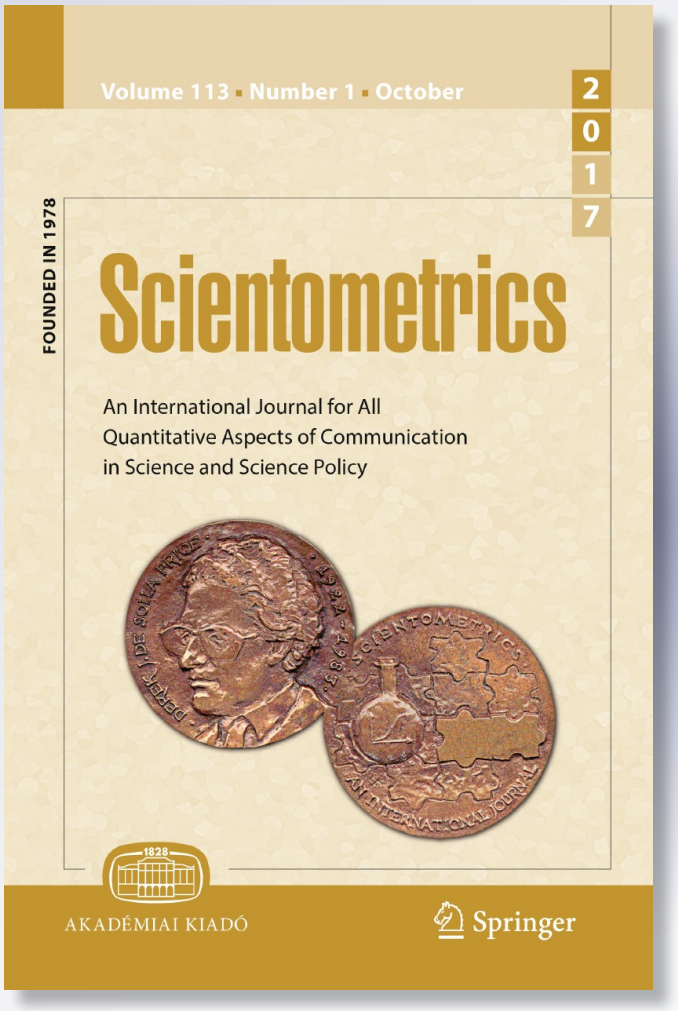
Difficulty of recruiting reviewers predicts review scores and editorial decisions at six journals of ecology and evolution

Charles W. Fox

Scientometrics
An International Journal for all Quantitative Aspects of the Science of Science, Communication in Science and Science Policy

ISSN 0138-9130
Volume 113
Number 1

Scientometrics (2017) 113:465-477
DOI 10.1007/s11192-017-2489-5



Springer

Scientometrics (2017) 113:465–477
DOI 10.1007/s11192-017-2489-5

Difficulty of recruiting reviewers predicts review scores and editorial decisions at six journals of ecology and evolution

Charles W. Fox[1]

Received: 15 May 2017 / Published online: 14 August 2017

Abstract Journal peer review relies on the willingness of researchers to volunteer their time to review manuscripts. However, editors often have difficulty recruiting reviewers, and this difficulty can vary quite substantially among manuscripts. This study examines whether the difficulty recruiting reviewers influences outcomes of the peer review process at six journals of ecology and evolution. The difficulty editors had recruiting reviewers varied substantially among papers, with editors successfully recruiting the first two people invited just 22% of the time, and being declined by two or more invitees for more than half (56%) of reviewed papers. Papers for which editors had more difficulty recruiting reviewers were more likely to be declined at all six journals, with an increase in the odds of acceptance ranging from a low of 3.5 ± 1.2% to a high of 17.3 ± 2.0% for each 10% increase in the proportion of reviewers agreeing to review. Papers for which editors had more difficulty recruiting reviewers were also reviewed less positively at all six journals, and this influence on review scores explained most but not all of the influence of recruitment difficulty on outcomes. Reviewers invited close together in sequence (without many declined invitations between them) were more consistent in the scores they submit than were reviewers invited more greatly separated in sequence, suggesting that editors recruit different kinds of reviewers early versus late in the reviewer invitation sequence. However, the scores submitted by later-recruited reviewers were not less predictive of the editor's decision than were scores of early-recruited reviewers. The influence of reviewer recruitment difficulty on decisions, although of small effect, should be considered among the diversity of variables that influence outcomes of the editorial and peer review process at academic journals.

Keywords Bias in peer review · Inter-rater reliability · Peer review · Referees

✉ Charles W. Fox
cfox@uky.edu

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

Introduction

The success of the peer review process relies on the willingness of researchers to review manuscripts, usually as volunteer service (i.e., unpaid). However, editors of many journals argue that they are finding reviewers increasingly difficult to recruit (Baveye and Trevors 2011; Goldman 2015; discussion in Albert et al. 2016). Most such reports are anecdotal, but some data supports this contention; for example, the proportion of review invitations that led to a submitted review declined substantially at four ecology journals, from 56% in 2003 to just 37% in 2015 (averaged across those four journals; Fox et al. 2017b), though reviewer recruitment success declined less or even stayed fairly constant at some other journals (Vines et al. 2010; Albert et al. 2016; Fox et al. 2017b). Within journals, the difficulty recruiting reviewers varies substantially across papers; for some papers editors successfully recruit the first two people they invite, but for other papers editors need to contact many more than two potential reviewers, occasionally even more than 10 or even 15 (Stamm et al. 2007; Lajtha and Baveye 2010; Baveye and Trevors 2011; Fox and Burns 2015), just to obtain two reviews. The reasons why prospective reviewers decline to review are diverse (Breuning et al. 2015; Willis 2016), but in some instances include an expectation that the paper will be low quality or uninteresting (Tite and Schroter 2007; Zaharie and Osoian 2016), likely signaled by aspects of the manuscript made available in the review invitation (paper title, author identities and/or abstract content).

