Scientometrics (2014) 101:717–735
DOI 10.1007/s11192-014-1354-z

The reviewer in the mirror: examining gendered and ethnicized notions of reciprocity in peer review

Bradford Demarest · Guo Freeman · Cassidy R. Sugimoto

Received: 9 October 2013 / Published online: 22 June 2014
© Akadémiai Kiadó, Budapest, Hungary 2014

Abstract Numerous studies have sought to uncover violations of objectivity and impartiality in peer review; however the notion of reciprocity has been absent in much of this discussion, particularly as it relates to gendered and ethnicized behaviors of peer review. The current study addresses this gap in research by investigating patterns of reciprocity (i.e., correspondences between patterns of recommendations received by authors and patterns of recommendations given by reviewers in the same social group) by perceived gender and ethnicity of reviewers and authors for submissions to the *Journal of the American Society for Information Science and Technology* from June 2009 to May 2011. The degree of reciprocity for each social group was examined by employing Monte Carlo resampling to extrapolate more robust patterns from the limited data available. We found that papers with female authors received more negative reviews than reviews for male authors. Reciprocity was suggested by the fact that female reviewers gave lower reviews than male reviewers. Reciprocity was also exhibited by ethnicity, although non-Western reviewers gave disproportionately more recommendations of major revision, while non-Western authors tended to receive more outright rejections. This study provides a novel theoretical and methodological basis for future studies on reciprocity in peer review.

Keywords Reciprocity · Peer review · JASIST · Scholarly communication · Monte Carlo resampling

B. Demarest (✉) · G. Freeman · C. R. Sugimoto
School of Informatics and Computing, Indiana University Bloomington, 1320 East 10th Street, Bloomington, USA
e-mail: bdemares@indiana.edu

G. Freeman
e-mail: guozhang@indiana.edu

C. R. Sugimoto
e-mail: sugimoto@indiana.edu

Mathematics subject classification C150 (Statistical Simulation Methods: General)
JEL classification 62F40 (Bootstrap, jackknife and other resampling methods)

Introduction

There is no dearth of studies on bias in peer review. Numerous studies have sought to display the violations to the idealized notions of objectivity and impartiality in peer review (e.g., Lee et al. 2013; Bornmann et al. 2008) (for reviews of the research on peer review, see Bornmann 2011; Weller 2002). Studies of extra-scientific social trends in academic journal submission acceptance have focused on bias related to social categories such as gender and ethnicity of submissions' authors or reviewers (e.g., Ernst and Kienbacher 1991; Link 1998; Wennerås and Wold 1997; Budden et al. 2008). However, the notion of reciprocity has been absent in much of this discussion, particularly as it relates to gendered and ethnicized behaviors of peer review. Reciprocity provides a more nuanced view of the peer review system as it evaluates not only how certain groups are reviewed, but also the relationship between how a group is reviewed and how that translates for reviewing behavior within that group. Simply, reciprocity examines whether scholars review in the way they are reviewed. This study seeks to examine potential reciprocity behaviors in peer review exhibited by gender and ethnicity, by analyzing rates of recommendations received by authors and given by reviewers between 2009 and 2011 in the *Journal of the American Society for Information Science and Technology* (hereafter *JASIST*). The objective of this paper is to conceptualize notions of reciprocity in peer review and, furthermore, to provide a systematic method for operationalizing studies of reciprocity.

Literature review

This research is informed theoretically by literature in psychology, sociology, and other social science domains; and empirically by scientometric studies. These bodies of literature will be briefly reviewed here, particularly in reference to dominant theoretical frameworks and studies examining social variables in peer review.

