Scientometrics (2017) 113:1113–1127
DOI 10.1007/s11192-017-2517-5

Language and socioeconomics predict geographic variation in peer review outcomes at an ecology journal

C. Sean Burns[1] · Charles W. Fox[2]

Received: 21 April 2017 / Published online: 11 September 2017

Abstract Papers submitted by scientists located in western nations generally fare better in the peer review process than do papers submitted by scientists from elsewhere. This paper examines geographic variation in peer review outcomes (whether a manuscript is sent for review, review scores obtained, and final decisions by editors) for 3529 submissions over a 4.5 year period at the journal *Functional Ecology*. In particular, we test whether geographic variation in language and socioeconomics are adequate to explain most or are all of this variation. There was no relationship between the geographic regions of handling editors and the decisions to send papers for review or invite revision, but there was substantial variation among author geographic locations; generally papers from first authors located in Oceania, the United States, and the United Kingdom fared better, and papers from first authors located in Africa, Asia, and Latin America fared worst. Language and the Human Development Index (HDI) explained the geographic variation in the proportion of papers sent for review, but socioeconomics alone (HDI) was the best predictor of mean review scores obtained by papers and whether authors were invited to submit a revision. Though we cannot exclude a role for editor and reviewer biases against authors based on their geographic location, variation in socioeconomics and language explain much of the variation in manuscript editorial and peer review outcomes among authors from different regions of the world.

Keywords Peer review · Language bias · Geographic bias · Socioeconomics · Human development index

✉ C. Sean Burns
sean.burns@uky.edu

Charles W. Fox
cfox@uky.edu

[1] School of Information Science, University of Kentucky, Lexington, KY 40506-0224, USA

[2] Department of Entomology, University of Kentucky, Lexington, KY 40546-0091, USA

Mathematics Subject Classification 62P25

JEL Classification C12 · C13 · C14

Introduction

Papers submitted to academic journals by scientists located in the United States (U.S.) and other western nations fare better in the peer review process than do papers submitted by scientists from elsewhere (Man et al. 2004), regardless of the national location of the reviewers evaluating their manuscripts (Link 1998). Much of this variation across nations may be due to differences in the scientific history and cultures of the regions, which are reflected in the quality of the manuscripts submitted by authors, and likely covary with investments in research (cf. King 2004).

Most scientific literature is published in English (Cronin 2009), and thus proficiency in English likely contributes to explaining variation across nations in publication success. Native English speakers likely have an advantage because they are writing in their first language, unlike non-native English speaking scientists (Cronin 2009; Clavero 2011). Even authors from highly-developed western but non-native English speaking nations (e.g., various European nations) often experience lower acceptance rates compared to those from English-speaking countries (Primack and Marrs 2008; Primack et al. 2009; Burgman et al. 2015). This suggests a language over a nationality bias, especially when the effects hold after controlling for national wealth (Treganza 2002) and because there may be no significant interaction between nationality and editorial role or nationality and a manuscript's outcome (Zhang 2012; Harris et al. 2015).

Other factors that vary globally likely also contribute to variation in success during peer review. In particular, various biases, conscious or otherwise, may influence editor or reviewer assessments independent of research or manuscript quality (Lee et al. 2013). For example, scientists may exhibit preferences during peer review for papers from authors who are located in the same geographical or national region of the journal's publishing location (Ernst and Kienbacher 1991; Opthof et al. 2002), regardless of the nationality of the reviewer (Daniel 1993). Alternatively, reviewers may exhibit biases against authors from their own geographic region; for example, Chinese reviewers were more likely to recommend rejection or less likely to recommend acceptance of papers by Chinese authors, compared to reviewers from other nations (Zhang 2012; Campos-Arceiz et al. 2015). Blinding reviews could reduce some of these biases, allowing researchers to test whether the differences in success of authors from different regions of the world are due to objective or non-objective evaluations of research and manuscript quality. Unfortunately, because the effect sizes may be small (Ross et al. 2006), if they exist (Justice et al. 1998), disentangling the roles of language, economic or other nationality effects from unconscious biases are difficult.

In previous papers, we examined how the editorial process has been influenced by the manuscript titles that authors create (Fox and Burns 2015), the gender of the authors (Fox et al. 2016a) or of the reviewers and editors (Fox et al. 2016b), and the role that author-suggested reviewers play in this process (Fox et al. 2017a). In this study, we examine how manuscript decisions and outcomes (whether papers are sent for review, review scores, and final decisions) vary according to handling editor geography and author geography, and whether language and socioecomics explain this variation. Specifically, we test the hypotheses that (1) the geographic location of handling editors and authors predict editorial and peer review outcomes, and that (2) the language (English versus other) and (3) the economic, health, and educational developments of an author's home country explain much or possibly all of the variation across geography in editorial and peer review outcomes.

Materials and methods

Our data comes from the journal *Functional Ecology*. The data was made available to us through the second author's (CWF) relationship with the British Ecological Society, which owns the journal, as Executive Editor of the journal, and was approved by the Society. The data is an uniquely high quality set of data that includes the complete set of standard research paper submissions to this specific journal over the period of our study, and so provides a near complete picture of the peer review process, from submission to decision, for one journal. Author names were not shared with the first author (CSB). The research was approved by the University of Kentucky's Institutional Review Board (IRB # 14-0570-P4S), and the anonymized data was deposited in the Dryad data repository for public access (Fox et al. 2015).

