the scorecards remained fixed across all conditions.

Table B1 presents descriptive statistics and frequencies for the sample used in study 2. Only participants who passed all of the screening checks in MTurk are included in the sample. Table B2 presents balance checks for study 2 across key respondent demographics, including gender, age, political ideology, and partisanship. Table B3 presents the regression coefficients for the main models for study 2 presented in the text. The models were estimated using ordinary least squares with interaction terms included for the treatment groups.

Figure B1. Example of Introductory Screen for Study 2 (Male Police Captain Condition)

Scott Schmidt is the police captain of the 5th police precinct in a city with a population of approximately 250,000 residents. As the captain, he is the highest-ranking officer in the precinct and is responsible for its performance. He has several support staff and oversees approximately 110 police officers and civilian law enforcement personnel. His precinct covers a jurisdiction with approximately 50,000 residents.

On the next screen you will be shown an annual precinct-level report card with information on the 5th precinct's performance on a number of aspects related to public safety.

For the purpose of assessment, the card reports green check marks for areas where the precinct has improved (✓) and red X's for areas where the precinct has worsened (✗).

Figure B2. Example Scorecard for Study 2 (Male Captain, High Performance Condition)

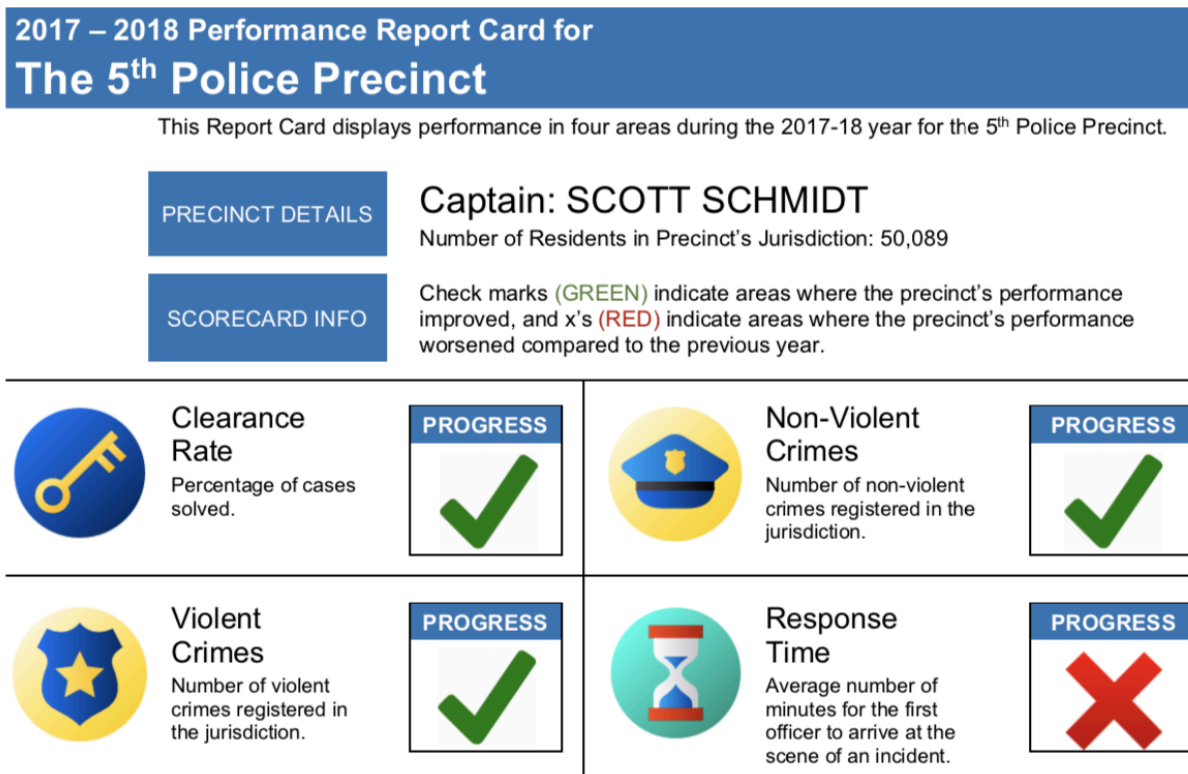


Table B1. Descriptive Statistics and Frequencies for Study 2

Variable	Operationalization	N	%	Min	Max	Mean	SD
<u><i>Dependent Variables</i></u>							
Captain Performance	<i>How would you rate Captain {Amy/Scott} Schmidt's performance?</i>	2,491		0	10	5.85	2.31
Captain Adequacy	<i>How would you rate Captain {Amy/Scott} Schmidt's adequacy?</i>	2,491		0	10	6.04	2.29
Captain Competence	<i>How would you rate Captain {Amy/Scott} Schmidt's competence?</i>	2,491		0	10	6.25	2.28
Precinct Performance	<i>How would you rate the 5th Precinct's overall performance?</i>	2,491		0	10	5.78	2.37
<u><i>Attention Checks</i></u>							
Gender Check	Indicator. Correct recall of the captain's gender (Correct = 1)	2,491		0	1	.92	.28
Performance Check	Indicator. Correct recall of number of improved areas (Correct = 1)	2,491		0	1	.87	.32
<u><i>Demographics</i></u>							
Gender	Respondent's gender (dichotomized)	2,423	100.0				
	"Male"	1,099	44.46				
	"Female"	1,373	55.54				
Age	Respondents' age (derived from birth year)	2,491		18	83	37.80	11.81
Education	Highest level of education	2,491	100.0	1	7	4.26	1.32
	"Less than high school"	18	.72				
	"High school graduate"	267	10.72				
	"Some college"	555	22.28				
	"2-year college"	270	44.56				
	"4-year college"	996	39.98				
	"Professional degree"	349	14.01				
	"Doctorate"	36	1.45				
Race/Ethnicity	Respondent's race or ethnicity (categories are not mutually exclusive)						
	"African American or Black"	312	12.53				
	"Asian American or Asian"	186	7.47				
	"Hispanic or Latino/Latina/Latinx"	154	6.18				
	"Middle Eastern or Arab"	7	.28				
	"Multiracial"	61	2.45				
	"Pacific Islander"	20	.80				
	"White or Caucasian"	1,832	73.54				
	"American Indian/Native American"	25	.72				
	"Other/Self-Description"	18	1.00				