Theoretical frameworks

Projective identification, a concept originating in the domain of psychoanalysis (Klein 1946), describes the way in which an individual ascribes or transfers qualities onto another individual. Projective identification is usually considered "an attempt to control, manipulate, or change the projected quality in the other person" (Hamilton 1990, p. 415) and has been classified as intrapsychic or interpersonal (Ogden 1997)—in the former case, one projects onto another, while in the latter case, an individual behaves under the influence of projections from others. In short, (a) a person experiences unacceptable or unpleasant feelings and displaces this experience onto another object; (b) the secondary recipient internalizes the transferred feelings and begins to think and behave in accordance with the projection; and (c) a relationship is created between the projector and recipient (Casement 1991; Goldstein 1991).

A similar body of research examines the concept of reciprocity, and more specifically the tendency of individuals to behave as they have been treated. Psychologists,

sociologists, economists and anthropologists have all hypothesized that reciprocity is an internalized social norm and a general tendency of human nature (e.g., Gouldner 1960; Homans 1961; Falk and Fischbacher 2006; Perugini et al. 2002; Thibaut and Kelley 1959). It is "the behavioral response to an action that is perceived as either kind or unkind" (Falk and Fischbacher 2006, p. 294). Reciprocity is governed by two related principles: (1) Behavior of homophily: one will react in the way in which one was treated; and (2) Purpose of equality: one will react in a similar way to which he has been treated to achieve equality. The distinction between homophily and equality is that the former describes the outward action, while the latter describes the actor's intention.

Projective identification and theories of reciprocity can be used to explain potential relationships between patterns of behavior in peer review: Reviewer A assigns a recommendation to Author B. The degree of criticism, the length of the review, the inclusion of references, and other characteristics of the review all serve to inform Author B's experience of being reviewed. This is internalized and, when Author B subsequently becomes a reviewer, they will express their understanding of a review onto Author C (enacting the behavior of homophily, due to the reviewer's intention to achieve equality), thus perpetuating the cycle and creating a reviewing norm for this community. It is important to note that this is a general rather than direct process of reciprocity: Author B will not necessarily return treatment to Reviewer A specifically as the identity of this reviewer is protected; however, the author will internalize the norms and apply the treatment to the next author she reviews.

The concept of community norming brings into question the social aspect of these behaviors. As Blau (1964) suggests, all exchanges are social and, as such, peers exert influence over each other. Peer effects are complicated by the complexity of operationalizing peer groups (Moretti 2011) and understanding the origins of the peer impact: namely, whether they are exerted (a) exogenously through predetermined and non-contemporary behaviors; (b) exerted endogenously through emerging and contemporary peers; or (c) influenced as a result of self-selection into a group on the basis of a preexisting and shared characteristic (Manski 1993). Each of these types of peer effect can be seen in the context of the peer review system: authors are typically building upon a non-contemporary standard for conducting peer review. They may receive anecdotal advice from a mentor on how to review, thus informing their notions. As soon as an author receives their first review they receive direct evidence of the nature of reviews (regardless of whether the review they receive is, in fact, a "normal" review). In addition, they may self-select into certain areas of studies or journals—these disciplines and journals may be populated with people with other similar characteristics or backgrounds, thus facilitating the norming process.

Social variables in peer review

Merton proposed four "institutional imperatives…taken to comprise the ethos of modern science" (Merton 1973, p. 270). One of these is universalism, which proposes that the "acceptance or rejection of claims entering the lists of science is not to depend on the personal or social attributes of their protagonist" (Merton 1973, p. 270). Yet, scientists do not always play by these rules, and violations of impartiality (Lee et al. 2013), including the influence of social stereotypes (e.g., Bornmann et al. 2008), have been documented in the review of scientific work. Attributes of authors and reviewers have been discussed as intervening variables (Bornmann and Daniel 2010), serving to reject a manuscript for attributes other than quality alone (Cronin and McKenzie 1992). Recommendation decisions (that is, whether the reviewer recommends acceptance, revision, or rejection of a

manuscript) have been examined against social variables to test for violations of impartiality in peer review.

