

Can Increasing Awareness of Gender Gaps in International Relations Help Close Them? Evidence from a Scholar Ranking Experiment

EMILY B. JACKSON 
Cornell University, USA

DANIEL MALINIAK 
William & Mary, USA

ERIC PARAJON 
University of North Carolina, Chapel Hill, USA

SUSAN PETERSON
William & Mary, USA

RYAN POWERS
The University of Georgia, USA

AND

MICHAEL J. TIERNEY
William & Mary, USA

Abstract: We report the results of a survey of international relations (IR) scholars on the use of an increasingly common policy designed to close recognition gaps in IR: gender balance in citation (GBC) statements. GBC statements remind and encourage authors submitting work to peer-reviewed outlets to consider the gender balance among the works they cite. We find that these policies enjoyed wide support among IR scholars in our sample countries soon after journals began instituting the policies, but women were more supportive than men of the policies. We also report the results of a question-order experiment that allows us to study how raising awareness of gender gaps in the IR discipline affects the proportion of women that scholars list among the most influential IR scholars in the last 20 years. The effects of exposure to the gender treatment vary, however, by respondents' gender and whether respondents teach in the United States. The treatment effects were much larger for women than for men in the United States, but the reverse was true outside the United States.

Resumen: Presentamos los resultados de una encuesta a académicos de las RRII sobre el uso de una política, cada vez más común, diseñada para cerrar las brechas de reconocimiento en las RRII: el equilibrio de género en las declaraciones de citas bibliográficas (GBC, por sus siglas en inglés). Las declaraciones de GBC recuerdan y animan a los autores que envían trabajos a medios revisados por pares a considerar el equilibrio de género entre los trabajos que citan. Hallamos que estas políticas disfrutaron de un amplio apoyo entre los académicos de las RRII en nuestros países de muestra desde poco después de que las publicaciones comenzaran a instituir estas políticas, pero también que las mujeres apoyaron estas políticas más que los hombres. También presentamos los resultados de un experimento de orden de preguntas que nos permite estudiar cómo la mayor

sensibilización sobre las brechas de género en la disciplina de las RRII afecta a la proporción de mujeres que los académicos enumeran entre los académicos de RRII más influyentes en los últimos 20 años. Sin embargo, los efectos de estar expuestos al tratamiento de género varían en función del género de los encuestados y de si los encuestados enseñan o no en los Estados Unidos. Los efectos de este tratamiento fueron mucho mayores para las mujeres que para los hombres en los Estados Unidos, pero fuera de los Estados Unidos sucedió lo contrario.

Résumé: Nous présentons les résultats d'une étude de chercheurs en relations internationales (RI) sur la généralisation de l'utilisation d'une politique visant à resserrer les écarts de reconnaissance entre les genres en RI : les déclarations d'équilibre des genres dans les citations (EGC). Ces déclarations sensibilisent les auteurs publiés par les revues spécialisées à l'équilibre des genres et les encouragent à lui accorder une attention particulière lorsqu'ils citent des travaux. Selon nos observations, ces politiques ont rapidement reçu un accueil positif chez les chercheurs en RI des pays de l'échantillon. Nous notons toutefois que les femmes y étaient davantage favorables que les hommes. Nous présentons également les résultats d'une expérience sur l'ordre des questions posées. Nous pouvons ainsi analyser les effets d'une sensibilisation croissante aux écarts entre les genres en RI sur la proportion de chercheuses citées par leurs pairs pour leur influence dans la discipline ces 20 dernières années. Néanmoins, les effets de l'exposition au traitement du genre varient en fonction du sexe de la personne interrogée et du fait qu'elle enseigne ou non aux États-Unis. Aux États-Unis, les effets du traitement étaient bien plus importants pour les femmes que les hommes, mais nous avons observé la relation inverse dans le reste du monde.

Keywords: gender, citations, journals, networks, academia, diversity & inclusion, survey experiment

Palabras clave: Género, citas bibliográficas, publicaciones, redes, mundo académico, diversidad e inclusión, experimento de encuesta

Mots clés: Genre, citations, revues, réseaux, monde universitaire, diversité et inclusion, expérience de sondage

Introduction

Recent research and commentary on the discipline of political science and its sub-field, international relations (IR), suggests that women are cited at disproportionately lower rates than their male colleagues (e.g., [Maliniak, Powers, and Walter 2013](#); [Mitchell et al. 2013a](#); [Dion and Mitchell 2020](#); [Roberts, Stewart, and Nielsen 2020](#)), appear on syllabi less frequently than they should based on their representation in the discipline (e.g., [Colgan 2017](#); [Maliniak et al. 2018](#); [Hardt et al. 2019](#); [Smith et al. 2020](#)), and receive less recognition for their research than is warranted by the quality of that work ([Dion, Summer, and Mitchell 2018](#)). Much of the work on these gender gaps ends with calls for changes to tenure standards, parental leave policies, journal citation policies, and a range of other actions that might reshape the composition of the field and its core practices (e.g., [Hancock, Baum, and Breuning 2013](#); [Mershon and Walsh 2015](#); [Beaulieu and Searles 2016](#); [Shames and Wise 2017](#); [Windsor and Crawford 2020](#)). In response, professional organizations, journals, and networks of scholars have launched new initiatives designed to increase awareness

and recognition of scholars who are members of underrepresented groups. We have little systematic evidence, however, of the actual effects of such policies on the behavior of IR scholars as they communicate their work—and the work of others—to the discipline, students, and the broader public.

In this paper, we report the results of a survey of IR scholars in thirty-one countries on the use of an increasingly common, but sometimes controversial, policy designed to raise awareness of and close gender recognition gaps in IR: gender balance in citation (GBC) statements by journal editors that remind and encourage authors to consider the gender balance among the works they cite. Journal editors increasingly use GBC policy statements in the submission, revise and resubmit, and/or conditional acceptance stages of the peer-review process.

Because self-reporting of future behavior is not always aligned with actual future behavior and, in this context, might also be subject to significant social desirability bias (Krumpal 2013), we offer additional evidence from a question-order experiment. We asked respondents to list up to four scholars “whose work has had the greatest influence on the field of IR in the past 20 years.” We randomly varied whether respondents received a set of questions related to GBC policy statements meant to address gender gaps before or after completing the scholar ranking task, an activity that is comparable to selecting citations. Our research design allows us to provide new descriptive information on attitudes toward GBC policy statements and to provide causally identified answers to two questions related to their effectiveness: (1) Do GBC statements broaden the range of scholars that individual authors think about when trying to identify important works? and (2) is the effect of GBC statements conditional on the gender (or other demographic characteristics) of the respondent?

We find that GBC statements enjoyed broad support among IR scholars, even a few years after journals began implementing these policies, but women were more supportive than men of these policies. A plurality of scholars believe that the adoption of such policies would cause them to increase the frequency with which they cite female scholars, but again, women are more likely than men to say GBC policies would lead them to change their behavior. The results of our experiment demonstrate, moreover, that reminding respondents about GBC policies and asking them to think about how these policies might affect their behavior had observable impacts on their rankings of IR scholars. The proportion of female scholars listed by respondents who received the gender treatment was about 5 percentage points higher than it was for respondents in the control group, suggesting that GBC policies may work and ultimately may lead IR scholars to increase the citation of female scholars in their research publications. We find, however, that the effects of exposure to the gender treatment vary by gender¹ of respondent and whether respondents teach in the United States. The treatment effects we observed were much larger for women than they were for men in the United States, but the reverse was true outside the United States. Among US-based women, the treatment effect was about 11 percentage points, while among men it was just 3 percentage points. For scholars based outside the United States, the treatment produced about a 6 percentage point increase among male respondents as compared to a statistically insignificant 1 percentage point increase for female respondents. This research yields new insights into the

¹Here and throughout the paper, we use the term “gender” synonymously with “sex” in large part because discussions of so-called “gender gaps” and “gender citation policies” use the term in the same way. We recognize that “gender” is widely understood to mean socially constructed identities and behaviors, and we agree with this usage. Our study does not attempt to measure scholars’ perceptions and evaluations of masculinities and femininities; rather, it engages with “men” and “women” as biological categories. We choose to employ the term “gender” since it is common practice in the literature on inclusion and gender equality. We choose to focus on “men” and “women” for the same reason and because the vast majority of our sample identifies along these lines. Nevertheless, we want to emphasize that we do not assume that gender is related to biological sex in an essentialist way. For a more detailed discussion of these terms and their usage in IR literatures, see Ellerby (2017), Sjöberg, Kadera, and Thies (2018), Cohen and Karim (2022), and Karim and Hill (2024).

structure of demographic biases in our discipline and suggests that efforts to raise awareness of these biases can shape scholars' behavior.