Employment	Employment Status	2,491	100.00				
	"Full-time"	1,419	56.97				
	"Part-time"	255	10.24				
	"Unemployed, looking for work"	140	5.62				
	"Unemployed, not looking for work"	131	5.26				
	"Self-employed"	299	12.00				
	"Retired"	87	3.49				
	"Student"	98	3.93				
	"Currently unable to work"	62	2.49				
Income	Yearly household income in \$10,000 brackets	2,491		1	13	6.41	3.43
Marital Status	Indicator: Single, never married (Yes = 1)	2,491		0	1	.37	.48
Liberal-Conservative	<i>Where would you place yourself among the following categories?</i>	2,439		1	7	3.57	1.76
	"Extremely liberal"	301	12.34				
	"Liberal"	559	22.92				
	"Slightly liberal"	329	13.49				
	"Moderate; middle of the road"	540	22.14				
	"Slightly conservative"	264	10.82				
	"Conservative"	309	12.67				
	"Extremely conservative"	137	5.62				
Democrat-Republican	<i>In politics today, how do you consider yourself?</i>	2,464	100.00	1	5	3.35	1.54
	"Republican"	492	20.44				
	"Independent, but lean Republican"	262	10.88				
	"Independent, with no party preference"	416	17.28				
	"Independent, but lean Democrat"	390	16.20				
	"Democrat"	847	35.19				
Worked in Law Enforcement	Indicator: Ever worked in law enforcement (Yes = 1, No = 2)	2,491		1	2	1.94	.24
Family in Law Enforcement	Indicator: Family member works in law enforcement (Yes = 1, No = 2)	2,499		1	2	1.86	.34

Table B2. Treatment Group Balance Across Key Demographic Variables for Study 2

DV	Gender Condition	Performance Condition			Test for Mean Differences		
		Low (3)	Moderate (4)	High (5)	3 vs. 4	3 vs. 5	4 vs. 5
<i>Rater Gender</i>	Male (1)	0.59	0.54	0.53	NS	.06+	NS

	Female (2)	0.57	0.54	0.57	NS	NS	NS
<i>Test for Mean Differences</i>	1 vs. 2	NS	NS	NS			
<i>Rater Age</i>	Male (1)	37.6 2	38.18	37.6 0	NS	NS	NS
	Female (2)	38.1 2	37.09	38.2 2	NS	NS	NS
<i>Test for Mean Differences</i>	1 vs. 2	NS	NS	NS			
<i>Liberal-Conservative Orientation</i>	Male (1)	3.52	3.54	3.68	NS	NS	NS
	Female (2)	3.69	3.51	3.46	NS	.23+	NS
<i>Test for Mean Differences</i>	1 vs. 2	NS	NS	.22+			
<i>Republican-Democrat Orientation</i>	Male (1)	3.46	3.32	3.26	NS	.18+	NS
	Female (2)	3.18	3.39	3.47	NS	NS	NS
<i>Test for Mean Differences</i>	1 vs. 2	.27*	NS	-.20 +			

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. NS = Not statistically significant. Estimates are from two-tailed *t* tests for mean group differences.

Table B3. Regression Coefficients for Study 2 (Police Context)

	(1) Performanc e	(2) Adequacy	(3) Competence	(4) Precinct Perf.
Female captain	0.155 (0.117)	0.034 (0.118)	0.163 (0.119)	0.147 (0.119)
Moderate performance	2.089*** (0.116)	2.021*** (0.117)	1.951*** (0.118)	2.157*** (0.118)
High performance	3.894*** (0.113)	3.740*** (0.115)	3.609*** (0.116)	3.914*** (0.116)
Female captain x Moderate performance	-0.166 (0.167)	0.015 (0.169)	0.064 (0.171)	-0.179 (0.171)
Female captain x High performance	-0.197 (0.166)	-0.114 (0.168)	-0.071 (0.170)	-0.004 (0.170)
Constant	3.850*** (0.080)	4.134*** (0.081)	4.337*** (0.082)	3.727*** (0.082)

