To Name or Not to Name: The Effect of Changing Author Gender on Peer Review

ROBYN M. BORSUK, LONNIE W. AARSSEN, AMBER E. BUDDEN, JULIA KORICHEVA, ROOSA LEIMU, TOM TREGENZA, AND CHRISTOPHER J. LORTIE

The peer review model is one of the most important tools used in science to assess the relative merit of research. We manipulated a published article to reflect one of the following four author designations: female, male, initial, and no name provided. This article was then reviewed by referees of both genders at various stages of scientific training. Name changing did not influence acceptance rates or quality ratings. Undergraduate referees were less critical than graduate students or postdoctoral researchers, independent of gender. However, female postdoctoral researchers were the most critical referees: Their rejection rates were the highest and quality ratings the lowest, regardless of the author name provided. Contrary to previous reports in the literature, there was no evidence of same-gender preferences. This study strongly suggests that female postgraduate biologists may apply different expectations to peer review.

Keywords: bias, graduate students, gender, name changing, peer review

The peer-review process is dependent on the largely unpaid participation of editors, referees, and authors in order to function. At any point in this process, however, it is possible that the assessment of a manuscript will be based on something other than actual scientific merit (Lortie et al. 2007). Errors in assessment, or "bias," can occur in relation to attributes of the referee or to attributes of the paper. In the first instance, scientists with more experience tend to be more critical in their evaluations (Nylenna et al. 1994) and have been shown within the medical field to produce lower-quality reviews (Kliewer et al. 2005). Male referees also tend to recommend either outright acceptance or rejection, whereas females more frequently recommend revisions (Davo et al. 2003). Second, attributes associated with a paper, such as the number of authors or authors' nationality, can influence the relative assessment of merit (Lloyd 1990, Tregenza 2002, Leimu and Koricheva 2005a, 2005b). Research has also shown the gender of authors to be an important factor in assessments of quality. For example, articles with females as first author were more likely to receive a lower rating, particularly when the author was not previously known in their field (Paludi and Bauer 1983, Lloyd 1990). Recent research does suggest a paradigm shift in attitudes toward author gender (Leimu and Koricheva 2005b, Borsuk et al. 2009), although the extent of such shifts has not been tested experimentally. At the very least, the sensitivity of author name and gender within the peer-review model should be tested to ensure that the general assumptions about the model's objectivity are well founded.

An important issue facing scientists is the potential for the gender of authors to inadvertently affect how a study is perceived (Schubert and Sinha 2004). Importantly, both editor and referee populations are predominantly male (Dickersin et al. 1998, Davo et al. 2003, Dalton 2006). Additionally, referee performance varies by gender, with males generally taking more time to return manuscripts (Davo et al. 2003). Studies have shown male and female referees in medicine to be partial toward authors of their own gender (Davo et al. 2003, Kliewer et al. 2005). These demographic differences could generate differences in the treatment of manuscripts under review.

In this manipulative experiment, we offered a single ecological article to a broad population of potential referees, including biology undergraduates, graduate students, postdoctoral researchers, and faculty researchers. The only variation in the article concerned the name of the author: In some instances we supplied a male name, in others a female name, and in yet others we gave only initials or no name at all. We tested the following hypotheses using this design: (a)

BioScience 59: 985–989. ISSN 0006-3568, electronic ISSN 1525-3244. doi:10.1525/bio.2009.59.11.10

author name: All else being equal (it is the exact same paper with only the name changed), author designation should not influence The recommendation to publish (null H_0); (b) referee attributes: The gender and education of the referee influence the degree of criticality when reviewing (Nylenna et al. 1994, Davo et al. 2003, Kliewer et al. 2005); (c) same gender preferences: Referees are more likely to accept papers written by authors of their own gender (Tregenza 2002). To the best of our knowledge, this is the first study to explore whether author name changes the perception of a biological study.

Sampling design

We used the publication article "Zebra Mussels Decrease Burrowing Ability and Growth of a Native Snail, *Campeloma decisum*" (published in *Hydrobiologia*; Van Appledorn et al. 2007) to test whether author name affects reviews. We selected this particular article because of its general scientific characteristics, including its short length (three pages), clear figures, and direct and uncomplicated text, as well as its design as an experimental manipulation. Importantly, this article is a representative publication for *Hydrobiologia,* having received the same number of citations as the mean for that journal in 2007 (ISI Web of Science, mean = two citations for articles in 2007). We thus rated this paper as acceptable and clearly publishable, and as such, an appropriate candidate for manipulation.