In the remainder of this paper, we review recent work on gender recognition gaps in IR and efforts to address these gaps, highlighting the fact that few of these efforts have been evaluated for their effectiveness. We then outline our research design before summarizing the results of our survey and experiment. Finally, in the conclusion we summarize our findings and assess various strategies for closing the gender recognition gap in IR. Importantly, these findings suggest the need for future research and related efforts to reduce recognition gaps related to other identities, such as race, ethnicity, and country of origin.

Gender Recognition Gaps in IR

Over the past 20 years, research has proliferated on representation and diversity in the academy in general and political science and IR in particular. Many studies show that women systematically lag behind men or are underrepresented according to traditional metrics of success in political science and IR. Together, these studies suggest the existence of a recognition gap between the work done by female IR scholars and the recognition of that work within the discipline.² In this section, we briefly summarize research on the existence of such a gap, as well as strategies for closing it.

Numerous studies highlight the gulf between men and women on many traditional metrics within the IR discipline. Despite increasing participation in Ph.D. programs, women are less likely than their male counterparts to achieve tenure in political science and IR departments (Maliniak et al. 2008; Hesli, Lee, and Mitchell 2012; Hancock, Baum, and Breuning 2013). This disparity is likely related to several other gender-based differences: Women have lower publication rates, especially in top-ranked journals (Breuning and Sanders 2007; Teele and Thelen 2017; Djupe, Smith, and Sokhey 2019); devote more time to service activities (Turner 2002; Mitchell and Hesli 2013); receive systematically lower teaching evaluations (Martin 2016); specialize in different and differently valued subfields (Leahey 2006); and bear unequal shares of household labor (Suitor, Mecom, and Feld 2001; Antecol, Bedard, and Stearns 2018; Fattore 2019). These gendered effects were more pronounced during the COVID-19 pandemic, during which female political scientists reported lower perceived levels of productivity and a greater nonresearch-related workload while devoting more hours to childcare and domestic responsibilities compared to their male counterparts (Shalaby, Allam, and Butturoff 2021). Individually and collectively, such factors produce a "leaky pipeline" in which women disproportionately leave the discipline before achieving academic tenure and promotion (Ysseldyk et al. 2019).

Recent literature on underrepresentation also focuses on different citation rates between men and women. This work finds that political science and IR publications by women are cited less often than comparable publications by men (Kadera 2013; Maliniak, Powers, and Walter 2013; Dion, Sumner, and Mitchell 2018; Dion and Mitchell 2020). Mitchell, Lange, and Brus (2013) find that women are more likely than men and mixed-gendered author teams to cite publications by women. Additionally, as Maliniak, Powers, and Walter (2013) find, women cite themselves less often than men do. Since citation counts tend to be self-reinforcing, and because self-citation has been linked to higher citation counts by others, IR research by women is less frequently cited (Fowler and Aksnes 2007).

²We borrow the term "recognition gap" from Michèle Lamont's 2017 presidential address to the American Sociological Association, where she defined "recognition gaps" as "disparities in worth and cultural membership between groups in a society", and the act of recognition as the affirmation and acknowledgment of positive social worth (Lamont 2018).

Female IR scholars are systematically underrepresented in the classroom as well. Based on the rate at which women publish, instructors in US graduate political science programs assign disproportionately less scholarship authored by women (Sumner 2018; Hardt et al. 2019; Smith et al. 2020). The underrepresentation of women among political science and IR faculty compounds the phenomenon of underassignment of publications written by women. Men are less likely to assign readings by women, and women are more reluctant to assign their own research, exacerbating the problem of women's representation on the syllabi used to train future political scientists (Colgan 2017). Further, evidence suggests that gender gaps in the discipline translate to student behavior, as male undergraduate students are less likely to cite female scholars (Liu, Devine, and Gauder 2020).

The problem may be particularly acute in some subfields. Phull, Ciflikli, and Meibauer (2019) find, for example, that men dominate reading lists for security courses, a historically male-dominated field, while other subfields, such as human rights and environmental politics, are more gender inclusive (Hoagland et al. 2020; Rublee et al. 2020). Shames and Wise (2017) similarly report that the underrepresentation of women in the political methodology subfield is “not gender-neutral” but rather a product of historical as well as individual biases.

Numerous strategies to address lower levels of output and recognition among women in political science and IR have grown up alongside research on the gender gap. Networks like Women Also Know Stuff highlight the expertise of female scholars (Beaulieu et al. 2017). Journeys in World Politics brings together junior and senior women in IR for annual workshops that include both research presentations and professional development opportunities. At larger conferences, individuals and groups of scholars also organize to encourage equal representation and discourage the organization of “manels”, especially in research areas in which there are prominent women scholars (Humphreys 2018).

Formal policies that encourage the equal representation of women in citations are among the more high-profile and potentially controversial approaches to closing the gender gap in the discipline. Such policies began a decade ago. *International Studies Quarterly* (ISQ), the flagship International Studies Association (ISA) journal, now prompts authors with the following statement on its submission form: “Before submitting a manuscript, authors should ensure that it conforms to the highest standards of proper attribution. We strongly recommend that authors check their references to ensure inclusion of authors from disadvantaged groups. ISQ is committed to ensuring that scholars receive appropriate intellectual acknowledgment regardless of race, gender, class, professional standing, or other categorical attributes” (“ISQ Guidelines and Policies”, 2023). The *European Journal of International Relations* (EJIR) draws potential authors’ “attention to research showing gendered and regional biases in citation practices and ask[s] authors to bear this in mind when preparing their manuscripts for submission to EJIR” (“Manuscript Submission Guidelines”, 2023).

Other journals and professional organizations include diversity statements in their mission statements or submission guidelines. For example, according to its “Scope and Aims” statement, the *American Political Science Review* “aims to represent the diversity of subfields, geographic areas of study, identities, methods, and approaches that are encompassed by our broad and pluralistic discipline” (“APSR Submission Guidelines”, 2023).

In total, we identified nine influential peer-reviewed journals that publish IR research that, as of this writing in October 2022, encouraged authors at some point in the peer-review process to consider the author’s gender for the work they cite. In addition to ISQ and EJIR, noted above, we found similar language on the websites of *International Organization*, *Security Studies*, *International Studies Review*, *Review of International Political Economy*, *Review of International Studies*, *International Theory*, and *International Interactions*. Among the 18 other journals that are commonly listed by respondents to the TRIP faculty survey as publishing important peer-reviewed

work on IR, we found no evidence of GBC encouragement policies, although four journals—American Journal of Peace Research, APSR, International Affairs, and International Studies Perspectives—address diversity and inclusion in their mission statements or submission guidelines.³

To date, despite the use of GBC policies by at least nine major peer-reviewed journals in the IR discipline, there has been no study of whether and to what extent these citation policies and other efforts to raise awareness of gender definition gaps actually work to reduce the magnitude of such gaps. In the narrowest sense, answering this question might entail studying whether, how, and when GBC policies change citation behavior. Citation balance statements, for example, may encourage authors to preemptively improve the gender balance of the citations in their manuscripts before submission, after being told to do so by reviewers, or after receiving a conditional acceptance from editors. Learning the precise answers to when and how GBC statements affect citation outcomes would be potentially illuminating, but it also would largely miss the broader and more important question of whether the discipline welcomes policies designed to raise awareness of gender recognition gaps and/or whether such policies productively affect how the field recognizes and rewards high-quality research in general. In this paper, we offer the first answers to both of these questions.