R-squared	0.460	0.440	0.420	0.463
Observations	2491	2491	2491	2491

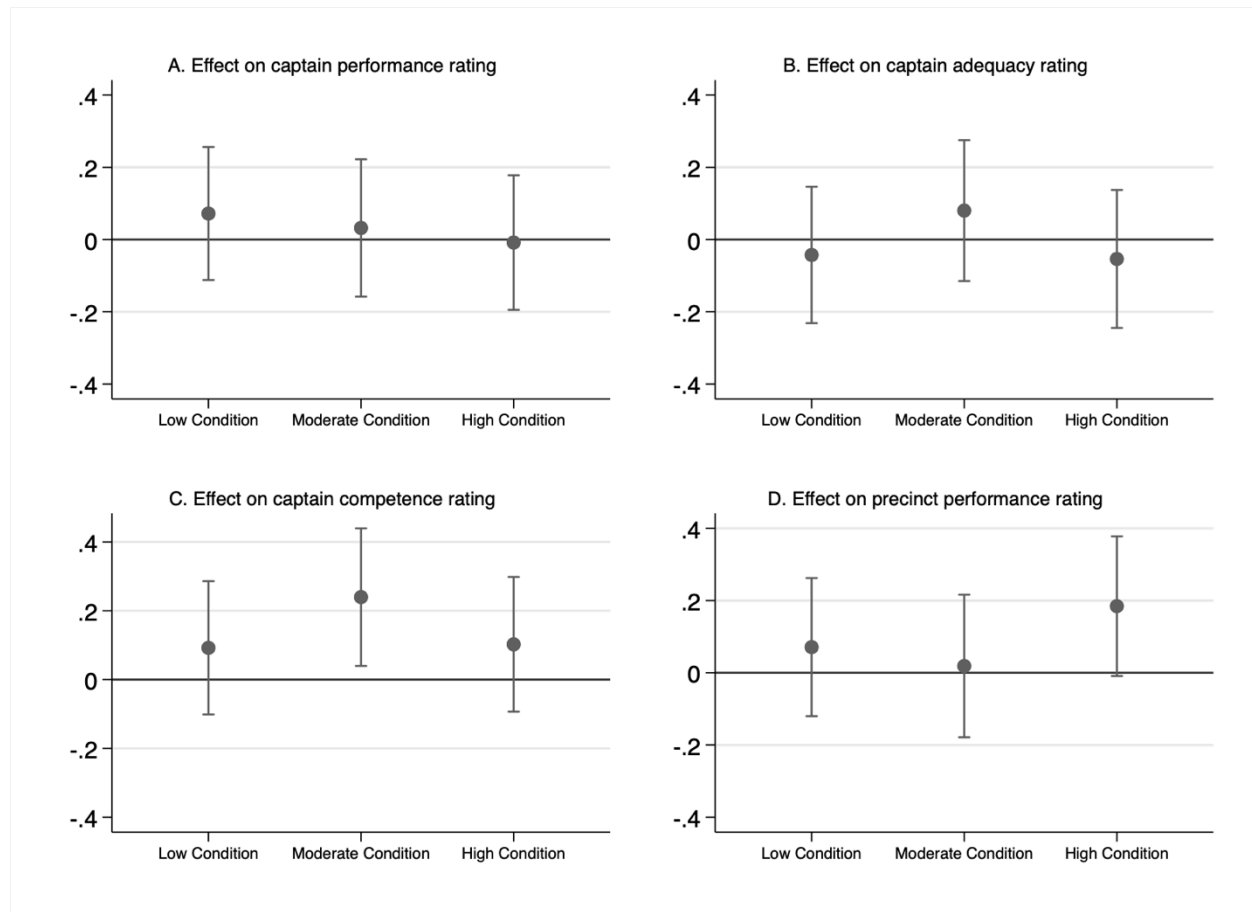
Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.

Robustness Checks for Study 2 (Police Context): Including Control Variables

Figure B3 presents the average marginal effects of the gender treatment across each performance treatment condition when control variables are included in the regression models. The main results are robust to the inclusion of multiple variables controlling for respondent demographics and attention checks. The average marginal effect of the captain's gender continues to be positive and statistically significant (p-value = 0.049) in terms of captain competence in the moderate performance condition, as shown in panel C of Figure B3. The marginal effect of gender also approaches, but does not reach, conventional levels of statistical significance in terms of precinct performance in the high performance condition (p-value = 0.117), as shown in panel D of Figure B3.

Table B4 presents the regression coefficients for the models including control variables for Study 2.

Figure B3. Average marginal effects of gender treatment for study 2 with controls



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table B4.

Table B4. Regression Coefficients for Study 2 with Controls

	(1) Performanc e	(2) Adequacy	(3) Competence	(4) Precinct Perf.
Female captain	0.072 (0.112)	-0.043 (0.115)	0.092 (0.118)	0.071 (0.116)
Moderate performance	2.063*** (0.111)	2.002*** (0.114)	1.936*** (0.117)	2.134*** (0.115)
High performance	3.836*** (0.109)	3.687*** (0.112)	3.549*** (0.115)	3.853*** (0.113)
Female captain x Moderate performance	-0.040 (0.161)	0.123 (0.165)	0.147 (0.169)	-0.052 (0.167)
Female captain x High performance	-0.080	-0.011	0.010	0.113

	(0.160)	(0.164)	(0.168)	(0.166)
Captain gender correct	0.011	-0.019	-0.081	-0.215+
	(0.123)	(0.126)	(0.129)	(0.128)
Performance check correct	-1.138***	-0.984***	-0.829***	-1.061***
	(0.115)	(0.118)	(0.121)	(0.119)
Female respondent	0.026	-0.023	0.048	-0.034
	(0.068)	(0.070)	(0.072)	(0.071)
Age	-0.004	-0.007+	-0.008*	-0.006
	(0.003)	(0.004)	(0.004)	(0.004)
Education	0.064*	0.065*	0.052+	0.061*
	(0.028)	(0.028)	(0.029)	(0.029)
African American or Black	0.051	0.030	0.079	0.076
	(0.171)	(0.176)	(0.180)	(0.178)
Asian American or Asian	-0.326+	-0.253	-0.278	-0.134
	(0.190)	(0.195)	(0.199)	(0.197)
Hispanic or Latino/Latina/Latinx	-0.032	-0.113	-0.143	0.038
	(0.174)	(0.178)	(0.183)	(0.180)
Middle Eastern or Arab	-0.733	-0.058	-0.948	0.146
	(0.618)	(0.634)	(0.650)	(0.642)
Multiracial	-0.299	-0.164	0.102	-0.001
	(0.229)	(0.235)	(0.241)	(0.238)
Pacific Islander	0.027	0.061	-0.044	-0.051
	(0.386)	(0.396)	(0.406)	(0.401)
White or Caucasian	-0.223	-0.148	-0.128	-0.072
	(0.159)	(0.163)	(0.167)	(0.165)
Self-identify or Other	0.549	0.683	0.385	0.383
	(0.472)	(0.484)	(0.497)	(0.490)
American Indian or Native American	-0.033	-0.095	-0.074	-0.033
	(0.323)	(0.331)	(0.340)	(0.335)
Employed part-time in an organization	0.125	0.163	-0.025	0.051
	(0.115)	(0.118)	(0.121)	(0.119)
Unemployed - looking for work	-0.203	-0.187	-0.162	-0.233
	(0.151)	(0.155)	(0.159)	(0.157)
Unemployed - not looking for work	-0.060	-0.147	-0.105	-0.185
	(0.154)	(0.158)	(0.162)	(0.160)
Retired	0.178	0.263	0.229	0.106
	(0.203)	(0.208)	(0.213)	(0.211)
Student	-0.160	-0.204	-0.243	-0.152
	(0.179)	(0.184)	(0.188)	(0.186)
Currently unable to work	0.152	0.419+	0.139	0.070
	(0.226)	(0.231)	(0.237)	(0.234)
Self-employed	-0.113	-0.093	-0.153	-0.182
	(0.108)	(0.111)	(0.113)	(0.112)
Income category	-0.009	-0.000	0.002	-0.011

