The Scandinavian
Journal of Economics

Scand. J. of Economics 114(2), 500–519, 2012
DOI: 10.1111/j.1467-9442.2011.01690.x

Discrimination in Scientific Review: A Natural Field Experiment on Blind versus Non-Blind Reviews*

Fredrik Carlsson

University of Gothenburg, SE-405 30 Gothenburg, Sweden
fredrik.carlsson@economics.gu.se

Åsa Löfgren

University of Gothenburg, SE-405 30 Gothenburg, Sweden
asa.lofgren@economics.gu.se

Thomas Sterner

University of Gothenburg, SE-405 30 Gothenburg, Sweden
thomas.sterner@economics.gu.se

Abstract

Using papers submitted to an international conference on economics held in Sweden in 2008, we analyze how gender, as well as other characteristics of the authors and reviewers, affects the grading of these papers by the reviewers. Correcting for other variables, including the country and research field, as well as the academic level of the author, we focus on the difference in grades between blind and non-blind review treatments. We find that non-blind reviewing has little effect, and there is no significant evidence of gender discrimination. Furthermore, we do not find any significant difference between the average grading by female and male reviewers.

Keywords: Gender discrimination; review

JEL classification: *C*93; *J*16

I. Introduction

Many occupations that were once exclusively male domains are today either dominated by women or have a more or less equal gender distribution. This applies to lawyers, medical doctors, veterinarians, politicians, and in some countries even clergymen. In academia, this is true for many social sciences, but economics still appears to be a bastion of male dominance. In their recent paper, Jonung and Ståhlberg (2008) point to the limited number of female economists, particularly at the top level. They look at

*We would like to thank three anonymous reviewers for their insightful comments. We would also like to thank Daniela Andrén, Katarina Katz, and the seminar participants at Karlstad University and Orebro University for comments. Financial support from Formas and Sida to the Environmental Economics Unit at the University of Gothenburg is acknowledged.

five countries – Australia, Canada, Sweden, the UK, and the US – and find that, in spite of some progress in the last few decades (e.g., approximately one-third of all PhDs nowadays are female in the countries studied), the figures are still low compared to other disciplines. This is particularly true at the level of full professorship, where the female representation ranges from 5 to 9 percent in the countries studied. The importance of identifying potential discrimination in economics is strengthened by the fact that the economics profession seems to be an outlier, with fewer women in top positions compared to other social sciences.

In the economics literature, considerable effort has been devoted to test for the existence of discrimination in general, and gender discrimination in particular. For example, see Cain (1986) for an overview of labor-market discrimination and see Kunze (2008) for a recent review of the empirical literature on the gender wage gap. One of the reasons put forward for gender discrimination is simply that men are preferred in recruitment and promotion. For example, in academia, Ginther and Hayes (1999) find that while gender differences in salaries (in the humanities) can, to a large extent, be explained by academic degree, there are also significant gender differences in promotion to tenure. That is, women are given tenure to a lower degree than men (controlling for individual characteristics). This is corroborated by McDowell *et al*. (1999), who find that the promotion prospects in economics are lower for women than for men (with comparable merits). Because the number of papers published by a researcher is an important determinant for promotion, it seems that the review process might be an important source of discrimination. However, it should be noted that the observed outcome could have many different explanations other than the fact that men prefer men, such as genetic differences in mathematical skills (Jones, 2008) and differences in preferences (Hakim, 2008).[1]

There are also studies concerned with the nature of discrimination. There can be different reasons for discrimination. "Preference-based" discrimination, for example, is a situation in which someone evaluating a woman for a job simply does not like women. "Statistical discrimination" describes a situation in which underlying factors or expectations explain the discrimination, for example, if women on average have lower productivity. List (2004) tests for the nature of discrimination by employing a number of experiments. He concludes that the observed discrimination of minority groups is explained by statistical discrimination rather than preference-based discrimination. However, Parsons *et al*. (2011) show that the expectation of

[1] Jonung and Ståhlberg (2008) sparked a lively debate on the potential causes, consequences, and remedies (see Hakim, 2008; Johnson, 2008; Jones, 2008; McCloskey, 2008; May, 2008; Jonung and Ståhlberg, 2009). There is also experimental evidence on the differences between men and women that could explain the difference in outcomes (e.g., Croson and Gneezy, 2009).

discrimination affects the productivity of a minority group negatively. Assuming that discrimination is the difference in the market outcome (e.g., wage) controlled for productivity, this could have the effect of understating the magnitude of preference-based discrimination (referring to the discrimination as statistical discrimination).

In this paper, we investigate whether or not the judgment of paper quality is affected by the gender of the author(s). For example, Goldin and Rouse (2000) find that the implementation of blind auditions in the US had a significant and large effect on the number of female musicians employed. However, the data necessary to test the discrimination hypothesis are generally difficult to access. As organizers of a large conference, we had the opportunity to conduct a natural field experiment[2] on gender discrimination in the review process by testing for the effects of blind and non-blind review processes on paper grades. The main advantages of a natural field experiment are that the researcher can control the conditions in which the decisions are made and that the subjects are unaware that they are participating in an experiment (e.g., List, 2008). While it would have been interesting to test for the nature of potential discrimination, this was not possible in this setting.

