Received: 5 November 2022 | Accepted: 9 December 2022

DOI: 10.1111/1365-2435.14259

Functional Ecology BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal

Charles W. Fox[1] | Jennifer Meyer[2] | Emilie Aimé[2,3]

[1]Department of Entomology, University of Kentucky, Lexington, Kentucky, USA

[2]British Ecological Society, London, UK

[3]Royal Entomological Society, St Albans, UK

Correspondence
Charles W. Fox
Email: cfox@uky.edu

Funding information
British Ecological Society

Abstract

1. There is substantial evidence that systemic biases influence the scholarly peer review process. Many scholars have advocated for double-blind peer review (also known as double-anonymous review) to reduce these biases. However, the effectiveness of double-blind peer review in eliminating biases is uncertain because few randomized trials have manipulated blinding of author identities for journal submissions and those that have are generally small or provide few insights on how it influences reviewer biases.
2. In 2019, *Functional Ecology* began a large, randomized trial, using real manuscript submissions, to evaluate the various consequences of shifting to double-blind peer review. Research papers submitted to the journal were randomly assigned to be reviewed with author identities blinded to reviewers (double-blind review) or with authors identified to reviewers (single-blind review). In this paper, we explore the effect of blinding on the outcomes of peer review, examining reviewer ratings and editorial decisions, and ask whether author gender and/or location mediate the effects of review type.
3. Double-blind review reduced the average success of manuscripts in peer review; papers reviewed with author identities blinded received on average lower ratings from reviewers and were less likely to be invited for revision or resubmission. However, the effect of review treatment varied with the author's location.
4. Papers with first authors residing in countries with a higher human development index (HDI) and/or higher average English proficiency fared much better than those from countries with a lower HDI and lower English proficiency, but only when author identities were known to reviewers; outcomes were similar between demographic groups when author identities were not known to reviewers.
5. Blinding author identities had no effect on gender differences in reviewer ratings or editor decisions.
6. Our data provide strong evidence that authors from higher income and/or English-speaking countries receive significant benefits (a large positive bias) to being identified to reviewers during the peer review process and that anonymizing author-identities (e.g. double-blind review) reduces this bias, making the peer review process more equitable. We suggest that offering optional blinding of

author identities, as some journals allow, is unlikely to substantially reduce the biases that exist because authors from higher-income and English-speaking countries are the least likely to choose to be reviewed with their identity anonymized.

KEYWORDS
double-anonymous peer review, gender bias, prestige bias, single-blind peer review, unconscious bias

1 | INTRODUCTION

The assessment of research by experts in the subject, commonly called peer review, is the primary means by which granting agencies and journal editors assess the quality and importance of scholarly research. Yet there is a widespread perception among researchers that systemic biases influence the outcomes of peer review. In particular, there is concern that the gender, nationality, location or reputation of authors may influence how manuscripts are assessed by reviewers due to unconscious, or sometimes conscious, biases (Lee et al., 2013). Of all of the possible biases that can occur during peer review, bias against women has received the most attention, although studies are inconsistent in their outcomes. Some find that papers authored by women are rated lower or have lower acceptance rates in academic journals (Fox et al., 2019; Murray et al., 2018; Walker et al., 2015; and references therein), but others have failed to find gender differences in peer review outcomes, or have even found that papers authored by women perform better than those authored by men (Lerback & Hanson, 2017; Squazzoni et al., 2021). Other biases, such as favouring papers by authors from the same country or that speak the same language as the reviewer (Murray et al., 2018), favouring authors from higher-income (Demarest et al., 2014; Harris, Macinko, et al., 2017; Harris, Marti, et al., 2017; Kowal et al., 2022; Saposnik et al., 2014) or English-speaking countries (Ross et al., 2006; Saposnik et al., 2014), discriminating based on author race (Nakamura et al., 2021), favouring papers by authors that are from prestigious institutions (Blank, 1991; Tomkins et al., 2017) or with prestigious reputations (Huber et al., 2022; Okike et al., 2016), favouring authors that are senior (Pleskac et al., 2021), or disfavouring newcomers to a discipline (Seeber & Bacchelli, 2017), can all distort the quality and fairness of peer review.

To reduce the influence of potential biases on peer review, many scholars have advocated for double-blind peer review. Double-blind review (sometimes called double-anonymous review) is review in which author identities are blinded to the reviewers (and sometimes blinded to editors), as opposed to single-blind review, which is more traditional in life science disciplines, where reviewers know the identities of the authors (Jubb, 2016; Shoham & Pitman, 2021). In both models, reviewers are usually anonymous to authors, although most journals allow reviewers to identify themselves if they choose (Fox, 2021). Surveys commonly show that double-blind review is the preferred model of peer review among scholars across a wide diversity of disciplines (e.g. Ho et al., 2013; Jagsi et al., 2014; Mulligan et al., 2013; Regehr & Bordage, 2006; Rowley & Sbaffi, 2018; Taylor & Francis, 2015), and a growing number of journals are adopting this model of peer review. Among ecology and evolutionary biology journals, double-blind peer review has taken primarily two forms, with journals either requiring that author identities be masked from reviewers for all papers (e.g. *Evolution*, *Journal of Evolutionary Biology*, *Oikos*, *Conservation Biology* and *Diversity and Distributions*, among others), or allowing authors to choose whether their identity is disclosed to reviewers (e.g. *The American Naturalist* and *Nature Ecology & Evolution*). However, the effectiveness of double-blind peer review in eliminating reviewer biases in scholarly journals is uncertain, partly because few randomized trials have manipulated the blinding of author identities on journal submissions. Of the trials that have been performed, most examine few or no author characteristics (McNutt et al., 1990; van Rooyen et al., 1998) or have sample sizes that are too small to have any confidence in negative results (Alam et al., 2011; Blank, 1991; Fisher et al., 1994; Godlee et al., 1998; Justice et al., 1998). The paucity of large-scale manipulative studies directly testing the effects of blinding in a scholarly journal context leaves us uncertain how switching to double-blind peer review will influence the review process at journals, and whether it will reduce the expression of conscious and unconscious biases by reviewers and editors.

In September 2019, the British Ecological Society journal *Functional Ecology* began a randomized trial to evaluate the various consequences of shifting to double-blind peer review, with the ultimate goal of providing data to help journals, including *Functional Ecology*, make informed decisions about the consequences of adopting a single- or double-blind peer review model (Fox, 2019; Fox et al., 2019). Every research paper submitted to the journal over 3 years (n = 3739 submissions, 1432 sent for peer review) was randomly assigned to be reviewed either with author identities blinded to reviewers (double-blind review) or with authors identified to reviewers (single-blind review). The dataset includes a large diversity of variables about each paper, including the names and addresses of all authors, the complete review history of each paper (including details on each of the editors who handled the paper, and details on all invited reviewers whether they agreed to review or not). In this paper, we explore the first results from this peer review trial focusing on the outcomes of peer review (reviewer ratings and editorial decisions). Future papers will focus on additional aspects of peer review, including success in recruiting reviewers, other aspects of reviewer performance, the ability of reviewers to de-anonymize authors, and prestige bias.

The goals of this paper are to test (a) whether blinding of author identities influenced the outcome of the peer review process, for example, whether review treatment affected the ratings given to papers by reviewers and/or the decisions of editors and (b) whether the effects of review treatment were influenced by demographic traits of an author, specifically an author's gender, language proficiency and/or the economic development status of their country of residence. We show that double-blind review reduced the average success of manuscripts; papers with author identities blinded were rated lower by reviewers and had lower acceptance rates. However, this reduction in success when reviewed double-blind was largely due to a reduction in quite substantial positive biases that benefit authors from wealthy and English-speaking countries. Blinding author identities had no effect on gender differences in reviewer ratings or editor decisions.

2 | MATERIALS AND METHODS

2.1 | The experiment

Beginning on 5 September 2019, all papers submitted to the journal *Functional Ecology* were required to be submitted as if they would be reviewed with the author identities blinded from reviewers during peer review. Specifically, authors were required to submit their title page separately from the rest of the document, with all information that could identify authors included only on the title page. This included (a) all author names, addresses, institutional affiliations and email addresses, (b) acknowledgements (including references to grants and funding), (c) a conflict-of-interest statement, (d) an author contributions statement and (e) the data accessibility statement (if it included identifiable information). Adherence to this was checked by the journal editorial office before sending manuscripts for initial assessment by editors. Details of the instructions given to authors are included in the *Supplemental Material*. In brief, 'All manuscripts should take care not to obviously identify the authors in the main text and supporting information'. This includes citing previous work without indicating that it belongs to the authors, for example, we request that authors cite their previous work as 'in a previous study by' rather than 'in our previous study'. Although this instruction to authors was clear, in-text citations were not checked in detail by the editorial office. Reviewers were limited to seeing the system-generated PDF (which included supplemental material) and not the original author-uploaded files, to prevent the identification of authors through file metadata.

Papers were randomly assigned to be reviewed either with the author identities known to reviewers (single-blind review) or with the author identities unknown to reviewers (double-blind review). Papers were assigned to treatments randomly, based only on the manuscript number that was assigned by the manuscript tracking system ScholarOne Manuscripts as the authors began the submission process. The authors, editors and journal administrative staff had no influence over which papers were assigned to which treatment. Papers assigned to the single-blind and double-blind treatments were handled identically at all stages, except that (a) letters inviting prospective reviewers identified the authors for papers assigned to the single-blind treatment but not in the double-blind treatment and (b) the PDF of the manuscript made available to reviewers included the title page (and thus author details, acknowledgements and author contribution statements) when reviewed single-blind but not when reviewed double blind. Author identities were *not* blinded to editors in either treatment; this study only blinded author identities to reviewers in the double-blind treatment.