Names of the original authors were removed and replaced by four author designations: no name, initial (J. Thompson), male (David Thompson), and female (Catherine Thompson). The surname was identical for each author treatment to ensure that any potential differences in manuscript evaluation could be attributed only to author gender. We formatted the paper as a manuscript and distributed it in March 2007 to a survey population including undergraduate students (naive, with no reviewing experience), graduate students (master's students and PhD candidates), postdoctoral fellows, and faculty. To poll the undergraduates, four biology classes totaling 269 students from York University, Canada, were selected and tested in class. We also posted the manuscript and questionnaire online (*www.yorku.ca/surveys*) to expand this survey population and to potentially include graduate students, postdoctoral researchers, and more senior scientists. We sent direct e-mails to the faculty of three Canadian and three British universities to advertise the survey, and also posted the requests on two listservs (ECOLOG and EvolDir). Assignment to the four manuscript versions was randomized using an online algorithm, and the survey was posted for eight months. All respondents were informed that the researchers were interested in exploring whether knowledge of a topic influences assessment, and were not informed of the changes in author name. We devised a seven-question survey using a 5-point Likert scale (figure 1). The name, gender, age, and career stage of each respondent was also recorded. The York University Human Protocol Research Committee approved this research.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. I have prior knowledge on the subject of this article	❍	❍	❍	❍	❍
2. This article is suitable for publication	❍	❍	❍	❍	❍
3. The title captured my attention	❍	❍	❍	❍	❍
4. The abstract is useful	❍	❍	❍	❍	❍
5. The graphs are useful, appropriate and clear	❍	❍	❍	❍	❍
6. The concepts this article is trying to convey understandable	❍	❍	❍	❍	❍
7. This topic merits scientific research	❍	❍	❍	❍	❍

Figure 1. The Likert scale used in this study.

Statistical analyses

Prior to analysis, we removed respondents who did not indicate gender or education. As there were virtually no faculty respondents ($n = 2$), this group was also removed. With exclusions and incompletes, the final data set totaled 989 respondents—230 in class and 759 online. For all subsequent analyses reported herein, master's and PhD candidates were grouped together as "graduate students" since preliminary analyses did not detect significant differences between these two groups. We tested for differences arising from the mechanism of survey presentation by analyzing undergraduate data from in-class and online surveys. All ratings were converted to an ordinal numerical score, with 1 being "strongly disagree" and 5 being "strongly agree."

We used chi-square (χ^2) statistics to test whether the respondent population varied by gender. As a first approximation, we used a 4×2 contingency analysis to test whether name designation (male, female, initials, or none) influenced recommendations (accept or reject, with data on the statement "The article is suitable for publication" converted to binary). We also tested the decision to publish directly using the full data reported in the statement on suitability for publication (figure 1). We calculated a net "quality" score by summing the five response variables that directly addressed manuscript quality (figure 1, statements 2–6). This approach allowed us to assess whether the results were sensitive to the type of question, and whether the respondents were consistent (quality and recommendation should coincide).

We used a general linear model (GLM) to analyze net quality scores with education, gender of respondent, and author assignment as independent factors. An additional

GLM with the same factors was used to analyze the dependent variable "decision to publish." We used independent sample *t* tests to test for a difference between author gender and respondent gender for both acceptance rates and quality scores. In all cases, alpha was set at 0.05, tests were two-tailed, and data met parametric assumptions. Tukey and students' *t* post-hoc tests were used to compare specific groups where appropriate. We used SPSS version 16.0.1 and JMP version 5.1.1 for all analyses.

Results

There were a total of 1031 respondents both online and in-class, with 269 biology undergraduate students completing the in-class experiment and 762 individuals completing the online version. The respondent population was entirely composed of biologists. The author-designated test manuscripts were equally distributed to each group ($\chi^2_1 = 1.751$, $p = 0.626$). There were significantly more female (62 percent) than male respondents (38 percent) ($\chi^2_1 = 58.490$, $p < 0.001$), and there was no gender difference by education ($\chi^2_2 = 0.579$, $p = 0.748$). There were significantly more undergraduate online respondents than any other group ($\chi^2_3 = 1667.874$, $p < 0.001$; postdoctoral researchers 3.3 percent, graduate students 10.3 percent, undergraduate students 86.4 percent). Undergraduate respondents who completed the in-class version of the survey were more likely than their online counterparts to reject the manuscript ($F_{1988} = 283.323$, $p < 0.001$) and give it a lower quality rating ($F_{1985} = 914.321$, $p < 0.001$).