The review process is not a new concern in economics, but attention has mostly focused on journal review. In the mid-1980s, the American Economic Association's Committee on the Status of Women in the Economics Profession was concerned that non-blind reviewing would reduce the acceptance rates for female authors. Ferber and Teiman (1980) reported that female acceptance rates were higher with blind reviewing, yet they did not control for other potentially relevant variables. This led to a well-known experiment, starting in 1987, in which Blank (1991) evaluated the effects of blind reviewing in the *American Economic Review*. The experiment found that blind reviewing led to comparatively harsher overall judgments. (Actually, this result only applied to male reviewers; the effect was the opposite for female reviewers, but their small number made the behavior of male reviewers dominant.) It was also found that blind reviewing made little difference to authors from the top five universities or from low-ranked

[2] The differences between conventional laboratory experiments, artifactual field experiments, framed field experiments, and natural field experiments are thoroughly discussed in Harrison and List (2004). They define the different types of experiments as follows: "a conventional lab experiment is one that employs a standard subject pool of students, an abstract framing, and an imposed set of rules; an artifactual field experiment is the same as a conventional lab experiment but with a non-standard subject pool; a framed field experiment is the same as an artifactual field experiment but with field context in either the commodity, task, or information set that the subjects can use; a natural field experiment is the same as a framed field experiment but where the environment is one where the subjects naturally undertake these tasks and where the subjects do not know that they are in an experiment" (pp.1013 – 1014).

universities or colleges. However, those from institutions just below the top (ranked 6–50) were indeed adversely hit by blind reviewing. These authors were, in other words, positively discriminated via non-blind reviewing. However, Blank (1991) did not find significant gender discrimination.

Yet, there are a few drawbacks to the experiment by Blank (1991). One of the drawbacks, which is also mentioned by Blank in her paper, is that the experiment was publicly known. This might have had some influence on the behavior of the authors, reviewers, and editors involved (for whom the process was not blind, for obvious reasons). Although a number of changes have been made since the 1980s, the issue of blind reviewing is still a widely discussed topic. However, the speed of computers and the widespread use of Internet search engines have made it much easier for reviews to identify authors, if they so desire. Blank (1991) has already raised the issue of whether reviewing really is blind, by showing that almost half (45 percent) of the reviewers in her study could successfully identify authors (by recognition or guesswork) even if the review process was designed to be blind. If there is discrimination against some individuals or groups, this could, of course, be bad news. The fact that reviewers can easily uncover the identity of authors makes it more difficult and less rewarding to test for discrimination.

While our experiment (described in detail in Section II) is somewhat similar in design to that of Blank (1991), there are also a number of differences. First, reviewers were asked to provide a grade from D (1) to A (4) rather than a review report, which facilitates the statistical analysis. Second, no reviewer knew that the review process for the conference was designed to be a natural field experiment. Third, we have access to relevant background data on authors and reviewers, such as gender, academic degree, and affiliation. Fourth, reviewing for a conference is a faster and less thorough procedure than reviewing for a journal. For this reason, the role of prejudices and discrimination might be stronger than when there is careful scientific consideration of merit. Fifth, conference papers are potentially more difficult to find on the Internet than, for example, working papers.[3] Finally, we have a fairly large (although far from equal) number of female reviewers.

[3] We did not have the capacity to control the availability at the time of the experiment, and we did not ask the reviewers afterwards whether they searched for the identity of the authors on the Internet. However, in order to investigate whether the papers were available at the time of the review, we checked 20 of the blind reviewers two years after the experiment. This can be done with Google and restricting searches to hits available as of February 28, 2008. We checked the availability of all the papers that 20 of the reviewers had to review (a total of 239 papers). Only 31 percent of the papers were available on the Internet via Google at the time of the review. Moreover, many of the papers were made available at the end of February. Thus, papers for a conference are available to a much lesser extent on the Internet compared to papers submitted to journals. In addition, we made an informal investigation

There could also be differences between reviewing for a journal and reviewing for a conference. For example, for a conference, reviewers can consider who they want to interact with and meet at a conference. However, we informed the reviewers that we, as organizers, wanted to select the best papers and that they should grade the papers with this in mind (the instructions that were given to the reviewers are provided in Appendix A). The reviewers were also aware of the fact that there were many more papers submitted than could be accepted.

The rest of the paper is organized as follows. In Section II, we describe the experiment and the review process. In Section III, we present the data. In Section IV, we present the results of the econometric analysis. We conclude in Section V.

II. Experiment

The experiment was conducted at the European Association of Environmental and Resource Economics (EAERE) annual conference, held in Gothenburg, Sweden, in June 2008. Submission of papers was open between November 15, 2007 and February 1, 2008. The Webmeets (http://www.webmeets.com) software was used for the whole submission and review system, in which submitters of papers had to register themselves (including co-authors). In addition, authors were asked to submit an anonymous version of their paper.[4]

We had a list of approximately 170 potential reviewers, consisting of both people who had previously served as reviewers and a number of new reviewers recruited through the Association (mainly because the number of submissions was much larger than expected). Out of the 170 potential reviewers, 155 completed the task. Among the reviewers, 60 percent were full professors and 17 percent were females. Among the female reviewers,

---

of 25 randomly selected environmental economists at the World Congress of Environmental Economists in Montreal (June 2010). They were asked whether they would Google paper titles to find out the identity of "blind" authors. Out of 25, there were six who said they might do this sometimes as a journal referee, but only one who said he/she would do it when rating conference papers. However, even though we emphasized that we wanted honest answers, it is possible that these numbers represent an underestimation if this is a behavior that people engage in but view as morally questionable. The very clear message we received was that most referees do not feel they have the time to Google each conference paper they have to rate.