Functional Ecology uses *ScholarOne Manuscripts* (previously *Manuscript Central*) to manage manuscript submissions and the peer review process (Fox et al. 2016b). We used *ScholarOne* data that were extracted on 19 December 2014 for all 'standard' papers submitted to *Functional Ecology* between 1 January 2010 and 30 June 2014 (inclusive). 'Standard' papers include all typical research studies (empirical or theoretical), and exclude review papers, commentaries, perspectives, editorials and other types of papers not considered typical research manuscripts. All papers examined had completed the editorial process at the time the data were extracted from *ScholarOne*. This time period (1 January 2010–30 June 2014) covers 3529 submissions and 14,248 total non-unique authors located in 100 nations across seven geographic regions. As in Fox et al. (2016b), we categorized regions using the M.49 area codes for continental regions, as defined by the United Nations' Statistical Commission (unstats.un.org), with two exceptions: (1) we divided the Americas into Latin America, including Mexico and other countries south of the United States, and North America, to reflect linguistic differences, and (2) we partitioned out the United Kingdom from Europe. We made the latter change because *Functional Ecology* is owned by a British learned society (the *British Ecological Society*) and is heavily contributed to by British authors and handling editors, relative to other regions.

A primary objective of this study is to test whether the language and/or socioeconomic status of authors affects peer review outcomes and explains much or all of the variation in outcomes across regions of the world. We tested for language effects by inferring language efficacy from geographic data (cf. Treganza 2002; Ross et al. 2006; Lee et al. 2013). We used the *CIA World Handbook* (2016) to assign languages to individual countries. Since the *World Handbook* lists language in rank order by dominance, we selected the first language listed. However, if English was listed as an official or a common language, despite the rank order, we selected English (for Cameroon, Ghana, India, Namibia, Rwanda, and Sri Lanka) since universities in, and thus authors from, these countries are likely proficient in English. We limited our analysis to a test of English versus not English; this is because the journal (*Functional Ecology*) is published in English and we are especially interested in whether being a proficient speaker in the language of the manuscript influences publication success. French Guiana and Martinique were not listed in the

CIA World Handbook, so we used *Wikipedia* to identify the major language (French in both cases).

To test for socioeconomic effects, we used the Human Development Index (HDI) from the United Nations Development Programme (2015). The HDI is a nation-centric measurement of individuals' capabilities with respect to health, education, and wealth (Zambrano 2014). It attempts to be a comprehensive measure of a nation's well-being and the capabilities of its peoples, and is meant to reflect not just gross national income (GNI) but also life expectancy and education. Since the index may range from zero to one, we rescaled the index to be from zero to 100 so that a one unit change in the index provides a useful interpretation of effect sizes (see Hosmer and Lemeshow 1989, p. 56). The HDI has seen limited use in the scholarly communication literature and results are inconsistent; however, although the HDI does not perfectly explain research output (quantity of publications) across all nations (Uthman et al. 2014), authors from nations with higher HDIs generally submit more papers than authors from nations with lower indexes (Keiser 2004; but see Fayaz-Bakhsh and Mousavi 2015) suggesting that HDI covaries with research investment.

The analyses were conducted in the *R Programming Language* (R Core Team 2016). Aside from the base library, data management and analyses were aided by packages *aod* (Lesnoff and Lancelot 2012), *car* (Fox and Weisberg 2011), *dplyr* (Wickham and Francois 2016), *ggplot2* (Wickham 2009), *Hmisc* (Harrell Jr. 2016) *MASS* (Venables and Ripley 2002), *plyr* (Wickham 2011), *pROC* (Robin et al. 2011), *reshape2* (Wickham 2007), *lme4* (Bates et al. 2015), and *vcd* (Meyer et al. 2016). All *R* code is available at the first author's GitHub account.[1]

Analyses

Each manuscript represents a single observation and contains a number of variables about authors (including national location, geographic region, and order of authorship), the peer review process (including whether papers were sent for review and mean review score), and the final decision (including whether the paper was rejected versus invited for revision). Like our previous studies (Fox et al. 2016a, b, 2017a) we primarily apply logistic regressions on binary dependent variables but here we also apply ordered logistic regression on binned review scores (Hosmer and Lemeshow 1989; Hilbe 2009; King 2008; Osborne 2008; Pedhazur 1997; Field et al. 2012). For logistic regressions, our binary responses include whether papers were sent for peer review (yes/no) and the decision on the paper (reject/not reject), and we use the geographic location of first authors in Europe as our reference group for comparing effect sizes (though choice of the reference group does not influence significance of the main effects). We used Europe as a reference because it is comprised of nations that speak a multitude of languages that vary in the size and health of their economies, and are either more western or eastern. Our ordinal dependent variable is analyzed by binning the mean review scores into three categories: low, middle, and high. Mean review scores range between 1 and 4, and low values indicate better papers (see Fox et al. 2016a for more details).

Our analysis is of the geographic regions and the national locations of handling editors or of first authors. First authors in *Functional Ecology* (as per the norms in ecology) are generally the main contributor to a paper, with last authors commonly being the head of a lab or research group (i.e., a guiding, senior author). First authors in our data are also generally the corresponding authors (77%) or share the same geographic region as

[1] https://github.com/cseanburns/peer_review_geography.

corresponding authors (Cramer's V: 0.965), and/or are the submitting authors (81%) or share the same geographic region as submitting authors (Cramer's V: 0.965).