Author gender had no effects on rejection rates for online referees ($\chi^2_6 = 3.195$, $p = 0.784$; male 25.44 percent, female 26.75 percent, initial 24.56 percent, no name 23.25 percent; figure 2). There were no author gender differences in rejection rate ($F_{1225} = 1.773$, $p = 0.184$) or quality rating ($F_{1222} = 0.5607$, $p = 0.455$) for the in-class undergraduate respondents. Referees were not more likely to accept authors of their own gender (female, $t_{267} = -1.088$, $p = 0.278$; male, $t_{119} = 1.508$, $p = 0.134$; figure 3), nor did they rate their own gender more favorably (female, $t_{232} = -0.775$, $p = 0.439$; male, $t_{154} = 0.647$, $p = 0.519$).

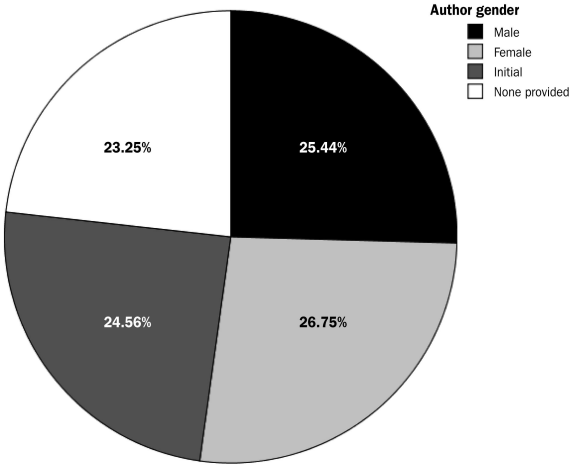

Figure 2. The relative distribution of acceptance rates of an experimentally manipulated author gender for a single manuscript. A total of 456 biologists were tested.

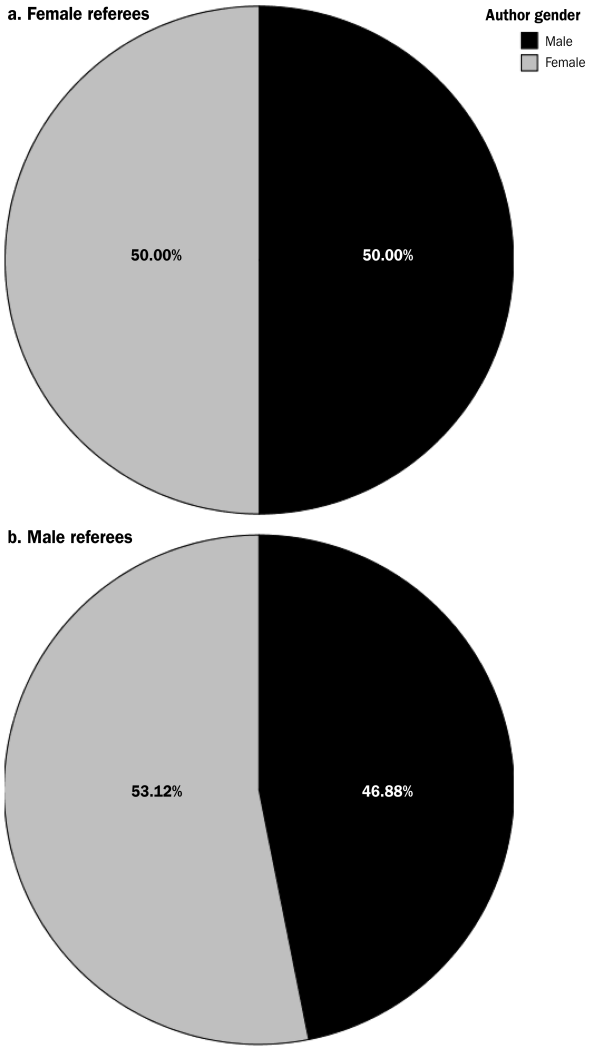

Figure 3. A comparison of the acceptance rate by referee gender by author gender of a single manuscript experimentally manipulated for author name. A total of 238 biologists were tested.