	(0.011)	(0.011)	(0.011)	(0.011)
Single or unmarried	-0.124	-0.154+	-0.088	-0.122
	(0.079)	(0.081)	(0.083)	(0.082)
Political ideology	-0.006	-0.008	0.039	-0.010
	(0.029)	(0.030)	(0.030)	(0.030)
Political partisanship	-0.092**	-0.058+	-0.024	-0.085*
	(0.033)	(0.034)	(0.035)	(0.035)
Worked in law enforcement	-0.687***	-0.768***	-0.586***	-0.568***
	(0.162)	(0.167)	(0.171)	(0.169)
Family in law enforcement	-0.390***	-0.307**	-0.297**	-0.347**
	(0.107)	(0.110)	(0.113)	(0.112)
Constant	7.464***	7.537***	7.052***	7.156***
	(0.403)	(0.413)	(0.424)	(0.418)
R-squared	0.534	0.503	0.468	0.523
Observations	2367	2367	2367	2367

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.

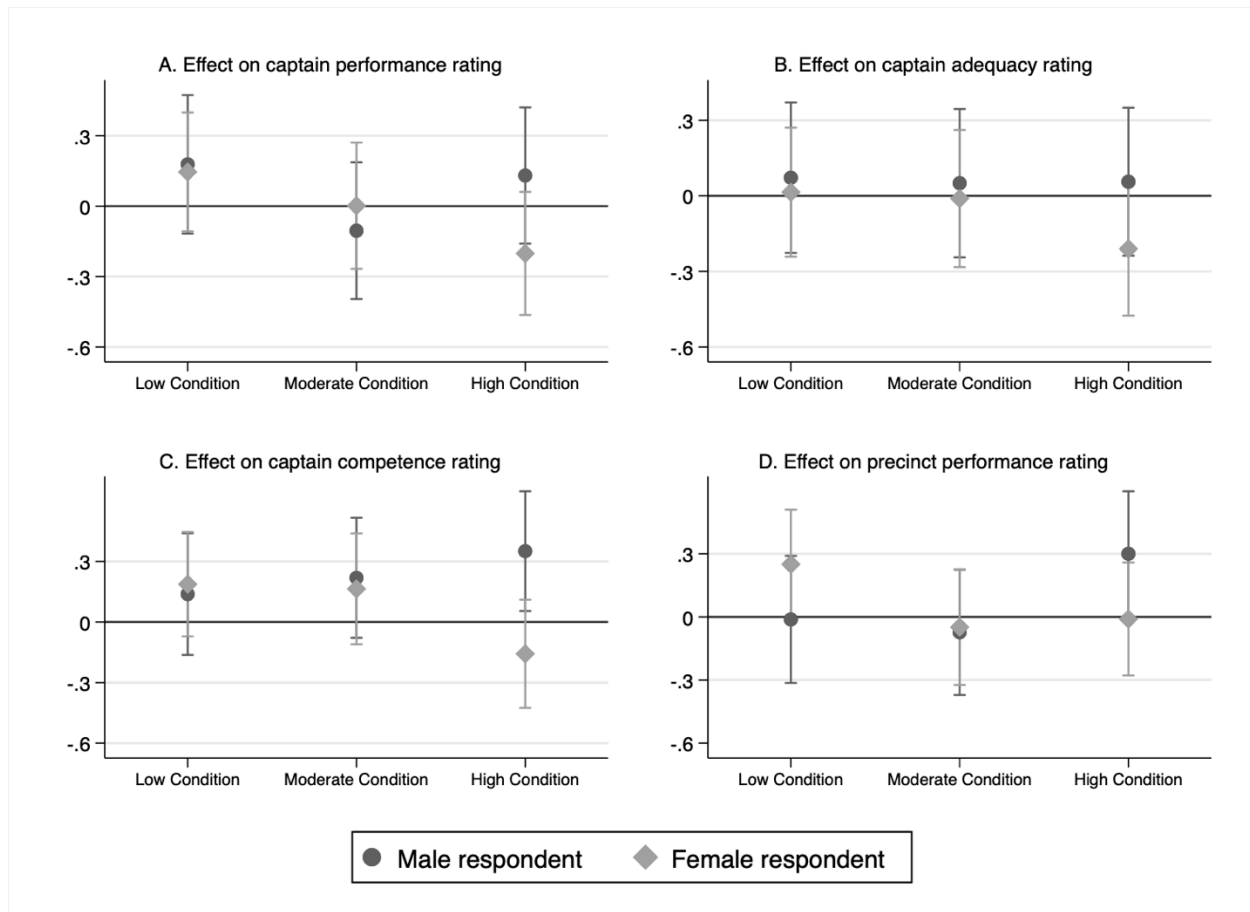
Robustness Checks for Study 2 (Police Context): Heterogeneous Treatment Effects

Next, we explore potential heterogeneous treatment effects across respondent gender, age, partisanship, and political ideology. Some individuals, such as males, older respondents, Republicans, and ideologically conservative respondents, might be more easily swayed by gender stereotypes, and thus more likely to perceive incongruities between women and management roles. To examine whether the treatment conditions have different effects on different groups of respondents (i.e., heterogeneous treatment effects), we include three-way interaction terms in each of the regression models that interact the two treatment groups with each of these respondent demographics.