Research Design

We use data from the multicountry 2017 Teaching, Research, and International Policy (TRIP) Faculty Survey. This data was collected at a particularly propitious time to evaluate the effectiveness of GBC policies. In 2013, when Maliniak, Powers, and Walter published their seminar article on the citation gap in IR journals, GBC policies were rare. By 2017, however, five of the top twelve IR journals had adopted such a policy, providing a good opportunity to review IR scholars' views on those policies and study the potential impact of GBS policies.⁴

For this experiment, we sent our survey to all faculty members at four-year colleges and universities in thirty-one countries who conduct research or teach courses on transborder political issues and who are appointed in a department of political science, IR, or public policy.⁵ We identified a total of 13,246 individuals in the thirty-one countries who met our criteria for inclusion. Of these individuals, 3,731 answered at least part of our survey, yielding a response rate of about 28 percent. The survey, which was in the field in October and November 2017, asked scholars about their research, the discipline of IR, and contemporary foreign and international policy issues.

In what follows, we often distinguish our results in the US survey from those among scholars in the other thirty-one countries we surveyed rather than report country-by-country results. We do so for several reasons. First, much of the debate and discussion of gender recognition gaps has occurred in IR and political science journals in the United States, so it is important to discern whether and to what

³Our complete set of influential journals includes those that are consistently said to publish research that has had the "greatest influence on the way IR scholars think about international relations" by respondents to the TRIP Faculty Survey (Maliniak et al. 2014, 2017). In October 2022, a team of research assistants examined the websites of each journal, looking for evidence of a formal or informal policy encouraging authors to be mindful of the gender citation gap when submitting manuscripts for review or at later stages of the review process, whether in the submission guidelines or editorial policies. Our findings are based on the publicly available information on the journals' websites in October 2022. We did not go through the full submission process or contact the editors of these journals. See Appendix A for complete results.

⁴Two more journals had adopted general diversity statements by 2017.

⁵We conducted this GBC experiment on IR scholars in Argentina, Australia, Brazil, Canada, Chile, Colombia, Denmark, Finland, France, Germany, Hong Kong, Ireland, Israel, Italy, Japan, Mexico, the Netherlands, New Zealand, Norway, the Philippines, Poland, Singapore, South Africa, South Korea, Sweden, Switzerland, Taiwan, Turkey, Ukraine, the United Kingdom, and the United States. Overall response rates per country are listed in [table A2](#).

extent US scholars differ in their support for and responsiveness to GBC policies. Second, and most importantly, US-based IR scholars make up a disproportionately large percentage of the population of IR scholars worldwide. In the 2017 TRIP Faculty Survey, for example, US scholars comprised 37 percent of the faculty we surveyed and 41 percent of those who responded to our survey across all thirty-one countries. Third, while there is evidence of insularity in many national IR communities, the size and highly insular nature of the US community—particularly in terms of US-based scholars' views on who the top scholars are—compel us to look at this community separately (Maliniak et al. 2018).⁶

The 2017 survey included a set of questions on GBC policies. We first reminded respondents about the adoption of citation balance statements meant to address gender recognition gaps by a number of prominent peer-reviewed journals publishing IR research. We then asked respondents about their views on such policies and whether the policies would change their citation behavior. As the first comprehensive assessment of support for these policies among IR scholars, the responses to these questions, which we describe below, are interesting in and of themselves, but the questions also served as treatments in a question-order experiment.

In the experiment, we randomly varied whether respondents received a set of questions related to GBC policy statements meant to address gender gaps before being asked to list up to four scholars “whose work has had the greatest influence on the field of IR in the past 20 years.” After the ranking task, we asked those respondents who had not yet answered the GBC policy questions to do so, giving us complete data on attitudes toward GBC policies and ranking behavior for all respondents. By randomizing whether respondents received questions about GBC policies just before or just after our scholar ranking question, we are able to identify the effect of raising awareness of recognition gaps on a scholar recognition task. Because we randomized when respondents received questions about GBC policies by journals attempting to address gender-based recognition gaps, and we asked about respondents' gender,⁷ we are also able to identify whether efforts to raise awareness have differential effects for women as well as the extent to which these efforts have spillover effects.⁸ Before discussing the results of the experiment, we summarize the descriptive results of our questions on attitudes toward GBC policies.

Results

Attitudes Toward GBC Policy Statements

We begin by reporting the results of our questions on attitudes toward GBC statements meant to address gender recognition gaps. We provided respondents with a brief description of work on demographic gaps in scholarly recognition and the role that GBC statements might play in addressing them. This preamble read:

A number of recent studies have highlighted the possible underrepresentation of female scholars in international relations, as reflected in article citations and graduate syllabi, and as a share of tenured and tenure-track faculty. Several prominent journals report taking steps to ensure that scholars receive appropriate intellectual

⁶We acknowledge that the survey experiment and results we report here do not address the disparities in race, ethnicity, or nationality created by the insularity of national IR communities and the lack of diversity within the discipline. We look forward to evaluating interventions to address these problems in our future research.

⁷We asked respondents whether they identified as “female”, “male”, or “transgender or other”, and we also included the response option “prefer not to answer.”

⁸We programmed the survey to ensure that people could not go back and change their answers. To do this, we divided the survey into segments; once respondents completed a section and hit the “Next” button, their data was captured and they could not revisit earlier questions.

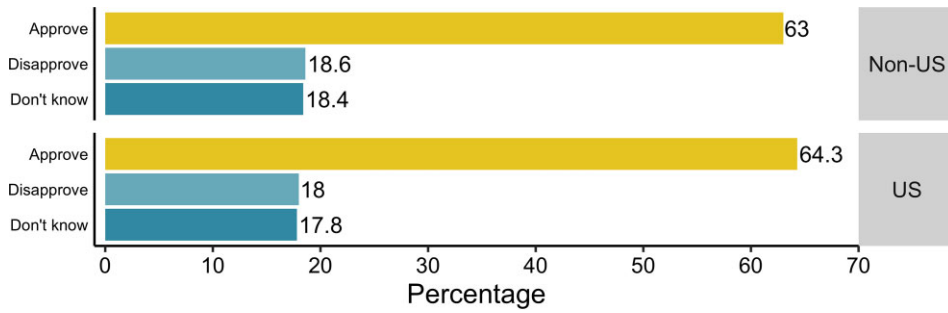


Figure 1. Do you approve of this citation policy?

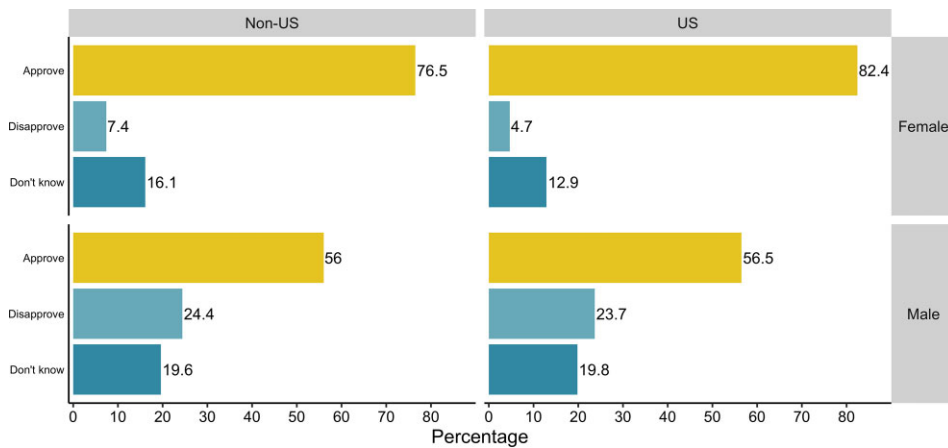


Figure 2. Do you approve of this gender citation policy? Broken down by gender.

acknowledgment regardless of their gender. To this end, these journal editors ask authors to pay particular attention to this issue by citing overlooked authors and literatures.

Next, we asked respondents two questions about the GBC policy. The first question read, “Do you approve or disapprove of this policy?” Response options included, “approve”, “disapprove”, or “don’t know.”

Our results show broad support for the use of GBC policies to address gender recognition gaps. Nearly two-thirds of all respondents said they approve of a GBC policy statement. A significant minority of respondents expressed uncertainty, however; about 18 percent said they “don’t know” whether they approve of GBC statements. The results are nearly identical for US and non-US respondents. We summarize the approval rates in figure 1. In figure 2, we present results on the approval of GBC policies broken down by respondents’ gender and whether they are based in the United States. We find that women overwhelmingly approve of the use of GBC policies. Support among their male colleagues is considerably lower, although men still favor GBC policies by more than two to one.⁹ We also see that the high level of “don’t know” responses in the aggregated results is driven, in part, by male respondents’ answers. About 20 percent of male respondents selected the “don’t know”

⁹ Our findings are broadly consistent with Fattore (2019, 59), which finds that “most men (65.7 percent) and women (81.8 percent) responded that they either already do check for gender balance in their citations or are open to doing so in the future.”