We follow our geographic analysis with analyses that include three sets of models with the language and socioeconomic terms added to the geographic models. The goal is to explore whether language and socioeconomic data reduce the variation among geographic regions. We examine how language or socioeconomic data reduce the variation explained by first author geography of (a) the proportion of papers sent for review, (b) mean review score, and (c) the final decision made by the editor. We first fit all possible combinations of our independent variables, with the caveat that the models include the geographic region of first authors as the baseline. The best-fit models were selected based on the lowest AIC and by examining whether likelihood ratios of the additional terms resulted in a significant improvement over the baseline models that included only the geographic regions of first authors (Burnham et al. 2011).

Results

The journal relies on handling editors (Associate Editors) to choose reviewers and to recommend decisions to the senior Editors. Handling editors were located in all geographic regions, but scientists from Asia, Africa and Latin America are under-represented on the editorial board of *Functional Ecology* relative to the proportion of papers received ($X^2(6) = 1206.7, p < 0.001$) or sent for review ($X^2(6) = 568.5, p < 0.001$) from all seven regions. Handling editors located in North America manage the majority (54.2%) of papers sent for review. This is followed by handling editors located in European nations (19.2% of papers), the United Kingdom (11.1%), and Oceania (9.5%). Handling editors located in Asia, Africa, and Latin America managed collectively just 6% of papers submitted to the journal (4.5, 1.2 and 0.3%, respectively), whereas first authors located in these regions contributed 14.8% of manuscripts (9.1, 1.2, and 4.5%). Scientists located in Europe were also more highly represented as first authors than as handling editors by a ratio of 1.89:1. Scientists from Oceania were represented about equally as authors and editors (1.02:1), whereas those from North America and the United Kingdom were overrepresented on the editorial board relative to manuscripts submitted from these regions (author: editor ratios of 0.58:1 and 0.69:1, respectively).

A substantial proportion (35%) of papers that are declined without review are declined by Senior Editors before they are given to handling editors for consideration. Of 2299 submissions that were assigned to handling editors, 1771 (77%) were sent for review and 528 were declined before review. In a mixed effects model with handling editor IDs added as a random effect, the proportion sent for review did not vary among editor locations (sent_for_review (yes/no) $\sim$ handling_editor_geog + (1| handling_editor); $X^2(5) = 5.9$, $p = 0.32$; handling editors located in Latin America were dropped from this analysis due to the insufficient number of manuscripts handled by editors located there).

The probability that a paper was sent for review by the handling editor varied among the geographical regions of the first author. Manuscripts submitted from first authors located in North America were more likely to be sent for review and manuscripts with first authors located in Africa, Asia, and Latin America were less likely to be sent for review [sent_for_review(yes/no) $\sim$ first_auth_geog; $X^2(6) = 133.6$, $p(>X^2) < 0.001$; Table 1], compared to papers with first authors located in Europe (our reference point). At each extreme, papers with first authors located in North America were 46% more likely (OR = 1.46,

Table 1 Logistic regression testing whether the probability that papers were sent for review is influenced by first author geography

Region	Estimate (SE)	OR	95% CI	*P*
Intercept	0.005 (0.057)			
Africa	−0.770 (0.277)	0.463	[0.264, 0.785]	0.005
Asia	−0.822 (0.112)	0.439	[0.352, 0.546]	0.000
Latin America	−0.371 (0.159)	0.690	[0.504, 0.941]	0.020
North America	0.379 (0.088)	1.462	[1.229, 1.739]	0.000
Oceania	0.234 (0.130)	1.263	[0.979, 1.632]	0.073
United Kingdom	0.255 (0.255)	1.291	[0.974, 1.715]	0.077

Papers with first authors located in Africa, Asia, and Latin America were less likely to have papers sent for review, and papers with first authors in North America were more likely to have papers sent for review. Results were not statistically significant at alpha = 0.05 for papers with first authors from Oceania or the United Kingdom

OR odds ratio, *CI* confidence interval (exponentiated); Region equals location of first author. $N = 3407$, and 103 observations were removed because of missing geographical data for first authors. The model is sent_for_review(no/yes) ~ first_auth_geog. Odds ratios are relative to Europe and the reference level for sent_for review is *no*

$p < 0.001$) to have a paper sent for review over papers with first authors located in Europe (the reference group), whereas papers with first authors located in Asia were just over half as likely to be sent for review (56%; OR = 0.44) over papers with first authors located in Europe.

Handling editors use review scores and reviewer comments as the main criteria for deciding whether a manuscript should be published. Reviewer comments cannot easily be quantified (and are unavailable to us) but review scores are available. These scores are on a scale of 1 (best) to 4 (worst), with almost all papers receiving more than one (usually two but occasionally more) review scores. For *Functional Ecology* the distribution of mean review scores (averaged across reviewers) per paper is multimodal, so we binned these review scores into three categories: low, middle, and high, where values $1 \leq x_{\mathrm{low}} \leq 2$ (strong score, $n = 650$), values $2 < x_{\mathrm{middle}} \leq 3$ (fair score, $n = 716$), and values $3 < x_{\mathrm{high}} \leq 4$ equal high (poor score, $n = 334$).