2.2 | Dataset

The data analysed and presented here was extracted from ScholarOne Manuscripts on 15 September 2022 and thus includes just over 3 years of submissions to the journal. It includes all research manuscripts submitted from the start of the trial up to that date. The dataset excludes all non-research manuscripts, for example, Perspectives, Reviews, Editorials, Commentaries and other non-research papers, since these were not part of the trial. We also consider only the first submission of a manuscript; invited revisions are excluded. Manuscripts submitted in the final 2 months of the date range considered were often still undergoing peer review when this dataset was extracted. Thus, the analyses consider a slightly different subset of papers depending on the stage of the review process the paper is at. The trial is ongoing at the time this is being written, so an expanded dataset will be available at a later date.

2.3 | Permits and permissions

The datasets analysed here include personal identifiers. It is thus essential to maintain the confidentiality of all participants. The dataset provided online (Dryad; datadryad.org) was anonymized to maintain the privacy of authors. As a consequence, some of the analyses reported here cannot be repeated exactly as reported in this paper using the published dataset. This project was reviewed and certified by the University of Kentucky Institutional Review Board (IRB Number 51706).

2.4 | Author traits

We examine three biographical features of authors that may interact with review treatment to predict the outcomes of the peer review process—author gender and the language proficiency and economic development status of the country in which they live. We focus on the traits of first authors in this paper but note that results for corresponding authors are qualitatively the same. This is not unexpected because a large majority of first authors in *Functional Ecology*, and ecology more broadly, are also corresponding authors of their paper (generally 75%–85%; Fox et al., 2018).

We assigned gender to authors and reviewers using an online database of given names, genderize.io, using GENDERIZER (Wais, 2016) as in previous studies (details in Fox et al., 2016; Fox & Paine, 2019). The database returns the most likely gender for each given name based on a dataset of more than 114 million data entries covering the entire world. Genderize.io has been demonstrated to perform very well for names in western countries (Karimi et al., 2016) and has recently expanded its coverage of non-Western names. We were able to assign a gender to all except 1% (39 of 3739) of first authors.

Authors and reviewers are required to self-identify the country of their affiliation when creating an account for the journal in *ScholarOne Manuscripts*. We used these author-entered countries as the best estimate of each author's and/or reviewer's location at the time they submitted to or reviewed for the journal. We used the United Nations Development Programme's (UNDP) measure of Human Development Index (HDI) to determine the level of development of the author's country of residence (United Nations Development Programme, 2020). HDI is calculated from indices of three dimensions of human development: life expectancy, education and gross national income per capita. It is highly correlated with measures of GDP per capita across countries (Szigeti et al., 2013) but is a broader measure of economic and societal well-being that includes investment in health and education (Deb, 2015). 64% of all submissions were from countries with HDI > 0.80, which the UNDP considers to be *Very High* human development.

We do not know the language fluency of authors or reviewers in our dataset. However, we can predict how likely an author is to be proficient in English based on the country in which they reside when they interact with the journal. We thus categorized an author's English proficiency in two ways; (a) mean total TOEFL scores (iBT Tests, 2020) for the author's country of residence and (b) whether English is the most common and/or an official language of their country of residence (as in Clavero, 2011). The TOEFL is a test of English-language proficiency required by many graduate schools in English-speaking countries. Whether English is the most common or an official language was determined from the online version of the *CIA World Factbook* (The World Factbook 2017). Both metrics of English-language proficiency are imperfect because many authors are bilingual from childhood, many who learn English as a second language are excellent at writing in English, many researchers get an education in English even if English is uncommon in their country of residence (and their English generally improves substantially after taking the TOEFL and immigrating to an English-language country), and researchers move among countries of different languages and thus may reside at the time they interact with the journal in a country that does not correctly indicate their native language.

2.5 | Reviewer recommendations

When submitting their review to the journal, reviewers are asked to rate a paper on a four-point scale (1–4). The journal provides guidance to the meaning of each rating, as follows (note: the scale used on the review submission form considers 1 the best possible rating but for this paper, we have reversed the order of the ratings so that higher values are better because this is more intuitive for readers):

4. An extremely novel paper that is in the top 10% of all papers you have read in the broader field of ecology.
3. A strong contribution to the broader field of ecology.
2. Solid work, but largely confirmatory.
1. Weak or flawed, or not of enough importance and general interest for Functional Ecology.

Although individual ratings are on a four-point scale, the average of the multiple ratings (multiple reviews) assigned to each paper (i.e. mean rating per paper) is a continuous variable that deviates only slightly from a normal distribution (skewness = −0.11, kurtosis = −0.68, both of which are well within the range considered adequate for use of normal distribution-based statistics by Curran et al., 1996).

2.6 | Editor decisions

The first decision editors make about each submission is whether to send the paper for review, which has two steps: first, a Senior Editor decides whether to assign a paper to an Associate Editor and then the Associate Editor decides whether to send the paper for peer review. We lump these into a single decision, whether the paper was sent for peer review. After peer review, the Associate Editor considers the reviewers' comments and recommends either that the paper be invited for revision or rejected. Senior Editors review these recommendations and only a Senior Editor can invite a paper for revision or resubmission, but it is unusual for the Senior Editors to make a different decision than recommended by the Associate Editor. We thus consider only the final decision made on a paper when evaluating the outcomes of peer review.

There are three decisions a Senior Editor can make: they can (A) invite a revision of the manuscript (nearly all of these are accepted for publication), (B) reject the manuscript but invite submission of a new manuscript pending addition of new data or other substantial changes (often called 'declined without prejudice' at other journals; many of these, but not all, are eventually published in the journal), or (C) reject the manuscript with no option for resubmission. Here we consider both A and B to be positive outcomes and consider C a negative outcome because C is the only decision that does not allow authors to submit a revised manuscript. However, analyses considering only A as a positive outcome and both B and C as negative outcomes reach the same conclusions as presented in this paper (analyses presented in Supplemental Material Table S1).

2.7 | Analyses

Most of the response variables examined in this study are binary, for example, whether a paper is sent for review [yes/no] or whether a resubmission is invited [yes/no]. The effects of the review treatment (single-blind vs. double-blind) and author biographical factors (gender, language, HDI) were tested using logistic regressions of the form *ResponseVariable*[1,0] = *Treatment*[SingleBlind/DoubleBlind] or *ResponseVariable*[1,0] = *AuthorCharacteristics* + *Treatment*[SingleBlind/DoubleBlind] + *Interactions* (SAS version 9.4, Proc Glimmix, link = logit and dist = binomial; SAS Institute Inc., 2016), with each manuscript counting as a single data point.

Variation in review scores (the rating given to papers by reviewers, averaged among reviewers) was examined using linear models (SAS Proc Mixed; SAS Institute Inc., 2016) structured similarly to the logistic regressions described above.

3 | RESULTS

In total, 1837 papers assigned to the single-blind treatment and 1852 assigned to the double-blind treatment had decisions made on them at the time we extracted these data (an additional 50 papers were still undergoing editorial or peer review). The small difference in sample size between treatments is because some submissions are reviews, commentaries or other non-research papers; those were assigned numbers by ScholarOne but are not included in our analyses.

During the period of this study, 40.2% of papers were sent for review. Of papers that were reviewed, 34.3% were invited for revision, 16.7% were declined without prejudice (and thus allowed to resubmit), and 48.9% were rejected without an option to resubmit. Thus, of all papers that were submitted, 79.4% were rejected with no option to resubmit, 13.8% were invited to submit a revision and 6.7% were declined without prejudice (and thus allowed to resubmit), for a total of 20.5% of submissions getting positive decisions (henceforth described as being invited to resubmit).

3.1 | Papers reviewed double-blind fare worse than papers reviewed single-blind

The proportion of papers sent by editors for peer review did not differ between the single-blind (37.9%) and double-blind (39.3%) treatments (*PaperReviewed*[yes/no] = *Treatment*[single-blind/double-blind]; $X_1^2 = 0.79$, $p = 0.38$). This is not surprising because author identities are not blinded to editors and thus the two review treatments are effectively the same until the paper is sent for peer review.

Significant treatment differences arose during peer review. Papers sent for peer review got lower ratings from reviewers, averaged across all papers, when reviewed with the author identities blinded (double-blind treatment; $N = 708$) than when the author identities were known to the reviewers (single-blind treatment, $N = 674$) (*ReviewScore*[mean per paper] = *Treatment*; $F_{1,1380} = 7.04$, $p = 0.008$). This difference in review scores led to a substantial difference in outcomes between papers reviewed single- versus double-blind; papers reviewed single-blind were 24.2% more likely to be invited to resubmit a revision compared to those reviewed double-blind (38.3 ± SE 1.9 vs. 30.8 ± 1.8%; $X_1^2 = 8.4$, $p = 0.004$) and were 15.2% more likely to have an overall positive outcome (either invited to submit a revision or invited be resubmit [if rejected]; 54.0 ± 1.9 vs. 46.8 ± 1.9%; $X_1^2 = 7.0$, $p = 0.008$). These treatment differences in final decisions were largely due to the treatment difference in review scores; when review score was included as a covariate in the model for editor decisions, only review score ($X_1^2 > 355.4$, $p < 0.001$) and not review treatment ($X_1^2 < 0.20$, $p > 0.66$) predicted the final decision for both revisions invited and resubmissions invited (model: *PositiveOutcome*[yes/no] = *Treatment* + *ReviewScore*[covariate] + *Interaction*). There was no evidence that the relationship between review scores end editor decisions differed between review treatments (the treatment-by-review score interactions were not significant for either revisions or resubmissions invited; $X_1^2 < 0.06$, $p > 0.81$ for each).