Regardless of the actual reasons why invitees decline to review, difficulty recruiting reviewers could influence outcomes of the peer review process. In particular, editors may be more likely to decline papers for which they had greater difficulty recruiting reviews, possibly because they consciously assume that difficulty recruiting reviewers signals poor manuscript quality, or because such difficulties frustrate or annoy the editor leading to a subconscious bias against the manuscript. However, studies that examine how reviewer recruitment influences peer review outcomes are mixed. The difficulty that editors had recruiting reviewers was unrelated to manuscript acceptance rates at the journal *Radiology* (Kallmes et al. 2017). In contrast, the probability that a manuscript was rejected increased with the number of prospective reviewers that declined the invitation to review at an ecology journal, *Functional Ecology* (Fox and Burns 2015) and at the biomedical journal *Head & Face Medicine* (Stamm et al. 2007). However, those studies did not examine whether this effect was mediated by the scores given to papers by reviewers and so did not explore whether the effect was mediated by manuscript quality.

The purpose of this study is to: (1) examine whether difficulty recruiting reviewers predicts the fate of submitted manuscripts at six journals in ecology and evolution (24,325 reviewed manuscripts), (2) test whether this influence is mediated by reviewer assessments of manuscript quality or importance (peer review scores), or whether editors are more likely to decline manuscripts for which they have difficulty recruiting reviewers after accounting for the influence of review scores, and (3) test whether difficulty recruiting reviewers predicts eventual citation numbers for papers in one of the focal journals.

Methods

The dataset

All six journals examined here use *ScholarOne Manuscripts* to manage submissions and peer review. The dataset includes peer review data for all manuscripts submitted between 1

January 2003 and 30 June 2015 for *Functional Ecology*, *J Animal Ecology*, *J Applied Ecology* and *J Ecology*, between 13 August 2009 and 30 June 2015 for *Methods in Ecology and Evolution* (this journal received its first ever submission on 13 August 2009), and between 20 May 2007 and 31 December 2015 for *Evolution* (*Evolution* began using *ScholarOne Manuscripts* to manage submissions in May 2007). The dataset includes only standard research papers (called a "Research Article" at *Methods in Ecol Evol*, an "Original Article" at *Evolution*, and a "Standard Paper" at the other journals); it excludes review papers, commentaries, perspectives, editorials, brief communications and other types of papers not considered typical full-length research manuscripts. I consider only the first submission of a paper; revisions and/or resubmissions were excluded, even if sent for a second round of peer review, since these papers commonly go to the same reviewers who reviewed the first submission. The dataset includes 113,687 reviewer invitations that yielded 51,984 reviews for 24,325 standard research papers that were sent for review.

For one of the six journals, *Functional Ecology*, citation data from Web of Science was extracted in December 2014 for papers that were published in the journal between 2005 (the first year for which I have peer review data for most published papers) and 2013. Citation data were merged with the peer review data. Citation counts are an imperfect measure of manuscript impact because they do not capture many uses of manuscripts (e.g., by practitioners; Stremersch et al. 2007) and can covary with features of a manuscript unrelated to it's significance or quality (Mingers and Xu 2010). However, citations do covary with article downloads (Perneger 2004; Moed and Halevi 2016; Vaughan et al. 2017) and other metrics of influence (Mingers and Xu 2010; Thelwall et al. 2013).

Variables in the dataset

For each manuscript that fits the criteria defined above, I have information on whether the paper was assigned to an associate editor, whether it was sent for peer review, when each selected reviewer was invited, the fate of all review invitations (including both whether they responded to the email and whether they agreed to review), whether and when they submitted a review, the review score, and the final decision on the manuscript. Prospective reviewers are provided the name of the manuscript, identity of the authors, and abstract of the paper in the invitation email.

The journals considered here vary in the reviewer scoring categories they use. With the assistance of the journal editorial teams, scoring systems for each journal were converted to a point scale (with high scores being better). I then averaged scores *within* manuscripts and standardized these scores to their overall mean and standard deviation (i.e., to mean of 0 and standard deviation of 1) within journals and years (because scoring criteria vary among journals and years), such that all review scores are on the same scale. Of course, review scores alone do not capture the entirety of a reviewer's assessment of a manuscript—reviewers also provide (often very thorough) written comments that contribute important information considered by editors when making their decisions (Bornmann et al. 2010). Review scores also lump together reviewer opinions of manuscript quality, novelty, and relevance to the journal, but they do predict editor decisions on manuscripts.

The decisions made by editors also vary among journals. Though all journals have the standard "reject" versus "revision invited" outcomes, some make frequent use of "reject with resubmission invited" (commonly called "reject without prejudice") as a more strict form of revision invitation, rather than, e.g., requesting major revision. Since resubmitted papers generally were re-reviewed and a significant proportion of these were eventually declined, papers that were rejected with revision invited are counted as having been

rejected for all analyses of editorial decisions. Alternative analyses including these papers as having been invited for revision do not change any conclusions.