[4] Not all submissions complied fully with the instructions. A small number were not anonymous, in which case the author information was simply removed by our staff. Furthermore, some people did not enter all the necessary information during submission. In these cases, we again complemented the information (mainly by accessing the authors' homepages). In addition, we checked all the information on all authors by accessing the department, institution, and personal homepages.

30 percent were full professors. Many of the reviewers were from North America (29 percent) or from Europe (55 percent), but there were also reviewers from Africa, Asia, and Latin America. By February 11, 2008, all reviewers had been assigned and the papers were sent out. Each paper was assigned to two reviewers. The deadline for the review was February 29. One advantage of the Webmeets software is that it assigns reviewers to papers based on systematic subject codes (either JEL or special EAERE codes). We made a few changes to the proposed assignments, mainly to even out the number of papers sent to each reviewer. The reviewers were asked to grade each paper from D (1) to A (4). The instructions given to the reviewers regarding grading were as follows.

(A) I would definitely accept: very good paper.
(B) I would probably accept: good paper.
(C) I might accept: OK paper.
(D) I don't think this paper can be accepted.

The reviewers were randomly assigned as either blind or non-blind reviewers (and hence no reviewers were assigned both blind and non-blind papers). This was done before assigning any papers to them. The reviewers received an e-mail message with information about the process and how to access the papers.[5] Those in the non-blind group could see the names and affiliations of all authors when they accessed the paper and on the review sheet, while the blind group could only see the title of the paper. The grades were reported online.

III. Data

We received 1,074 submissions, and 940 of these were sent out for review. Each submission received at least one review, and 825 received two. This leaves us with 1,765 observed grades. The average grade was 2.6. Out of the 940 papers, 483 were accepted to the conference.[6] Table 1 presents descriptive statistics for the whole sample, and for the blind and non-blind subsamples.

It is not at all straightforward to see how we can distinguish between female and male papers. Given the richness of our data, we decided not to limit the analysis only to papers with and without female authors. First of all, we test the three different definitions discussed by Blank (1991): (i) at least one author is female; (ii) the first (primary) author is female; (iii) all

[5] The complete letter to the reviewers is provided in Appendix A.
[6] The average grade for the accepted papers was 3.27 and the average grade for the rejected papers was 1.95.

Table 1. *Variable definitions and mean values*

Variable	Description	Whole sample ($n = 1765$)	Non-blind ($n = 895$)	Blind ($n = 870$)
Grade	Grade 1–4 (4 best)	2.579	2.618	2.539
		(1.05)	(1.06)	(1.04)
Non-blind	= 1 if non-blind	0.507		
At least one female	= 1 if at least one author is female	0.417	0.421	0.413
Only female	= 1 if only female authors	0.131	0.131	0.131
First author female	= 1 if first author is female	0.276	0.285	0.267
Only professors	= 1 if all authors are professors	0.056	0.060	0.051
Only students	= 1 if all authors are students	0.184	0.180	0.187
No. authors	Number of authors	2.003	1.990	2.017
		(1.1)	(1.1)	(1.1)
Student	= 1 if any author is a PhD student	0.465	0.466	0.464
PhD	= 1 if any author is a PhD	0.713	0.713	0.713
Professor	= 1 if any author is a professor	0.299	0.293	0.305
Female student	= 1 if any author is a female student	0.178	0.171	0.185
Female PhD	= 1 if any author is a female PhD	0.266	0.274	0.257
Female professor	= 1 if any author is a female professor	0.020	0.025	0.016
Male student	= 1 if any author is a male student	0.313	0.321	0.305
Male PhD	= 1 if any author is a male PhD	0.560	0.546	0.574
Male professor	= 1 if any author is a male professor	0.286	0.277	0.295
Female reviewer	= 1 if reviewer is female	0.174	0.161	0.186
Professor reviewer	= 1 if reviewer is a professor	0.603	0.570	0.637
Europe	= 1 if any author affiliation is in Europe	0.577	0.559	0.597
North America	= 1 if any author aff. is in North America	0.198	0.200	0.197
Developing countries	= 1 if any author aff. is in dev countries	0.198	0.212	0.184
Eastern Europe	= 1 if any author aff. is in E. Europe	0.034	0.038	0.030
Asia	= 1 if any author aff. is in Asia	0.032	0.036	0.028
Latin America	= 1 if any author aff. is in Latin America	0.033	0.036	0.030
Oceania	= 1 if any author aff. is in Oceania	0.027	0.028	0.026
Pollution control	= 1 if paper concerns pollution control	0.248	0.220	0.276
Valuation	= 1 if paper concerns environment valuation	0.221	0.206	0.237
Resources	= 1 if paper concerns resources and ecosystem	0.659	0.686	0.632
Growth	= 1 if paper concerns growth and env.	0.155	0.165	0.145
International issues	= 1 if paper concerns int. env. issues	0.176	0.173	0.179
Firm	= 1 if paper concerns env. and the firm	0.041	0.045	0.038
Risk	= 1 if paper concerns risk and uncertainty	0.076	0.086	0.067
Agriculture	= 1 if paper concerns agriculture	0.121	0.139	0.102
Extensions	= 1 if paper concerns extensions of theory	0.239	0.251	0.226