Postdoctoral researchers and graduate students were, however, more likely to reject the manuscript than undergraduate students ($F_{2750} = 15.340$, $p < 0.001$; figure 4). Importantly, female respondents generally rated manuscripts significantly lower than did male respondents ($F_{1750} = 6.732$, $p = 0.010$), but this difference was driven primarily by female postdoctoral researchers, who were more critical than any other group ($F_{2750} = 4.290$, $p = 0.014$; figure 5).

Discussion

Peer review is generally assumed to be a fair and objective method for disseminating credible biological research. Nonetheless, simple differences between referees in experience, familiarity with authors, or gender of either the author or the referee can introduce other considerations into the relative

assignment of merit—not necessarily mistakes, but rather tendencies. These tendencies can nonetheless be important to a potential author with a manuscript in review at a journal if the decision is borderline or if other manuscripts of comparable merit are under consideration. In this experiment, we tested the hypotheses that (a) author gender does not influence assessment, (b) different tendencies exist between groups of referees, and (c) same-gender preferences influence assessment. In the first instance, name changing did not affect the recommendation to publish in any of the groups tested, which suggests that referees at all stages of experience in biology detect merit. As predicted in the second instance, attributes of the referee pool significantly influenced the relative assessment of the same manuscript—that is, graduate students and postdoctoral researchers were more critical. However, female referees also tended to be more critical in general, and interestingly, female postdoctoral researchers were the most critical group tested. Finally, no same-gender preferences were detected in reviewing. Admittedly, this is the first experimental study of its kind and it has been tested on only one publication; nonetheless, there was no evidence of gender discrimination by author name, which is an extremely positive finding for biologists. The different tendencies in the likelihood of rejection by female referees does suggest, however, that not all individuals approach peer review using the same set of criteria.

Scientific merit identified a priori by each journal as a set of specific and unique criteria should be used in evaluating manuscripts; in principle, this concept is simple, but in application it can be more difficult. For example, the actual merit of a particular study described in a manuscript can be influenced by the referee's perception of numerous elements of the paper, including communication style, clarity, quality of figures, and so on. It is reasonable to suggest that these elements do not necessarily relate to the paper's scientific merit, but they can nonetheless reduce the referee's capacity to detect merit (Lortie et al. 2007). Less obviously, however, articles with only initials reported can be rated more highly than those authored by females (Paludi and Bauer 1983), and sociological experiments using manipulated, gender-specific, or unique names have detected biases in relative assessment (Anderson and Schmitt 1990, Kasof 1993). We did not detect such tendencies. We hope this preliminary study reflects positive changes within the biological field in tandem with the overall increase in the number of female biologists. Most important, this study does not imply that the peer-review process is perfect, nor that all published papers within a particular journal are of equally high quality. Rather, it demonstrates that potential referees at all levels of training are, on average, assessing the manuscript and not the name of the author. The differences between male and female referees, however, could be very important, and future research should

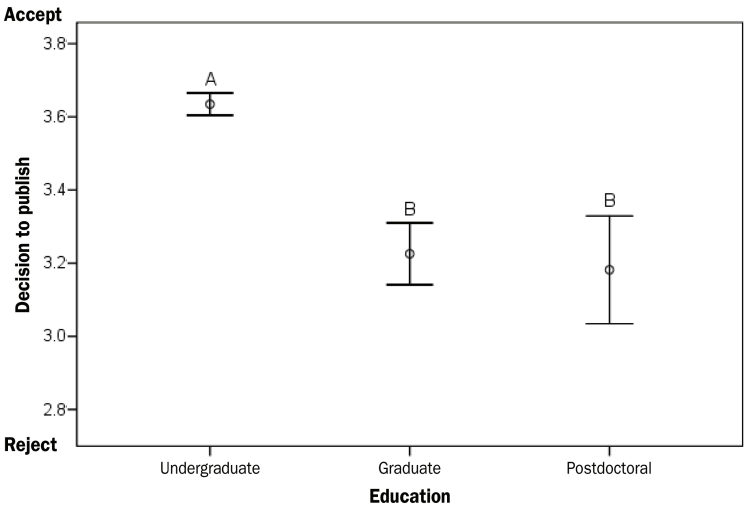

Figure 4. The mean score of "decision to publish" recommended by referees of a single manuscript. We tested 759 biologists, including undergraduates, graduates, and postdocs on a Likert scale from 1 to 5 (1 = negative and 5 = positive); however, for purposes of visual clarity, "decision to publish" was inverted to show responses from negative to positive (i.e., low scores = reject, high scores = accept). Shown is the mean ±1 standard error. Different letters denote significantly different Tukey post-hoc contrasts ($p < 0.05$*).*