However, there is considerable ambiguity in the results. In studying the effect of nationality on peer review, Ernst and Kienbacher (1991) demonstrated that national affiliation of journals is strongly associated with the national affiliation of the authors in a given journal, suggesting a bias in favor of national authors. Yet, in a study of economics journals, Oswald (2008) found no national bias, although some institutional biases were present. Link (1998) examined a year of reviews for a gastroenterology journal and concluded that reviewers from the US and outside the US evaluate non-US papers similarly and evaluate papers submitted by U.S. authors more favorably. However, Marsh et al. (2008) found the opposite: U.S. authors gave high ratings on grant reviews for those outside of the country. We are a long way from a consensus on this topic given the small sample sizes and disciplinary homogeneity of these studies.

The issue of gender is similarly complicated territory. Given the gender gap in STEM fields (Wennerås and Wold 1997; Budden et al. 2008; Lariviere et al. 2013), "the prevailing assumption has been that men are overall more favorably treated than women" (Lee et al. 2013, p. 7). This assumption been reinforced by a number of studies (e.g., Bornmann et al. 2007, 2009, 2010; Marsh et al. 2009; Wennerås and Wold 1997). However, other studies have shown that female authors are not disproportionately rejected in single-blind review journals (e.g., Gilbert et al. 1994; *Nature Neuroscience* 2006; Valkonen and Brooks 2011; Borsuk et al. 2009). Additionally, studies have shown that moving to a double-blind review system does not increase acceptance rates for female authors (Blank 1991). In a recent meta-analysis of the literature, Ceci and Williams (2011, p. 3157) suggested that claims of gender bias in peer review "are no longer valid". A quantitative meta-analysis of peer review in grant applications came to the same conclusion: noting a lack of gender effect across countries, disciplines, and publication years (Marsh et al. 2009).

Research questions

While there is substantial literature on reciprocity and a number of studies investigating social bias in peer review, there is little literature demonstrating the interaction between reciprocity and social bias in the review of scholarly manuscripts. This work, therefore, will investigate the degrees of reciprocity in peer review, by examining two social variables (i.e., gender and nationality). We propose, therefore, a set of research questions to first establish a baseline for our analysis (RQs1–4) and then to examine notions of reciprocity (RQs5–6):

RQ1: Are female-authored manuscripts assigned recommendations in different proportions than male-authored manuscripts?
H1: Female-authored manuscripts will be assigned a higher proportion of rejection and major revision, and a lower proportion of acceptance and minor revision than male-authored manuscripts (based on Wennerås and Wold 1997).
RQ2: Do female reviewers assign recommendations in different proportions than do male reviewers?
H2: Female reviewers will assign a higher proportion of rejection and major revision, and a lower proportion of acceptance and minor revisions than male reviewers (based on Borsuk et al. 2009; Wing et al. 2010).

RQ3: Are Westerner-authored manuscripts assigned recommendations in different proportions than non-Westerner authored ones?
H3: Westerner-authored manuscripts will be assigned a higher proportion of acceptance and a lower proportion of rejection than non-Westerner authored ones (based on Bornmann 2011; Oswald 2008).
RQ4: Do Westerner reviewers assign different proportions of recommendations than do non-Westerners?
H4: Westerner reviewers will assign a higher proportion of acceptance and a lower proportion of rejection than non-Westerners (based on Marsh et al. 2008).
In addition to these questions based in previous research, we also pose two open research questions for which we have no hypotheses, due to the lack of previous research in this area:
RQ5: What differences in degree and kind of reciprocity exist for the proportions of recommendation levels assigned by gender (i.e., in what ways do male and female reviewers as groups each attend to and depart from reciprocal recommendations)?
RQ6: What differences in degree and kind of reciprocity exist for the proportions of recommendation levels assigned by ethnicity (i.e., in what ways do Western and non-Western reviewers as groups each attend to and depart from reciprocal recommendations)?

