Gender in the *International Studies* Quarterly Review Process

Dani K. Nedal, Georgetown University and Massachusetts Institute of Technology Daniel H. Nexon, Georgetown University

e have good reasons to believe—not only from our own experience and observation, but also from recent studies-that women are underrepresented in syllabi (Colgan 2017), citations (Maliniak, Powers, and Walter 2013), faculty rosters (Shen 2013), and, most relevant to this forum, the pages of our leading journals (Mathews and Andersen 2001; Breuning and Sanders 2007; Evans and Moulder 2011; Williams et al. 2015).

While we see considerable variation in the number of female authors in each issue of International Studies Quarterly (ISQ), women publish in the journal at a rate lower than we would expect given the number of research-active female scholars in international studies. The December 2017 issue has 26 authors, six of whom are women. Only two articles are single authored by a woman and only one is coauthored only by women. The pooled average across the last three years is a little less lopsided, but not by much. Only about 36% of articles accepted in 2013-16 had at least one female author, only around 11% were authored solely by women and about 5% were coauthored by all-female teams.

Looking only at ultimate outcomes, however, only provides part of the picture. For example, are there gaps in representation or biases at work in other stages of the editorial process that contribute to the underrepresentation of female authors in "top" journals? Is this pattern reflected in the pool of submissions? Or both? In this article, we use data from our online submissions system—supplemented by hand coding conducted by a variety of editorial and undergraduate assistants—to examine patterns that might suggest bias in ISQ's peer-review process. Only one model suggests evidence of any such bias (though in the opposite direction), which implies that the "gender gap" in publication rates is mostly-if not entirely-a function of differential rates in submissions. That is, that larger structural factors (and perhaps expectations of bias at ISQ) account for the "gender gap." But we caution that such analysis is preliminary, based on blunt aggregate data, and therefore of limited value.

OVERVIEW: THE EDITORIAL PROCESS AT ISQ

ISQ uses a large team of editors, and the journal has seen editorial turnover over the years. The lead editor oversees all decisions and is ultimately responsible for the editorial process. A team of senior editors, which currently includes one woman and three men, enjoys formal "voting rights" and access to the administrative functions of the online system. The lead and senior editors also handle a proportionately larger number of manuscripts than other editors.

A team of associate editors, currently including three women (down from four as of January 1, 2018) and seven men, is responsible for shepherding a subset of manuscripts. In cases where the lead or a senior editor acts as a "shepherd," they do so without the direct assistance of an associate editor. When associate editors act as shepherd, they always work with a senior editor.

Nonetheless, in practice, any editor may weigh in on a manuscript when it comes to making a decision, and editors often request input from specific colleagues on a variety of matters related to manuscripts and referees. All members of the editorial team participate, if they want to, in policy discussions and by serving on ad hoc committees dealing with matters such as appeals.

Our process creates multiple opportunities for differential treatment of manuscripts. These include variation in the number of reviewers consulted to the number of editors involved. The "opt in" character of a lot of ISQ's decision making creates even more room for practical variation. For example, we use a "single veto" rule for overruling recommendations not to send a manuscript for external review (desk rejections). But, often, very few editors weigh in on these decisions.

Moreover, International Studies Association (ISA) policy emphasizes that referee reports are merely "advisory." The number of manuscripts that ISQ receives means that editors sometimes decline to pursue submissions that garner revise and resubmit (or *more* favorable) recommendations from referees. Thus, fairly subjective assessments about "contribution" and "scope of revisions necessary for publication" guide decision making. This creates even more space for conscious and unconscious bias to shape editorial decisions.

Such concerns led to discussions about moving to "tripe-blind review," in which decision-making editors do not know the identity of authors. Our process is already partially triple blind, in that only the editors assigned to a manuscript always know the identities of authors. The shared spreadsheet, draft decision letters, and PDFs of the manuscripts posted along with the draft decision letters, do not contain identifying information. But shepherding editors and the lead editor—who tracks the status of manuscripts in the administrative system—know the identity of authors. Since their evaluations matter the most to the fate of a manuscript, this partial anonymity is of limited value in checking conscious and unconscious bias associated with the identity of authors.