For the rest of our analyses, we consider both an invited revision and an invited resubmission ('reject with resubmission invited') as the positive outcome, 'Resubmission invited'. Analyses considering only 'revision invited' as a positive outcome (excluding 'reject with resubmission invited') are very similar and do not change any conclusions; those results are presented in the supplemental material, in Table S1 and Figure S1.

3.2 | Author gender and the outcomes of peer review

A primary objective of this study is to ask whether blinding author identities influence peer review differently for papers submitted by male versus female authors. In brief, we find that papers with female first authors were equally likely to be sent for peer review (Figure 1a), received slightly (but statistically significantly) higher review scores from reviewers (Figure 1b) and were equally likely to be invited for resubmission (if reviewed) (Figure 1c), compared to papers with male first authors. We found no evidence that peer review treatment (single- vs. double-blind) differentially influenced the review of male versus female-authored papers at any stage of the process (the non-significant treatment-by-gender interactions in Figure 1).

Because logistic regression analyses have low power for detecting interactions, and because our conclusions in this paper may have important consequences for journal policy decisions, we explore these results more thoroughly to confirm that our conclusion—that there is no interaction between review treatment and author gender—is justified. We focus on the gender of first authors but note that analyses for gender of corresponding authors reach the same conclusions.

3.2.1 | Proportion of papers sent for review

The proportion of papers sent for peer review by editors did not differ between male and female first authors, and the

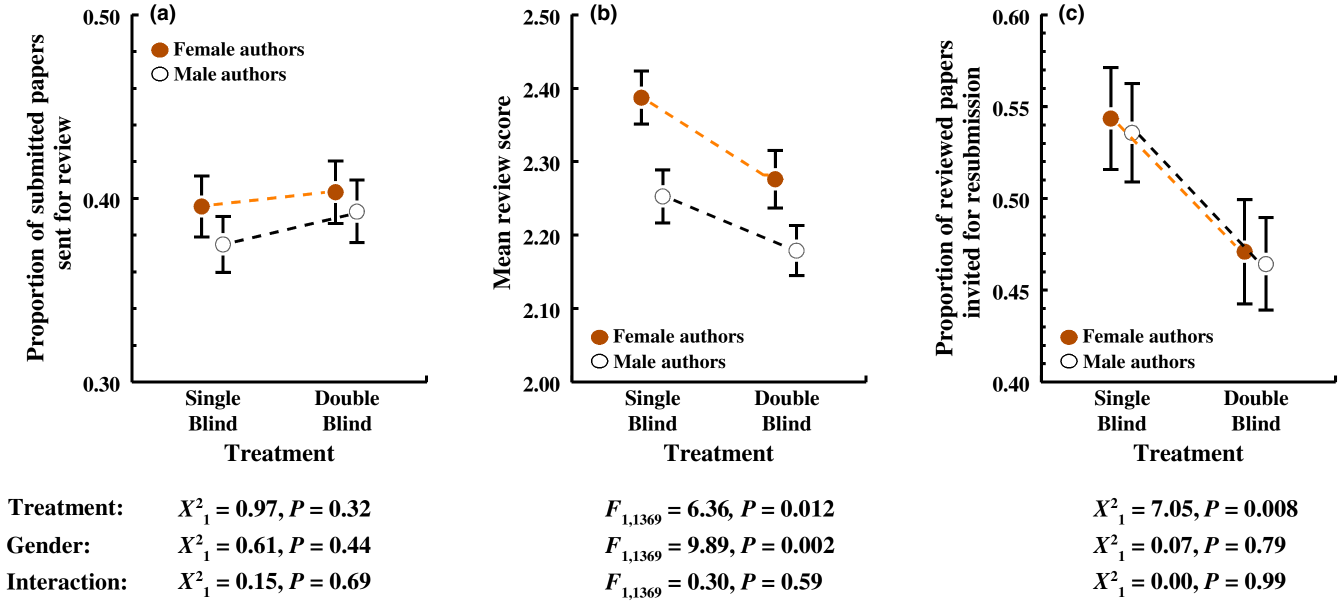


FIGURE 1 Double-blind review does not differentially affect the review of papers by male versus female first authors. (a) The proportion of submitted papers sent for peer review; (b) The mean scores given to papers by reviewers (higher is better); (c) The proportion of reviewed papers invited for resubmission (includes both revision invitations and papers rejected with an invitation to resubmit). The statistical models (logistic regressions in A and C, linear model in B) are as follows: *ResponseVariable* = *Treatment* + *AuthorGender*[male/female] + *Interaction*.

treatment-by-gender interaction was non-significant in the logistic regression (statistics presented in Figure 1a). Linear contrasts bolster the conclusion that the effect of treatment did not vary with author gender—there was no evidence that treatment affected whether papers were sent for review for either gender (linear contrasts comparing treatments, $t_{3671} = 0.87$, $p = 0.38$ and $t_{3671} = 0.26$, $p = 0.79$, for male and female authors, respectively), and no evidence of a significant gender difference in the proportion of papers sent for peer review in either treatment (linear contrasts, $t_{3671} = 0.97$, $p = 0.33$ and 0.42, $p > 0.67$, for the single and double blind treatments, respectively).

3.2.2 | Review scores and post-review editorial decisions

Of papers that were sent for peer review, those with female first authors received slightly higher review scores, on average across papers and treatments (Figure 1b), compared to papers with male first authors. The difference in review scores between female and male authors was fairly small (2.33 ± 0.03 vs. 2.21 ± 0.03 [averaged across treatments]) and did not translate into a significant difference in how likely they were invited to resubmit (Figure 1c). There was also no evidence that treatment influenced the gender difference in review scores or the gender difference in the probability an author was invited to resubmit (non-significant interactions between treatment and gender in Figure 1b,c).

The conclusion that there is no significant interaction between treatment and gender is bolstered by the treatment-specific effect sizes. Review scores for papers authored by women were on average 6.1% higher than review scores for papers authored by men in the single-blind treatment (linear contrasts, $t_{1369} = 2.59$, $p = 0.01$), nearly the same as the difference (4.5% higher) in the double-blind treatment ($t_{1369} = 1.85$, $p = 0.06$). Similarly, papers authored by women were on average 1.4% more likely than men to be invited to resubmit in the single-blind treatment (linear contrasts, $t_{1418} = 0.20$, $p = 0.84$), identical to the difference (1.4%) in the double-blind treatment ($t_{1418} = 0.18$, $p = 0.86$).

If we add review score (as a covariate) to the model predicting the probability that a reviewed paper is invited for resubmission (*ResubmissionInvited*[yes|no] = *Treatment* + *AuthorGender*[male/female] + *ReviewScore*[covariate] + *Interaction*), review score is the variable that overwhelmingly predicts whether a resubmission was invited or not (*ReviewScore*: $X_1^2 = 355.7$, $p < 0.001$). In this analysis (with review score as a covariate) we see evidence that papers authored by men (relative to those authored by women) are invited for resubmission slightly more often than expected based on their review score (AuthorGender: $X_1^2 = 5.3$, $p = 0.02$), but the effects of review treatment and its interaction with gender are both non-significant ($X_1^2 < 1.32$, $p > 0.25$ for each).

3.3 | Author location predicts the outcomes of peer review

In contrast to our results for gender, our data suggest that peer-review model (single- vs. double-blind peer review) *does* differentially influence the outcomes of peer review for authors from high versus low-income countries and English versus non-English-speaking countries. Specifically, our data suggest that authors from high-HDI and/or English-speaking countries benefit during peer review from being identified as authors (single-blind review) and lose those benefits when their identities are blinded during peer review (double-blind review) (Figures 2b,c and 3b,c). In contrast, papers submitted by authors from low-HDI and/or non-English countries perform similarly across review treatments. We elaborate on these results below.

3.3.1 | Main effects of economic status (HDI)

HDI (United Nations human development index) is measured quantitatively and so we present analyses treating HDI as a continuous variable in Table 1. However, interactions between categorical and continuous variables in logit models are not analogous to interactions in linear models (Ai & Norton, 2003; Norton et al., 2004)—the interaction effect is conditional on the independent variables—and thus they are difficult to interpret and visualize. We thus focus our discussion here on a categorical description of HDI, with HDI > 0.80 classified as *Very High* human development (following the definitions of the United Nations Development Programme) (64% of all submissions). The results treating HDI as a categorical variable match those for HDI as a continuous variable but are easier to discuss and visualize.