Audience

This study has both narrow and broad implications, impacting several different kinds of audiences. In terms of the specific findings generated from *JASIST* in the current study, the information science community stands to benefit in knowing about the specific social patterns that exhibit in recent cycles of manuscript submission at a leading journal in the field (Nisonger and Davis 2005; Manzari 2013). This can have implications insofar as it might suggest persistent issues regarding cultural standards of academic writing and cultural biases toward certain kinds of authors, research, or styles (this study remains agnostic to the specific kinds of trends or preferences these patterns suggest, choosing instead to invite investigation of the particular motivations for recommendations to future research). The novel method proposal for studying deriving levels of reciprocity can be applied by scientometric researchers to other fields and communities. Our objective in this study is not to examine a single reviewer and how their experiences have shaped their behaviors, but rather to examine—on the macro scale—how groups perform in potentially conditioned ways. We provide a method that does not require intimate knowledge of each individual in the system, but allows for the use of proxies and, thereby, large-scale analysis. More broadly, this study's proposed focus on reciprocity is useful for scholarly communities interested in how the wider social milieu affects the peer review process and by extension scholarly communication overall; thus, scholars of the sociology of science and scholarly communication stand to benefit, as do the editorial entities of academic journals that employ the peer review process.

Methods

Data

The entire population of reviews submitted between June 2009 and May 2011 to *JASIST* was included in this study. At the time of the study, *JASIST* used a single-blind peer-review process (i.e., the reviewers were aware of the authors' names, but reviewer names were not

provided to the authors). There were a total of 927 reviews (comments on the manuscripts) associated with 450 unique manuscripts. Each review was accompanied by a recommendation: "accept", "accept with minor revisions" (hereafter simply "minor revision"), "accept with major revision" (hereafter "major revision"), and "reject". Each manuscript was associated with an author or set of authors and each review was associated with a named reviewer.

Name-gender and name-ethnicity association

Full name information and institutional affiliations were used to identify unique names. In total, there were 1,393 unique author and reviewer names associated with the dataset. Of these, 148 names (or 10.6 % of the names) were verified as being both authors and reviewers, 781 (56.1 %) were authors only, and 464 (33.3 %) were reviewers only. In coding names by gender and ethnicity, we chose to employ perceived rather than personally declared values, to simulate a reviewer's perception of the author (and in the case of reviewers, to simulate their previous perception as authors, that might affect their own recommendation patterns).

Names were coded for perceived gender using data derived from the U.S. Census Bureau Decennial Census in combination with the 1990 Post-Enumeration Survey ("Genealogy Data", 1990). Data about the frequency of first names per reported sex were derived from 6.3 million census records cross-referenced by the Census Bureau between the two surveys; these frequencies were then reported tabularly (by the Census Bureau) for the 1219 most frequently occurring male first names and the 4275 most frequently occurring female first names. Based on the resulting lists, if a first name appeared in one list but not both, the gender of the list in which the name appeared was assigned. In cases where the name appeared on both gendered lists, the first name in each record in our dataset was coded as the gender with which they were more highly associated (e.g. the name "Pat" is ranked 387th for females and 457th for males, and so would be coded as female in all cases). This logic accounted for 742 of the unique records (431 males, 311 females), or 53.3 % of the total data.

Names that did not appear on either of the gendered lists of the 1990 Census data were then determined using the online tool "Baby Name Guesser" (Peters 2013). The "Baby Name Guesser" algorithm determines gender associations for names by counting co-occurrences of names with gendered terms and abbreviations (e.g., a name that repeatedly follows "Mr." would be considered to be more likely male than female). Using this secondary coding approach, an additional 415 names were coded as male, and an additional 163 were coded as female, resulting in total counts of 846 perceived male records (or 60.7 %), 474 perceived female records (or 34 %), and 73 (or 5.2 %) uncategorized records.