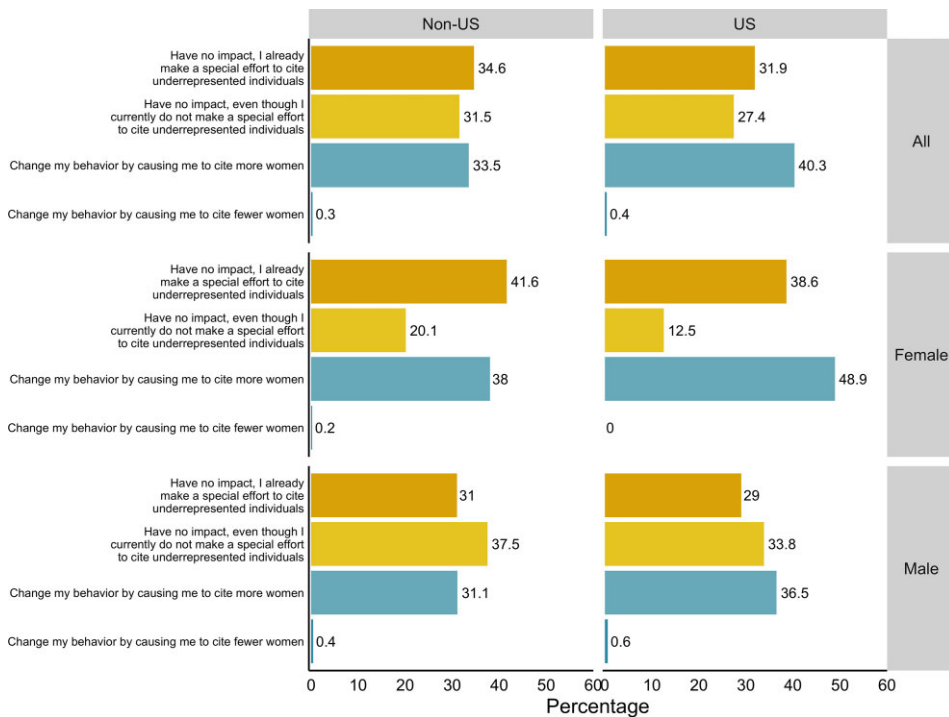


Figure 3. If you received such a reminder from a journal editor in the review process, this would:

option compared to about 13 percent of female respondents.¹⁰ The results are consistent for scholars based in the United States and those based in other countries.

In addition to asking whether respondents approved of the use of GBC policies, we asked these IR scholars how a reminder from a journal editor about gender (and other demographic) recognition gaps might affect their citation behavior. The question read, “If you received such a reminder from a journal editor in the review process, this would. . .” Respondents could select among the following response options:

- Change my behavior by causing me to cite more women.
- Change my behavior by causing me to cite fewer women.
- Have no impact on my behavior because I already make a special effort to cite underrepresented individuals.
- Have no impact on my behavior even though I currently do not make a special effort to cite underrepresented individuals.

Again, we find fairly high support for GBC policies designed to increase representation of women, where the largest single response category for both men and women suggests that such policies would change the respondent’s behavior by causing them to cite more women. We also find significant differences, however, between male and female scholars on this question. [Figure 3](#) displays those results broken down by gender and whether the respondent is US-based. Female respondents were more likely than their male counterparts to respond that they

¹⁰ This pattern is strikingly different than that found in most surveys on political issues, including previous TRIP surveys, in which women are more likely to select the “don’t know” response option. See, e.g., [Atkeson and Rapoport \(2003\)](#), [Mondak and Anderson \(2004\)](#), and [Merriman-Goldring, Paulson-Smith, and Peterson \(2017\)](#).

would change their behavior to cite more women after reading the preamble about GBC policies. Male respondents also were much more likely to respond that a policy would “[h]ave no impact on my behavior even though I do not make a special effort to cite women,” with around 34 percent of male respondents selecting this option compared to only 13 percent of women based in the United States and 20 percent of female respondents outside the United States.

Since this survey question also included a response option allowing respondents to say they would “[c]hange [their] behavior to cite fewer women,” the results provide a partial measure of a possible backlash effect against a GBC policy. No female scholars selected this response option in the United States, and less than one percent of male respondents said they would cite fewer women in response to an editor’s reminder about the GBC policy. While men are more likely overall to report that GBC policies would not change their behavior, we also find significant differences between female respondents who work in the United States and their counterparts in the other countries in our study; US-based women are considerably more likely to say they would change their citation practice if they received a reminder from a journal editor.

Evidence from Question Ordering Experiment

Thus far, we have seen that the use of one increasingly common method of bringing attention to the existence of gender recognition gaps in our discipline, GBC policies, enjoys broad support among IR scholars and, according to self-reported data, the use of these statements is likely to have either null or positive effects on scholars’ citation behavior. Although this kind of data is useful, it is potentially subject to social desirability bias and, of course, survey responses about likely future behavior may not always be aligned with actual future behavior (Powell 2013; Findley et al. 2017; Boas, Hidalgo, and Melo 2019). With these concerns in mind, we now examine the results of our question-order experiments, which use actual behavior (ranking scholarly contributions) as the dependent variable.

As we explain above, we assigned each respondent to a treatment or control group. In the treatment group, respondents read information and answered questions about GBC policies meant to address gender recognition gaps before being asked to name up to four of the most influential IR scholars in the last 20 years. In the control group, respondents answered no questions related to gender gaps before being asked to complete the influential scholars question. Thanks to this randomization scheme, we can test whether reminding scholars about recognition gaps in our discipline changes the gender balance among the scholars they list as having the greatest influence on IR research over the last 20 years.

Although we do not look directly at scholars’ citations, our choice of dependent variable provides a useful proxy for at least two reasons. First, the act of choosing the names of scholars who have had the greatest influence in the field is not dissimilar to the act of choosing which works to cite in an article. Kristensen’s (2018) three models of citation behavior in the IR literature closely align with how we imagine most respondents might complete our name generator task. His normative model, in which citations are “rewards” for quality work, most closely fits the wording of our survey question. Nevertheless, Kristensen’s symbolic theory—that citations are used as signposts for certain positions or ideas—and his constructivist theory—that citations are appeals to authorities—both clearly mirror the process that some respondents might use to generate a list of the top four scholars. The work of Roberts, Stewart, and Nielsen (2020) is instructive here as well. They find that the gender-citation gap is widest among highly cited articles. Asking respondents to identify the most influential scholars in the field is similar to measuring who gets cited most frequently as a way to conceptualize influence in the field.

Second, we believe this measure of our dependent variable represents a hard test of whether GBC policies and other efforts to raise awareness of gender recognition

Table 1. Question ordering experiment descriptive results

Sample	Condition	Number of respondents	Total names	Female (percentage of total names)	Unique female names
US-based scholars	Treatment	381	1,467	310 (21.1)	78
	Control	398	1,518	246 (16.2)	57
Non-US-based scholars	Treatment	292	1,132	240 (21.2)	76
	Control	286	1,127	179 (15.9)	54

can shape how individual scholars recognize and reward high-quality research. The names that scholars list as having the greatest influence on the IR field are likely the products of socialization in the field. The TRIP survey has included the question on scholars with the greatest influence in the last 20 years on each of its six surveys since 2004, and respondents are remarkably consistent in their responses (Maliniak et al. 2011; Maliniak et al. 2018). Over the six waves of the US version of the survey, eight of the fifteen top scholars have never changed, and since 2011, eleven of the top fifteen have remained the same. With relatively little movement over more than a decade in the IR faculty's views of the top IR scholars, our dependent variable appears to be one on which most scholars have established opinions that should be difficult to move with any informational treatment.