Papers submitted by authors from countries with a *Very High* HDI were 68% more likely to be sent for peer review than were papers by authors from countries with lower HDI (45.1 ± 1.0% vs. 26.8 ± 1.2% of papers were sent for peer review) (Figure 2a). Of papers sent for review, those by authors from higher HDI countries received higher review scores (Figure 2b) and were 16% more likely to be invited for resubmission than were papers from countries with HDI < 0.80 (52.3 ± 1.6% vs. 45.0 ± 2.7%) (Figure 2c). If we add the mean review score (per paper) as a covariate in the model for resubmission invitations the large difference between higher versus lower HDI countries disappears (the effect of HDI becomes non-significant; $X_1^2 = 0.01$, $p = 0.92$) and only review score remains a significant predictor of whether a resubmission was invited ($X_1^2 = 353.7$, $p < 0.001$), suggesting that the large influence of HDI on editorial decisions was mediated by papers from low-HDI countries obtaining lower review scores.

3.3.2 | Interactions between treatment and author HDI

The most striking result we find is that the influence of the HDI for an author's country of residence on scores given to papers by reviewers differed between the two review treatments (significant interaction between treatment and HDI whether HDI is treated categorically [Figure 2b] or continuously [Table 1]). Specifically, review scores were on average higher for researchers from countries classified as having a *Very High* HDI relative to researchers from lower HDI countries in the single-blind treatment (linear contrasts, $t_{1378} = 4.48$, $p < 0.001$) but not in the double-blind treatment ($t_{1378} = 1.21$, $p = 0.23$). Likewise, the proportion of authors invited to resubmit was higher for researchers from high HDI countries in

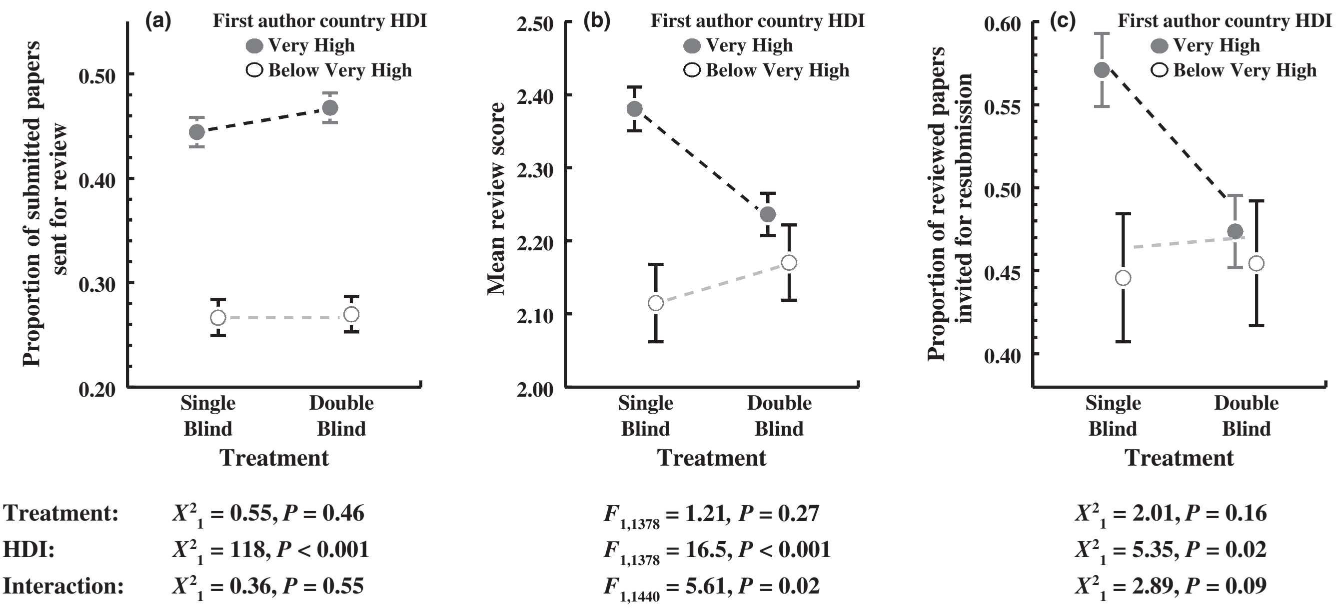
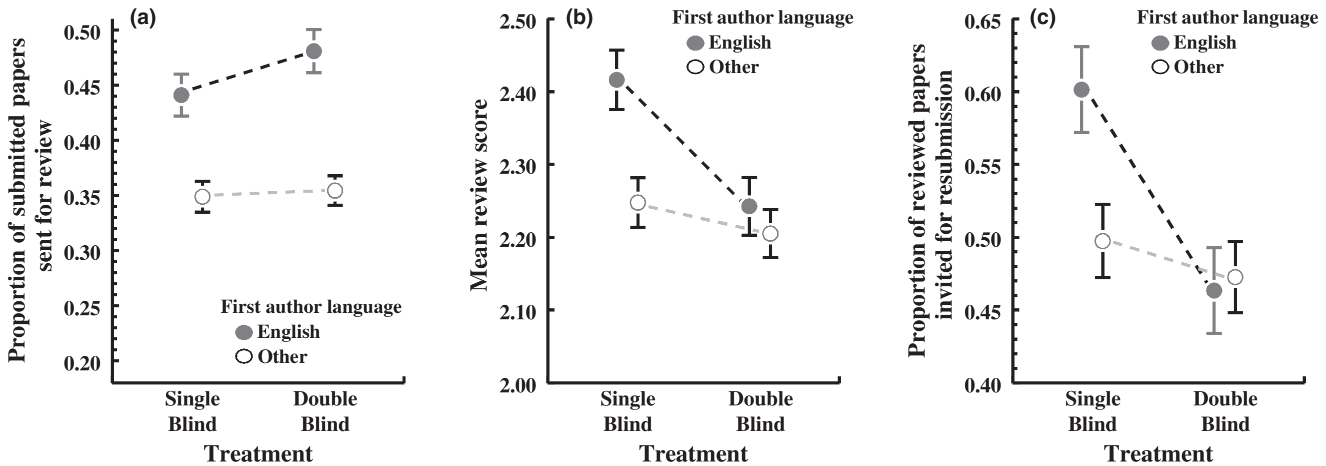


FIGURE 2 The difference in manuscript outcomes for papers submitted by authors from countries with an HDI classified as *Very High* by the United Nations Development Programme (2020) versus countries with an HDI less than *Very High*. (a) The proportion of submitted papers sent for peer review; (b) The mean scores given to papers by reviewers (higher is better); (c) The proportion of reviewed papers invited for resubmission (includes both revision invitations and papers rejected with an invitation to resubmit). Analyses of HDI as a continuous variable are presented in Table 1. The statistical models (logistic regressions in A and C, linear model in B) are as follows: *ResponseVariable* = *Treatment*[SingleBlind/DoubleBlind] + *HDI*[*VeryHigh/Other*] + *Interaction*.



	(a)	(b)	(c)
Treatment:	$X^2_1 = 1.82, P = 0.18$	$F_{1,1378} = 8.38, P = 0.004$	$X^2_1 = 8.87, P = 0.003$
Language:	$X^2_1 = 40.1, P < 0.001$	$F_{1,1378} = 6.34, P = 0.01$	$X^2_1 = 3.03, P = 0.08$
Interaction:	$X^2_1 = 1.16, P = 0.28$	$F_{1,1440} = 2.48, P = 0.12$	$X^2_1 = 4.32, P = 0.04$

FIGURE 3 The difference in manuscript outcomes for papers submitted by authors from English- versus non-English-speaking countries. (a) The proportion of submitted papers sent for peer review; (b) The mean scores given to papers by reviewers (higher is better); (c) The proportion of reviewed papers invited for resubmission (includes both revision invitations and papers rejected with an invitation to resubmit). We categorized an author as being fluent in English if English is the most common and/or an official language of their country of residence. Analyses for English proficiency as a continuous variable (TOEFL scores) are presented in Table 1. The statistical models (logistic regressions in A and C, linear model in B) are as follows: *ResponseVariable* = *Treatment*[SingleBlind/DoubleBlind] + *Language*[*English*/*Other*] + *Interaction*.

TABLE 1 The influence of first author HDI (the UN Human Development Index of the author's country of residence; continuous variable) and peer review treatment (single- vs. double-blind) on the outcomes of peer review. The statistical models (logistic regressions for *Paper sent for review* and *Resubmission invited*, linear model for *Review score*) are: *ResponseVariable* = *Treatment*[SingleBlind/DoubleBlind] + *HDI*[*continuous*] + *Interaction*, *with AuthorCountry* included as a random effect. HDI is transformed as log[1-HDI] to create a more even spread of values

	Paper sent for review	Review score (reviewed papers only)	Resubmission invited (if reviewed)
Treatment	$X_1^2 = 0.02$, $p = 0.88$	$F_{1,1325} = 1.93$, $p = 0.16$	$X_1^2 = 1.37$, $p = 0.24$
HDI (continuous)	$X_1^2 = 44.4$, $p < 0.001$	$F_{1,1325} = 19.3$, $p < 0.001$	$X_1^2 = 7.46$, $p = 0.006$
Interaction	$X_1^2 = 0.2$, $p = 0.66$	$F_{1,1325} = 4.32$, $p = 0.04$	$X_1^2 = 3.37$, $p = 0.07$

the single-blind treatment (28% higher; 57.1 ± 2.2 vs. 44.6 ± 3.9%; linear contrasts, $t_{1378} = 2.79$, $p = 0.005$) but not in the double-blind treatment (only 4% higher; 47.4 ± 2.2 vs. 45.5 ± 3.8%; $t_{1378} = 0.44$, $p = 0.66$), although the interactions were not statistically significant in the model for resubmissions invited ($p = 0.07$ for continuous HDI [Table 1] and $p = 0.09$ for categorical HDI [Figure 2c]).