Statistical analyses

Many of the variables examined here are binary; e.g., the reviewer agreed to review [yes/no], or a revision was invited for a manuscript [yes/no]. These variables were thus analyzed using logistic regression (SAS Proc logistic with link = logit). For non-binary variables, such as the time it took reviewers to respond to the review invitation, the time it took for them to submit their review, and their review score, I used general linear models (SAS Proc GLM). All analyses were of the form $DependentVariable = Year + IndependentVariables + TwoWayInteractions$. Year is included because reviewer response rates (Fox et al. 2017) and manuscript outcomes (Fox et al. 2016) have varied substantially over time for most of these journals. For all analyses each manuscript provided a single data point (with one exception; for the analysis of the relationship between time and review score, each review was treated as a single data point). Further details are described as necessary as results are presented.

To test whether the number of citations a paper receives is predicted by the difficulty editors had recruiting reviews (for *Functional Ecology*), I used linear models with $\log_{10}(Citations + 1) = PublicationYear + IndependentVariables$.

Results

Over the period for which data on reviewer invitations are available, these journals have invited an average of 4.73 ± (SEM) 0.16 prospective reviewers per paper to obtain on average 2.26 ± 0.04 agreed reviewers and 2.14 ± 0.03 reviews per paper (means were averaged first across papers, then across years within journals, then across journals; 94.7% of agreed reviewers submitted a review). However, the number of review invitations sent per paper has varied among journals and over time (Fig. 1a). This variation in invitations is much greater than that for the number of reviewers that agreed to review, which averages slightly over 2 per paper and varied only slightly among journals (from a low average of 2.15 to a high of 2.28; $F_{5,24297} = 72.8$, $P < 0.001$) and over time ($F_{1,24297} = 4.82$, $P = 0.02$). More importantly, the number of invitations sent per paper varied because the proportion of invited reviewers that agree to review varies among journals and has declined over time for most journals (Fig. 1b). I thus include main effects of both *Journal* and *Year* in most of the analyses presented below. [A more thorough examination of reviewer response rates is presented in Fox et al. 2017a (for all six journals) and Fox et al. 2016a, b, 2017b (for *Functional Ecology*)].

Within journals, the number of invitees that decline to review varied substantially among papers. Only 21.7% of papers (averaged across journals and years) had all invitees agree to review (i.e., no one declined the review invitation), with a low of just 14.4% for *J Appl Ecol* to a high of 26.0% for *Evolution* (Fig. 2). In contrast, 56% of papers had two or more invitees decline the review invitation, 27% of papers had four or more invitees decline, and 12% of papers had six or more invitees decline, though these percentages were all higher in later years than in early years.

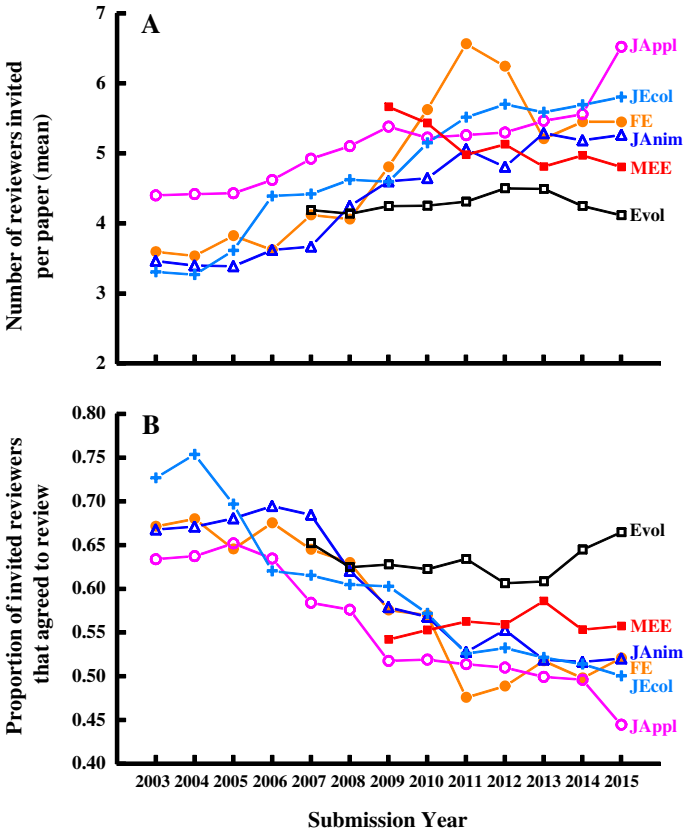


Fig. 1 The **a** mean number of prospective reviewers invited per manuscript, and **b** proportion of invited reviewers that agreed to review for six journals in ecology and evolution. The number of reviewers invited has varied among journals and over time (analysis of variance; model: *NumberOfReviewers* = *Year* + *Journal* + *Interaction*; *Year*: $F_{1,24313} = 240.9$, $P < 0.001$; *Journal*: $F_{5,24313} = 50.8$, $P < 0.001$; *Interaction*: $F_{5,24313} = 50.9$, $P < 0.001$), in large part because the proportion of invited reviewers that agree to review varies among journals and has declined over time (logistic regression; model: *ReviewersAgreed/ReviewersInvited* = *Year* + *Journal* + *Interaction*; *Year*: $\chi^2_1 = 372.4$, $P < 0.001$; *Journal*: $\chi^2_5 = 26.5$, $P < 0.001$; *Interaction*: $\chi^2_5 = 26.6$, $P < 0.001$)