Our dependent variable is the percentage of female scholars listed by each respondent. We estimate the percentage of female scholars listed by each respondent for the treatment group relative to the control group. To facilitate this test, we code each individual response to the influential scholars question for gender. Because responses were open-ended, we manually matched responses to known individuals. When necessary, we made reasonable guesses as to whom our respondents were referring. If a respondent offered "Keohane", "Bob Keohane", "Robert Keohane", or even "Kohane", for example, we matched that to the standardized value of "Robert O. Keohane." With a standardized list of influential scholars in hand, we then conducted web searches to identify the gender of each individual.¹¹

At the aggregate level, we had a similar number of respondents in our control (398 US-based scholars, 286 in the international sample) and treatment (381 US, 292 non-US) groups, and each respondent was asked to list up to four influential IR scholars.¹² We present descriptive results in table 1. For scholars based in the United States, this produced 1,518 names from members of the control group, of which 246 (16.2 percent) were women; the treatment group generated 1,467 names, of which 310 (21.1 percent) were women. Among non-US-based scholars, the control group gave 1,107 names, of which 179 (15.9 percent) were women; the treatment group produced 1,132 names, of which 240 (21.2 percent) were women. Additionally, the gap between the treatment and control groups persists even if we look only at the number of unique names produced by each group rather than the total number of names. In the US and non-US samples, the control group

¹¹ We used photos and names listed on faculty websites to code for gender. Although some individuals identify with genders that are inconsistent with gender stereotypes associated with their outward appearance or name, our approach here ensures that we have a measure of gender based on how each scholar is likely to be perceived by other scholars in the discipline. This measurement strategy is consistent with those taken in studies of race-based discrimination. See, for example, Butler and Broockman (2011).

¹² Note that we did not block randomize on demographic variables. Following Mutz, Pemantle, and Pham (2019), we used computer-generated responses on our online survey platform to verify that our randomization tool was programmed correctly. In addition, we confirmed that there was no differential attrition between treatment and control. For the first test, we used Qualtrics to generate 2,504 test responses that were assigned to treatment and received a randomly generated gender. When regressed on gender, treatment has no statistically significant relationship, confirming our randomization scheme was programmed correctly. We also compared the rates of attrition between treatment and control and found no relationship with treatment and either the choice to respond to the ranking question or the number of names given to the ranking question. Additionally, gender was unrelated to non-response or number of responses in both treatment groups. See our discussion in Appendix B.

Table 2. Overall gender experiment results

Treatment Group	Average percentage of responses that are female (US only)	Average percentage of responses that are female (non-US)
Treatment	20.95	21.79
Control	15.97	17.00
Difference of means	4.98 (95percent CI: 1.78– 8.18)	4.80 (95percent CI: 0.90–8.70)
<i>p</i> -value	0.002	0.016

Table 3. Gender experiment results (female respondents only)

Treatment group	Average percentage of responses that are female (US only)	Average percentage of responses that are female (non-US)
Treatment	35.77	29.35
Control	25.21	28.33
Difference of means	10.56 (95percent CI: 3.74–17.38)	1.02 (95percent CI: –7.66 to 9.70)
<i>p</i> -value	0.003	0.817

listed fifty-seven and fifty-four distinct female scholars, respectively, while the group receiving the treatment listed seventy-eight (US) and seventy-six (non-US) unique female scholars. In sum, our gender treatment resulted in a 30.2 percent increase among US scholars and a 33.3 percent increase among non-US scholars in the total proportion of female names mentioned in the ranking task.

We now move to a more formal test of the treatment effects. To construct our gender composition dependent variable, we coded responses that we could identify as female with a 1, while all other values (male or no information) we coded as a 0 (zero). For each respondent, we generated a gender composition score by dividing the number of women that a respondent listed as among the top four scholars by the total number of individuals that the respondent listed for this question. The score ranges from 0 to 1, with 1 indicating that the individual listed only women. We average these scores within the treatment group and subtract them from the same value within the control group to estimate the causal effect of raising awareness of a gender recognition gap in IR on a task in which scholars are asked to recognize the influence of their colleagues on the discipline.

We present the results of this analysis in [table 2](#). We find that the treatment had a measurable effect on the relative balance of male and female scholars listed by our respondents as being the most influential in the IR field. The proportion of female scholars listed was about 5 percentage points higher (95 percent CI: 1.78, 8.18) among those who received the GBC treatment than it was among respondents in the control group, a result consistent in both the US-based sample and the non-US sample.¹³

Importantly, however, when we subset by respondent gender (see [tables 3](#) and [4](#)), we find that the effects vary by the self-reported gender of respondents and whether the respondent is currently based in the United States. For US-based women, the gender treatment produced a 10.56 percentage point increase (95 percent CI: 3.74–17.38) in the proportion of female scholars listed as influential, compared

¹³ At first glance, this may seem to indicate that a much smaller percentage of respondents changed their behavior than thought they would in the hypothetical question (see [figure 3](#)), but it is necessary to remember that each respondent could choose up to four scholars in their answer to the question about the most influential scholars. What might this mean in real terms? If out of a hypothetical sample of 100 people, 64 individuals included one woman on their list of scholars, then we would observe 64 women out of 400 names or 16 percent total (which is the actual value of the control for the United States). If 40 of the 100 people increased the number of women they listed from 0 to 1 or 1 to 2 (40.3 percent of our respondents self-reported that they would increase the number of women they cite in response to GBC policies), we would observe 104 women out of 400 names or 26 percent (which is not far off the observed result in the treatment group of 21 percent).

Table 4. Gender experiment results (male respondents only)

Treatment Group	Average percentage of responses that are female (US only)	Average percentage of responses that are female (non-US)
Treatment	14.55	17.36
Control	11.72	12.83
Difference of means	2.84 (95percent CI: -0.27 to 5.95)	4.53 (95percent CI: 0.02-9.04)
<i>p</i> -value	0.07	0.049

to a smaller treatment effect of 2.84 percentage points (95 percent CI: -0.27 to 5.95) among male respondents. The treatment effect here is approximately 4 times larger among women than among men. When we look to scholars based outside the United States, these results are flipped; the treatment produces a 4.53 percentage point (95 percent CI: 0.02-9.04) increase among male respondents as compared to a statistically insignificant 1.02 percentage point (95 percent CI: -7.66 to 9.70) increase for female respondents.

The results also show that the baseline rate in the control condition among male respondents for listing women scholars (about 12 percent) was much lower than that of female respondents (about 25 percent) in the United States. This gender gap is even larger among non-US-based scholars, among whom the baseline for male respondents for listing women scholars (about 13 percent) was similar to US-based male scholars; it was lower by around 15 percent, however, than that among our female respondents (about 28 percent) based outside the United States. Regardless of where they are based, women are more likely than men to list women as among the most influential scholars in the discipline.

Discussion and Implications

Can raising awareness of gender recognition gaps help close them? At the outset, we highlighted a number of recent studies documenting pervasive gender imbalances in our discipline, and we noted a number of high-profile efforts to narrow or close these gaps. Many of these efforts seek to address gender recognition gaps by raising awareness of their existence and suggesting direct and practical steps to address them. We study one set of initiatives, GBC policies, to better understand their reception among IR scholars and their potential efficacy.

We find that GBC policies enjoy broad support among IR scholars. Large majorities of scholars self-report that they “support” such policies, and they expect that, if and when they receive such guidance from journal editors, they will cite more female authors. IR scholars, it seems, are ready and willing to narrow the gender citation gap. Their enthusiasm for these policies may be encouraging, but it remains to be seen whether policies designed to raise awareness of, and thereby close, gender recognition gaps will have broad effects on how our discipline recognizes and rewards high-quality research. This question is particularly salient since there are significant differences between men and women in their level of support for GBC policies and their anticipated behavior. Although we see no evidence that suggests a backlash against these policies among men, male scholars are less willing to endorse the policies outright.

The results of our experiment provide more reliable evidence of the likely behavioral implications of GBC policies and other efforts to raise awareness of gender recognition gaps. We find that a mild intervention just prior to our question about which scholars have had the “greatest influence” on the discipline increased the proportion of women that respondents listed. Again, while this is encouraging, we also showed that the treatment effects differed rather substantially by gender and geographic context. While our ranking exercise is not a precise analog to citing others’ work in a peer-reviewed setting, we argue that it has much in common with the

citation process. Among other things, citations are statements about which works are influential, and so, in a narrow sense, our results suggest that GBC policies may work to decrease gender citation gaps. In terms of the real-world impact, our results may be especially compelling in the long run. Citations are cumulative and self-reinforcing; as such, our finding that a gender equality intervention increases the overall proportion of women scholars mentioned by 30 percent may have a much larger downstream effect.¹⁴

More broadly, we take this as evidence that efforts to raise awareness about gender recognition gaps may have important effects on the way that scholars think about, communicate, and reward high-quality research. On average, such policies increase the share of women recognized as influential and increase the number of unique women mentioned. Citations are an important way in which gender gaps manifest within the academy, but they are not the only one. Invited talks, invitations from journalists to comment on current affairs, hiring, promotion, retention, and more are all likely affected by similar gender recognition gaps. Our results suggest that efforts to raise awareness of such gaps may cause IR scholars, on average, to reevaluate how they evaluate and recognize research quality.