Importantly, our results do *not* suggest that papers submitted by authors from low-HDI countries are negatively impacted by single-blind peer review; papers from low-HDI countries performed similarly (obtain similar review scores and are similarly likely to be invited to resubmit) whether reviewed single-blind or double-blind (linear contrasts, $t_{1378} = 0.73$, $p = 0.47$ and $t_{1378} = 0.16$, $p = 0.87$ for review scores and resubmission invited, respectively). Instead, our results suggest that authors from high HDI countries get a significant boost (get higher review scores and are more likely to be invited for revision) when being reviewed single-blind compared to double-blind (linear contrasts, $t_{1378} = 3.49$, $p < 0.001$ and $t_{1378} = 3.13$, $p = 0.002$). In other words, the bias we have detected appears to be a positive bias towards authors from high-income countries when their identities are known by authors, rather than a negative bias against authors from low-income countries. The bias appears to be entirely at the reviewer stage—when including review scores in the analysis for resubmissions invited (the decision made by editors), only the effect of review score is significant ($X_1^2 = 353.7$, $p < 0.001$) and all linear contrasts between groups become non-significant ($t_{1377} < 1.01$, $p > 0.31$ for each).

3.3.3 | Main effect of author language

Our study examined two metrics of English language proficiency, TOEFL scores from an author's country of residence (Table 2) and whether the author's country of residence has English as the most common or official language (Figure 3). As noted above, interactions

TABLE 2 The influence of first author TOEFL scores (mean total TOEFL score of the author's country of residence; continuous variable) and peer review treatment (single- vs. double-blind) on the outcomes of peer review. The statistical models (logistic regressions for *Paper sent for review* and *Resubmission invited*, linear model for *Review score*) are as follows: *ResponseVariable* = *Treatment*[SingleBlind/DoubleBlind] + *TOEFL*[*continuous*] + *Interaction*, *with AuthorCountry* included as a random effect. Note that the main effects of Treatment and TOEFL, and their interaction, become non-significant in all models if HDI is included as a covariate; in each case, HDI is significant and the only variable that significantly explains the main effects and interactions

	Paper sent for review	Review score (reviewed papers only)	Resubmission invited (if reviewed)
Treatment	$X_1^2 = 0.08$, $p = 0.77$	$F_{1,1331} = 3.77$, $p = 0.05$	$X_1^2 = 1.64$, $p = 0.20$
TOEFL (continuous)	$X_1^2 = 76.3$, $p < 0.001$	$F_{1,1325} = 9.56$, $p = 0.002$	$X_1^2 = 5.67$, $p = 0.02$
Interaction	$X_1^2 = 0.06$, $p = 0.80$	$F_{1,1325} = 4.34$, $p = 0.04$	$X_1^2 = 2.01$, $p = 0.16$

in logistic regression are difficult to interpret and visualize when one of the independent variables is continuous. We thus focus our discussion on the categorical description of English proficiency (Figure 3) but note that the results and conclusions for TOEFL scores (Table 2) are consistent.

Papers submitted by authors from English-speaking countries (35% of submissions) were 31% more likely to be sent for peer review than were papers by authors from non-English-speaking countries (45.5 ± 1.4 vs. 34.9 ± 1.4%) (Figure 3a). However, English-speaking countries also have high HDI and, if we include the HDI of an author's country of residence as a covariate (continuous variable) in the analysis, the language effect becomes non-significant ($X_1^2 = 1.54$, $p = 0.21$) while the HDI effect is highly significant ($X_1^2 = 42.4$, $p < 0.001$).

Of papers sent for review, those from English-speaking countries received higher review scores (Figure 3b) and were 10% more likely to be invited to resubmit (53.3 ± 2.1% vs. 48.5 ± 1.7%) (Figure 3c). When HDI was included as a covariate in the models testing for the effect of author language, the main effect of English proficiency became non-significant for both review scores ($F_{1,1326} = 0.71$, $p = 0.40$) and the probability that a resubmission was invited ($X_1^2 = 0.69$, $p = 0.41$).

3.3.4 | Interactions between treatment and author English proficiency

The influence of author language on scores given to papers by reviewers (Figure 3b) and the proportion of papers invited for resubmission (Figure 3c) showed generally the same pattern as that for author HDI (Figure 2). For review scores, we see a significant interaction between treatment and the TOEFL scores of an author's country of residence (Table 2), but this treatment-by-language interaction is not significant when we treat English proficiency categorically ($p = 0.12$; Figure 3b). In contrast, there was a significant interaction between treatment and language for whether a resubmission was invited when English proficiency was treated as a categorical variable (Figure 3c) but not for TOEFL scores (Table 2). Comparing linear contrasts, we see that mean review scores were on average higher for researchers from English-speaking countries in the single-blind treatment (linear contrasts, $t_{1378} = 2.86$, $p = 0.004$) but not in the double-blind treatment ($t_{1378} = 0.68$, $p = 0.50$) (Figure 3b). Similarly, the proportion of authors invited to resubmit their paper was higher for researchers from English-speaking countries in the single-blind treatment (21% higher; 60.1 ± 2.9 vs. 49.8 ± 2.5%; linear contrasts, $t_{1378} = 2.66$, $p = 0.008$) but not in the double-blind treatment (for which they were 2% lower; 46.3 ± 2.9 vs. 47.3 ± 2.4%; $t_{1378} = 0.24$, $p = 0.81$) (Figure 3c). As with HDI, our results do not suggest that papers submitted by authors from non-English countries are negatively impacted by single-blind peer review; papers from non-English countries performed similarly whether reviewed single-blind or double-blind (linear contrasts, $t_{1378} = 1.03$, $p = 0.30$ and $t_{1378} = 0.71$, $p = 0.48$ for review scores and resubmissions invited, respectively). Instead, our results suggest that authors from English-speaking countries get a boost when being reviewed single-blind compared to double-blind (linear contrasts, $t_{1378} = 2.90$, $p = 0.004$ and $t_{1378} = 3.27$, $p = 0.001$ for review scores and resubmissions invited, respectively).

We noted above that when HDI is included as a covariate in the linear regression (review score) or logistic regression (resubmission invited) the main effect of author language becomes non-significant. However, we continue to detect evidence of an interaction between author language and treatment for the probability a resubmission was invited (treatment-by-language interaction: $X_1^2 = 4.58$, $p = 0.03$), suggesting that HDI alone is not enough to explain the influence of author language. The pattern of language effects is as described above. Specifically, we continue to detect a significant difference in the proportion of papers invited for resubmission between review treatments for English-speaking authors (contrast: $t = 3.28$, $p = 0.001$) but not for non-English-speaking authors ($t = 0.62$, $p = 0.54$). We also continue to detect a significant difference between English versus non-English-speaking authors in the single-blind treatment ($t = 2.03$, $p = 0.04$) but not in the double-blind treatment ($t = 0.84$, $p = 0.40$). However, the treatment-by-language interaction and all contrasts become non-significant when review score is also included in the model; only review score ($t = 351.9$, $p = 0.40$) explains the variation among papers in whether a resubmission was invited in this full model.

4 | DISCUSSION

The goals of this study were to test whether blinding of author identities influenced the outcome of the peer review process (e.g.

whether review treatment affected the ratings given to papers by reviewers and/or affected the decisions of editors), and test whether the effects of review treatment were influenced by demographic traits of an author, specifically an author's gender, language, or the economic development status (HDI) of their country of residence. Our key results are that: (1) papers reviewed with author identities blinded (double-blind review) received lower review scores and were less likely to be invited for resubmission compared to papers reviewed with the author identities known to reviewers (single-blind review); (2) this reduction in success was greater for authors in wealthier (higher HDI) and/or English-speaking countries (papers submitted by authors from high HDI and/or English-speaking countries had better outcomes when reviewed single-blind than when reviewed double-blind, whereas there was no detectable difference in outcomes between review treatments when authors were from low HDI or non-English-speaking countries); and (3) there was no evidence that the influence of double-blind review differed between male and female authors.

4.1 | Double-blind review reduces success of authors from high-income countries

Many studies have demonstrated that papers submitted by authors from low-income countries fare worse during journal peer review than do papers submitted by authors from higher-income countries. The degree to which this is due to those papers being lower quality, versus editors and reviewers changing how they evaluate papers based on an author's location, has largely been unclear, although it is often speculated, and sometimes demonstrated experimentally (Harris, Marti, et al., 2017), that biases of editors and reviewers account for at least some of the difference. At *Functional Ecology*, we find that papers from authors in lower-income countries are much more likely to be declined before being sent for review, get lower review scores and are more likely to be declined after peer review. Our results suggest that much of this difference is due to biases of reviewers. However, what's particularly striking in our results is that reviewers do not appear to exhibit a bias against authors from low-income countries, at least not at the peer review stage. Instead, reviewers appear to exhibit a bias favouring authors from high-income countries. If reviewers were exhibiting a bias against authors from low-income countries we would expect papers from low-income countries to perform worse when reviewed single-blind than when reviewed double-blind. We did not observe this; papers by authors from low-income countries performed similarly across treatments. Instead, papers from authors in high-income countries appear to get a positive benefit from having their identities (names, institutions, home countries) known to reviewers, and they lose this benefit when their identities are masked.