Does reviewer recruitment predict peer review and editorial decisions

The difficulty that editors had obtaining reviews predicted the final fate of the manuscript for these six journals; papers for which the proportion of invitees that agreed to review was higher were more likely to be invited for revision (Fig. 3a; Logistic regression; model: *RevisionInvited* = *Year* + *Journal* + *ProportionAgreed* + *Journal*-x-*ProportionAgreed Interaction*; *Year*: $\chi^2_1 = 30.4$, $P < 0.001$; *Journal*: $\chi^2_5 = 98.1$, $P < 0.001$; *ProportionAgreed:* $\chi^2_1 = 251.1$, $P < 0.001$; *Interaction*: $\chi^2_5 = 87.6$, $P < 0.001$; adding a quadratic term for *ProportionAgreed* does not improve the fit of the model, $\Delta AIC = 1.9$). The magnitude of the *ProportionAgreed* effect varied among journals, but separate analyses for each journal indicate that the relationship is significantly positive for all six journals ($\chi^2_1 > 7.9$, $P < 0.03$ for each journal), with an increase in the odds of acceptance ranging from a low of $3.5 \pm 1.2\%$ for each 10% increase in the proportion of reviewers that agreed

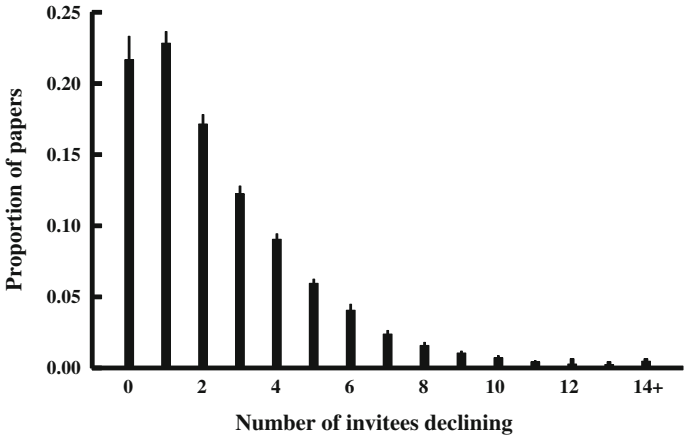


Fig. 2 The distribution of declined review invitations (mean ± SEM), of invitees that responded to the invitation (excludes non-responses), averaged across six journals in ecology and evolution over the total period for which data are available (the average number of declines per paper has increased over time; see Fig. 1b). The maximum number of declines for any paper was 25

to review (for *Functional Ecology*) to a high of 17.3 ± 2.0% for each 10% increase in the proportion of reviewers agreeing to review (for *Methods in Ecology and Evolution*).

An alternative way to visualize the magnitude of these effects is to calculate the reviewer agreement rate for papers that were rejected versus invited for revision (Fig. 3b). Though the effect size varies among journals, the pattern is consistent—papers that were declined from the journals had more difficulty obtaining reviews, with just 56.8 ± 1.6% of reviewers agreeing to review, compared to 63.0 ± 1.4% for papers with a positive outcome (averaged across journals and years). In numbers of reviewers: papers that were reviewed but eventually rejected had, on average, 4.9 ± 0.2 people invited to review, compared to 4.4 ± 0.1 for those papers that had a positive outcome.

The difficulty editors have in recruiting reviews is also predictive of review scores (Fig. 4; *ReviewScore* (standardized by journal and year to a mean of 0 and standard deviation of 1, with high scores being better) = *Year* + *Journal* + *ProportionAgreed* + *Journal*-x-*ProportionAgreed Interaction*; *ProportionAgreed:* $F_{1,23198} = 213.6$, $P < 0.001$; *Interaction*: $F_{5,23198} = 14.9$, $P < 0.001$). The relationship between reviewer recruitment (proportion of invitees that agreed to review) and review scores is significantly positive ($P < 0.05$) for all journals, with the slopes ranging between a very low of 0.12 ± 0.06 for *Functional Ecology* to a high of 0.87 ± 0.09 for *Methods in Ecology and Evolution* (Fig. 4).

Given that review scores covary (among papers) with the difficulty in recruiting reviewers, it is possible that the observed influence of reviewer recruitment difficulty on editor decisions for reviewed papers is mediated by review scores. Indeed, review score is the largest factor affecting editor decisions, but the influence of reviewer recruitment difficulty remains detectable even after accounting for review scores; when I fit a model sequentially adding *ProportionAgreed* after review scores, *ProportionAgreed* remains a significant predictor of manuscript outcome, though the variance explained is much smaller than that of review scores (*RevisionInvited* = *Year* + *Journal* + *ReviewScore* + *ProportionAgreed* + *Interactions*; *ReviewScore*: $\chi^2_1 = 3554.2$, $P < 0.001$; *ProportionAgreed:* $\chi^2_1 = 71.2$, $P < 0.001$). When separate models were fit for each journal, *ProportionAgreed*