Note: We report standard deviations in brackets for variables that are not dichotomous.

authors are female. In addition, each author of a paper is classified as one of the following: (i) a student; (ii) a PhD but not a full professor; (iii) a professor. For each academic category, we create a dummy variable equal to one if at least one author falls into the respective category. In almost 46 percent of the observations, there is at least one student author, in 67 percent there is at least one PhD (but not full professor), and in 36 percent there is at least one author who is a professor. In order to allow testing for differences between male and female authors, we also create separate categories for male and female authors for each academic category. The distribution of these variables shows that there are roughly as many female as male students in our sample and almost twice as many male as female PhDs. Only 2 percent of the papers have at least one female professor as co-author, while 29 percent have at least one male professor as co-author. Thus, for this group, we observe the usual gender pattern with equality at the graduate level but great inequality at higher levels.

We also control for two other characteristics: (i) the country of the author's affiliation; (ii) EAERE subject code. Countries are classified into seven categories; each dummy variable is equal to one if at least one author is from the respective category. EAERE codes[7] are used to classify the observations into different research fields; note that each paper can fit into several fields.

As can be seen by visually inspecting Table 1, it is hard to find differences in characteristics between the blind and non-blind samples. Using chi-square or Wilcoxon–Mann–Whitney tests, we cannot reject the hypothesis of equal distributions between the two subsamples for any of the characteristics.

IV. Results

We begin by reporting the grades for the whole sample and the two subsamples split by various variables. There are a total of 1,764 grades for the papers and the grades range from 1 or D (definitely reject) to 4 or A (definitely accept). Table 2 presents the mean grades; the distribution of grades for the whole sample and by author gender and academic level can be found in Appendix B.

There are no significant overall differences in grades for any of the three definitions of a female-authored paper. Also, there are no significant differences in this respect between the two treatments, apart from the finding that the average grade for papers co-authored by female PhDs is higher in the non-blind treatment; the difference in distributions is significant using a

[7] EAERE subject codes were more finely disaggregated than JEL codes. Some examples are shown in Table 1 as subject areas.

Table 2. *Grades for whole sample and by review treatment*

	All reviewers		Non-blind		Blind	
	Grade	No. obs.	Grade	No. obs.	Grade	No. obs.
Whole sample	2.58	1765	2.62	895	2.54	870
By author gender						
At least one female	2.58	736	2.62	377	2.54	359
Only female	2.21	231	2.24	117	2.18	114
First author female	2.51	487	2.57	255	2.44	232
Only male	2.58	1029	2.62	518	2.54	511
By gender and degree						
Female students	2.44	314	2.43	153	2.45	161
Male students	2.43	552	2.50	287	2.36	265
Female PhDs	2.67	469	2.76	245	2.57	224
Male PhDs	2.65	988	2.70	489	2.61	499
Female professors	2.94	36	2.86	22	3.07	14
Male professors	2.92	505	2.98	248	2.86	257
By reviewer gender						
Female reviewers	2.59	306	2.70	144	2.50	162
Male reviewers	2.58	1457	2.60	750	2.55	707

Wilcoxon–Mann–Whitney test (p-value $= 0.045$). This would, if anything, be an example of positive, and not negative, discrimination of women. The result is not confirmed in the multivariate regression analysis though (see below), and it is thus not worth further comment. However, there are some differences between different categories of papers. For example, papers with students as authors received lower grades (in both the blind and non-blind treatments) than papers with PhD authors, and PhDs received lower grades than professors. The only other effect that is of interest, and which is perhaps surprising, is that the mean grade for papers authored only by females is substantially lower than the grades for other papers. However, the main explanation for this difference is that very few of these papers include a female professor as a co-author.[8] If we look at papers with at least one female author, then only 3 percent of the papers with only female authors include a professor (who by definition is a female professor), while for the other papers 45 percent include at least one professor (male or female). The number of papers with only female authors is also lower than for other papers. For papers with only female authors, the average number of authors is 1.4, while for other papers with at least one female author it is 2.9. As we see later in the analysis, there is a positive correlation between the number of authors and the paper grade.

In order to analyze the difference between the blind and non-blind reviewers in more detail and to control for other characteristics, we estimate

[8] Thus, there is no similar effect for papers authored only by males.

ordered probit models, where the standard errors are corrected for clustering at the paper level.[9] As discussed, there are many ways of defining a female-authored paper, but here we only present the models using the definition reported in Blank (1991): a female-authored paper is a paper where at least one author is female. All the results (available upon request) hold if we use the two other definitions (the first author is female, and all authors are female).