At the same time, the treatment effects we observed were much larger for women than they were for men in the United States, while the reverse was true outside the United States. This may be the result of variation in the gender composition of various IR subfields,¹⁵ gendered professional mentoring networks, sexist assessments of research quality, and/or a variety of other gendered processes and outcomes in the profession. It may be that the differences in training and canonical works and approaches that are less prominent in American IR are already part of the curriculum of non-US-based scholars. Women outside the United States may show no response to the treatment because they already include female scholars in their account of the most influential scholars in the field, although we find only a small, statistically insignificant difference between the control groups for US and non-US-based women. Although our experiment was not designed to tease out mechanisms, it clearly reveals treatment effects that are conditional on the gender of respondents and larger contextual factors.

What does this imply for the effectiveness of awareness-raising efforts? Our results show that such efforts constitute an important but, by no means, sufficient intervention. To the extent that the gendered pattern of results we recovered is the result of gendered professional networks, then efforts to promote research by women in networks composed predominantly of men may be among the most effective interventions. This is what many efforts, like *Women Also Know Stuff*, are already doing.

These results also raise questions about recognition gaps in IR for other underrepresented groups. The historical exclusion and continued underrepresentation of people of color in IR is particularly concerning given the lack of attention to critical perspectives and the roots of our discipline in white supremacy (Zvobgo and Loken 2020). As Zvobgo and Loken (2020) note, there is a great deal of work to be done to make the academy more accessible and increase the presence of historically excluded groups. Our results suggest that citation policies may be one way to increase visibility. Scholars from the Global South and people of color within the academy in the United States and elsewhere may benefit from these kinds of interventions. Further research should test whether the GBC approach is feasible for these communities of scholars. In other work, we are investigating the effectiveness of GBC policies in comparative perspective—assessing similar policies aimed at closing recognition gaps between scholars of the global north and south.¹⁶ If similar dynamics obtain, those in the “in group”—the global north—are unlikely to be

¹⁴ We thank Mathis Lohaus for emphasizing this point.

¹⁵ Tables C1 and C2 display our analysis in which we include a control variable for the respondent's subfield. Even including this (and other control variables), we still find an effect for our gender treatment.

¹⁶ See Lohaus and Wemheuer-Vogelaar (2021) for an overview of geographic diversity in IR journals.

moved by analogous recognition exercises. Significant work remains to be done, but the results from this experiment suggest that “Western IR” is likely to have a hard time confronting and closing recognition gaps that stem from bias against scholars from the Global South.

Acknowledgments

For their support as research assistants and data analysts, we thank Vera Choo, Aidan Donovan, Moira Johnson, Lauren Lanzalotto, and Mary Trimble. For comments on earlier versions of this paper, we thank Amitav Acharya, Mathis Lohaus, Gustav Meibauer, Sarah Mitchell, Thomas Risse, Kathryn Sikkink, Wiebke Wemheuer-Vogelaar, and Antje Wiener.

Funding

This research was supported by the Carnegie Corporation of New York, grant numbers B-8914 (2015-2017) and G-17-55327 (2018-2020).

References

- ANTECOL, HEATHER, KELLY BEDARD, AND JENNA STEARNS. 2018. “Equal but Inequitable: Who Benefits from Gender-neutral Tenure Clock Stopping Policies?” *American Economic Review* 108 (9): 2420–41.
- AMERICAN POLITICAL SCIENCE ASSOCIATION. 2023. “APSR Submission Guidelines.” *American Political Science Association*. Accessed January 4, 2023. <https://www.apsanet.org/APSR-Submission-Guidelines>.
- ATKESON, LONNA RAE, AND RONALD B. RAPOPORT. 2003. “The More Things Change the More They Stay the Same: Examining Gender Differences in Political Attitude Expression, 1952–2000.” *Public Opinion Quarterly* 67 (4): 495–521.
- BEAULIEU, EMILY, AMBER E. BOYDSTUN, NADIA E. BROWN, KIM YI DIONNE, ANDRA GILLESPIE, SAMARA KLAR, YANNA KRUPNIKOV, MELISSA R. MICHELSON, KATHLEEN SEARLES, AND CHRISTINA WOLBRECHT. 2017. “Women also Know Stuff: Meta-level Mentoring to Battle Gender Bias in Political Science.” *PS: Political Science & Politics* 50 (3): 779–83.
- BEAULIEU, EMILY, AND KATHLEEN SEARLES. 2016. “Disrupting Implicit Bias: Crowdsourced Database Highlights Women Experts in the Social Sciences #WomenAlsoKnowStuff.” *Impact of Social Sciences Blog* (10 March 2016). <https://blogs.lse.ac.uk/impactofsocialsciences/2016/03/10/disrupting-implicit-bias-crowdsourced-database-women-experts/>.
- BOAS, TAYLOR C., F. DANIEL HIDALGO, AND MARCUS ANDRÉ MELO. 2019. “Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil.” *American Journal of Political Science* 63 (2): 385–400.
- BREUNING, MARIJKE, AND KATHRYN SANDERS. 2007. “Gender and Journal Authorship in Eight Prestigious Political Science Journals.” *PS: Political Science & Politics* 40 (2): 347–51.
- BUTLER, DANIEL M., AND DAVID E. BROOCKMAN. 2011. “Do Politicians Racially Discriminate against Constituents? A Field Experiment on state Legislators.” *American Journal of Political Science* 55 (3): 463–77.
- COHEN, DARA KAY, AND SABRINA M. KARIM. 2022. “Does More Equality for Women Mean Less War? Rethinking Sex and Gender Inequality and Political Violence.” *International Organization* 76 (2): 414–44.
- COLGAN, JEFF. 2017. “Gender Bias in International Relations Graduate Education? New Evidence from Syllabi.” *PS: Political Science & Politics* 50 (2): 456–60.
- DION, MICHELLE L., AND SARA McLAUGHLIN MITCHELL. 2020. “How Many Citations to Women is ‘Enough’? Estimates of Gender Representation in Political Science.” *PS: Political Science & Politics* 53 (1): 107–13.
- DION, MICHELLE L., JANE LAWRENCE SUMNER, AND SARA McLAUGHLIN MITCHELL. 2018. “Gendered Citation Patterns across Political Science and Social Science Methodology Fields.” *Political Analysis* 26 (3): 312–27.
- DJUPE, PAUL A., AMY ERICA SMITH, AND ANAND EDWARD SOKHEY. 2019. “Explaining Gender in the Journals: How Submission Practices Affect Publication Patterns in Political Science.” *PS: Political Science & Politics* 52 (1): 71–77.
- ELLERBY, KARA. 2017. *No Shortcut to Change: An Unlikely Path to a More Gender Equitable World*. New York, NY: NYU Press.
- EUROPEAN JOURNAL OF INTERNATIONAL RELATIONS EDITORIAL BOARD. 2023. “Manuscript Submission Guidelines.” *European Journal of International Relations* SAGE Journals. Accessed January 4, 2023. <https://journals.sagepub.com/author-instructions/EJT>.