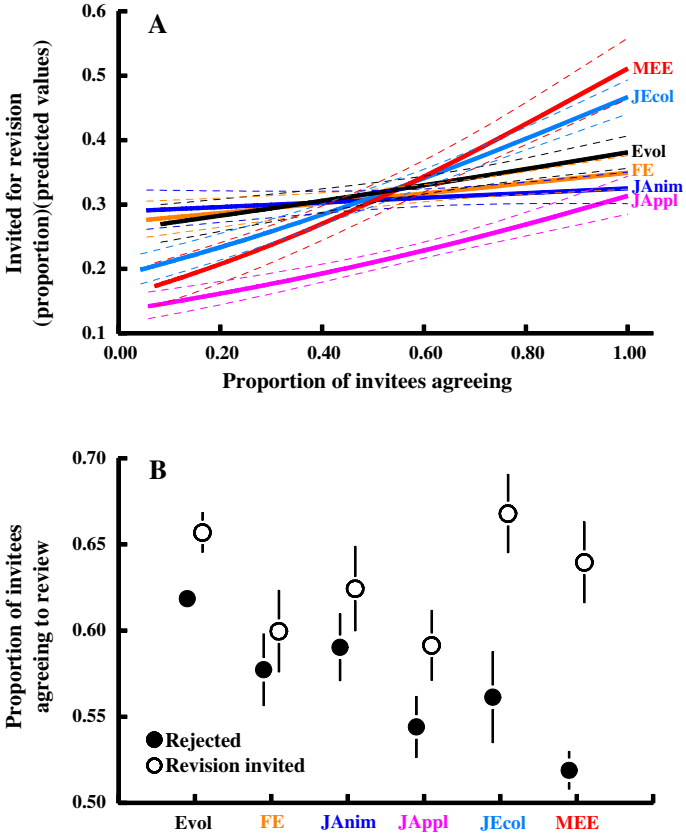


Fig. 3 **a** The relationship between manuscript outcome (revision invited) and the proportion of invitees that agree to review (predicted relationship (*solid lines*), removing the among-year variation, plus 95% confidence interval (*dashed lines*)). **b** The proportion of invited reviewers that agree to review (mean ± SEM) for papers that were rejected from the journal (*filled circle*) versus those invited for revision (*empty circle*). Averages were first calculated across papers within years, and then across years within journals

remains significant for every journal, but the variance explained is generally small ($4.1 < \chi^2_1 < 70.9$; $P < 0.05$ for each journal).

One possible explanation for why papers are more likely to be declined if reviewers are difficult to recruit, after accounting for average review scores, is that editors need to resort to inviting reviewers whose opinions they value less and are thus more likely to disregard. The correlation between the review scores submitted by early and later-recruited reviewers does decline with the number of unsuccessful review invitations sent between their recruitment (Fig. 5). However, there was no evidence that earlier-recruited reviewers are more or less positive than later-recruited reviewers ($t_{2464} = 1.28$, $P = 0.20$), and no evidence that the scores submitted by later-recruited reviewers are less predictive of the editor's decision (model: *RevisionInvited* = *Year* + *Journal* + *EarlyReviewerScore* + *LateReviewerScore*; *EarlyReviewerScore*, Odds Ratio: 3.14 [95% CI: 2.75–3.62], $\chi^2_1 = 267.9$, $P < 0.001$; *LateReviewerScore*, Odds Ratio:4.05 [95% CI: 3.51–4.70], $\chi^2_1 = 355.9$, $P < 0.001$); the trend is instead opposite of that predicted by the hypothesis

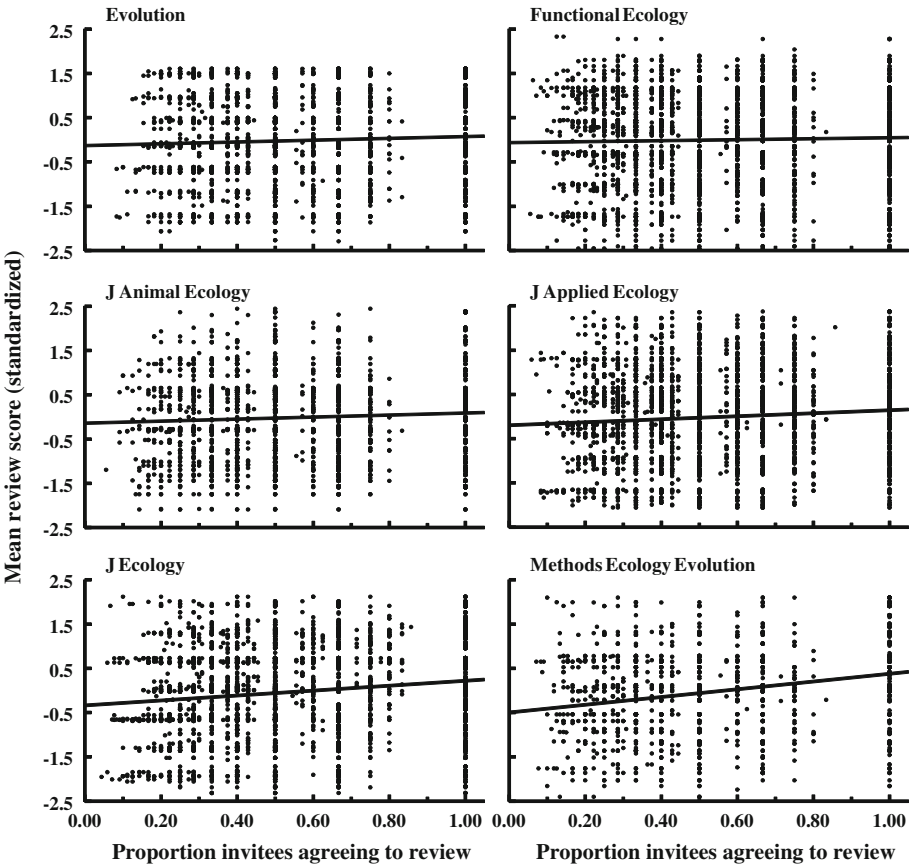


Fig. 4 The relationship between the proportion of invitees that agree to review and manuscript review scores (mean per paper; higher is better) for six journals in ecology and evolution. Review scores are standardized within journals and years (to a mean of 0 and standard deviation of 1) and, for the figure, the variance among years is removed