In the end, we determined that fully triple-blind review is probably not a practical option without substantial costs in terms of administrative oversight. Instead, we decided to, on a trial basis, blind editors during the desk rejection phase of the process. This created a quasi-experiment. As we discuss later, we did see a "gender gap" effect from this shift. Prior to nonblind desk rejection decisions, manuscripts with female authors received fewer desk rejections—in percentage terms—than those authored by men. Afterwards, the difference in this stage of evaluation lost statistical significance. This suggests bias arising out of a policy commitment to address gender disparities. At the same time, the change did not alter ultimate acceptance or rejection. How we interpret this effect depends on how we view the tradeoffs at

academic production by women. Another possible answer to the question of underrepresentation relates to what kinds of scholarship is submitted to *ISQ* and where it originates, and how gender gaps vary across geography and substantive areas of research.

Reports from some of the other journals—especially high-volume, high-reputation journals—involved in this project also find that the "gender gap" begins in the pipeline. Structural factors may be impeding or hampering academic production by women.

stake, such as "wasted time" for authors versus the putative value of referee, rather than only editorial, feedback.

SUBMISSIONS

One useful place to start is the submission pool. As shown in tables 1 and 2, the underrepresentation problem starts at the submission stage, with only 36% of manuscripts having at least one female author, and only 21% being authored exclusively by women (individually or coauthoring).

This first cut at the data suggests that while women are underrepresented in the pages of the journal relative to their overall distribution in the population, the gap *might be* largely explained by a disproportion at the submission stage.

One important avenue of research is, therefore, why might it be the case that women self-select out of submitting to *ISQ* or other journals. Reports from some of the other journals—especially high-volume, high-reputation journals—involved in this project also find that the "gender gap" begins in the pipeline. Structural factors may be impeding or hampering

Table 1 All Submissions, 2013–2017

Any Female Author	N	%
0	1,793	64.34
1	993	35.66

Table 2 All Submissions, 2013–2017

	Ratio of Female to Male	N	%
All Male	0	1,793	64.4
	0.2	7	0.2
	0.25	19	0.7
	0.333	80	2.9
	0.5	254	9.1
	0.6	2	0.1
	0.666	30	1.1
	0.75	8	0.3
	0.833	1	0.04
All Female	1	590	21.2

For example, the most recent ISA data suggests that women are no longer a minority in the student body and among junior faculty globally. But we do not have data on the distribution within the countries that contribute the most submissions, the United States and United Kingdom (over 60% of submissions).

ISA section membership data does suggest, however, that certain subfields within *ISQ*'s scope are more generally unbalanced than others. In particular, the largest ISA section, and the issue area that responds for the largest number of *ISQ* submissions, International Security, is among the most male-dominated (around 64% of registered members of the International Security Studies Section are men—see table 3).

This examination of the sample of submissions and the broader population of scholars who are potential submitters helps explain the pattern of underrepresentation and suggests that the main causes of the gap precede the editorial process. However, this does not, in itself, mean that there are no systematic differences in how men and women experience the editorial process. In other words, gender of the authors may still be a factor in the editorial decisions at different stages of the process. In the next section, we turn to some evidence from statistical models of editorial decisions.

MODELING DECISIONS

Modeling this type of process is challenging and all conclusions drawn should be taken with the requisite grain of salt. There are important omitted variables, namely the "quality" of submissions. These cannot practically be measured outside of the decisions and recommendations themselves. But they may "objectively" differ systematically depending on the gender of the author(s)—due, for example, to systematic differences with regards to self-confidence and risk-acceptance. Another way to put it is to think of it as a self-selection stage missing from the model.

Table 3 **Gender Gap in International Security**

	lı	International Security		
Any Female Author	0	1	Total	
0	59.65%	66.35%	1,386	
1	40.35%	33.65%	830	
Total	(1,129)	(957)	(2,216)	
Diff. of means <i>T</i> -statistic: 3.	23 p < 0.001			

Working with the data at hand-ISQ submissions data pooled from October 2013 to October 2017—we are able to find a few patterns of likely interest. The models in this article simplify the editorial process slightly. The unit of observation is the manuscript. Each manuscript may go through several iterations and rounds of decisions. Instead of considering the result of each iteration and each individual decision, we focus on the first editorial decision (whether or not to send a manuscript out for peer review) and the final decision (after all is said and done, is the piece accepted?). For many manuscripts, the final decision is the first decision after initial peer review

than those without (very unlikely that the difference, if there is one, favors manuscripts without female authors).2