The U.S. Census Bureau states that "although censuses are a source of genealogical information, the Census Bureau does not provide these data," ("Genealogy main"); thus *perceived* ethnicity was identified using a combination of multiple online databases ("Last name meanings and origins, 1990"; "Surname database"; "Family name history"). Results from "Last name meanings and origins" are based on *Dictionary of American family names* published by Oxford University Press in 2006, which contains more than 70,000 of the most commonly occurring surnames in the United States, giving their comparative frequencies, linguistic and historical explanations, selected associated forenames, and occasional genealogical notes. "Surname database" contains 49,352 last names, and is based on a database of surname origins over the last 20 years run by Name Origin Research Group. "Family name history" is provided by Family Education Network, which

Table 1 Mappings of surname origins to regions

Perceived cultural origin	Categories of regions
English, Dutch, Irish, Scottish, Finnish, Greek, Italian, Spanish, Portuguese, Hebrew, Jewish, etc.	Western
Russian, Slovak, Hungarian, Polish	Eastern European
Chinese, Japanese, Korean	East Asian
Indian, Thai, Philippine	South Asian
Arabic, Muslim, Iran, Turkish	Middle East
Nigerian	African

was launched in 1996. Its name database is based on lexicographer Patrick Hanks' (1989) work, *A Dictionary of Surnames*, which was published by Oxford University Press and contains 100,000 entries coming from throughout Europe, Australia, North America, and the British Isles.

Surnames were checked across all three databases and an aggregate list of possible origins was generated. These were then grouped into regional categories as derived from inter-coder consensus, and largely reflect geographic origins of the last century or so, with the exception of Jewish and Hebrew; due to the multi-regional historical range of these cultural groups, these cultural origins were grouped under Western, reflecting the large Jewish populations of North America and Western Europe as well as the cultural affinity of Israel with the West. While ideally a survey of academics' perceptions of cultural grouping by region would be conducted, reaching a consensus among coders allows for some substantiation of these groups based on members of the disciplinary culture under study. Table 1 supplies a comprehensive list of perceived cultural origins mapped to regions. The modal region was selected for final coding. For example, "Jacob" was described as a Jewish, English, German, Portuguese, French, Dutch, or southern Indian across the three databases. It was therefore coded as Western (rather than South Asian). If a name only appeared in one database, the information from this database was used.

Using this coding approach, 861 names were coded as Western (61.8 %), 38 were Eastern Europe (2.7 %), 357 were East Asian (25.6 %), 61 were South Asian (4.4 %), 71 were Middle East (5.1 %), and 5 were African. Given the low occurrence of non-Western names, these were later merged to Western and non-Western for analysis.

Analysis

Manuscript reviews were then divided into two sub-groups: one containing single-authored papers (297 reviews) and one containing multiple-authored papers (639 reviews). Due to the added complexity of multiple-authored papers, two measurements were used for this subset. The first was a measure of presence and absence of categories—for example, categories such as no Western author versus one or more Western author and no female versus one or more female author. The second measure was of categorical majorities; for example, majority of Western authors versus a majority of non-Western authors and majority of female authors versus a majority of male authors. Pearson's Chi squared test was used to investigate potential inequalities in the distribution of gender and ethnicity across review recommendation levels (accept, minor revision, major revision, and reject), and Cramer's V values for each χ^2 value were derived to present effect sizes in each case.

Due to the large number of comparisons tested for significance in the current study, we applied a Bonferroni correction to our interpretation of the significance of Pearson's Chi squared test P values, i.e., $\alpha = 0.05/40$ (given five tests for gender and five for ethnicity, for single authors, majorities of multiple authors, presence/absence in multiple authors, and reviewers). This yielded a conservative threshold for significance which no Pearson's Chi squared results surpassed. However, as the sample size would need to be much greater to generate significant results, we orient our interpretation of results toward Crmaer's V values.

Ideally one would measure the level of reciprocity by collecting the recommendations assigned to manuscripts for each member of a group under study (e.g., female authors), as well as the recommendation levels assigned by these same authors in their capacity as reviewers (to manuscripts authored by males and females). However, analyzing only matching authors and reviewers would greatly diminish the sample size. Therefore, we applied Monte Carlo resampling without replacement to identify sampling cohorts. Specifically, for each combination of demographic group and frame of analysis, 1,000 iterations of the following sampling and analysis were conducted, with distributions of the resulting statistics then taken and analyzed in aggregate.