that editors are more likely to disregard the later-recruited reviewer. For this analysis I included reviewers recruited with five or more declines between them ($n = 2465$ papers) because some of the journals being analyzed (e.g., *Functional Ecology*) request that handling editors submit at least six prospective reviewer names for newly assigned manuscripts, such that reviewers invited with five or more declines between them would usually be from different batches of editor-selected reviewers. Changing this threshold influences the parameter estimates but there is no threshold for which the scores submitted by early recruited reviewers are more predictive of outcomes than are scores submitted by later-recruited reviewers.

Another variable that may influence editor opinions of a paper, irrespective of manuscript quality, is the delay experienced in recruiting reviewers and obtaining reviews. However, there is no evidence that the final decision on a manuscript was predicted by either the mean or maximum (per paper) number of days it took reviewers to respond to the review request (of those that agreed to review) or the mean or maximum number of days it took reviewers to submit their review (logistic regression, model:

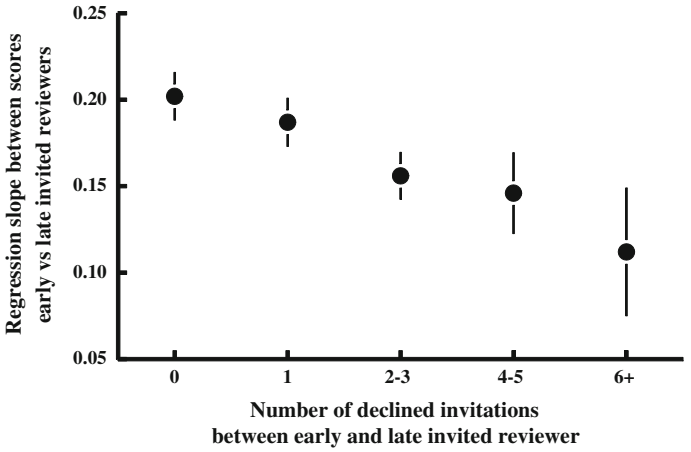


Fig. 5 The slope of the relationship (regression coefficient ± standard error) between the scores submitted by earlier versus later invited reviewers; this relationship is lower for reviewers invited further apart in sequence. Slopes are calculated from the model $ReviewScore_{LaterInvited} = Journal + ReviewScore_{EarlierInvited}$, run separately for each bin (data are binned along the x-axis for presentation purposes; the full model treating *NumberOfDeclines* as a continuous variable is: $ReviewScore_{LaterInvited} = Journal + ReviewScore_{EarlierInvited} + NumberOfDeclines + ReviewScore_{EarlierInvited}{*}NumberOfDeclines$, with *Journal* as a fixed effect and *ReviewScore* and *NumberOfDeclines* as continuous variables. $ReviewScore_{EarlierInvite}$: $F_{1,23202} = 4033.1$, $P < 0.001$; *NumberOfDeclines*: $F_{1,23202} = 20.5$, $P < 0.001$; *Interaction*: $F_{1,23202} = 1371.6$, $P < 0.001$

$RevisionInvited = Year + Journal + Time + Journal{*}Time$ *interaction*; $\chi^2_1 < 1.95$; $P > 0.16$ for each). There is also no evidence that the length of time reviewers take to respond to the initial review invitation (given they agreed to review) or the time they took to submit their review was predictive of the score they gave to papers (analysis of variance, $P > 0.24$ for each term).

Does reviewer recruitment predict citations?

For the subset of reviewed papers that were accepted and thus published by the journal *Functional Ecology* and for which I have citation data ($n = 1043$ standard papers), there is no evidence that higher reviewer recruitment success was predictive of increased citations; the slope was opposite the predicted direction and not significantly different from zero (model: $\log_{10}(Citations + 1) = PublicationYear + ProportionAgreed$; *ProportionAgreed*: slope: -0.17 ± 0.09, $F_{1,1040} = 3.41$, $P = 0.07$). When added to this model, peer review scores (standardized by year to a mean of 0 and standard deviation of (1) also failed to predict citations received by manuscripts (slope: 0.00 ± 0.02, $F_{1,1039} = 0.01$, $P = 0.93$) and their inclusion in the model had little effect on the slope and statistical non-significance of the reviewer recruitment effect (slope: -0.17 ± 0.09, $F_{1,1039} = 3.26$, $P = 0.07$).

Discussion

Most people that decline invitations to review manuscripts for academic journals do so because they are too busy (Tite and Schroter 2007; Ware and Monkman 2008; Breuning et al. 2015; Willis 2016) or because the topic does not match their expertise (Tite and

Schroter 2007; Breuning et al. 2015; Willis 2016). However, a modest proportion of prospective reviewers acknowledge that they decline to review a manuscript if they expect it to be low quality (Tite and Schroter 2007; Ware and Monkman 2008; Sense About Science 2009). Editors should thus have greater difficulty finding reviewers when manuscripts are lower quality. In this study of six journals of ecology and evolution, I find support for this; papers for which editors had more difficulty recruiting reviewers (a) were scored as lower quality and/or less significant papers by reviewers, and (b) were more likely to be declined by the journal after review, compared to papers for which reviewers were more easily recruited. Though the magnitude of these effects varied among journals, and were small for some journals, both relationships were significant for all six journals.