But what of manuscripts that are sent out for review? To investigate those, we look first at the initial decisions they receive. Most first decisions after review are either rejections (75%, coded o) or revise and resubmit decisions (22%, coded 1), with few conditional acceptances and acceptances (2%, coded 2).3

As we see in table 5, among manuscripts sent out for initial review, Any Female Author has no noticeable effect on likelihood of getting positive decision (revise and resubmit or conditional acceptance/acceptance).4

Namely, before triple blind, manuscripts with female authors were less likely to receive desk rejections (36% vs. 44% without female authors) and therefore more likely to proceed to peer review. Under blinded screening that difference ceases to be statistically significant.

(decline or, extremely rarely, accept). For others, those that receive initial revise and resubmit decisions, there are intermediate stages between the first and final editorial decision. We condense those stages.

In analyzing the initial decision stage (whether or not to send a manuscript out for peer review), we leverage a quasi-experiment created by the introduction, in May 2016, of triple-blind review. As noted previously, under the original procedures for the peerreview process at ISQ, the senior editor or associate editors assigned to handle a given manuscript would see the names of the authors when they "claimed" the manuscript, and therefore were not "blind" when making the decision of whether or not to send the manuscript out for peer review. Under current rules, the names of the authors are not revealed to the relevant editors until after the final decision about whether or not to send a manuscript for external review.

If editors are subject to implicit or explicit biases regarding the gender of authors, the introduction of blinded desk screening should, in principle, mitigate these biases. Table 4 presents the results from a logit model of the initial decision. The dependent variable is coded as o if the manuscript is sent out for review and 1 if it receives an editorial rejection (aka desk rejection). We interact the variable Any Female Author¹ and the variable indicating whether the manuscript was handled under "double blind" (0) or "triple blind" (1). We control for the methods employed in the manuscripts, the substantive areas of research, the country of origin of the submitting author, whether the manuscript is coauthored, and whether the manuscript was submitted by graduate student(s). Negative coefficients mean that manuscripts are less likely to be desk rejected.

We find evidence that the introduction of blinded editorial screening changes the effect of the gender of the authors. Namely, before triple blind, manuscripts with female authors were less likely to receive desk rejections (36% vs. 44% without female authors) and therefore more likely to proceed to peer review. Under blinded screening that difference ceases to be statistically significant. Figure 1 plots the coefficients for the model. While the coefficient for Any Female Author does not reach conventional levels of significance (cannot reject the null that there is no difference), it seems that even under blinded desk screening, manuscripts with female authors are more likely to fare better

We also don't find significant differences for manuscripts with particular methodological or substantive orientations. The only variables that show consistent results are Coauthored, Anglophone Author, which are associated with more positive decisions (i.e., coauthored papers do better than single-authored papers), and

Table 4
Triple Blind, Gender, and Desk Rejections

Any Female Author Double Blind	-0.30** (0.13)
Triple Blind	0.08 (0.12)
Coauthored	-0.18* (0.10)
Uses Formal	-0.99** [*] (0.25)
Uses Quantitative	-0.43*** (0.10)
International Security	-0.42*** (0.10)
International Political Economy	0.02 (0.12)
Foreign Policy	0.15 (0.11)
Human Rights	0.01 (0.14)
International Relations Theory	-0.30** [*] (0.11)
Methodology	0.38** (0.16)
Anglo Submitting Author	-0.73*** (0.10)
Student Author	0.32*** (0.10)
Constant	0.15 (0.27)
Observations	2,205
Pseudo R ²	0.066

Coefficient Plot for Table 4

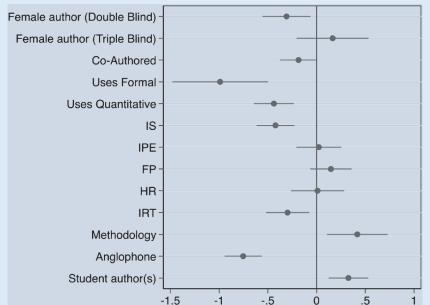


Table 5
First Decision after Peer Review

Any Female Author	-0.04 (0.13)
Coauthored	0.31** (0.13)
Uses Formal	0.21 (0.28)
Uses Quantitative	0.13 (0.15)
International Security	0.13 (0.14)
International Political Economy	0.23 (0.16)
Foreign Policy	-0.23 (0.16)
Human Rights	0.05 (0.20)
International Relations Theory	0.33** (0.16)
Methodology	0.22 (0.22)
Anglophone Author	0.79*** (0.16)
Student Author	-0.44*** (0.16)
Constant	-2.74*** (0.21)
Observations	2,202
Pseudo R ²	0.029

Student Author, which is associated with less positive decisions.