The proportional odds of receiving fair scores (the middle bin) or poor scores (the high bin) versus strong scores (the low bin) varied among geographic regions (ordinal logistic regression with first authors located in Europe as the baseline; mean_ review_score(low/middle/high) ~ first_auth_geog; $X^2(6) = 29.75$, $p < 0.000$). Papers with first authors located in Asia were more likely to receive poor scores than fair or strong scores, whereas the reverse was true for papers with first authors located in North America. Papers with first authors located in Asia were 51% more likely to receive poorer peer review scores, whereas papers with first authors located in North America and the United Kingdom were 22.2 and 39.2%, respectively, less likely to receive fair scores or poor scores (versus strong scores), compared to papers with first authors located in Europe, with the remaining regions not significantly different from Europe (Table 2). We tested whether the variation in review scores among first author geographic regions could be explained by the location of peer reviewers or by the interaction between author and reviewer regions (e.g., if reviewers are showing either positive or negative bias towards authors of specific regions). We found that review scores were not explained by peer reviewer location ($X^2(6) = 4.94$,

Table 2 Ordinal logistic regression testing whether mean review scores (per paper) varied among first author geographic locations

Region	Estimate (SE)	OR	95% CI	P	OR ≥ 2	OR ≥ 3
Africa	0.513 (0.408)	1.671	[0.749, 3.739]	0.209	1.386	−1.099
Asia	0.413 (0.169)	1.512	[1.086, 2.106]	0.014	0.916	−0.885
Latin America	0.397 (0.228)	1.487	[0.951, 2.326]	0.081	0.853	−0.853
North America	−0.251 (0.110)	0.778	[0.627, 0.965]	0.023	0.339	−1.683
Oceania	−0.171 (0.164)	0.843	[0.611, 1.163]	0.299	0.380	−1.504
United Kingdom	−0.498 (0.185)	0.608	[0.421, 0.872]	0.007	0.000	−1.647

Papers with first authors located in Asia were more likely to receive poor scores than fair or strong scores, and the reverse was true for papers with first authors located in North America and the United Kingdom

OR odds ratio, *CI* confidence interval (exponentiated); Region equals location of first author. $N = 1700$ cases due to missing data for mean review score on six cases and missing data for first author geography on 65 cases. The model is mean_review_score(low/middle/high) ~ first_auth_geog. Proportional odds ratios are relative to Europe and the reference level for the mean review score is *low*

$p = 0.55$) or its interaction with first author location ($X^2(34) = 44.2$, $p = 0.11$). In this model, only first author region was predictive of review scores ($X^2(6) = 36.8$, $p < 0.001$; overall model: review_score(1–4) ~ author_region + reviewer_region + author_region:reviewer_region).

We found no detectable relationship between the geographic region of the handling editor and the final outcome of the editorial process (whether or not a revision is invited for a paper, of papers sent for review) [mixed model with handling editor ID as a random effect: paper_rejected (yes/no) ~ handling_editor_geog + (1| handling_editor); $X^2(6) = 2.77$, $p = 0.84$; Table 3]. However, final outcomes were variable among geographic regions of the first author. Compared to papers sent for review with first authors located in Europe, authors were more likely to be invited to revise their paper if the first authors were located in North America, Oceania, or the United Kingdom (ORs > 1.0, $p < 0.05$); at the extreme, first authors located in the United Kingdom were 1.76 times more likely to have a paper accepted than were first authors located in Europe. However, while the overall model was significant (logistic regression: paper_rejected (yes/no) ~ first_auth_geog; $X^2(6) = 26.2$, $p < 0.001$; Table 4), we found no significant difference between papers with first authors located in Africa, Asia, or Latin America and the reference group, Europe.

Of all multi-authored papers submitted, papers were more likely to be authored by people located within a single geographic region ($n = 2419$, 71%) or a single nation ($n = 2106$, 62%) than they were to be authored by people across multiple geographic regions ($n = 987$, 29%) or multiple nations ($n = 1300$, 38%). Papers with authors located in multiple geographic regions or nations were more likely to be sent for review (two-way contingency table: $X^2(1) = 23.93$, $p < 0.001$, $X^2(1) = 24.36$, $p < 0.001$, respectively), but, of those that were sent for review, there was no significant difference in the proportion of papers rejected between papers written by authors from single versus multiple geographic regions or nations (two-way contingency table: $X^2(1) = 0.14$, $p = 0.705$, $X^2(1) = 1.90$, $p = 0.168$, respectively).

Table 3 Logistic regression testing whether manuscript rejection frequencies varies with the geographic location of handling editors

Region	Estimate (SE)	OR	95% CI	*P*
Intercept	−0.907 (0.153)			
Africa	0.195 (0.577)	1.215	[0.300, 3.764]	0.735
Asia	−0.127 (0.348)	0.881	[0.446, 1.742]	0.716
Latin America	−0.508 (1.183)	0.602	[0.059, 6.123]	0.668
North America	0.076 (0.176)	1.079	[0.763, 1.524]	0.668
Oceania	0.220 (0.255)	1.246	[0.755, 2.054]	0.389
United Kingdom	0.300 (0.244)	1.349	[0.837, 2.174]	0.219

No significant results were found

OR odds ratio, *CI* confidence interval (exponentiated); Region equals location of handling editor. $N = 1771$ cases. The model is paper_rejected(yes/no) ~ handling_editor_geog + (1| handling_editor), where handling_editor equals the identification number of the handling editor as a random effect. Odds ratios are relative to Europe and the reference level for paper_rejected is *yes*

Tests of whether language and socioeconomics predict editorial and peer review outcomes

The proportion of papers sent for review, their mean review scores, and final outcome of the peer review process all varied among the geographic regions of their first authors. In general, papers from North America, Oceania, and the United Kingdom performed better and papers from Africa, Asia, and Latin America performed worse. Here we tested whether variation in language and/or socioeconomics of the authors' home countries can contribute to explaining the geographic variation in manuscript outcomes. Specifically, we tested whether (a) English language (because *Functional Ecology* is published in English) or (b) socioeconomic status (which have higher per capita resource allocation to research and/or education) are adequate to explain much or all of the observed geographic variation in manuscript outcomes. To test whether language and socioeconomic status explain geographic variation in manuscript outcomes, we first built models including an intercept plus first author geography; in each case, author geography was statistically significant. We then tested whether (a) the addition of language and/or socioeconomics improved the fit of the model, and (b) whether the effect of author geography became non-significant.