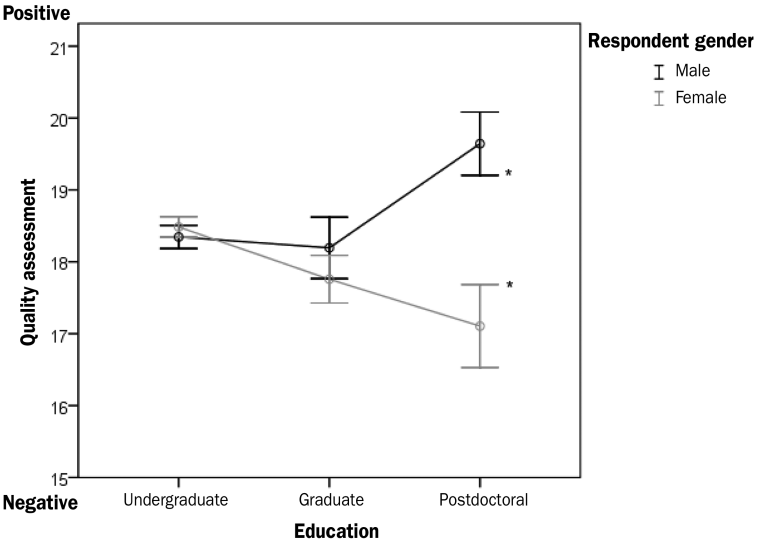

Figure 5. The mean "quality rating" score provided by referees of a single manuscript. We tested 759 references, and included males and females at one of three career stages. A Likert scale from 1 to 5 was used, with 1 being negative and 5 being positive; however, for purposes of visual clarity, scores were inverted to show responses from negative to positive (i.e., low scores = reject, high scores = accept). Shown is the mean ±1 standard error. Asterisks denote significantly different Student's t *post-hoc analyses (*$p < 0.05$*).*

focus on identifying why differences may exist. It would be easy to speculate that female biologists are more critical because they were subjected to more critical experiences in science, but this is conjecture and may be incorrect in the contemporary culture. More extensive tests of these ideas should examine whether provision of criteria before review is important for standardizing reviews, and whether faculty researchers also differ by gender.

Postdoctoral researchers and graduate students were more likely than undergraduates to reject the manuscript. This finding supports related research in medicine that found experienced referees to be more critical (Nylenna et al. 1994), and senior referees to be more likely to rank manuscripts as less important (Kliewer et al. 2005). We propose that a comparative experiment involving faculty researchers would determine whether these differences persist in biology. However, there was an alarming lack of interest or a nonresponse bias from the senior researchers we contacted through the two largest listservs in the disciplines of ecology and evolution. There are several possible explanations for reviews becoming increasingly critical with greater experience, such as more junior scientists are less familiar with the expectations associated with successful publication, or, less positively, referees generally become harder to impress as they become older and more experienced (Kliewer et al. 2004), or that elements associated with "doing science" increases our capacity to critically review the research of our peers. In all instances, it is evident that we should view this variation as a positive opportunity to enhance the peer-review system, and introduce formal training early in the education and research pipeline.

Gender issues in science are complicated by many factors changing concurrently. Nonetheless, we propose that considerations of gender and peer review are crucial because the assessment of scientists relies heavily on their publication records (Barres 2006). Since experienced referees have been shown to rate articles written by females less favorably than those written by males (Lloyd 1990), and since females frequently do not attain the same level of seniority as males (Dickersin et al. 1998, Knapp 2005), these conditions can set the stage for males and females to approach the peer-review system differently. We found no evidence for differential rating on the basis of author gender, nor did we find evidence of same gender preferences; in both instances, our findings contradict previous studies (Paludi and Bauer 1983, Lloyd 1990). Although this is encouraging, it is still only a first step toward understanding the importance of author gender—particularly since we found that female referees were more critical. We propose the novel (yet perhaps obvious to editors) explanation that the gender and career stage of the available referee population are important, and due consideration of that fact is needed to ensure fair reviews. Moreover, referee training for biologists, explicit and transparent protocols for selection of referees, and journal-specific criteria for referees and authors could improve the process of peer review.