Another way of looking at this is to ask if the manuscripts that pass editorial review eventually get accepted. While for an initial round of reviews getting a revise and resubmit decision is undeniably a positive result, roughly 23% of revised manuscripts still get rejected in a second or third round, and ultimately what matters is getting accepted for publication. To this effect, we ran a logit model with the dependent variable coded as o if the manuscript is rejected and 1 if it is accepted. Manuscripts that had outstanding revise and resubmits at the time of the data collection were excluded, as were manuscripts that never made it to peer review. The results, presented in table 6, again show no appreciable difference in performance between manuscripts with and without female authors.

There remains, however, a question of whether these steps in the editorial process should be modeled separately, or if so doing introduces biases. In fact, we *know* that these

decisions are interdependent. Generally speaking, editors are more likely to send out for review manuscripts that they think are likely to get positive reviews (those that are unlikely to survive the review process receive editorial rejections). While editors may not be perfect predictors, this still means that the same variables that predict selection into peer review should also be predictors of the ultimate outcome. It also, importantly, means that if some unobserved factor influences selection of some individuals into the sample (peer review), and that same omitted variable influences the outcome in a way related to the original variable, the coefficient for that variable in the outcome model will be biased.

In this case, if manuscripts without female authors are less likely to be sent for review, there may be something about those that make it out to review that will make them also more likely to get accepted. Conversely, if some manuscripts with female authors make it out for review for any reason other than their inherent quality relative to those without female authors, we should expect those additional manuscripts to fail the second stage at higher rates than other manuscripts with and without female authors. In other words, we might expect a *negative* correlation between Any Female Author and acceptance once selection into peer review is accounted for.

Because selection models are inherently hard to estimate and are subject to bias and inconsistency depending on the viability of the so-called exclusion restriction and the degree of correlation between the errors in both stages, we estimate this process in a few different ways.

In this case we have what should on its face be a semi-decent exclusionary variable, triple blind, which we know increased the rate of desk rejections and seems to have mitigated the positive effect of female authorship, but should not, in principle, affect the ultimate outcome of individual manuscripts. Rather than report all of the results of each model, we report here in table 7 the top-line results from a Heckman probit and Anne Sartori's selection

estimator (Sartori 2003), which was designed specifically to do away with the necessity of an exclusion restriction (and therefore serves as a good test for the viability of the restriction). The Heckman was estimated with and without the exclusion restriction (triple blind).

make it more likely to be accepted, not less). This lends support to the use of Sartori's selection model and the prioritization of the results obtained therewith. In that model, we find no effect of author gender on the outcome of the submission conditional on it being sent for review.

It is possible that male and female scholars have systematically different attitudes toward manuscripts, that male and female reviewers get selected according to different criteria, or that reviewer gender correlates with other factors such as method or substantive issue-area, which in turn correlate with different propensities for positive recommendations.

The takeaway is that the results are not robust. Once we account for selection into peer review with a Heckman probit, manuscripts *without* female authors seem slightly more likely to receive positive decisions. However, that might just be a statistical artifact from the bias introduced into the Heckman model by the absence of an appropriate exclusion restriction. This is extremely likely given that the Heckman model returns a highly negative correlation between the error term in the two models (the rho statistic is -.99), which is extremely unlikely given all we know about how the process actually works (i.e., overall, things that make a manuscript more likely to be sent out for review should also

Table 6 Final Decision, Reject (0) or Accept (1)?