Because we wish to explore a number of interactions between variables, we present five different models (see Table 3). The full set of marginal effects for models 2 and 5 is presented in Appendix C. There are four categories in the model corresponding to the four grades.[10]

The first model shows the difference in grades between the blind and non-blind treatments and the test of an overall discrimination of female-authored papers. The conclusion is that there is no sign of gender discrimination, the interaction term is highly insignificant, and the coefficient is small. This test is essentially the same as the non-parametric test because we do not control for other characteristics of the papers. In the second model, we therefore include variables controlling for academic degree, paper subject area, and country of affiliation. However, this has no effect on the significance of the interaction term between the non-blind treatment and female-authored papers. The results reveal that there is a significant difference in grades between papers with students as co-authors and other papers, and papers with only students as authors receive even lower grades. As shown in Appendix C, the corresponding marginal effects are non-negligible. For example, a paper with at least one student as co-author has a 7-percentage-point lower probability of receiving the highest grade. In the third model, therefore, we add a test of difference between the blind and non-blind treatments for the academic level of authors. The results show that there is no observed discrimination of papers authored only by students; papers authored only by professors do not receive a significantly higher grade in the non-blind treatment. In the fourth model, we add reviewer characteristics. In particular, we include reviewer gender, whether the reviewer is a full professor, and whether the reviewer is from the same country as at least one author. Female reviewers do not give higher or lower grades than male reviewers; they do not give higher or lower grades to female-authored papers in the non-blind treatment. The coefficient of a professor being a reviewer is almost significant at the 10 percent level, and the coefficient is negative, suggesting that on average they give lower

[9] The results in terms of significance are basically the same in a model without correction of standard errors for clustering, and the results in terms of sign and significance are the same in a standard OLS model.

[10] We have also estimated models for the restricted sample in which each paper had one blind and one non-blind reviewer, and all the main results still hold.

Table 3. *Ordered probit models with paper grade as dependent variable (grades 1–4, where 4 is the highest)*

	Model 1		Model 2		Model 3		Model 4		Model 5	
Variable	Coeff.	*p*-value	Coeff.	*p*-value	Coeff.	*p*-value	Coeff.	*p*-value	Coeff.	*p*-value
Author academic level and gender										
At least one female author	0.002	0.985	−0.121	0.154	−0.085	0.327	−0.079	0.364		
Student			−0.239	0.005	−0.353	0.003	−0.351	0.004		
PhD			−0.138	0.340	−0.250	0.187	−0.243	0.200		
Professor			0.092	0.279	0.139	0.200	0.150	0.169		
Female student									−0.204	0.111
Female PhD									−0.141	0.248
Female professor									0.189	0.464
Male student									−0.172	0.146
Male PhD									0.063	0.673
Male professor									0.215	0.061
Only professors			0.173	0.315	0.068	0.787	0.117	0.638	0.358	0.109
Only students			−0.273	0.068	−0.190	0.325	−0.187	0.332	−0.073	0.647
No. authors			0.114	0.003	0.098	0.062	0.100	0.061	0.062	0.282
Reviewer characteristics										
Female reviewer							0.027	0.785	−0.172	0.861
Professor reviewer							−0.137	0.118	−0.132	0.135
Review paper from own country							0.104	0.367	0.102	0.376
Treatment effects										
Non-blind	0.082	0.245	0.114	0.100	−0.141	0.604	−0.129	0.648	−0.043	0.842
Non-blind × at least one female author	0.002	0.985	0.010	0.926	−0.062	0.589	−0.042	0.721		
Non-blind × student					0.209	0.175	0.214	0.166		
Non-blind × PhD					0.220	0.417	0.221	0.416		
Non-blind × professor					−0.084	0.592	−0.067	0.674		
Non-blind × female student									0.057	0.732
Non-blind × female PhD									0.206	0.245
Non-blind × female professor									−0.242	0.382
Non-blind × male student									0.211	0.170
Non-blind × male PhD									0.106	0.613
Non-blind × male professor									−0.045	0.790
Non-blind × only professors					0.210	0.538	0.169	0.632	0.089	0.759
Non-blind × only students					−0.158	0.562	−0.144	0.593	−0.180	0.424
Non-blind × no. authors					0.036	0.623	0.038	0.605	0.026	0.752
Non-blind × female reviewer							0.127	0.446	0.068	0.634
Non-blind × female rev. × at least 1 fem. auth.							−0.114	0.602		
Non-blind × professor reviewer							−0.078	0.506	−0.083	0.481
Non-blind × review paper from own country							−0.133	0.457	−0.144	0.425
Controls for paper area and country	No		Yes		Yes		Yes		Yes	

grades than other reviewers. Reviewers do not give higher or lower grades to an author from their own country. None of the interaction terms with the non-blind treatment is significant.

Finally, in the last model we use more information about the authors, because we estimate separate effects for students, PhDs, and professors for both females and males. Again, we find no signs of discrimination. All the interaction terms with the non-blind treatment are highly insignificant. We also test for discrimination of authors from various countries or continents, because some people might have a prejudice against authors from countries from which there are fewer well-known economists. Indeed, there is evidence that fewer papers are accepted from some regions, but we do not find any significant evidence of discrimination by author region (i.e., the difference between the blind and non-blind treatments). This strengthens the impression that referees appear to undertake their task without prejudice or discrimination.

V. Discussion

Our study provides a contribution to the scarce body of literature that empirically measures discrimination in the review process. Our results confirm the conclusion of Blank (1991) that there is no gender discrimination (i.e., there is no significant difference between blind and non-blind reviews with respect to the grades of female- and male-authored papers). Furthermore, we do not find any significant difference between the average grading by female and male reviewers. Perhaps, as expected, papers with students as co-authors receive lower grades than other papers, yet this is true for both the blind and the non-blind treatments.