The model that provided the most significant improvement over the baseline model (intercept plus author geography only) in predicting whether manuscripts were sent for review includes both language (English versus not English) and Human Development Index terms (Table 5a; Model 4). First author geography remained a significant term in this best fit model indicating that, although language and HDI are predictive of whether a manuscript is sent for review, these terms are not adequate to explain all of the observed variation among geographic regions.

For mean review scores, the baseline model with first author geography is statistically significant ($X^2(6) = 29.75$, $p < 0.001$). Adding language does not significantly improve this model (Likelihood ratio $X^2(1)$ 1.69, $p = 0.193$). Adding HDI to the baseline model, however, improves the model (Likelihood ratio $X^2(1) = 10.17$, $p = 0.001$). First author geography is not statistically significant (Likelihood ratio $X^2(6) = 9.07$, $p = 0.170$) when HDI is added to the baseline model, indicating that variation among countries in HDI can

Table 4 Logistic regression testing whether manuscript rejection frequency varies according to first author geographic location

Region	Estimate (SE)	OR	95% CI	*P*
Intercept	−0.923 (0.090)			
Africa	−0.176 (0.524)	0.839	[0.270, 2.202]	0.738
Asia	−0.346 (0.214)	0.707	[0.460, 1.065]	0.105
Latin America	−0.496 (0.301)	0.609	[0.326, 1.071]	0.100
North America	0.282 (0.127)	1.326	[1.034, 1.702]	0.026
Oceania	0.416 (0.184)	1.515	[1.054, 2.167]	0.024
United Kingdom	0.568 (0.199)	1.765	[1.192, 2.601]	0.004

Papers with first authors located in North America, Oceania, and the United Kingdom were more likely to have papers not rejected

OR odds ratio, *CI* confidence interval (exponentiated); Region equals location of first author. $N = 1706$ due to missing data for first author geographies on 65 cases. The model is paper_rejected(yes/no) $\sim$ first_auth_geog. Odds ratios are relative to Europe and the reference level for paper_rejected is *yes*

explain the variation among mean review scores among geographic regions (Table 5b; Model 3).

This same result holds true for final decisions of papers sent for review. The baseline model with first author geography alone is statistically significant ($X^2(6) = 26.5$, $p < 0.001$), and adding language does not significantly improve the fit of the model (Likelihood ratio $X^2(1) = 2.64$, $p = 0.10$). However, adding HDI to the baseline model results in a statistically significant improvement (Likelihood ratio $X^2(1) = 7.67$, $p = 0.006$). When HDI is added to the baseline model, first author geography becomes non-significant (Likelihood ratio $X^2(6) = 11.0$, $p = 0.09$), indicating again that HDI is adequate to explain the variation in final decision among first author geographic regions (Table 5c; Model 3). However, when we add the mean review score as a covariate, review scores alone, without English or the Human Development Index, provided the best fit for explaining final decisions (Table 5d; Model 1). And, notably, neither first author geography nor HDI contribute to the variation in final decisions (rejection versus revision invited) when mean review scores are included (Likelihood ratio $X^2(6) = 10.6$, $p = 0.101$ for first author geography for Table 5d; Model 1)(though, as noted above, HDI predicts peer review scores).

Discussion

In this study, we examined (a) the degree to which author and handling editor geographic locations predict outcomes at three separate steps of the scientific peer review process, whether papers are sent for review, mean review scores of reviewed papers, and final decisions, and (b) the extent to which language and socioeconomics can explain the observed geographic variation in outcomes.

Scientists in the United States and the United Kingdom are overrepresented among handling editors (relative to the distribution of papers received from the various geographic regions of the world). However, there was little variation among editors from different geographic regions in the frequency of different decisions they made on papers (whether they sent papers for review or invited revision after review). In contrast, papers submitted

Table 5 Four sets of models testing three dependent variables: whether papers were sent for review, binned mean review scores, and whether papers were paper rejected

	Model	AUC	AIC	X^2	*Df*	*P*
5a: sent for review, n = 3425 cases						
Model 1	First_author_geog	0.6023	4628.4	133.58	6	0.000
Model 2	First_author_geo + English	0.6023	4619.2	11.24	1	0.001
Model 3	First_author_geo + HDI_100	0.6149	4610.5	19.90	1	0.000
Model 4	**First_author_geog + English + HDI_100**	**0.6152**	**4605.7**	**26.70**	**2**	**0.000**
5b: binned mean review scores, n = 1770 cases						
Model 1	First_author_geog	–	3561.34	29.75	6	0.000
Model 2	First_author_geo + English	–	3561.65	1.69	1	0.193
Model 3	**First_author_geo + HDI_100**	–	**3553.17**	**10.17**	**1**	**0.001**
Model 4	First_author_geog + English + HDI_100	–	3553.68	11.65	2	0.003
5c: paper rejected, n = 1770 cases						
Model 1	First_auth_geog	0.573	2100.4	26.46	6	0.000
Model 2	First_auth_geog + English	0.576	2099.8	2.64	1	0.104
Model 3	**First_auth_geog + HDI_100**	**0.583**	**2094.7**	**7.67**	**1**	**0.006**
Model 4	First_auth_geog + English + HDI_100	0.585	2093.1	11.36	2	0.003
5d: paper rejected, n = 1770 cases						
Model 1	**First_auth_geog + mean_review_score**	**0.877**	**1368.4**	**733.99**	**1**	**0.000**
Model 2	First_auth_geog + English + mean_review_score	0.877	1370.0	734.46	2	0.000
Model 3	First_auth_geog + HDI_100 + mean_review_score	0.877	1369.8	734.62	2	0.000
Model 4	First_auth_geog + English + HDI_100 + mean_review_score	0.877	1371.3	735.15	3	0.000