The most likely explanation for the negative relationships between reviewer recruitment success and both peer review scores and editor decisions is that prospective reviewers are detecting signals of manuscript quality in the invitation email. None of these journals used double blind review during the period studied here, and so prospective reviewers were provided the name of the manuscript, identity of the authors, and abstract of the paper when invited. However, after accounting for variation in manuscript quality by including peer review scores in our model, editors continue to be more likely to reject papers for which they had trouble recruiting reviewers, suggesting that variation in manuscript quality alone is inadequate to explain the observed relationships. I suggest two possible explanations for this. First, when too many invitees decline to review editors may need to resort to inviting less qualified reviewers with whose opinion they might agree less often. However, this hypothesis is not supported in the current dataset; the relative influence on editorial decisions of the scores submitted by earlier-invited versus later-invited reviewers did not differ, and the trend (albeit not significant) was for later-recruited reviewers to have more rather than less influence on editor decisions. This might be because editors turn to professional colleagues, of whom they can request favors, or to reviewers they know to be reliable and trustworthy (although not necessarily most appropriate for the specific paper), when they have trouble finding reviewers. Regardless of the explanation, the results are inconsistent with the hypothesis that editors weigh the opinions of late-recruited reviewers less than those of their preferred (early-invited) reviewers.

An alternative hypothesis is that editors may be biased, although subtly, against papers for which they have difficulty recruiting reviewers, either because they believe that difficulty recruiting reviewers is more informative about manuscript quality than it actually is or because they become annoyed at or frustrated by such papers. Editors are necessarily aware of the number of reviewers that have declined because they are either regularly prompted by the editorial office to submit new prospective reviewer names (e.g., at *Functional Ecology*) or must themselves log into the journal manuscript management system to invite new prospective reviewers. It seems reasonable that the additional workload necessary to manage such manuscripts—the need to repeatedly identify new reviewers—will influence, however slightly, an editor's disposition towards the paper. Unfortunately, I cannot test this hypothesis with the data available, but anecdotal evidence (my discussions with editors) suggests that both of these factors occur; some editors do interpret difficulty recruiting reviews as evidence that papers are low quality and/or uninteresting, and they do get frustrated trying to find reviewers for such papers.

Estimates of inter-rater agreement—the degree to which independent assessors score a document similarly—tend to be low for journal and grant peer review (Campos-Arceiz et al. 2015; reviewed in Bornmann et al. 2010a), as observed in the current study, and reviewers generally comment on different aspects of a manuscript (e.g., Fiske and Fogg 1990). Though many studies have quantified inter-rater agreement, few have examined

features of manuscripts, reviewers or editorial processes that influence its magnitude. Those few studies find little evidence that inter-rater agreement varies among disciplines or with the extent to which peer review is blinded, though estimates of inter-rater agreement do generally decline with sample size possibly suggesting a reporting bias in which low estimates are only reported when datasets are large (meta-analysis and review in Bornmann et al. 2010a). Inter-rater agreement is likely influenced in the current study by the high proportion of papers that are declined without review at these journals (reducing the variance in quality among papers) and by the fairly limited reviewer scoring categories they provide (scales between 1 and 4 or between 1 and 7, depending on the journal and year), but, most interestingly, reviewers invited close together in sequence tend to be more consistent in the scores they submit than are reviewers invited more greatly separated in sequence. The effect size was quite large; a drop of nearly half in the correlation between review scores at the extremes (no declined reviewers between invitations versus 6+ declined reviewers between invitations; Fig. 5). This suggests that editors are inviting different kinds of reviewers later in the invitation process than they are at the start. Most likely, editors are first attempting to recruit reviewers with the most appropriate expertise in the topic and/or methods presented in the paper, but resort to reviewers with less expertise in the area of the paper, or reviewers that are less experienced at peer review and/or less familiar with the journal and ratings scale (Sattler et al. 2015), later in the process. These hypotheses may be testable by investigating the details of reviewer history with these journals and asking whether prior reviewing experience and/or similarity in research expertise predict the consistency between review scores.

Despite detectable relationships between reviewer recruitment success and both peer review scores and editorial decisions, I found no evidence that the difficulty editors had recruiting reviewers was predictive of the number of citations that papers received for papers published in *Functional Ecology*, the one journal for which I have both citation and peer review data merged together. That reviewer recruitment success could predict citations seems unlikely given that the journals examined here all have low manuscript acceptance rates (*Functional Ecology* consistently declines >60% of reviewed papers, and >75% of all submissions [Fox and Burns 2015]) and thus the papers published reflect only that subset that were positively evaluated by reviewers and editors. However, the results for *Functional Ecology* contrast with that for the journal *Radiology*, for which papers received fewer citations when reviewers were harder to recruit, despite that journal being more selective (having a lower manuscript acceptance rate) than the journals evaluated here (Kallmes et al. 2017).

Conclusions

Papers for which editors have greater difficulty recruiting reviewers are more likely to be declined after peer review at six journals in ecology and evolution (though the magnitude of the relationship varies substantially among journals). Much of this relationship is explainable by variance in peer review scores—papers for which editors have greater difficulty recruiting reviewers obtain lower scores from reviewers. However, after accounting for peer review scores, editors remain more likely to decline papers for which they had trouble recruiting reviewers. This influence of reviewer recruitment difficulty, although of small effect, should be considered among the diversity of factors, beyond

quality and significance of the research being described, that influence outcomes of the editorial and peer review process at academic journals.