Any Female Author	-0.22 (0.17)
Coauthored	0.50*** (0.16)
Uses Formal	-0.03 (0.32)
Uses Quantitative	0.06 (0.20)
International Security	0.06 (0.18)
International Political Economy	0.46** (0.21)
Foreign Policy	-0.14 (0.20)
Human Rights	0.14 (0.25)
International Relations Theory	0.13 (0.21)
Methodology	0.40 (0.28)
Anglophone	0.69*** (0.21)
Student Author	-0.47** (0.20)
Constant	-2.45*** (0.29)
Observations	1,200
Pseudo R ²	0.033
Standard errors in parentheses $p < 0.10, p < 0.05, p < 0.01$	

REVIEWER GENDER

The above analysis focuses on the gender of the authors, but brackets the gender of reviewers. There are a few ways the gender of reviewers may affect the outcome of reviews. It is possible that male and female scholars have systematically different attitudes toward manuscripts, that male and female reviewers get selected according to different criteria, or that reviewer gender correlates with other factors such as method or substantive issue-area, which in turn correlate with different propensities for positive recommendations. Generally speaking, roughly 30% of reviews were submitted by female scholars, and 55% of manuscripts had at least one female reviewer. In some issue areas these proportions are different. For example, 44% of reviews submitted for manuscripts categorized under Human Rights were done by female scholars, and 73% of those manuscripts had at least one female reviewer. Manuscripts using formal methods are usually reviewed only by male scholars (64%).

Table 8 models reviewer recommendations for first-time submissions (i.e., excluding resubmissions). The unit of analysis is the reviewer for a given manuscript. The key dependent variable is whether the reviewer recommends a negative (reject) or positive decision (revise and resubmit, conditional accept, and accept). The key independent variable is the gender of the reviewer. In model 1 we treat the gender of the reviewer and the authors as independent. In model 2 we interact the two. For the sake of simplicity, we again present results for Any Female Author. The results are the same using other categorizations.

Table 9 changes the unit of analysis to the manuscript and asks whether having female reviewers affects the likelihood of positive *decisions* in first-round reviews. The key independent variables of interest are again whether any of the reviewers are women and whether any of the authors are women, first independently, and then interacted. We again find no significant effects for either. In robustness tests we test for the effects of number of female

Table 7
Selection Models

	Heckman (no exclusion)	Heckman (exclusion)	Sartori
Any Female Selection	0.188***	0.181***	0.167***
Outcome	-0.165***	-0.178***	-0.059
Rho	-0.99***	-0.99***	1 (assumed)

reviewers (controlling for overall number of reviewers) and the gender of the authors (all male, all female, mixed, as well as a more fine-grained measure of the ratio of male to female authors in a manuscript). We find no statistically significant effects for any of these measures.

CONCLUDING THOUGHTS

The crude analysis of this aggregate data suggests that the "gender gap" is not a consequence, at least at ISQ, of factors endogenous to the review process. Rather, it likely stems from broader factors trained in different countries and in different research traditions have different expectations about not just manuscripts, but how referees should evaluate manuscripts. And many editors believe that certain research communities are marked by an ethic of "supportiveness" toward one another, while others have a more "critical" culture of peer review.

As we noted inter alia above, solid analysis of the effects of various treatments-including gender-on the different aspects of the peer-review process proves incredibly challenging. Thus, in addition to observational statistical studies, we need to study the

Given that publishing in peer-reviewed journals is becoming one of the single most important sources of academic capital—and hence of status, employment, advancements, and compensation—we find this troubling. It strikes us as incumbent to better understand the biases and tradeoffs of peer review.

in the field, or from specific characteristics of the journal—the comparative lack of female editors (currently only one of our five senior editors is a woman, and three of the ten associate editors) and or other signals—that might discourage women from submitting. Thus, we are very interested in the comparative data generated by the broader inquiry represented by this forum. That the other contributors seem to find nearly identical results in their respective journals is somewhat encouraging, or discouraging, depending on how you look at it.

Political science in general, and international relations in particular, has lagged other fields in systematically analyzing the peer-review process. Given that publishing in peer-reviewed journals is becoming one of the single most important sources of academic capital—and hence of status, employment, advancements, and compensation—we find this troubling. It strikes us as incumbent to better understand the biases and tradeoffs of peer review.