All models include the first author geography as the baseline and then iteratively add English as a language variable and the Human Development Index (HDI) as a socioeconomic variable. The fourth set (5d) adds the mean review score as a covariate

For *Sent for Review*, the best fit model is Model 4, which includes both additional terms. For *Binned Mean Review Scores*, the best fit model is Model 3, which includes the addition of the Human Development Index as a term. For *Paper Rejected*, the best fit model is Model 3 under 5c, which includes language and the socioeconomic variable as added terms (however, English is marginally significant), and Model 1 under 5d, the latter which includes only the mean review score as a covariate or. The Chi square statistic for each set's Model 1 s report the difference between that model and the intercept only, but the Chi square differences for the remaining models show the overall Chi square improvement between the respective model and each Model 1

Bold represents the best fitting models per iteration by authors from different regions of the world fared quite differently; at the extreme, papers from Asia were less likely to be sent for peer review, and received poorer review scores when reviewed, whereas papers from the United States and/or the United Kingdom were most likely to be reviewed and received stronger review scores when reviewed. Most of this variation among geographic regions was predicted by language or socioeconomics; papers submitted by authors from non-English speaking countries, and those of low socioeconomic status, were more likely to be rejected by handling editors without peer review, whereas low socioeconomic status (but not language) predicted low review scores. Notably, although socioeconomics and language predicted whether papers were reviewed, and socioeconomics predicted review scores, once papers were sent for review we found that variance in review scores alone, and not geography, socioeconomics, or language, predicted the final fate of reviewed manuscripts.

Studies on reviewer characteristics have found mixed results. Both US and non-US reviewers rank US submissions more highly than they rank submissions from authors in other countries (Link 1998), but review ratings vary across all reviewers by region; that is, depending on the reviewer's nationality, reviews may be either overly harsh or overly lenient on submissions from authors from the reviewers' respective nations (Opthof et al. 2002) or whether the authors are competing for the same publication outlet (Blackburn and Hakel 2006). However, reviewer attributes may not influence review outcomes (Bornmann and Daniel 2007). This is in line with our findings, which do not show that the geographic region of the reviewer have an observable and consistent effect on the review scores received by manuscripts, averaged across submissions (we cannot tell if any specific individual paper experiences bias). And, despite having a geographically unrepresentative editorial board, with respect to the geographic distribution of submitting authors, we found no evidence that this affects editorial decisions; that is, the geographic location of the handling editors did not predict the decisions made by editors (whether papers were sent for review or declined after review) (cf. Hsiehchen and Espinoza 2016).

Since English is currently the dominant language of scientific discourse, papers written by non-native English speakers may face extra hurdles in passing successfully through the peer review process (Clavero 2011), and this may be due to issues related to the mechanics of writing rather than to the scientific merit of the work (Flowerdew 2001). Our findings do indicate that language is a predictive factor of whether a paper is sent for review, but not how a paper is scored by reviewers or whether it is accepted after review. Though we have no independent evaluation of the quality of manuscripts that are declined without review, we speculate that the large language effect observed at this stage reflects editors declining papers that are poorly written rather than of low research quality. The scientific community at large may wish to continue investigating ways to assist non-native English speaking scientists through the initial hurdle of getting a manuscript reviewed [e.g., see Cronin (2009)]. Although the peer review process has been shown to increase the quality of writing of submitted manuscripts (Goodman et al. 1994), the peer review process is already burdened and subject to various stresses related to its validity (Naik 2017) and to the recruitment of qualified reviewers (Fox et al. 2017b), and copywriting should not be its primary function.

Although we found that outcomes varied based on the first author's geographical region, and that language explained some of the author geographic variation in whether papers were sent for review, we also found that the socioeconomic status of the first author's affiliated country, as measured by the Human Development Index (HDI), explained much of the observed geographic variation in whether papers were sent for review, review scores, and final decisions. Few scholarly communication studies have incorporated the HDI in their analyses, but when they have, the HDI has been shown to explain the proportion of author submissions from nations with low and high indexes (Keiser 2004) and research productivity in the health literature (Uthman et al. 2014, although GDP was a stronger predictor, and HDI was not significant in their multivariate analysis).

Although non-native English speakers may face barriers to scientific publication (Cronin 2009; Man et al. 2004), the explanatory influence of HDI in our models indicates that the advantage shared by authors from western nations or from English-speaking regions (Burgman et al. 2015; Clavero 2011; Harris et al. 2015; Link 1998; Primack and Marrs 2008; Primack et al. 2009; Ross et al. 2006; Zhang 2012) are more complicated than language or national/geographic (as reduced to) factors alone. Our results indicate that issues related to the well-being of nations and to the cost for scientists who are located in nations with lower life expectancies at birth, schooling, and standard of living may have an impact on the ability to conduct and publish scientific studies. Although international scientific collaboration is growing (Witze 2016), developing nations face significant barriers that include competition with developed nations, level of infrastructure, quality and quantity of education, and investment in research and development (Gálvez et al. 2000; Waheed 2012). Future research on peer review specifically, and scientific activity more broadly, would benefit from exploring how socioeconomic issues put stress on the scientists who act within this system.