Acknowledgements The *British Ecological Society* and the *Society for the Study of Evolution* provided permission to access their databases for this peer review analysis. Katie Simmons assisted with extracting the reviewer database for *Evolution*, and Emilie Aimé, Christopher Grieves, Kate Harrison, Simon Hoggart, Jennifer Meyer, Erika Newton, Alice Plane, James Ross and Leila Walker extracted the reviewer databases for the *British Ecological Society* journals. Emilie Aimé, C. Sean Burns, Allyssa Kilanowski, Melise Lecheta, Josiah Ritchey and Boris Sauterey provided helpful comments on earlier drafts of this manuscript. This work was approved by the University of Kentucky's Institutional Review Board (IRB 15–0890). C. Fox is Executive Editor of one of the journals examined in this study (*Functional Ecology*).

References

Albert, A. Y., Gow, J. L., Cobra, A., & Vines, T. H. (2016). Is it becoming harder to secure reviewers for peer review? A test with data from five ecology journals. *Research Integrity and Peer Review, 1*(1), 14.

Baveye, P. C., & Trevors, J. T. (2011). How can we encourage peer-reviewing? *Water, Air, and Soil pollution, 214,* 1–3.

Bornmann, L., Mutz, R., & Daniel, H. D. (2010a). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PLoS ONE, 5*(12), e14331.

Bornmann, L., Weymuth, C., & Daniel, H. D. (2010b). A content analysis of referees' comments: How do comments on manuscripts rejected by a high-impact journal and later published in either a low-or high-impact journal differ? *Scientometrics, 83*(2), 493–506.

Breuning, M., Backstrom, J., Brannon, J., Gross, B. I., & Widmeier, M. (2015). Reviewer fatigue? Why scholars decline to review their peers' work. *PS: Political Science and Politics, 48*(04), 595–600.

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.

Fiske, D. W., & Fogg, L. F. (1990). But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. *American Psychologist, 45*(5), 591.

Fox, C. W., Albert, A. Y., & 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, 2*(1), 3.

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*(10), 1970–1980.

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*(1), 140–153.

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*(1), 126–139.

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, 31*(1), 270–280.

Goldman, H. V. (2015) The scarce peer reviewer and challenges journal editors face. http://www.editage.com/insights/the-scarce-peer-reviewer-and-challenges-journal-editors-face. Accessed 2 May 2017.

Kallmes, K. M., Brinjikji, W., Ahmed, A. T., & Kallmes, D. F. (2017). Difficulty in finding manuscript reviewers is not associated with manuscript acceptance rates: a study of the peer-review process at the journal Radiology. *Scientometrics, 111,* 971–978. doi:10.1007/s11192-017-2331-0.

Lajtha, K., & Baveye, P. C. (2010). How should we deal with the growing peer-review problem? *Biogeochemistry, 101,* 1–3.

Mingers, J., & Xu, F. (2010). The drivers of citations in management science journals. *European Journal of Operational Research, 205*(2), 422–430.

Moed, H. F., & Halevi, G. (2016). On full text download and citation distributions in scientific-scholarly journals. *Journal of the Association for Information Science and Technology, 67*(2), 412–431.

Perneger, T. V. (2004). Relation between online "hit counts" and subsequent citations: Prospective study of research papers in the BMJ. *BMJ, 329*(7465), 546–547.

Sattler, D. N., McKnight, P. E., Naney, L., & Mathis, R. (2015). Grant peer review: improving inter-rater reliability with training. *PLoS ONE, 10*(6), e0130450.

Sense About Science (2009). *Peer Review Survey 2009*. http://archive.senseaboutscience.org/pages/peer-review-survey-2009.html. Accessed 2 May 2017.

Stamm, T., Meyer, U., Wiesmann, H. P., Kleinheinz, J., Cehreli, M., & Cehreli, Z. C. (2007). A retrospective analysis of submissions, acceptance rate, open peer review operations, and prepublication bias of the multidisciplinary open access journal Head & Face Medicine. *Head and Face Medicine, 3*(1), 27.

Stremersch, S., Verniers, I., & Verhoef, P. C. (2007). The quest for citations: Drivers of article impact. *Journal of Marketing, 71*(3), 171–193.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS ONE, 8*(5), e64841.

Tite, L., & Schroter, S. (2007). Why do peer reviewers decline to review? A survey. *Journal of Epidemiology and Community Health, 61*(1), 9–12.

Vaughan, L., Tang, J., & Yang, R. (2017). Investigating disciplinary differences in the relationships between citations and downloads. *Scientometrics, 111*(3), 1533–1545.

Vines, T., Rieseberg, L., & Smith, H. (2010). No crisis in supply of peer reviewers. *Nature, 468*(7327), 1041.

Ware, M., & Monkman, M. (2008). *Peer Review in scholarly journals: An international study into the perspective of the scholarly community*. Bristol: Mark Ware Consulting.

Willis, M. (2016). Why do peer reviewers decline to review manuscripts? A study of reviewer invitation responses. *Learned Publishing, 29*(1), 5–7.

Zaharie, M. A., & Osoian, C. L. (2016). Peer review motivation frames: A qualitative approach. *European Management Journal, 34*(1), 69–79.