There is much work to be done. Together with Catherine Weaver, Jeff Colgan, and others we have begun to leverage our data for more granular insights. For example, our preliminary analysis suggests that untenured scholars tend to recommend rejection at higher rates than their tenured colleagues. This accords with the "folk wisdom" among many editors that such scholars have come of age in an era marked by more (perceived) selectivity at highly-ranked journals. Other "folk wisdom" holds that scholars

Table 8 **Reviewer Gender and Recommendations**

Female Reviewer	0.08 (0.08)	0.02 (0.10)
Any Female Author	0.05 (0.08)	0.00 (0.09)
Female Reviewer * Female Author		0.15 (0.16)
Observations	3,125	3,125
Pseudo R ²	0.001	0.001
Clustered standard errors in parentheses $p < 0.10, p < 0.05, p < 0.01$		

process using experimental, ethnographic, and other methods. For example, journals should consider conducting experiments,

1 ' J	0	1	,
such as randomly assigning the number	of review	ers to	manu-
scripts. Professional associations should	facilitate	lower	-stakes
•			
m [1]			
Table 9			

Any Female Reviewer	-0.10 (0.13)	0.03 (0.16)	0.04 (0.17)
Any Female Author	-0.06 (0.13)	0.15 (0.20)	0.07 (0.21)
Female Reviewer * Female Author		-0.36 (0.26)	-0.35 (0.28)
Coauthored			0.29** (0.14)
Uses Formal			-0.15 (0.29)
Uses Quantitative			-0.08 (0.17)
International Security			-0.07 (0.15)
International Political Economy			0.26 (0.18)
Foreign Policy			-0.16 (0.17)
Human Rights			0.14 (0.21)
International Relations Theory			0.20 (0.18)
Methodology			0.49*** (0.24)
Anglophone			0.46*** (0.17)
Student Author			-0.31* (0.17)
Observations	1,313	1,313	1,259
Pseudo R ²	0.000	0.002	0.0194

experiments in peer review, such as those not involving real decisions at journals.

At least on the observational side, *ISQ* has been collecting detailed demographic data on reviewers and authors. But many scholars decline to fill out our survey, and some even send angry emails attacking this effort. We think it is important for the profession to send a strong message in favor of the legwork necessary for sustained social-scientific inquiry into the editorial and peerreview processes. These shape not only our careers, but also the production of knowledge about politics and international affairs.

ACKNOWLEDGMENTS

This paper was originally prepared for "Gender in the Journals: Exploring Potential Biases in Editorial Processes," at the 2017 Annual Meeting of the American Political Science Association in San Francisco, California. We thank all participants for their feedback.

NOTES

1. In robustness tests, we reran the analysis using All Female Authors as the independent variable, as well as a trichotomous categorical variable (all female, all male, and mixed) and a continuous variable indicating the female-to-male ratio among authors (see table 2). The results suggest that, prior to triple blind, there is an effect for manuscripts having at least one female author, but no additional effects for having more than one, or only female authors. We repeated these robustness tests for all models presented below.

- It is still possible that as we collect more observations we will once again find statistically significant effects.
- 3. Because so few manuscripts get accepted off the bat, the results would be nearly identical using multinomial, ordinal, or standard logit models. We report the results of standard logit here for ease of interpretation and convenience.
- 4. The results are also nearly the same if we substitute "any" for "only female authors."

REFERENCES

- Breuning, Marijke, and Kathryn Sanders. 2007. "Gender and Journal Authorship in Eight Prestigious Political Science Journals." PS: Political Science & Politics 40 (2): 347–51.
- Colgan, Jeff. 2017. "Gender Bias in International Relations Graduate Education? New Evidence from Syllabi." *PS: Political Science & Politics* 50 (2): 456–60.
- Evans, Heather K., and A. Moulder. 2011. "Reflecting on a Decade of Women's Publications in Four Top Political Science Journals." *PS: Political Science & Politics* 44 (4): 793–798.
- Maliniak, Daniel, Ryan Powers, and Barbara F. Walter. 2013. "The Gender Citation Gap in International Relations." *International Organization* 67 (4): 889–922.
- Mathews, A. Lanethea, and Kristi Andersen. 2001. "A Gender Gap in Publishing? Women's Representation in Edited Political Science Books." *PS: Political Science & Politics* 34 (1): 143–47.
- Sartori, Anne E. 2003. "An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions." *Political Analysis* 11 (2): 111–38.
- Shen, Helen. 2013. "Mind the Gender Gap." Nature 495 (7439): 22.
- Williams, Helen, Stephen Bates, Laura Jenkins, Darcy Luke, and Kelly Rogers. 2015. "Gender and Journal Authorship: An Assessment of Articles Published by Women in Three Top British Political Science and International Relations Journals." European Political Science 14: 116–30.