Success at publishing one's research projects is essential to obtaining grants, employment, and other rewards in an academic environment. It is thus essential to understand the degree to which scientific publishing is impacted by editor or reviewer discrimination against papers written by specific types of authors, such as women or researchers from developing countries. Our data cannot inform us whether there are biases in editorial decisions made pre-review—we observe substantial variation based on author location in editorial rejection pre-review, but we have no independent metric quantifying manuscript quality that would allow us to distinguish the influence of variation in quality from that of editorial bias. We also cannot test to what degree peer review scores are objective and unbiased assessments of manuscript quality. However, we find it encouraging that, once a paper has been reviewed, the variable that best predicts the final editor decision is the review score, and neither author language, HDI or geographic location contribute to explaining the final decision. This result argues in favor of editorial decisions (at least at this final stage) being little influenced by aspects of the authors or manuscript beyond an assessment of manuscript quality as signaled by review scores.

Data accessibility

Data for this project are deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.37312 (Fox et al. 2015). The deposited data allow recreation of most results in the paper. However, because the data set contains information on human subjects, the data available at Dryad are anonymized and lack variables that allow parts of the data set to be de-anonymized. Please see the metadata accompanying the Dryad submission for additional details.

Acknowledgements We thank the British Ecological Society (BES), owners of the journal *Functional Ecology*, for permitting us to use their peer review database for this project. Brandi Frisby provided comments on an earlier draft of this paper. This work was reviewed and approved by the Internal Review Board at the University of Kentucky, IRB 14-0570-P4S.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01.

Blackburn, J. L., & Hakel, M. D. (2006). An examination of sources of peer-review bias. *Psychological Science, 17*(5), 378–385.

Bornmann, L., & Daniel, H.-D. (2007). Gatekeepers of science—Effects of external reviewers' attributes on the assessments of fellowship applications. *Journal of Informetrics, 1,* 83–91. doi:10.1016/j.joi.2006.09.005.

Burgman, M., Jarrad, F., & Main, E. (2015). Decreasing geographic bias in conservation biology. *Conservation Biology, 29*(5), 1255–1256. doi:10.1111/cobi.12589.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65,* 23–35. doi:10.1007/s00265-010-1029-6.

Campos-Arceiz, A., Primack, R. B., & Koh, L. P. (2015). Reviewer recommendations and editors' decisions for a conservation journal: Is it just a crapshoot? And do Chinese authors get a fair shot? *Biological Conservation, 186,* 22–27. doi:10.1016/j.biocon.2015.02.025.

CIA. (2016). *The world factbook*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html.

Clavero, M. (2011). Language bias in ecological studies. *Frontiers in Ecology and the Environment, 9,* 93–94. doi:10.1890/11.wb.001.

Cronin, B. (2009). Vernacular and vehicular language. *Journal of the American Society for Information Science and Technology, 60*(3), 433. doi:10.1002/asi.21010.

Daniel, H.-D. (1993). Fairness in manuscript evaluation (W. E. Russey, Trans.). In *Guardians of science: fairness and reliability of peer review* (pp. 29–46). Weinheim: VCH Verlagsgesellschaft.

Ernst, E., & Kienbacher, T. (1991). Chauvinism. *Nature, 352*(6336), 560.

Fayaz-Bakhsh, A., & Mousavi, A. (2015). Science growth and human development index in Iran. *Journal of Research in Medical Sciences, 20,* 1218.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Los Angeles: Sage.

Flowerdew, J. (2001). Attitudes of journal editors to nonnative speaker contributions. *TESOL Quarterly, 35*(1), 121–150.

Fox, C. W., Albert, A. Y. K., & Vines, T. H. (2017a). Recruitment of reviewers is becoming harder at some journals: A test of the influence of reviewer fatigue at six journals in ecology and evolution. *Research Integrity and Peer Review*. doi:10.1186/s41073-017-0027-x.

Fox, C. W., & Burns, C. S. (2015). The relationship between manuscript title structure and success: Editorial decisions and citation performance for an ecological journal. *Ecology and Evolution, 5,* 1970–1980. doi:10.1002/ece3.1480.

Fox, C. W., Burns, C. S., & Meyer, J. A. (2015). Data from: Editor and reviewer gender influence the peer review process but not peer review outcomes at an ecology journal. *Dryad Digital Repository*. doi:10.5061/dryad.5090r.

Fox, C. W., Burns, C. S., & Meyer, J. A. (2016a). Editor and reviewer gender influence the peer review process but not peer review outcomes at an ecology journal. *Functional Ecology, 30,* 140–153. doi:10.1111/1365-2435.12529.

Fox, C. W., Burns, C. S., Muncy, A. D., & Meyer, J. A. (2016b). Gender differences in patterns of authorship do not affect peer review outcomes at an ecology journal. *Functional Ecology, 30,* 126–139. doi:10.1111/1365-2435.12587.

Fox, C. W., Burns, C. S., Muncy, A. D., & Meyer, J. A. (2017b). Author-suggested reviewers: Gender differences and influences on the peer review process at an ecology journal. *Functional Ecology*. doi:10.1111/1365-2435.12665.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks: Sage.

Gálvez, A., Maqueda, M., Martínez-Bueno, M., & Valdivia, E. (2000). Scientific publication trends and the developing world: What can the volume of scientific articles tell us about scientific progress in various regions? *American Scientist, 88*(6), 526–533.