Each iteration began by resampling each of the two subgroups under study (either male and female, or Western and non-Western) without replacement from the overall study sample. If a sampled review had an author matching the specific category in the frame under study (e.g., female gender for single-authored papers) but a reviewer who does not match, the recommendation level of this sampled review was added to the distribution for authors. Similarly, if the sampled review had a reviewer belonging to the category (but no matching author), the associated recommendation level was added to the distribution for reviewers. In circumstances where both the author and the reviewer matched the category, a number between 0 and 1 was randomly generated, and if the random number exceeded a split value, the sampled review's recommendation level was added to the reviewer distribution (whereas if it did not exceed the split value, the recommendation level was added to the author distribution). The split value was derived from the proportion of authors and reviewers in a given subgroup [Author Count/(Author Count + Reviewer Count)].

To account for variations both in the sizes of author and reviewer distributions within studied categories and in the amount of overlap (the proportion of reviews with both an author and a reviewer belonging to the same category), a total subsample size was derived from the split value, the total number of reviews with author and reviewer from the subgroup being tested (referred to as the AuthRev Count), and the count of either authors or reviewers in the category being tested (choosing whichever was smaller). When there were fewer authors than reviewers from a given subgroup, this total subsampling size was derived as follows:

$$\text{Total subsample size} = \text{Author Count} - [\text{AuthRev Count} * (1 - \text{split value})]$$

When there were fewer reviewers than authors, the total subsample size was derived as follows:

$$\text{Total subsample size} = \text{Reviewer Count} - (\text{AuthRev Count} * \text{split value})$$

In order to allow for comparisons between subgroups (i.e., between male and female gender categories and Western and non-Western ethnic categories), the smaller of the two subgroups' total subsample size was used for both subgroups (e.g., the total subsample size for female single authors and reviewers was 69, while the total subsample size for male

single authors and reviewers was 116; thus a total subsample size of 69 was used for both subgroups). Table 2 provides the resulting n-values for subsamples in each analytic frame.

After accumulating subsamples in a given iteration, the subsamples' distributions of recommendation levels were then analyzed using Pearson's Chi squared test, comparing the distribution of author recommendations to the distribution of reviewer recommendations under the assumption that similarity between author and reviewer distributions would thus be quantified. The resulting χ^2 values (along with the P values) were in turn collected for all iterations, aggregated into distributions, and compared between subgroups. We then iteratively calculated sets of Cramer's V values based on the sets of χ^2 values, and due to non-normal distributions supply the median Cramer's V values as summary statistics of effect size for purposes of comparison.

Limitations

While the findings may suggest patterns proposed by the first set of hypotheses (as well as implying patterns in answer to the second, descriptive set of research questions), several limitations to the current study exist which curtail the certainty of such an assertion, including the size of the dataset (only 927 reviews and 1,393 unique names), the number of journals analyzed (only one journal), the basis of analysis for reviewers (e.g., perceived ethnicities and gender, rather than actual collected from curricula vitae and other background data), and ambiguity of where and how JASIST gets its pool of reviewers: JASIST reviewers, while selected by the editorial staff on the basis of subject matter expertise within the field, are not necessarily previously published authors in JASIST. Additionally, the current presentation of findings leaves aside any interactions between reviewer and submitting author categories (e.g., inter-gender vs. intra-gender effects). Further research is required to overcome these limitations, to support the theoretical foundations and to confirm empirical findings, using a larger dataset from diversified journals in LIS, as well as to interpret these differences in light of larger socio-cultural contexts. Ideally the expanded set of LIS journals would include those with double- as well as single-blind reviewing policies, thereby allowing a consideration of whether the double-blind review model might alleviate bias in peer review, or if not, whether empirical evidence might suggest that female and non-Western authors use different enough writing styles to be recognizable and penalized even when their work is blind reviewed. For that matter, in an ideal study of reciprocity individual scientists would be followed in their capacities as authors and reviewers, and their histories in each role taken into specific account. However in light of these limitations it is important to emphasize the strength and intention of the current study, namely that it serve as a case study and demonstration of the proposed method of analysis, a shot over the bows of this new angle on peer review research.