At least two types of biases could produce the pattern of outcomes we observed. The first is that authors may be biased towards papers by authors from their own country. *Functional Ecology* is owned by the *British Ecological Society* and published in the United Kingdom, with Senior Editors mostly located in high-income countries. Although the journal's editorial board and reviewers are substantially more international than is the Senior Editor team, both are dominated by ecologists in wealthier countries much more so than is the author population. If these researchers exhibit a bias in favour of authors from their own country or their general geographic region when reviewing papers, we would expect papers by authors in wealthier countries to be reviewed better when the author identities are known. However, we think this is unlikely to explain the results observed here; preliminary analyses by C. Fox (unpublished) of a different dataset (the data in Fox et al., 2016) suggest that bias in favour of authors from a reviewer's country of residence does exist but is larger for papers from lower-income countries.

An alternative explanation for the benefit that authors from high-income countries have when their identities are revealed is that they are benefitting from some form of prestige bias. Prestige bias comes in many forms and has many names—affiliation bias, reputation bias, institution bias and the Matthew effect—all of which describe biases in which reviewers favour papers by authors, from institutions, or from countries from which reviewers expect work to be high quality and/or significant (review in Lee et al., 2013). Reviewers may expect work from high-income countries to be of higher quality, a form of prestige bias. However, we imagine two other mechanisms by which prestige bias can generate a benefit for authors from high-income countries. One is that, because the majority of universities rated as top universities in the world are located in wealthier (and primarily western) countries (Tuesta et al., 2019), we expect that researchers globally will be more familiar with these universities and thus possibly give deference to authors from these universities when reviewing. Alternatively, deference to authors from high-income countries may be mostly from authors from high-income countries; given that the reviewer population of *Functional Ecology* is dominated by researchers from wealthier countries, they may be expected to be more familiar with prestigious institutions and researchers located in those countries and give deference to them when reviewing. Testing for prestige bias and whether it varies with reviewer location is beyond the scope of this paper but we expect to test these hypotheses in future analyses.

4.2 | Double-blind review and the success of authors of varying English-proficiency

Papers by authors residing in English-speaking countries outperformed papers by authors from non-English countries, but only when reviewed single-blind (similar in direction and effect sizes to that observed for HDI). Previous studies have found that papers submitted to journals by authors from countries with low English proficiency fare worse than those with high English proficiency (e.g. Burns & Fox, 2017; Ross et al., 2006; Saposnik et al., 2014; Tregenza, 2002; Walker et al., 2015), but these studies have been descriptive and were unable to distinguish direct effects of language from variation in manuscript quality or other factors that covary with author language among countries. English proficiency covaries with a country's economic status which could be the primary underlying factor influencing some of the relationships previously observed. However, after controlling for the effect of HDI of an author's country of residence, we still find that authors from countries where English is commonly spoken are more likely to have positive outcomes from the peer review process, compared to authors residing in countries where English is less common, but only when their identities are known to reviewers. As with the effect of HDI, this was a positive bias favouring authors from countries with higher English proficiency, rather than a negative bias against authors from countries with lower English proficiency.

That papers submitted by authors from countries with lower English proficiency perform just as well as those from English-speaking countries when reviewed double-blind was unexpected. The need to write in English is a significant disadvantage for authors who learn English as a second language (Clavero, 2011; Mudrak, 2013), especially when their primary language is more distant from English (Elder & Davies, 1998) or when they learn later in life (Qureshi, 2016). For these authors, writing is often more laborious than it is for a native English speaker (Hyland, 2016), although how much more laborious will certainly vary a lot among individual authors (Brereton & Cousins, 2022), and writing can also be a struggle for many native English-speakers. Authors writing in English as a second language tend to use simpler rhetorical structure, have smaller vocabularies (Flowerdew, 2019) and have more problems with word order, word choice and punctuation (Hyland, 2016 and references therein), although this also certainly varies substantially among individual authors. These differences in language use are often evident to reviewers, and evidence suggests that reviewers are more critical of non-standard English (Politzer-Ahles et al., 2020). Reviewers of scholarly papers also frequently note language problems in their reviews (Mungra & Webber, 2010; Flowerdew, 2019; C. Fox, personal observation). We thus expected papers by authors from non-English countries to perform more poorly compared to papers from authors in English-speaking countries even when their identities were blinded. But our data here suggest that is not the case, at least not for papers that were sent for peer review. Our personal experience is that language problems can sometimes be the primary reason a paper is not sent for review, accounting for part of the large effect of author's English proficiency on whether a paper is sent for review (Figure 3a). But our data suggest that, once a paper is sent for review, grammar and other writing problems are rarely the primary reason for rejection of the paper (Hyland, 2016; but see Andrew, 2020 and Flowerdew, 2019 for a different perspective).

4.3 | Double-blind review did not differentially affect papers by male and female authors

Most of the studies that have examined biases in scholarly publishing focus on gender. Results testing for biases in scholarly publishing vary a lot among studies, although some report that papers (and grants or abstracts) authored by women are reviewed more critically than the papers authored by men (e.g. Fox & Paine, 2019 and references therein). Surveys indicate that many scholars believe that publishing is biased against women, but the proportions that perceive this bias as substantial are generally low (Bacchelli & Beller, 2017; Rowley & Sbaffi, 2018; Taylor & Francis, 2015), although higher among women (Ho et al., 2013). This potential bias against papers authored by women is one of the biases commonly noted (among others) as a motivator for the adoption of double-blind peer review at the ecology journals that have adopted this model (e.g. Bronte, 2018; Darling, 2015; McGill et al., 2017; Nature, 2015), and some data suggest that women choose double-blind review (when there is an option) more often than do men (Bolnick, 2018), although this was not seen in another study (McGillivray & De Ranieri, 2018).

Our data show that double-blind review does not differentially affect papers authored by men and women. Papers submitted by female first authors received slightly higher ratings from reviewers compared to papers authored by male first authors, and there was a trend towards papers authored by women being more likely to be invited for resubmission, but we found no evidence that review treatment differentially affected papers by female and male authors. A previous randomized trial by Blank (1991) likewise concluded that double-blind review did not differentially affect papers authored by women and men, although sample sizes were too small to have confidence in the negative result (despite being a large study, only a small proportion of papers were by female authors). Our study is large enough, and the proportion of women high enough, that we have confidence in our conclusion that blinding of author identities does not affect gender differences in peer review outcomes.

4.4 | Implications for double-blind review

In our double-blind peer review trial we find compelling evidence that blinding author identities improves fairness during peer review. That authors from low income and/or non-English countries are aware of geographic and language biases is clear from perspectives presented in editorials and commentaries (e.g. Clavero, 2011) and surveys (Ho et al., 2013; Rowley & Sbaffi, 2018; Taylor & Francis, 2015). Authors from lower-income and non-English countries are also more likely to choose to be reviewed double-blind when given the option (McGillivray & De Ranieri, 2018). Unfortunately, our data suggest that choosing to be reviewed with your identity blinded (when binding is optional) will probably not benefit authors from low-income countries because the primary bias is not that authors from low-income countries are judged more harshly when their identities are known (relative to how they would be judged if their identities were unknown). Instead, authors from higher-income countries appear to be judged more positively when their identities are known, relative to how they would be judged when their identities are unknown. We suggest that offering optional blinding of author identities, as the *Nature* journals and *The American Naturalist* allow, is therefore unlikely to solve the actual biases that exist in peer review. Also, optional blinding may exacerbate other biases if reviewers interpret the choice to be reviewed double-blind as signalling by authors that they 'have something to hide' (Enserink, 2017). It is known that papers tend to fare poorly when authors choose to be reviewed double-blind, relative to those that choose single-blind, in journals that allow a choice (Bolnick, 2018; McGillivray & De Ranieri, 2018), but the data are not available to disentangle the specific causes of this difference. We argue that mandatory double-blind is almost certainly better than optional double-blind review; it will assuage author concerns that biases in peer review are being considered by the journal, it will at least partially address some of the known biases that have been demonstrated, and it avoids creating new biases, for example, against authors who signal a preference for double-blind.

One of the main arguments against double-blind review is that authors are often easily de-anonymized. Indeed, numerous studies demonstrate this, with between a low of 13% (O'Connor et al., 2016) to a high of 52% (Fisher et al., 1994) of reviewers speculating that they knew the identity of the authors (see also Isenberg et al., 2009; Jagsi et al., 2014; and references therein). At *Functional Ecology* we surveyed reviewers in the double-blind treatment (after peer review was completed) and asked if they could identify the authors; if they answered that they could, we asked them to identify the authors and compared their answers with the author list. The ability of reviewers to unmask blinded authors at *Functional Ecology* is on the high end of those reported for other studies; roughly half (53% in the first 2 years of the trial) were able to identify the authors, based on (from the most-common to least-common reason given) the research location, research topic, research organism, methods used and/or manuscript referencing (C. Fox and J. Meyer, unpublished data). Clearly, blinding of manuscripts remains a challenge to be solved for journals that adopt double-blind review. But it is notable that, despite this difficulty blinding manuscripts, we observed a large influence of review treatment on the influence of author demographic traits on peer review that should not be ignored.