- FATTORE, CHRISTINA. 2019. "Nevertheless, She Persisted: Women's Experiences and Perceptions within the International Studies Association." *International Studies Perspectives* 20 (1): 46–62.
- FINDLEY, MICHAEL G., ADAM S. HARRIS, HELEN V. MILNER, AND DANIEL L. NIELSON. 2017. "Who Controls Foreign Aid? Elite versus Public Perceptions of Donor Influence in Aid-dependent Uganda." *International Organization* 71 (4): 633–63.
- FOWLER, JAMES, AND DAG AKSNES. 2007. "Does Self-Citation Pay?" *Scientometrics* 72 (3): 427–37.
- HANCOCK, KATHLEEN J., MATTHEW A. BAUM, AND MARIJKE BREUNING. 2013. "Women and Pre-Tenure Scholarly Productivity in International Studies: An Investigation into the Leaky Career Pipeline." *International Studies Perspectives* 14 (4): 507–27.
- HARDT, HEIDI, HANNAH JUNE KIM, AMY ERICA SMITH, AND PHILIPPE MEISTER. 2019. "The Gender Readings Gap in Political Science Graduate Training." *The Journal of Politics* 81 (4): 1528–32.
- HESLI, VICKI L., JAE MOOK LEE, AND SARA McLAUGHLIN MITCHELL. 2012. "Predicting Rank Attainment in Political Science: What Else Besides Publications Affects Promotion?" *PS: Political Science & Politics* 45 (3): 475–92.
- HOAGLAND, JACK, AMY OAKES, ERIC PARAJON, AND SUSAN PETERSON. 2020. "The Blind Men and the Elephant: Comparing the Study of International Security across Journals." *Security Studies* 29 (3): 393–433.
- HUMPHREYS, MACARTAN. 2018. "Gender Discrimination in Political Science: when Men Are Poor Allies." Committee on the Status of Women in the Profession. Accessed January 18, 2018. <http://web.apsanet.org/cswp/gender-discrimination-in-political-science-and-the-problem-of-poor-allies/>.
- INTERNATIONAL STUDIES QUARTERLY EDITORIAL BOARD. 2023. "ISQ Guidelines and Policies," *International Studies Quarterly*. Accessed January 4, 2023. https://academic.oup.com/isq/pages/General_Instructions.
- KADERA, KELLY M. 2013. "The Social Underpinnings of Women's Worth in the Study of World Politics: Culture, Leader Emergence, and Coauthorship." *International Studies Perspectives* 14 (4): 463–75.
- KARIM, SABRINA M., AND DANNY HILL JR. 2024. *Positioning Women in Conflict Studies: How Women's Status Affects Political Violence*. Oxford University Press, forthcoming.
- KRISTENSEN, PETER MARCUS. 2018. "International Relations at the End: A Sociological Autopsy." *International Studies Quarterly* 62 (2): 245–59.
- KRUMPAL, IVAR. 2013. "Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review." *Quality & Quantity* 47 (4): 2025–47.
- LAMONT, MICHÈLE. 2018. "Addressing Recognition Gaps: Destigmatization and the Reduction of Inequality." *American Sociological Review* 83 (3): 419–44.
- LEAHEY, ERIN. 2006. "Gender Differences in Productivity: Research Specialization as a Missing Link." *Gender & Society* 20 (6): 754–80.
- LIU, LI-YIN, CHRISTOPHER J. DEVINE, AND HEIDI GAUDER. 2020. "The Gender Citation Gap in Undergraduate Student Research: Evidence from the Political Science Classroom." *PS: Political Science & Politics* 53 (4): 729–33.
- LOHAUS, MATHIS, AND WIEBKE WEMHEUER-VOGELAAR. 2021. "Who Publishes Where? Exploring the Geographic Diversity of Global IR Journals." *International Studies Review* 23 (3): 645–69.
- McLAUGHLIN MITCHELL, SARA, AND VICKI L. HESLI. 2013. "Women Don't Ask? Women Don't Say no? Bargaining and Service in the Political Science Profession." *PS: Political Science & Politics* 46 (2): 355–69.
- McLAUGHLIN MITCHELL, SARA, SAMANTHA LANGE, AND HOLLY BRUS. 2013. "Gendered Citation Patterns in International Relations Journals." *International Studies Perspectives* 14 (4): 485–92.
- MALINIAK, DANIEL, AMY OAKES, SUSAN PETERSON, AND MICHAEL J. TIERNEY. 2008. "Women in International Relations." *Politics & Gender* 4 (1): 122–44.
- . 2011. "International Relations in the US Academy." *International Studies Quarterly* 55 (2): 437–64.
- MALINIAK, DANIEL, SUSAN PETERSON, RYAN POWERS, AND MICHAEL J. TIERNEY. 2014. *TRIP 2014 Faculty Survey*. Williamsburg, VA: Institute for the Theory and Practice of International Relations. <https://trip.wm.edu/>.
- . 2017. "TRIP 2017 Faculty Survey." *Teaching, Research, and International Policy Project*. Williamsburg, VA: Global Research Institute. <https://trip.wm.edu/>.
- . 2018. "Is International Relations a Global Discipline? Hegemony, Insularity, and Diversity in the Field." *Security Studies* 27 (3): 448–84.
- MALINIAK, DANIEL, RYAN M. POWERS, AND BARBARA F. WALTER. 2013. "The Gender Citation Gap." In *APSA 2013 Annual Meeting Paper, American Political Science Association 2013 Annual Meeting*.
- MARTIN, LISA L. 2016. "Gender, Teaching Evaluations, and Professional Success in Political Science." *PS: Political Science & Politics* 49 (2): 313–9.
- MERRIMAN-GOLDRING, RACHEL, KADEN PAULSON-SMITH, AND SUSAN PETERSON. 2017. "Gender and the Expression of Attitudes about International Relations," *Paper presented to the Annual Meeting of the International Studies Association*, Baltimore.

- MERSHON, CAROL, AND DENISE WALSH. 2015. "How Political Science Can Be More Diverse: Introduction." *PS: Political Science & Politics* 48 (3): 441–4.
- MONDAK, JEFFERY J., AND MARY R. ANDERSON. 2004. "The Knowledge Gap: A Reexamination of Gender-Based Differences in Political Knowledge." *The Journal of Politics* 66 (2): 492–512.
- MUTZ, DIANA C., ROBIN PEMANTLE, AND PHILIP PHAM. 2019. "The Perils of Balance Testing in Experimental Design: Messy Analyses of Clean Data." *The American Statistician* 73 (1): 32–42.
- PHULL, KIRAN, GOKHAN CIFLIKLI, AND GUSTAV MEIBAUER. 2019. "Gender and Bias in the International Relations Curriculum: Insights from Reading Lists." *European Journal of International Relations* 25 (2): 383–407.
- POWELL, DOUGLAS A. 2013. "Surveys Suck: Consumer Preferences when Purchasing Genetically Engineered Foods." *GM Crops & Food* 4 (3): 195–201.
- ROBERTS, MARGARET E., BRANDON M. STEWART, AND RICHARD A. NIELSEN. 2020. "Adjusting for Confounding with Text Matching." *American Journal of Political Science* 64 (4): 887–903.
- RUBLEE, MARIA ROST, EMILY B. JACKSON, ERIC PARAJON, SUSAN PETERSON, AND CONSTANCE DUNCOMBE. 2020. "Do You Feel Welcome? Gendered Experiences in International Security Studies." *Journal of Global Security Studies* 5 (1): 216–26.
- SHALABY, MARWA, NERMIN ALLAM, AND GAIL J. BUTTORFF. 2021. "Leveling the Field: Gender Inequity in Academia during COVID-19." *PS: Political Science & Politics* 54 (4): 661–7.
- SHAMES, SHAUNA L., AND TESS WISE. 2017. "Gender, Diversity, and Methods in Political Science: A Theory of Selection and Survival Biases." *PS: Political Science & Politics* 50 (3): 811–23.
- SJOBERG, LAURA, KELLY KADERA, AND CAMERON G. THIES. 2018. "Reevaluating Gender and IR Scholarship: Moving beyond Reiter's Dichotomies toward Effective Synergies." *Journal of Conflict Resolution* 62 (4): 848–70.
- SMITH, AMY ERICA, HEIDI HARDT, PHILIPPE MEISTER, AND HANNAH JUNE KIM. 2020. "Gender, Race, Age, and National Origin Predict Whether Faculty Assign Female-authored Readings in Graduate Syllabi." *PS: Political Science & Politics* 53 (1): 100–6.
- SUITOR, J. JILL, DOROTHY MECOM, AND ILANA S. FELD. 2001. "Gender, Household Labor, and Scholarly Productivity among University Professors." *Gender Issues* 19 (4): 50–67.
- SUMNER, JANE LAWRENCE. 2018. "The Gender Balance Assessment Tool (GBAT): A Web-based Tool for Estimating Gender Balance in Syllabi and Bibliographies." *PS: Political Science & Politics* 51 (2): 396–400.
- TEELE, DAWN LANGAN, AND KATHLEEN THELEN. 2017. "Gender in the Journals: Publication Patterns in Political Science." *PS: Political Science & Politics* 50 (2): 433–47.
- TURNER, CAROLINE SOTELLO VIERNES. 2002. "Women of Color in Academe: Living with Multiple Marginality." *The Journal of Higher Education* 73 (1): 74–93.
- WINDSOR, LEAH C., AND KERRY F. CRAWFORD. 2020. "Best Practices for Normalizing Parents in the Academy: Higher-and Lower-order Processes and Women and Parents' Success." *PS: Political Science & Politics* 53 (2): 275–80.
- YSSLEDYK, RENATE, KATHARINE H. GREENAWAY, ELENA HASSINGER, SARAH ZUTRAUEN, JANA LINTZ, MAYA P. BHATIA, MARGARET FRYE, ELSE STARKENBURG, AND VERA TAI. 2019. "A Leak in the Academic Pipeline: Identity and Health among Postdoctoral Women." *Frontiers in Psychology* 10: 1297.
- ZVOBGO, KELEBOGILE, AND MEREDITH LOKEN. 2020. "Why Race Matters in International Relations." *Foreign Policy* 237: 11–13.