We first examine whether male and female respondents react differently to male and female public managers. Figure B4 shows the average marginal effects of the captain's gender across each performance condition for both male and female respondents. The figure shows that, in general, female respondents and male respondents evaluate male and female police captains similarly. The effect of the captain's gender remains statistically indistinguishable from zero, except in a few cases. In particular, the average marginal effect of the captain's gender is positive and statistically significant for male respondents in the high performance condition for two of the four dependent variables. Male respondents in the high performance condition evaluate the female captain higher than the male captain in terms of competence (p-value = 0.051), as shown in panel C in Figure B4, and in terms of precinct performance (p-value = 0.096), as shown in panel D in Figure B4.

Table B5 presents regression coefficients for the models that include interactions by respondent gender.

Figure B4. Average marginal effects of gender treatment for study 2 by respondent gender



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table B5.

Next, we examine whether the gender treatment has different effects across respondent age. Figure B5 presents the average marginal effects of the captain's gender for each performance condition across respondent age. The gender treatment remains statistically insignificant across respondent age, except in a few cases. The average marginal effect of the captain's gender is positive and statistically significant in terms of captain performance for 20-year-old (p -value = 0.080) and 30-year-old respondents (p -value = 0.072) in the low performance condition, as shown in panel A of Figure B5, and positive and statistically significant in terms of captain competence for 30-year-old respondents in the moderate performance condition (p -value = 0.060), as shown in panel C of Figure B5. These findings suggest that younger respondents might evaluate female captains more favorably compared to male captains on certain dimensions under conditions of low or moderate performance.

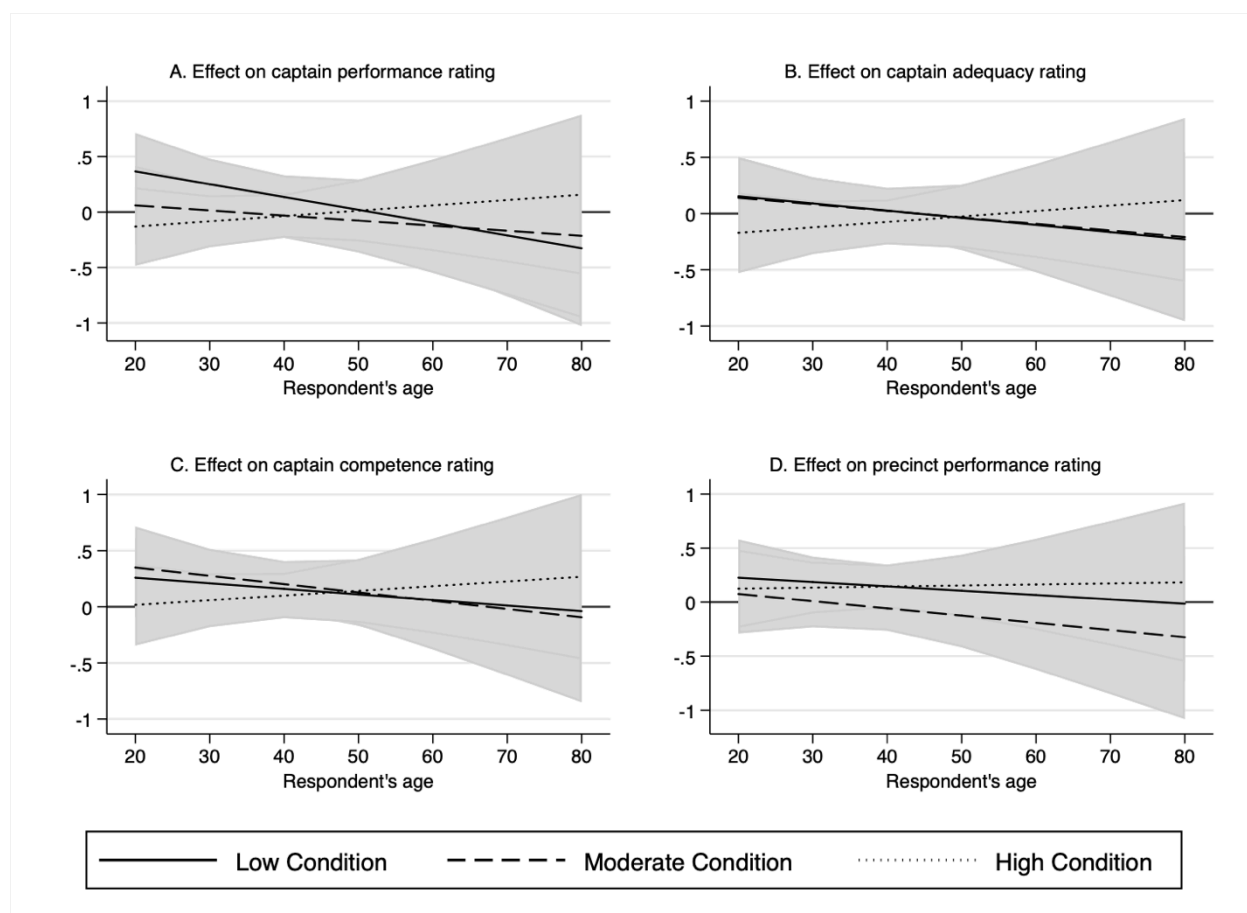
Table B6 presents regression coefficients for the models that include interactions by respondent age.