AUTHOR CONTRIBUTIONS

Charles W. Fox conceived and designed the experiment, with critical input from Jennifer Meyer and Emilie Aimé. Charles W. Fox and Emilie Aimé wrote the proposal to obtain approval and funding from the British Ecological Society. Jennifer Meyer and Emilie Aimé implemented the workflows in *ScholarOne Manuscripts* and managed the day-to-day flow of manuscripts through peer review. Charles W. Fox analysed the data and wrote the manuscript. All authors discussed the results, contributed critically to the manuscript and gave final approval for publication.

ACKNOWLEDGEMENTS

We are immensely grateful to the British Ecological Society (BES), and especially the members of the Publications Committee of the BES, for permitting us to run this experiment and providing the funding necessary to support the editorial staff to execute it. We also thank the many members of the BES' journal editorial office in London for their help managing the journal during this experiment, especially since this work was in addition to the many disruptions caused by the covid pandemic. The editorial office is very collaborative across journals and we cannot know everyone who touched this project in some way, but a few people have been particularly impactful, including Andrea Baier, Catherine Hill, Rowena Gordon and Frank Harris. We also thank the many editors of *Functional Ecology* for being participants in this trial; these include the Senior Editors during this trial (Lara Ferry, Katie Field, Alan Knapp, Enrico Rezende, Emma Sayer and Ken Thompson) and the more than 100 Associate Editors that handled papers for the journal during this experiment. Savannah Piper and Josiah Ritchey helped with various tasks behind the scenes at the University of Kentucky. William Bryan, Ruth Bryan and Enrico Rezende translated our abstract to Spanish, and Huijie Qiao translated our abstract to Chinese. Ruth Bryan, Jim des Lauriers, Lara Ferry, Allyssa Kilanowski, Bob O'Hara, Savannah Piper, Ian Thornhill and Tom Tregenza provided comments on earlier versions of this manuscript.

FUNDING INFORMATION

This project was funded by the British Ecological Society.

CONFLICTS OF INTEREST

Charles Fox was Executive Editor of *Functional Ecology* when this experiment was initiated and was Senior Editor of the journal during the entire experiment. Emilie Aimé was the Managing Editor of *Functional Ecology* when this experiment was initiated. Jenny Meyers was Assistant Editor of *Functional Ecology* when this experiment was initiated and is currently Managing Editor of the journal.

DATA AVAILABILITY STATEMENT

An anonymized version of the dataset is available from the Dryad Digital Repository https://doi.org/10.5061/dryad.m63xsj466 (Fox et al., 2023).

ORCID

Charles W. Fox https://orcid.org/0000-0002-7545-7967
Jennifer Meyer https://orcid.org/0000-0002-0366-6348
Emilie Aimé https://orcid.org/0000-0003-2290-584X

REFERENCES

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123–129. https://doi.org/10.1016/S0165-1765(03)00032-6

Alam, M., Kim, N. A., Havey, J., Rademaker, A., Ratner, D., Tregre, B., West, D. P., & Coleman, W. P., III. (2011). Blinded vs. unblinded peer review of manuscripts submitted to a dermatology journal: A randomized multi-rater study. *British Journal of Dermatology*, *165*(3), 563–567. https://doi.org/10.1111/j.1365-2133.2011.10432.x

Andrew, N. R. (2020). Design flaws and poor language: Two key reasons why manuscripts get rejected from austral ecology across all countries between 2017 and 2020. *Austral Ecology*, *45*, 505–509. https://doi.org/10.1111/aec.12908

Bacchelli, A., & Beller, M. (2017). Double-blind review in software engineering venues: The community's perspective. In *2017 IEEE/*

ACM 39th international conference on software engineering companion (ICSE-C) (pp. 385–396). IEEE. https://doi.org/10.1109/ICSE-C.2017.49
Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the American economic review. *The American Economic Review*, *81*, 1041–1067. https://www.jstor.org/stable/2006906
Bolnick, D. I. (2018). Abiding in the midst of ignorance. *Blog post*. https://ecoevoevoeco.blogspot.com/2018/12/abiding-in-midst-of-ignorance.html
Brereton, P., & Cousins, E. Y. (2022). Practitioners respond to John Flowerdew's 'the linguistic disadvantage of scholars who write in English as an additional language: Myth or reality'. *Language Teaching*, *55*(1), 145–148. https://doi.org/10.1017/S0261444821000033
Bronte, D. (2018). *Oikos goes double blind*. *Oikos blog*. https://www.oikosjournal.org/blog/oikos-goes-double-blind
Burns, C. S., & Fox, C. W. (2017). Language and socioeconomics predict geographic variation in peer review outcomes at an ecology journal. *Scientometrics*, *113*(2), 1113–1127. https://doi.org/10.1007/s11192-017-2517-5
Clavero, M. (2011). Unfortunately, linguistic injustice matters. *Trends in Ecology & Evolution*, *26*, 146–157. https://doi.org/10.1016/j.tree.2011.01.011
Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16
Darling, E. S. (2015). Use of double-blind peer review to increase author diversity. *Conservation Biology*, *29*(1), 297–299. https://www.jstor.org/stable/24481603
Deb, S. (2015). *Gap between GDP and HDI: Are the rich country experiences different from the poor?* In *Proceedings of the IARIW-OECD special conference*, Paris, France, 16–17 August 2015. http://old.iariw.org/papers/2015/deb.pdf
Demarest, B., Freeman, G., & Sugimoto, C. R. (2014). The reviewer in the mirror: Examining gendered and ethnicized notions of reciprocity in peer review. *Scientometrics*, *101*(1), 717–735. https://doi.org/10.1007/s11192-014-1354-z
Elder, C., & Davies, A. (1998). Performance on ESL examinations: Is there a language distance effect? *Language and Education*, *12*(1), 1–17. https://doi.org/10.1007/s11192-014-1354-z
Enserink, M. (2017). *Few authors choose anonymous peer review, massive study of nature journals shows*. *Science blog*. https://doi.org/10.1126/science.aaq0322
Fisher, M., Friedman, S. B., & Strauss, B. (1994). The effects of blinding on acceptance of research papers by peer review. *JAMA*, *272*(2), 143–146. https://doi.org/10.1001/jama.1994.03520020069019
Flowerdew, J. (2019). The linguistic disadvantage of scholars who write in English as an additional language: Myth or reality. *Language Teaching*, *52*(2), 249–260. https://doi.org/10.1017/S0261444819000041
Fox, C. (2019). Single vs double-blind peer review: An experiment. *Functional Ecologists: A Blog for the People Behind the Research*. https://functionalecologists.com/2019/09/02/2295/
Fox, C. W. (2021). Which peer reviewers voluntarily reveal their identity to authors? Insights into the consequences of open-identities peer review. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1961), 20211399. https://doi.org/10.1098/rspb.2021.1399
Fox, C. W., Burns, C. S., & Meyer, J. A. (2016). Editor and reviewer gender influence the peer review process but not peer review outcomes at an ecology journal. *Functional Ecology*, *30*(1), 140–153. https://doi.org/10.1111/1365-2435.12529
Fox, C. W., Meyer, J. A., & Aimé, E. (2023). Data from: Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.m63xsj466
Fox, C. W., & Paine, C. T. (2019). Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution. *Ecology and Evolution*, *9*(6), 3599–3619. https://doi.org/10.1002/ece3.4993
Fox, C. W., Ritchey, J. P., & Paine, C. T. (2018). Patterns of authorship in ecology and evolution: First, last, and corresponding authorship vary with gender and geography. *Ecology and Evolution*, *8*(23), 11492–11507. https://doi.org/10.1002/ece3.4584
Fox, C. W., Thompson, K., Knapp, A., Ferry, L. A., Rezende, E. L., Aimé, E., & Meyer, J. (2019). Double-blind peer review–an experiment. *Functional Ecology*, *33*(1), 4–6. https://doi.org/10.1111/1365-2435.13269
Godlee, F., Gale, C. R., & Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: A randomized controlled trial. *JAMA*, *280*(3), 237–240. https://doi.org/10.1001/jama.280.3.237
Harris, M., Macinko, J., Jimenez, G., & Mullachery, P. (2017). Measuring the bias against low-income country research: An implicit association test. *Globalization and Health*, *13*(1), 1–9. https://doi.org/10.1186/s12992-017-0304-y
Harris, M., Marti, J., Watt, H., Bhatti, Y., Macinko, J., & Darzi, A. W. (2017). Explicit bias toward high-income-country research: A randomized, blinded, crossover experiment of English clinicians. *Health Affairs*, *36*(11), 1997–2004. https://doi.org/10.1377/hlthaff.2017.0773
Ho, R. C. M., Mak, K. K., Tao, R., Lu, Y., Day, J. R., & Pan, F. (2013). Views on the peer review system of biomedical journals: An online survey of academics from high-ranking universities. *BMC Medical Research Methodology*, *13*(1), 1–15. https://doi.org/10.1186/1471-2288-13-74
Huber, J., Inoua, S., Kerschbamer, R., König-Kersting, C., Palan, S., & Smith, V. L. (2022). Nobel and novice: Author prominence affects peer review. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(41), e2205779119. https://doi.org/10.1073/pnas.2205779119
Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58–69. https://doi.org/10.1016/j.jslw.2016.01.005
iBT Tests. (2020). TOEFL iBT: Test and score data summary 2020. https://www.ets.org/toefl
Isenberg, S. J., Sanchez, E., & Zafran, K. C. (2009). The effect of masking manuscripts for the peer-review process of an ophthalmic journal. *British Journal of Ophthalmology*, *93*(7), 881–884. https://doi.org/10.1136/bjo.2008.151886
Jagsi, R., Bennett, K. E., Griffith, K. A., DeCastro, R., Grace, C., Holliday, E., & Zietman, A. L. (2014). Attitudes toward blinding of peer review and perceptions of efficacy within a small biomedical specialty. International journal of radiation oncology* biology*. *Physics*, *89*(5), 940–946. https://doi.org/10.1016/j.ijrobp.2014.04.021
Jubb, M. (2016). Peer review: The current landscape and future trends. *Learned Publishing*, *29*(1), 13–21. https://doi.org/10.1002/leap.1008
Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A., Rennie, D., & Investigators, P. (1998). Does masking author identity improve peer review quality?: A randomized controlled trial. *JAMA*, *280*(3), 240–242. https://doi.org/10.1001/jama.280.3.240
Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M., & Strohmaier, M. (2016). Inferring gender from names on the web: A comparative evaluation of gender detection methods. In *Proceedings of the 25th international conference companion on world wide web* (pp. 53–54). https://doi.org/10.1145/2872518.2889385
Kowal, M., Sorokowski, P., Kulczycki, E., & Żelaźniewicz, A. (2022). The impact of geographical bias when judging scientific studies. *Scientometrics*, *127*(1), 265–273. https://doi.org/10.1007/s11192-021-04176-7
Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*(1), 2–17. https://doi.org/10.1002/asi.22784