Acknowledgments

This research was conducted as part of the Role of Publication-related Biases in Ecology Working Group supported by the National Center for Ecological Analysis and Synthesis, a center funded by the National Science Foundation (grant no. DEB-0072909).

References cited

Anderson T, Schmitt PR. 1990. Unique first names in male and female psychiatric inpatients. Journal of Social Psychology 130: 835–837.

Barres BA. 2006. Does gender matter? Nature 442: 133–136.

Borsuk RM, Budden AE, Leimu R, Aarson LW, Lortie CJ. 2009. The influence of author gender, national language and number of authors on citation rate in ecology. Open Ecology 2: 25–28.

Dalton R. 2006. Societies spurn women editors. Nature 440: 974–975.

Davo MD, Vives C, Alvarez-Dardet C. 2003. Why are women underused in the JECH peer review process? Journal of Epidemiology and Community Health 57: 936–937.

Dickersin K, Fredman L, Flegal KM, Scott JD, Crawley B. 1998. Is there a sex bias in choosing editors? Epidemiology journals as an example. Journal of the American Medical Association 280: 260–264.

Kasof J. 1993. Sex bias in the naming of stimulus persons. Psychological Bulletin 113: 140–163.

Kliewer MA, DeLong DM, Freed K, Jenkins CB, Paulson EK, Provenzale JM. 2004. Peer review at the *American Journal of Roentgenology*: How reviewer and manuscript characteristics affected editorial decisions on 196 major papers. American Journal of Roentgenology 183: 1545–1550.

Kliewer MA, Freed KS, DeLong DM, Pickhardt PJ, Provenzale JM. 2005. Reviewing the reviewers: Comparison of review quality and reviewer characteristics at the *American Journal of Roentgenology*. American Journal of Roentgenology 184: 1731–1735.

Knapp S. 2005. A suitable job for a woman. Trends in Ecology and Evolution 20: 55–56.

Leimu R, Koricheva J. 2005a. Does scientific collaboration increase the impact of ecological articles? BioScience 55: 438–443.

———. 2005b. What determines the citation frequency of ecological papers? Trends in Ecology and Evolution 20: 28–32.

Lloyd ME. 1990. Gender factors in reviewer recommendations for manuscript publication. Journal of Applied Behavior Analysis 23: 539–543.

Lortie CJ, Aarssen LW, Budden AE, Koricheva JK, Leimu R, Tregenza T. 2007. Publication bias and merit in ecology. Oikos 116: 1247–1253.

Nylenna M, Riis P, Karlsson Y. 1994. Multiple blinded reviews of the same two manuscripts: Effects of referee characteristics and publication language. Journal of the American Medical Association 272: 149–151.

Paludi MA, Bauer WD. 1983. Goldberg revisited: What's in an author's name. Sex Roles 9: 387–390.

Schubert C, Sinha G. 2004. A lab of her own. Nature 10: 114–115.

Tregenza T. 2002. Gender bias in the refereeing process? Trends in Ecology and Evolution 17: 349–350.

Van Appledorn M, Lamb DA, Albalak K, Bach CE. 2007. Zebra mussels decrease burrowing ability and growth of a native snail, *Campeloma decisum*. Hydrobiologia 575: 441–445.

Robyn M. Borsuk (r.m.borsuk@gmail.com) and Christopher J. Lortie are with the Department of Biology at York University in Toronto, Canada. Lonnie W. Aarssen is with the Department of Biology at Queens University in Kingston, Canada. Amber E. Budden is with the National Center for Ecological Analysis and Synthesis in Santa Barbara, California. Julia Koricheva is with the School of Biological Sciences at Royal Holloway University of London in the United Kingdom. Roosa Leimu is with the Section of Ecology at the University of Turku in Finland. Tom Tregenza is with the Centre for Ecology and Conservation at the University of Exeter, Cornwall Campus, in Penryn, United Kingdom.