Table B5. Regression Coefficients for Study 2 with Interactions by Respondent Gender

	(1) Performanc e	(2) Adequacy	(3) Competence	(4) Precinct Perf.
Female captain	0.178 (0.179)	0.072 (0.181)	0.138 (0.183)	-0.012 (0.183)
Moderate performance	2.113*** (0.176)	1.978*** (0.178)	1.914*** (0.180)	1.952*** (0.180)
High performance	3.779*** (0.171)	3.527*** (0.173)	3.379*** (0.175)	3.549*** (0.175)
Female captain x Moderate performance	-0.283 (0.252)	-0.022 (0.255)	0.081 (0.258)	-0.061 (0.258)
Female captain x High performance	-0.048 (0.251)	-0.016 (0.255)	0.213 (0.257)	0.312 (0.257)
Female respondent	-0.089 (0.162)	-0.185 (0.164)	-0.126 (0.166)	-0.449** (0.166)
Female captain x Female respondent	-0.033 (0.236)	-0.058 (0.239)	0.049 (0.241)	0.262 (0.242)
Moderate performance x Female respondent	0.018 (0.234)	0.123 (0.237)	0.119 (0.239)	0.390 (0.239)
High performance x Female respondent	0.226 (0.229)	0.399+ (0.232)	0.440+ (0.234)	0.657** (0.234)
Female cap. x Moderate perf. x Female res.	0.139 (0.337)	-0.003 (0.341)	-0.104 (0.345)	-0.238 (0.345)
Female cap. x High perf. x Female res.	-0.300 (0.335)	-0.208 (0.339)	-0.558 (0.343)	-0.572+ (0.343)
Constant	3.903*** (0.124)	4.242*** (0.126)	4.414*** (0.127)	3.995*** (0.127)
R-squared	0.463	0.442	0.423	0.466
Observations	2472	2472	2472	2472

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.

Figure B5. Average marginal effects of gender treatment for study 2 by respondent age



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table B6.

Finally, we test for heterogeneous treatment effects across respondent partisanship and political ideology. Figure B6 presents the average marginal effects of the captain's gender for each performance condition across respondent partisanship. We measure partisanship using a 5-point scale, where 1 = Republican, 2 = Independent but lean Republican, 3 = Independent with no party preference, 4 = Independent but lean Democrat, and 5 = Democrat.

We find evidence that Democratic and Democratic-leaning respondents in the moderate performance condition evaluate female captains more favorably than male captains in terms of competence. The average marginal effect of the captain's gender is positive and statistically significant for captain competence for independents with no party preference (p-value = 0.062), independents who lean Democrat (p-value = 0.020), and Democrats (p-value = 0.034) in the moderate performance condition, as shown in panel C of Figure B6.

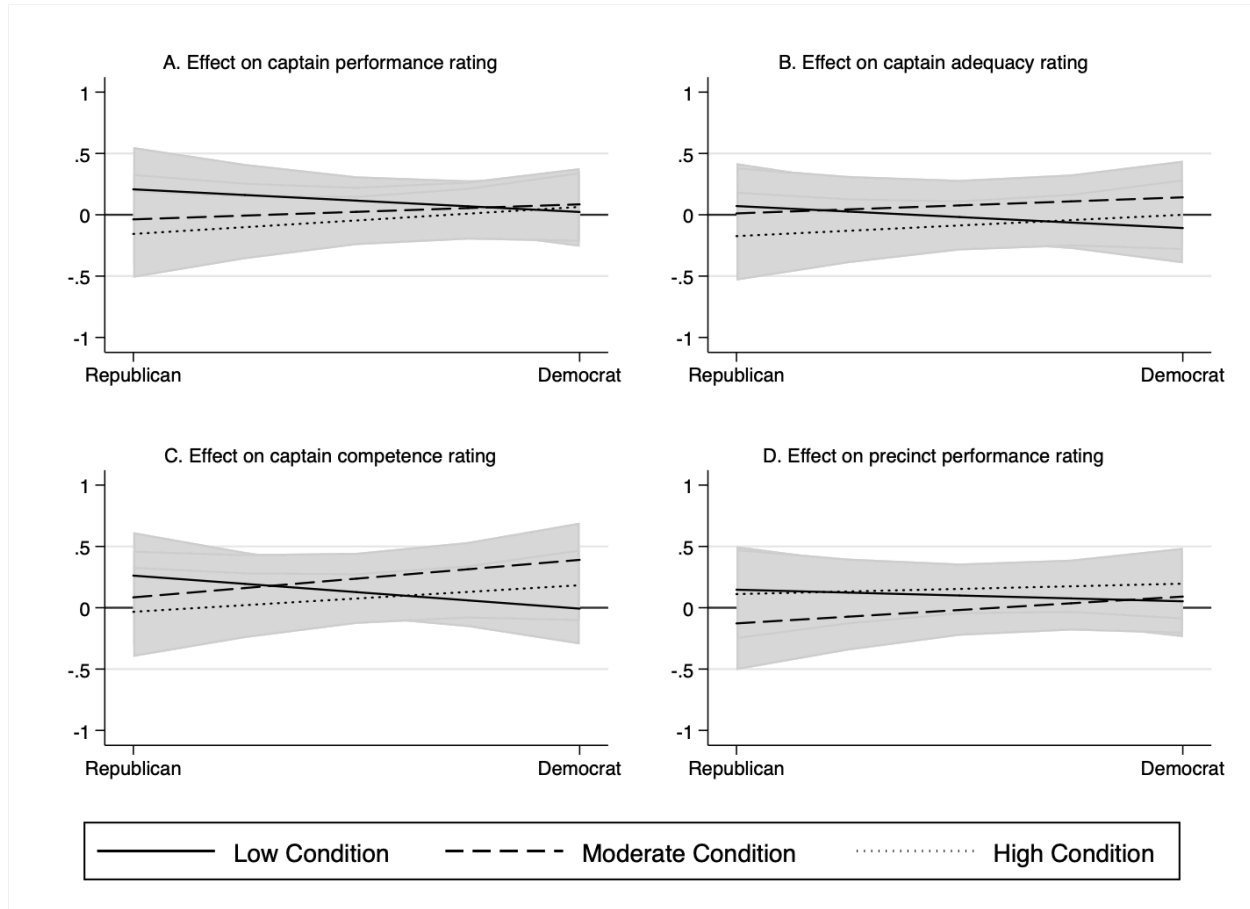
Table B7 presents regression coefficients for the models that include interactions by respondent partisanship.