Goodman, S. N., Berlin, J., Fletcher, S. W., & Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Annals of Internal Medicine, 121,* 11–21. doi:10.7326/0003-4819-121-1-199407010-00003.

Harris, M., Macinko, J., Jimenez, G., Mahfoud, M., & Anderson, C. (2015). Does a research article's country of origin affect perception of its quality and relevance? A national trial of US public health researchers. *British Medical Journal Open*. doi:10.1136/bmjopen-2015-008993.

Hilbe, J. M. (2009). *Logistic regression models*. Boca Raton: CRC Press.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.
Hsiehchen, D., & Espinoza, M. (2016). Detecting editorial bias in medical publishing. *Scientometrics, 106,* 453–456. doi:10.1007/s11192-015-1753-9.
Harrell, Jr, F. E. (2016). *Hmisc: Harrell miscellaneous*. R package version 4.0-2. https://CRAN.R-project.org/package=Hmisc.
Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A., & Rennie, D. (1998). Does masking author identity improve peer review quality? A randomized controlled trial. *JAMA, 280*(3), 240–242. doi:10.1001/jama.280.3.240.
Keiser, J. (2004). Representation of authors and editors from countries with difference human development indexes in the leading literature on tropical medicine: Survey of current evidence. *BMJ*. doi:10.1136/bmj.38069.518137.F6.
King, D. A. (2004). The scientific impact of nations. *Nature, 430,* 311–316. doi:10.1038/430311a.
King, J. E. (2008). Binary logistic regression. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 358–384). Chicago: Sage.
Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the Association for Information Science and Technology, 64*(1), 2–17. doi:10.1002/asi.22784.
Lesnoff, M., and Lancelot, R. (2012). *aod: Analysis of overdispersed data*. R package version 1.3. http://cran.r-project.org/package=aod.
Link, A. M. (1998). US and non-US submissions. *JAMA, 280,* 246–247. doi:10.1001/jama.280.3.246.
Man, J. P., Weinkauf, J. G., Tsang, M., & Sin, D. D. (2004). Why do some countries publish more than others? An international comparison of research funding, English proficiency and publication output in highly ranked general medical journals. *European Journal of Epidemiology, 19*(8), 811–817.
Meyer, D., Zeileis, A., and Hornick, K. (2016). *Vcd: Visualizing categorical data*. R package version 1.4-3.
Naik, G. (2017). Peer-review activists push psychology journals toward open data. *Nature*. doi:10.1038/nature.2017.21549.
Opthof, T., Coronel, R., & Janse, M. J. (2002). The significance of the peer review process against the background of bias: priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovascular Research, 56*(3), 339–346. doi:10.1016/S0008-6363(02)00712-5.
Osborne, J. W. (2008). Bringing balance and technical accuracy to reporting odds ratios and the results of logistic regression analysis. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 385–389). Chicago: Sage.
Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). South Melbourne: Thomson Learning.
Primack, R. B., Ellwood, E., Miller-Rushing, A. J., Marrs, R., & Mulligan, A. (2009). Do gender, nationality, or academic age affect review decisions? An analysis of submissions to the journal. *Biological Conservation, 142,* 2415–2418. doi:10.1016/j.biocon.2009.06.021.
Primack, R. B., & Marrs, R. (2008). Bias in the review process. *Biological Conservation, 141,* 2919–2920. doi:10.1016/j.biocon.2008.09.016.
R Core Team. (2016). R: A language and environment for statistical computing. (Version 3.3.1). *R foundation for statistical computing*. Available at http://www.R-project.org/.
Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. doi:10.1186/1471-2105-12-77.
Ross, J. S., Gross, C. P., Desai, M. M., Hong, Y., et al. (2006). Effect of blinded peer review on abstract acceptance. *JAMA, 295*(14), 1675–1680. doi:10.1001/jama.295.14.1675.
Treganza, T. (2002). Gender bias in the refereeing process? *Trends in Ecology and Evolution, 17*(8), 349–350. doi:10.1016/S0169-5347(02)02545-4.
United Nations Development Programme. (2015). *International human development indicators*. Retrieved from http://hdr.undp.org/en/countries.
Uthman, O. A., Wiysong, C. S., Ota, M. O., Nicol, M., Hussey, G. D., Ndumbe, P. M., et al. (2014). Increasing the value of health research in the WHO African Region beyond 2015—reflecting on the past, celebrating the present and building the future: a bibliometric analysis. *British Medical Journal Open, 5,* e006340. doi:10.1136/bmjopen-2014-006340.
Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.
Waheed, A. (2012). Why developing countries are lesser innovators. *International Journal of Social, Behavioral, Educational, Business and Industrial Engineering, 6*(7), 1686–1691.
Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software, 21*, 1–20. http://www.jstatsoft.org/v21/i12/.
Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software, 40*, 1–29. http://www.jstatsoft.org/v40/i01/.

Wickham, H., and Francois, R. (2016). *dplyr: A grammar of data manipulation*. R package version 0.5.0. https://CRAN.R-project.org/package=dplyr.

Witze, A. (2016). Research gets increasingly international. *Nature*. doi:10.1038/nature.2016.19198.

Zambrano, E. (2014). An axiomatization of the human development index. *Social Choice and Welfare, 42*(4), 853–872. doi:10.1007/s00355-013-0756-9.

Zhang, X. (2012). Effect of reviewer's origin on peer review: China vs. non-China. *Learned Publishing, 25*(4), 265–270. doi:10.1087/20120405.