One possible objection to our results is that environmental economics might be seen as a possible outlier in economics. The argument is that there are more women in fields such as environmental economics than in other more traditional fields of economics, and therefore gender discrimination might not be present or at least might not be very strong. In a study on females in environmental economics in the US and Canada, Bhattacharjee *et al*. (2007) find that departments that offer a graduate field in environmental economics have a higher share of women at the associate and assistant professor levels, although the differences are rather small and only significant for associate professors. They also find that women publish fewer articles and are cited less than men. At the same time, they find that women are well represented in the economics association and at editorial positions at the *Journal of Environmental Economics and Management* (the leading journal in environmental economics). However, our data tell another story when it comes to the distribution of academic levels among men and women. As reported earlier, there is a very uneven distribution of

male and female professors, and there are twice as many papers with male PhD co-authors as there are papers with female PhD co-authors.

Also, while working on this study, we became aware of the problems of using blind review as a mechanism for dealing with discrimination today, given the information availability on the Internet. As pointed out by reviewers, there is a possibility that the blind process in our study is not as blind as we would like to think. While we could not control for whether the "blind" reviewers tried to check the authors of the papers they received (which is a drawback of our study), we did investigate afterwards, for a subsample, whether the papers were available at the time of the review. It turned out that only 31 percent of the papers were available. Hence, while we cannot rule out the possibility that reviewers searched for the identity of the authors, our blind treatment seems to have been at least moderately blind. This is a clear advantage compared with the review process for a journal.

It is also interesting to note that, in both our study and that of Blank (1991), no evidence of discrimination is found when comparing blind and non-blind review processes. In contrast, Goldin and Rouse (2000) find that blind auditioning is an effective mechanism for dealing with discrimination (this implicitly implies that there is discrimination). However, in line with the results of Parsons *et al*. (2011), discrimination can affect the productivity of the affected group (i.e., women, in this instance). For example, in our setting, this could imply that women chose research areas where there are more women, or research areas that are considered "easier" (e.g., comparing empirical studies versus theoretical studies), because they expect to be more scrutinized than male colleagues. This would not show up as discrimination in our natural field experiment, and it could have the effect of understating the magnitude of preference-based discrimination. This should also be addressed in future research.

Finally, drawing on the experience of our experiment, we believe that future researchers who want to investigate the nature of the discrimination should consider some manipulation of the review process. In particular, they should vary the author identity and gender at the paper level (i.e., for the same paper, some reviewers should be told that the authors are female, while other reviewers should be told that they are male). In this way, it might be possible to test whether a potential discrimination is preference-based or statistical.

Appendix A. Instructions to Reviewers

Dear,

We have received a record number of submissions (over 1000, compared to 650 last year and 850 at the World Conference). This is great fun. It shows growing interest

in and awareness of the important challenges we face and it gives us the possibility to make 2008 a really good conference. It does, however, also hinge on your and our good work in selection. We are very keen to select the best papers and to put them into coherent sessions so as to build as good a program as possible. The large number of papers and the fact we are using double refereeing means that you will have to referee somewhat more papers. We have spent a week increasing the number of referees but still many of you will be asked to referee as many as 13 or 14 papers. (One of two of our very best friends may even have got 15).

We originally had the ambition of staying below 10 papers per person but it was not possible. I really want to thank you so much for the effort you will be putting into this. It is decisive for the conference.

Following is the information you will need in order to assess submissions for the 2008 Meeting of the European Association of Environmental and Resource Economists. I hope the evaluation system will be easy for you to use and will minimize the time required to process these papers. There is a complete online facility for the evaluation process. You will be able to view online the papers which have been assigned to you and to download the corresponding pdf files, conveniently packed in a zip file if you choose that option. To access the online evaluation facility, please go to http://www.webmeets.com/eaere/2008, and choose "About You" from the Browser window. Login using the User Profile you created when submitting your topics of interest earlier upon invitation to join the program committee. This will lead you to secure, online access of the Evaluation facility.

You will find instructions under the link "About You/Your Assigned Papers". The direct URL is www.webmeets.com/EAERE/2005/AboutYou/howto.asp Please read them carefully. Note especially that you must re-enter your password to update your grades. If you have any problems reading a given file, please report it using the bad file report tool found in the link "View Paper Details". If you experience any other technical difficulties using the system, please report them to support@webmeets.com and include a copy of any error messages that you may have received.

The Contributed Sessions program has slots for about 320–350 papers + a few dozen Posters. There is always some attrition, so we aim at accepting around 400 papers for contributed sessions. Given these figures, the target acceptance rate is about 40 percent. When you review the submissions, please provide a letter grade (A to C) as follows.

(A) I would definitely accept: very good paper.
(B) I would probably accept: good paper.
(C) I might accept: OK paper.
(D) I don't think this paper can be accepted.

For simplicity, we may assume that the overall distribution is such that there are OVERALL 25 percent of papers in each of the categories A, B, C, and D (naturally your own subset of referee papers does not necessarily have these exact proportions but I am assuming we all know something about sampling and distributions . . .). We will have two referees for each paper. We aim to accept all papers where both referees have an "A" and then take as many as we can fit in, depending on the grades the two

referees assign. We would also like you to suggest a session for each paper – there is an easy function for this on the website. (Please note in particular that there is a special category for Poster session.) Given the tight time constraints that we face for the organization of this conference, I would ask you to provide your grades before February 29. Please let me know immediately if you have any difficulty meeting this deadline.