Lerback, J., & Hanson, B. (2017). Journals invite too few women to referee. *Nature*, *541*(7638), 455–457. https://doi.org/10.1038/541455a

McGill, B. J., Dornelas, M., & Field, R. (2017). GEB goes double blind. *Global Ecology and Biogeography*, *26*(11), 1223–1224. https://doi.org/10.1111/geb.12674

McGillivray, B., & De Ranieri, E. (2018). Uptake and outcome of manuscripts in nature journals by review model and author characteristics. *Research Integrity and Peer Review*, *3*(1), 1–12. https://doi.org/10.1186/s41073-018-0049-z

McNutt, R. A., Evans, A. T., Fletcher, R. H., & Fletcher, S. W. (1990). The effects of blinding on the quality of peer review: A randomized trial. *JAMA*, *263*(10), 1371–1376. https://doi.org/10.1001/jama.1990.03440100079012

Mudrak, B. (2013). Understanding the needs of international authors. *Learned Publishing*, *26*(2), 139–147. https://doi.org/10.1087/20130212

Mulligan, A., Hall, L., & Raphael, E. (2013). Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology*, *64*(1), 132–161. https://doi.org/10.1002/asi.22798

Mungra, P., & Webber, P. (2010). Peer review process in medical research publications: Language and content comments. *English for Specific Purposes*, *29*(1), 43–53. https://doi.org/10.1016/j.esp.2009.07.002

Murray, D., Siler, K., Lariviére, V., Chan, W. M., Collings, A. M., Raymond, J., & Sugimoto, C. R. (2018). Gender and international diversity improves equity in peer review. *BioRxiv*, 400515. https://doi.org/10.1101/400515

Nakamura, R. K., Mann, L. S., Lindner, M. D., Braithwaite, J., Chen, M. C., Vancea, A., Byrnes, N., Durrant, V., & Reed, B. (2021). An experimental test of the effects of redacting grant applicant identifiers on peer review outcomes. *eLife*, *10*, e71368. https://doi.org/10.7554/eLife.71368

Nature. (2015). Nature journals offer double-blind review. *Nature*, *518*, 274. https://doi.org/10.1038/518274b

Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal*, *4*(2), 154–167. https://doi.org/10.1177/1536867X0400400206

O'Connor, E. E., Cousar, M., Lentini, J. A., Castillo, M., Halm, K., & Zeffiro, T. A. (2016). Efficacy of double-blind peer review in an imaging subspecialty journal. *American Journal of Neuroradiology*, *38*, 230–235. https://doi.org/10.3174/ajnr.A5017

Okike, K., Hug, K. T., Kocher, M. S., & Leopold, S. S. (2016). Single-blind vs double-blind peer review in the setting of author prestige. *JAMA*, *316*(12), 1315–1316. https://doi.org/10.1001/jama.2016.11014

Pleskac, T., Kyung, E., Chapman, G., & Urminsky, O. (2021). Blinded versus unblinded review: A field study comparing the equity and fairness of review processes. *PsyArXiv*. https://doi.org/10.31234/osf.io/q2tkw

Politzer-Ahles, S., Girolamo, T., & Ghali, S. (2020). Preliminary evidence of linguistic bias in academic reviewing. *Journal of English for Academic Purposes*, *47*, 100895. https://doi.org/10.1016/j.jeap.2020.100895

Qureshi, M. A. (2016). A meta-analysis: Age and second language grammar acquisition. *System*, *60*, 147–160. https://doi.org/10.1016/j.system.2016.06.001

Regehr, G., & Bordage, G. (2006). To blind or not to blind? What authors and reviewers prefer. *Medical Education*, *40*(9), 832–839. https://doi.org/10.1111/j.1365-2929.2006.02539.x

Ross, J. S., Gross, C. P., Desai, M. M., Hong, Y., Grant, A. O., Daniels, S. R., Hachinski, V. C., Gibbons, R. J., Gardner, H. K., & Krumholz, H. M. (2006). Effect of blinded peer review on abstract acceptance. *JAMA*, *295*(14), 1675–1680. https://doi.org/10.1001/jama.295.14.1675

Rowley, J., & Sbaffi, L. (2018). Academics' attitudes towards peer review in scholarly journals and the effect of role and discipline. *Journal of Information Science*, *44*(5), 644–657. https://doi.org/10.1177/0165551517740821

Saposnik, G., Ovbiagele, B., Raptis, S., Fisher, M., & Johnston, S. C. (2014). Effect of English proficiency and research funding on acceptance of submitted articles to stroke journal. *Stroke*, *45*(6), 1862–1868. https://doi.org/10.1161/STROKEAHA.114.005413

SAS Institute Inc. (2016). *SAS/SHARE 9.4: User's guide* (2nd ed.). SAS Institute Inc.

Seeber, M., & Bacchelli, A. (2017). Does single blind peer review hinder newcomers? *Scientometrics*, *113*(1), 567–585. https://doi.org/10.1007/s11192-017-2264-7

Shoham, N., & Pitman, A. (2021). Open versus blind peer review: Is anonymity better than transparency? *BJPsych Advances*, *27*(4), 247–254. https://doi.org/10.1192/bja.2020.61

Squazzoni, F., Bravo, G., Farjam, M., Marusic, A., Mehmani, B., Willis, M., Birukou, A., Dondio, P., & Grimaldo, F. (2021). Peer review and gender bias: A study on 145 scholarly journals. *Science Advances*, *7*(2), eabd0299. https://doi.org/10.1126/sciadv.abd0299

Szigeti, C., Tóth, G., Borzán, A., & Farkas, S. (2013). GDP alternatives and their correlations. *Journal of Environmental Sustainability*, *3*(3), 3–12. https://doi.org/10.14448/jes.03.0002

Taylor & Francis. (2015). *Peer review in 2015: A global view*. Taylor & Francis Group.

Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12708–12713. https://doi.org/10.1073/pnas.1707323114

Tregenza, T. (2002). Gender bias in the refereeing process? *Trends in Ecology & Evolution*, *17*(8), 349–350. https://doi.org/10.1016/S0169-5347(02)02545-4

Tuesta, E. F., Garcia-Zorita, C., Ayllon, R. R., & Sanz-Casado, E. (2019). Does a country/region's economic status affect its universities' presence in international rankings? *Journal of Data and Information Science*, *4*(2), 56–78. https://doi.org/10.2478/jdis-2019-0009

United Nations Development Programme. (2020). *Human development report 2020. The next frontier: Human development and the Anthropocene*. https://hdr.undp.org/content/human-development-report-2020

van Rooyen, S., Godlee, F., Evans, S., Smith, R., & Black, N. (1998). Effect of blinding and unmasking on the quality of peer review. *Journal of General Internal Medicine*, *14*(10), 622–624. https://doi.org/10.1001/jama.280.3.234

Wais, K. (2016). Gender prediction methods based on first names with genderizeR. *The R Journal*, *8*(1), 17–37. https://doi.org/10.32614/RJ-2016-002

Walker, R., Barros, B., Conejo, R., Neumann, K., & Telefont, M. (2015). Personal attributes of authors and reviewers, social bias and the outcomes of peer review: A case study. *F1000Research*, *4*. https://doi.org/10.12688/f1000research.6012.2

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Additional analyses and the guidelines provided to authors submitting to *Functional Ecology*.

How to cite this article: Fox, C. W., Meyer, J., & Aimé, E. (2023). Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Functional Ecology*, *37*, 1144–1157. https://doi.org/10.1111/1365-2435.14259