Table B6. Regression Coefficients for Study 2 with Interactions by Respondent Age

	(1) Performanc e	(2) Adequacy	(3) Competenc e	(4) Precinct Perf.
Female captain	0.598 (0.387)	0.281 (0.391)	0.359 (0.395)	0.307 (0.395)
Moderate performance	2.157*** (0.392)	1.974*** (0.396)	1.609*** (0.401)	1.937*** (0.401)
High performance	3.584*** (0.382)	3.233*** (0.385)	3.154*** (0.390)	3.377*** (0.390)
Female captain x Moderate performance	-0.446 (0.559)	-0.024 (0.564)	0.139 (0.571)	-0.098 (0.571)
Female captain x High performance	-0.825 (0.555)	-0.550 (0.561)	-0.425 (0.567)	-0.202 (0.568)
Age	-0.005 (0.007)	-0.009 (0.007)	-0.013+ (0.007)	-0.012+ (0.007)
Female captain x Respondent Age	-0.012 (0.010)	-0.006 (0.010)	-0.005 (0.010)	-0.004 (0.010)
Moderate performance x Respondent Age	-0.002 (0.010)	0.001 (0.010)	0.009 (0.010)	0.006 (0.010)
High performance x Respondent Age	0.008 (0.010)	0.013 (0.010)	0.012 (0.010)	0.014 (0.010)
Female cap. x Moderate perf. x Age	0.007 (0.014)	0.001 (0.014)	-0.002 (0.014)	-0.003 (0.014)
Female cap. x High perf. x Age	0.016 (0.014)	0.011 (0.014)	0.009 (0.014)	0.005 (0.014)
Constant	4.035*** (0.268)	4.459*** (0.271)	4.811*** (0.274)	4.183*** (0.274)
R-squared	0.462	0.443	0.423	0.465
Observations	2491	2491	2491	2491

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.

Figure B6. Average marginal effects of gender treatment for study 2 by partisanship



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table B7.

Figure B7 presents the average marginal effects of the captain's gender for each performance condition across respondents' political ideology. We measure political ideology using a 7-point scale where 1 = extremely liberal, 2 = liberal, 3 = slightly liberal, 4 = moderate or middle of the road, 5 = slightly conservative, 6 = conservative, and 7 = extremely conservative. The average marginal effect of the captain's gender is positive and statistically significant in terms of captain competence for liberal (p-value = 0.077) and slightly liberal (p-value = 0.062) respondents in the moderate performance condition, as shown in panel C of Figure B7. The gender treatment is also positive and (nearly) statistically significant for precinct performance, as shown in panel D of Figure B7, for extremely liberal (p-value = 0.103), liberal (p-value = 0.087), and slightly liberal (p-value = 0.105) respondents in the high performance condition and for slightly conservative (p-value = 0.110), conservative (p-value = 0.094), and extremely conservative (p-value = 0.096) respondents in the low performance condition. Both liberals in the high performance condition and conservatives in the low performance condition evaluate the female captain higher than the male captain in terms of overall precinct performance.

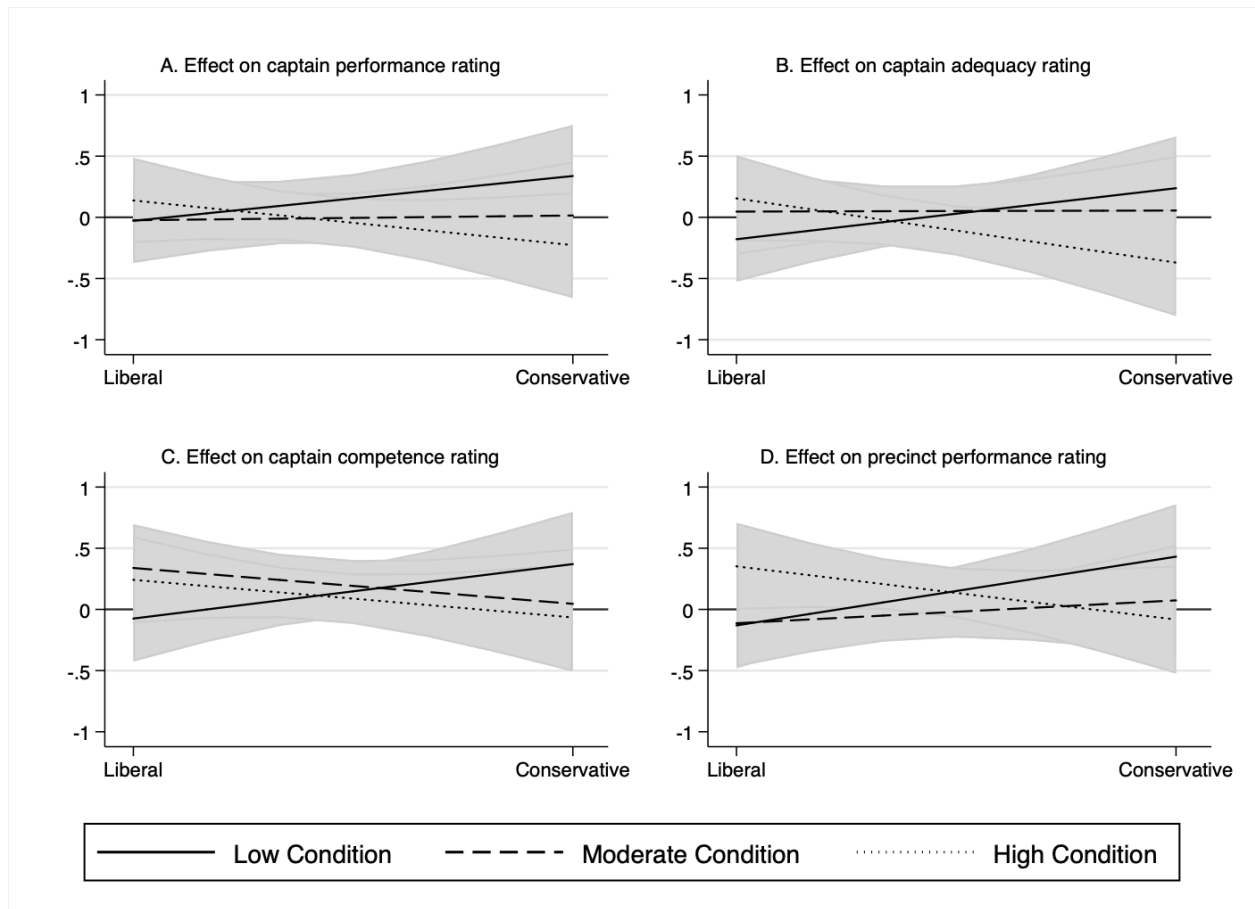
Table A8 presents regression coefficients for the models that include interactions by respondent political ideology.