Thank you, once again, for your help in this critical task for our conference.

My best wishes,

Thomas Sterner

President of the EAERE

Appendix B.

Table B1. *Distribution of grades for the whole sample and by author gender and academic level for the blind and non-blind treatments*

	Blind			Non-blind		
Grade	Frequency	Percent	Cumulative frequency	Frequency	Percent	Cumulative frequency
Whole sample						
1	172	19.77	19.77	166	18.55	18.55
2	246	28.28	48.05	243	27.15	45.70
3	263	30.23	78.28	253	28.27	73.97
4	189	21.72	100.00	233	26.03	100.00
Female students						
1	30	18.63	18.63	40	26.14	26.14
2	53	32.92	51.55	42	27.45	53.59
3	53	32.92	84.47	36	23.53	77.12
4	25	15.53	100.00	35	22.88	100.00
Female PhDs						
1	43	19.20	19.20	31	12.65	12.65
2	61	27.23	46.43	68	27.76	40.41
3	70	31.25	77.68	74	30.20	70.61
4	50	22.32	100.00	72	29.39	100.00
Female professors						
1	0	0	0	0	0	0
2	5	35.71	35.71	9	40.91	40.91
3	3	21.43	57.14	7	31.82	72.73
4	6	42.86	100.00	6	27.27	100.00
Male students						
1	74	27.92	27.92	59	20.56	20.56
2	66	24.91	52.83	84	29.27	49.83
3	80	30.19	83.02	85	29.62	79.44
4	45	16.98	100.00	59	20.56	100.00
Male PhDs						
1	85	17.03	17.03	77	15.75	15.75
2	143	28.66	45.69	132	26.99	42.74

Table B1. *Continued*

	Blind			Non-blind		
Grade	Frequency	Percent	Cumulative frequency	Frequency	Percent	Cumulative frequency
3	153	30.66	76.35	140	28.63	71.37
4	118	23.65	100.00	140	28.63	100.00
Male professors						
1	25	9.73	9.73	21	8.47	8.47
2	66	25.68	35.41	56	22.58	31.05
3	86	33.46	68.87	78	31.45	62.50
4	80	31.13	100.00	93	37.50	100.00

Appendix C.

Table C1. *Marginal effects for ordered probit models 2 and 5 (p-values in parenthesis)*

	Model 2				Model 5			
Variable	Grade = 1	Grade = 2	Grade = 3	Grade = 4	Grade = 1	Grade = 2	Grade = 3	Grade = 4
Author degree and gender								
At least one female author	0.030	0.018	−0.014	−0.034				
	(0.160)	(0.147)	(0.169)	(0.149)				
Student	0.059	0.036	−0.027	−0.068				
	(0.005)	(0.005)	(0.006)	(0.005)				
PhD	0.033	0.022	−0.014	−0.041				
	(0.327)	(0.355)	(0.301)	(0.351)				
Professor	−0.022	−0.014	0.010	0.027				
	(0.271)	(0.290)	(0.259)	(0.285)				
Female student					0.053	0.028	−0.026	−0.055
					(0.123)	(0.071)	(0.157)	(0.090)
Female PhD					0.036	0.021	−0.017	−0.039
					(0.098)	(0.220)	(0.283)	(0.223)
Female professor					−0.042	−0.032	0.016	0.058
					(0.060)	(0.500)	(0.293)	(0.491)
Male student					0.043	0.025	−0.020	−0.048
					(0.104)	(0.125)	(0.174)	(0.135)
Male PhD					−0.015	−0.010	0.007	0.018
					(0.056)	(0.672)	(0.676)	(0.672)
Male professor					−0.050	−0.035	0.021	0.064
					(0.000)	(0.073)	(0.035)	(0.069)
Paper characteristics								
Only professors	−0.039	−0.029	0.015	0.053	−0.074	−0.065	0.023	0.115
	(0.274)	(0.354)	(0.186)	(0.341)	(0.000)	(0.149)	(0.000)	(0.144)
Only students	0.073	0.036	−0.036	−0.073	0.018	0.011	−0.009	−0.020
	(0.092)	(0.027)	(0.112)	(0.047)	(0.105)	(0.663)	(0.689)	(0.668)