Results

Gender

Table 3 presents counts and proportions of recommendation levels along with Pearson's Chi squared test results and Cramer's V to measure effect sizes for single- and multiple-authored manuscripts by author and reviewer gender, along with total counts per gender category. Papers authored singly by females received proportionally more accepts and

Table 2 *N* values for subsamples (gender and ethnicity)

	Gender	Ethnicity
Single	69	43
Category present	129	81
Category absent	39	127
Category majority	73	111

major revisions than male single-authored papers and fewer minor revisions and rejections, although no comparisons were found to be significant baed on Pearson's Chi squared test.

A higher proportion of papers were assigned major revisions and rejections among those with at least one female author or primarily authored by women than those written without any women or primarily men. Furthermore, the difference in rates of acceptance as-is between papers with majority female authors versus male authors, while not found to be significant by Pearson's Chi squared test, did yield a Cramer's *V* value of 0.0982—the highest Cramer's *V* value of all gender comparisons. Regarding reviewer data, male reviewers assigned acceptances and minor revisions at a greater rate than female reviewers (and minor revisions and rejections at lesser rates). While the Pearson's Chi squared test comparing male and female reviewers' proportions of major revision did not find significant differences, the Cramer's *V* (0.0818) was the second highest of all gender tests.

Ethnicity

Table 4 presents counts and proportions of recommendation levels for single- and multiple-authored manuscripts by author and reviewer ethnicity, as well as Pearson's Chi squared test results and Cramer's *V* values. Papers authored singly by Westerners received a higher proportion of accepts, minor revisions, and major revisions, while those singly-authored by non-Westerners received a higher proportion of rejections as did multiply-authored papers written primarily by non-Western authors; Cramer's *V* values for acceptance (0.0859) and rejection (0.0841) comparisons between single authors by ethnicity suggest sizable effects. Papers with at least one Western author present received proportionally more accepts and minor revisions than those with no Western authors, while those with no Western authors received proportionally more major revisions and rejections, with relatively stronger effect size for the rejection comparison (0.0775). This same pattern plays out between papers with a majority of Western authors and those with a majority of non-Western authors, with rejection proportions generating the highest Cramer's *V* value for either gender or ethnicity comparisons (0.15). Turning to reviewers, Westerners assigned proportionally more acceptances, minor revisions, and rejections than did non-Western reviewers, while non-Western reviewers assigned proportionally more major revisions.

Reciprocity

Five different recommendation level groupings were tested. First, all four recommendation levels were tested individually against one another. Next, each recommendation level was tested against the other three grouped together (e.g., minor revision vs. accept, major revision, and reject combined). Due to the exceedingly low counts of recommendations to accept without revision, test results could not be generated for accept vs. non-accept; consequently, a three-category comparison was conducted among (1) acceptance combined

Table 3 Counts, percentages, Pearson's Chi squared test results, and Cramer's *V* by recommendation, for single- and multiple-authored paper gender measures