Table B7. Regression Coefficients for Study 2 with Interactions by Respondent Partisanship

	(1) Performance	(2) Adequacy	(3) Competence	(4) Precinct Perf.
Female captain	0.254 (0.274)	0.116 (0.278)	0.330 (0.282)	0.171 (0.282)
Moderate performance	1.877*** (0.281)	1.945*** (0.285)	2.023*** (0.289)	2.092*** (0.289)
High performance	3.564*** (0.274)	3.439*** (0.278)	3.378*** (0.282)	3.534*** (0.282)
Female captain x Moderate performance	-0.322 (0.403)	-0.138 (0.409)	-0.321 (0.415)	-0.353 (0.415)
Female captain x High performance	-0.465 (0.395)	-0.333 (0.401)	-0.417 (0.406)	-0.081 (0.406)
Political partisanship	-0.207*** (0.052)	-0.152** (0.053)	-0.129* (0.053)	-0.184*** (0.053)
Female captain x Partisanship	-0.046 (0.075)	-0.045 (0.076)	-0.067 (0.077)	-0.024 (0.077)
Moderate performance x Partisanship	0.053 (0.075)	0.013 (0.077)	-0.035 (0.078)	0.011 (0.078)
High performance x Partisanship	0.090 (0.074)	0.082 (0.075)	0.060 (0.076)	0.100 (0.076)
Female cap. x Moderate perf. x Partisanship	0.077 (0.110)	0.078 (0.111)	0.144 (0.113)	0.078 (0.113)
Female cap. x High perf. x Partisanship	0.101 (0.107)	0.088 (0.109)	0.121 (0.110)	0.045 (0.110)
Constant	4.570*** (0.196)	4.671*** (0.199)	4.801*** (0.202)	4.378*** (0.202)
R-squared	0.473	0.449	0.427	0.470
Observations	2407	2407	2407	2407

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.

Figure B7. Average marginal effects of gender treatment for study 2 by respondent ideology



Note: 90% confidence intervals are shown. Estimates were derived from models 1-4 presented in Table B8.

Table B8. Regression Coefficients for Study 2 with Interactions by Respondent Ideology

	(1) Performanc e	(2) Adequacy	(3) Competence	(4) Precinct Perf.
Female captain	-0.090 (0.266)	-0.248 (0.269)	-0.149 (0.272)	-0.224 (0.272)
Moderate performance	2.083*** (0.261)	1.919*** (0.264)	1.713*** (0.267)	2.004*** (0.267)
High performance	3.891*** (0.262)	3.708*** (0.265)	3.618*** (0.268)	3.739*** (0.268)
Female captain x Moderate performance	0.060 (0.380)	0.294 (0.384)	0.536 (0.389)	0.080 (0.389)
Female captain x High performance	0.288 (0.379)	0.490 (0.383)	0.442 (0.388)	0.649+ (0.388)
Political ideology	0.097* (0.046)	0.062 (0.046)	0.072 (0.047)	0.049 (0.047)
Female captain x Ideology	0.061 (0.066)	0.069 (0.067)	0.074 (0.068)	0.094 (0.068)
Moderate performance x Ideology	-0.002 (0.066)	0.024 (0.067)	0.062 (0.068)	0.034 (0.068)
High performance x Ideology	-0.009 (0.065)	-0.001 (0.066)	-0.016 (0.067)	0.035 (0.067)
Female cap. x Moderate perf. x Ideology	-0.055 (0.096)	-0.068 (0.097)	-0.123 (0.098)	-0.063 (0.098)
Female cap. x High perf. x Ideology	-0.122 (0.095)	-0.157 (0.096)	-0.125 (0.097)	-0.166+ (0.097)
Constant	3.522*** (0.179)	3.936*** (0.181)	4.113*** (0.183)	3.584*** (0.183)
R-squared	0.463	0.442	0.422	0.464
Observations	2439	2439	2439	2439

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001. Models estimated using ordinary least squares. Standard errors in parentheses.