Continued

Table C1. *Continued*

Variable	Model 2				Model 5			
	Grade = 1	Grade = 2	Grade = 3	Grade = 4	Grade = 1	Grade = 2	Grade = 3	Grade = 4
No. authors	−0.028	−0.017	0.013	0.033	−0.015	−0.010	0.007	0.018
	(0.003)	(0.004)	(0.005)	(0.003)	(0.012)	(0.284)	(0.286)	(0.281)
Reviewer characteristics								
Female reviewer					0.004	0.003	−0.002	−0.005
					(0.052)	(0.860)	(0.863)	(0.861)
Professor reviewer					0.032	0.021	−0.014	−0.038
					(0.073)	(0.145)	(0.122)	(0.474)
Review paper from own country					−0.024	−0.017	0.010	0.030
					(0.027)	(0.396)	(0.323)	(0.390)
Treatment effects								
Non-blind	−0.028	−0.017	0.013	0.033	0.010	0.007	−0.005	−0.012
	(0.101)	(0.101)	(0.103)	(0.101)	(0.113)	(0.842)	(0.841)	(0.842)
Non-blind × at least 1 female auth.	−0.002	−0.001	0.002	0.003				
	(0.926)	(0.926)	(0.926)	(0.926)				
Non-blind × female student					−0.014	−0.009	0.006	0.017
					(0.063)	(0.741)	(0.718)	(0.739)
Non-blind × female PhD					−0.046	−0.034	0.019	0.062
					(0.026)	(0.280)	(0.135)	(0.269)
Non-blind × female professor					0.066	0.030	−0.034	−0.063
					(0.228)	(0.243)	(0.457)	(0.326)
Non-blind × male student					−0.048	−0.035	0.019	0.064
					(0.016)	(0.200)	(0.084)	(0.191)
Non-blind × male PhD					−0.025	−0.017	0.011	0.031
					(0.070)	(0.623)	(0.590)	(0.619)
Non-blind × male professor					0.011	0.007	−0.005	−0.013
					(0.094)	(0.784)	(0.797)	(0.787)
Non-blind × only professors					−0.021	−0.015	0.009	0.027
					(0.107)	(0.770)	(0.728)	(0.766)
Non-blind × only students					0.047	0.025	−0.023	−0.048
					(0.170)	(0.353)	(0.479)	(0.392)

Continued

Table C1. *Continued*

Variable	Model 2				Model 5			
	Grade = 1	Grade = 2	Grade = 3	Grade = 4	Grade = 1	Grade = 2	Grade = 3	Grade = 4
Non-blind × no. authors					−0.006	−0.004	0.003	0.008
					(0.034)	(0.752)	(0.752)	(0.752)
Non-blind × female reviewer					−0.016	−0.011	0.007	0.020
					(0.048)	(0.646)	(0.607)	(0.642)
Non-blind × professor reviewer					0.021	0.012	−0.010	−0.023
					(0.079)	(0.469)	(0.499)	(0.474)
Non-blind × review from own country					0.037	0.020	−0.018	−0.039
					0.135	(0.366)	(0.478)	(0.397)

References

Bhattacharjee, S., Herriges, J., and Kling, C. (2007), The Status of Women in Environmental Economics, *Review of Environmental Economics and Policy 1*, 212–227.

Blank, R. M. (1991), The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from the *American Economic Review*, *American Economic Review 81* (5), 1041–1067.

Cain, G. (1986), The Economic Analysis of Labor Market Discrimination: A Survey, in O. Ashenfelter and R. Layard (eds.) *Handbook of Labor Economics*, Vol. I, Elsevier, Amsterdam, 693–785.

Croson, R. and Gneezy, U. (2009), Gender Differences in Preferences, *Journal of Economic Literature 47*, 448–474.

Ferber, M. and Teiman, M. (1980), Are Women at a Disadvantage in Publishing Journal Articles?, *Eastern Economic Journal 6*, 180–193.

Ginther, D. K. and Hayes, K. J. (1999), Gender Differences in Salary and Promotion in the Humanities, *American Economic Review 89* (2), 397–402.

Goldin, C. and Rouse, C. (2000), Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians, *American Economic Review 90* (4), 715–741.

Hakim, C. (2008), Diversity in Tastes, Values, and Preferences: Comment on Jonung and Ståhlberg, *Econ Journal Watch 5*, 204–218.

Harrison, G. W and List, J. A. (2004), Field Experiments, *Journal of Economic Literature XLII*, 1009–1055.

Johnson, J. A. (2008), Preferences Underlying Women's Choices in Academic Economics, *Econ Journal Watch 5*, 219–226.

Jones, G. (2008), What is the Right Number of Women? Hints and Puzzles from Cognitive Ability Research, *Econ Journal Watch 5*, 227–239.

Jonung, C. and Ståhlberg, A. C. (2009), Does Economics have a Gender?, *Econ Journal Watch 6*, 60–72.

Jonung, C. and Ståhlberg, A. C. (2008), Reaching the Top? On Gender Balance in the Economics Profession, *Econ Journal Watch 5*, 174–192.

Kunze, A. (2008), Gender Wage Gap Studies: Consistency and Decomposition, *Empirical Economics 35*, 63–76.

List, J. A. (2004), The Nature and Extent of Discrimination in the Marketplace: Evidence From the Field, *Quarterly Journal of Economics 119*, 49–89.

List, J. A. (2008), *Homo experimentalis* Evolves, *Science 321*, 207–208.

McCloskey, D. N. (2008), Mr. Max and the Substantial Errors of Manly Economics, *Econ Journal Watch 5*, 199–203.

McDowell, J. M., Singell L. D. Jr, and Ziliak, J. P., (1999), Cracks in the Glass Ceiling: Gender and Promotion in the Economics Profession, *American Economic Review 89* (2), 392–396.

May, A. M. (2008), On Gender Balance in the Economics Profession, *Econ Journal Watch 5*, 193–198.

Parsons, C. A., Sulaeman, J., Yates, M. C., and Hamermesh, D. S. (2011), Strike Three: Discrimination, Incentives and Evaluation, *American Economic Review 101*(4), 1410–1435.

First version submitted November 2009;
final version received September 2010.