Appendix A.

Table A1. Gender citation policies in IR journals

Type of policy	Journal
No statement or policy	British Journal of Political Science; Comparative Politics; Foreign Policy Analysis; Global Governance; International Relations; <i>International Security</i> ; <i>Journal of Conflict Resolution</i> ; <i>Journal of Peace Research</i> ; <i>Journal of Politics</i> ; <i>Millennium</i> ; <i>Political Science Quarterly</i> ; <i>Review of International Organizations</i> ; <i>Survival</i> ; <i>World Politics</i>

Table A1. Continued

Type of policy	Journal
General diversity statement in journal mission statement or editorial letter	<i>American Journal of Political Science; American Political Science Review; International Affairs; International Studies Perspectives</i>
Citation-specific policy that asks authors to “be mindful”, or “check” or “review their references”	<i>European Journal of International Relations; International Interactions; International Organization; International Studies Quarterly; International Studies Review; International Theory; Review of International Political Economy; Review of International Studies; Security Studies</i>

Table A2. Overall number of responses by country

Country	Responses (<i>n</i>)	Response rate (percent)
Argentina	43	38.39
Australia	137	43.39
Brazil	94	25.2
Canada	196	31.68
Chile	13	33.33
Colombia	33	33
Denmark	70	32.14
Finland	20	32.78
France	37	14.91
Germany	125	21.42
Hong Kong	11	28.2
Ireland	26	45.31
Israel	69	45.51
Italy	66	45.03
Japan	86	19.86
Mexico	41	14.69
Netherlands	80	34.87
New Zealand	30	50.76
Norway	56	20.06
Philippines	41	60.56
Poland	112	25.9
Singapore	23	35.29
South Africa	10	18.86
South Korea	31	10.5
Sweden	93	29.09
Switzerland	18	15.92
Taiwan	48	24.75
Turkey	173	22.19
Ukraine	97	54.8
United Kingdom	281	21.19
United States	1552	31.77

Appendix B. Tests of randomization scheme, differential attrition across gender, and effect of treatment on the number of names listed

Table B1. Using computer-generated test data to show that our randomization scheme was correctly programmed and implemented

Treatment group	Prop. female respondents
Treatment	0.463
Control	0.461
Difference of means	0.002 (95 percent CI: -0.076 to 0.080)
<i>p</i> -value	0.953

Table B2. There is no evidence of differential attrition across gender of respondent

	Non-response (1 = no answer)	
	(1)	(2)
Treatment (reference = control)		
Gender treatment	0.04 (0.03)	0.04 (0.03)
Gender (reference = male)		
Female		0.03 (0.04)
Treatment interacted with gender (reference = male)		
Gender treatment: female		-0.01 (0.06)
Constant	0.19*** (0.02)	0.18*** (0.02)
<i>N</i>	995	995

*** $p < 0.01$.

Table B3. Assignment to treatment does not appear to affect the number of names listed by respondents

	Sum of names listed	
	(1)	(2)
Treatment (reference = control)		
Gender treatment	0.11 (0.10)	0.12 (0.12)
Gender (reference = male)		
Female		0.09 (0.16)
Treatment interacted with gender (reference = male)		
Gender treatment: female		0.01 (0.23)
Constant	0.92*** (0.07)	0.88*** (0.09)
<i>N</i>	995	995

*** $p < 0.01$.

Appendix C. Regression results

Model 1: Includes the gender treatment, and several demographic variables: the respondents gender, subfield, paradigm, and rank.

Model 2: Includes the treatment, the same demographics variables, and an interaction term between the respondents gender and the treatment.

To create the reference category Subfield: Other we collapsed the following relatively smaller subfields:

Comparative Foreign Policy, Development Studies, European Studies/European Integration, Global Civil Society, History of the international relations discipline, Human Security, International Relations of a particular region/country, International Relations Theory, International/Global Ethics, International/Global Health, International/Global History, International/Global Environmental Politics, Gender in IR, Religion in IR, International Law, Other, and respondents whose subfield does not explicitly fall within IR.

Table C1. Gender treatment including demographic controls: US

	(1)	(2)
Treatment (reference = control)	4.30 ^{***} (1.47)	10.17 ^{***} (2.69)
Gender (reference = female)	-14.18 ^{***} (1.65)	-10.15 ^{***} (2.25)
Rank: associate professor (reference = assistant professor)	-2.93 (2.01)	-3.21 (2.00)
Rank: emeritus	-13.99 ^{***} (4.66)	-14.43 ^{***} (4.64)
Rank: full professor	-2.75 (1.97)	-3.03 (1.97)
Rank: other	-6.89 ^{***} (2.56)	-7.22 ^{***} (2.55)
Subfield: human rights (reference = other)	17.53 ^{***} (4.12)	16.98 ^{***} (4.11)
Subfield: international organization(s)	4.17 (3.51)	4.21 (3.50)
Subfield: international/global political economy	2.80 (2.48)	2.83 (2.47)
Subfield: international/global security	1.16 (1.84)	1.12 (1.83)
Subfield: US foreign policy	-3.02 (2.77)	-3.17 (2.75)
Paradigm: liberalism (reference = constructivism)	-7.03 ^{***} (2.47)	-7.17 ^{***} (2.46)
Paradigm: non-paradigmatic	-5.13 ^{**} (2.12)	-5.17 ^{**} (2.11)
Paradigm: other	-3.47 (2.56)	-3.74 (2.56)
Paradigm: realism	-13.01 ^{***} (2.42)	-13.27 ^{***} (2.42)
Treatment interacted with gender (reference = female)		-8.36 ^{***} (3.20)
Constant	33.30 ^{***} (2.50)	30.88 ^{***} (2.65)
N	756	756

** $p < 0.05$; *** $p < 0.01$.

Table C2. Gender treatment including demographic controls: non-US

	(1)	(2)
Treatment (reference = control)	5.39** (2.17)	1.48 (3.76)
Gender (reference = female)	-14.52*** (2.33)	-17.38*** (3.23)
Rank: associate professor (reference = assistant professor)	5.15* (2.74)	5.00* (2.74)
Rank: emeritus	-5.58 (7.85)	-6.51 (7.87)
Rank: full professor	1.65 (3.06)	1.71 (3.06)
Rank: other	-4.29 (3.56)	-4.34 (3.56)
Subfield: human rights (reference = other)	-1.30 (6.63)	-1.53 (6.63)
Subfield: international organization(s)	7.27 (5.62)	7.01 (5.61)
Subfield: international/global political economy	1.41 (3.57)	1.47 (3.57)
Subfield: international/global security	-1.38 (2.91)	-1.45 (2.91)
Subfield: US foreign policy	-12.78*** (3.99)	-12.97*** (3.99)
Paradigm: liberalism (reference = constructivism)	-7.48* (3.94)	-7.51* (3.94)
Paradigm: non-paradigmatic	-3.03 (3.07)	-3.34 (3.07)
Paradigm: other	3.50 (2.81)	3.20 (2.82)
Paradigm: realism	-17.43*** (3.87)	-17.56*** (3.87)
Treatment interacted with gender (reference = female)		5.82 (4.58)
Constant	29.46*** (3.24)	31.63*** (3.66)
<i>N</i>	394	394

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.