				Accept	Minor revision	Major revision	Reject	Total (count only)	All Rec Lvls (Acc + Min, Maj, Rej)
Authors	Single-authored papers	Author gender	Male	8.46 % (17)	28.86 % (58)	33.83 % (68)	28.86 % (58)	201	
			Female	12.09 % (11)	24.18 % (22)	36.26 % (33)	27.47 % (25)	91	
			Cramer's *V*	0.0570	0.0486	0.0234	0.0143		0.0719
			χ^2 (*P* value)	$\chi^2 = 0.95$ ($P = 0.3292$)	$\chi^2 = 0.69$ ($P = 0.4062$)	$\chi^2 = 0.16$ ($P = 0.6892$)	$\chi^2 = 0.06$ ($P = 0.8065$)		$\chi^2 = 1.51$ ($P = 0.6797$)
	Multiple-authored papers	Female authors (present/absent)	Absent	8.97 % (21)	31.2 % (73)	32.05 % (75)	27.78 % (65)	234	
			Present	6.92 % (27)	28.97 % (113)	35.64 % (139)	28.46 % (111)	390	
			Cramer's *V*	0.0373	0.0233	0.0367	0.0069		0.0373
			χ^2 (*P* value)	$\chi^2 = 0.87$ ($P = 0.3510$)	$\chi^2 = 0.34$ ($P = 0.5541$)	$\chi^2 = 0.84$ ($P = 0.3594$)	$\chi^2 = 0.03$ ($P = 0.8625$)		$\chi^2 = 0.87$ ($P = 0.6558$)
		Majority gender	Male	7.85 % (27)	32.27 % (111)	33.72 % (116)	26.16 % (90)	344	
			Female	2.02 % (2)	29.3 % (29)	40.4 % (40)	28.3 % (28)	99	
			Cramer's *V*	0.0982	0.0265	0.0584	0.0202		0.1095
			χ^2 (*P* value)	$\chi^2 = 4.27$ ($P = 0.0388$)	$\chi^2 = 0.31$ ($P = 0.5777$)	$\chi^2 = 1.51$ ($P = 0.2191$)	$\chi^2 = 0.18$ ($P = 0.6714$)		$\chi^2 = 5.31$ ($P = 0.1505$)
Reviewers	Reviewers - all papers	Reviewer gender	Male	9.57 % (53)	30.69 % (170)	31.41 % (174)	28.34 % (157)	554	
			Female	6.12 % (20)	25.99 % (85)	39.45 % (129)	28.44 % (93)	327	
			Cramer's *V*	0.0605	0.0500	0.0818	0.0011		0.0976
			χ^2 (*P* value)	3.22 ($P = 0.0727$)	2.2 ($P = 0.1380$)	5.89 ($P = 0.0152$)	0.001 ($P = 0.9748$)		8.39 ($P = 0.0386$)

Bolded/shaded numbers denote statistically significant results given $\alpha = 0.05$, and Bonferroni correction

Table 4 Counts, percentages, Pearson's Chi squared test results, and Cramer's *V* by recommendation, for single- and multiple-authored paper gender measures

				Accept	Minor revision	Major revision	Reject	Total (count only)	All Rec Lvls (Acc + Min, Maj, Rej)
Authors	Single-authored papers	Author ethnicity	Western	11.9 % (21)	27.1 % (48)	33.9 % (60)	27.1 % (48)	177	
			Non-Western	6.8 % (8)	26.3 % (31)	33 % (39)	33.9 % (42)	120	
			Cramer's *V*	0.0859	0.0142	0.0142	0.0841		0.1090
			χ^2 (*P* value)	$\chi^2 = 2.19$ ($P = 0.1389$)	$\chi^2 = 0.06$ ($P = 0.8065$)	$\chi^2 = 0.06$ ($P = 0.8065$)	$\chi^2 = 2.10$ ($P = 0.1473$)		$\chi^2 = 3.53$ ($P = 0.3168$)
	Multiple-Authored papers	Western authors (Present/Absent)	Present	38 (9.97 %)	153 (32.08 %)	164 (34.38 %)	122 (25.58 %)	477	
			Absent	11 (6.83 %)	39 (24.22 %)	56 (34.78 %)	55 (34.16 %)	161	
			Cramer's V	0.0173	0.0692	0.0037	0.0775		0.0704
			χ^2 (*P* value)	$\chi^2 = 0.22$ ($P = 0.6390$)	$\chi^2 = 3.53$ ($P = 0.0603$)	$\chi^2 = 0.01$ ($P = 0.9203$)	$\chi^2 = 4.43$ ($P = 0.0353$)		$\chi^2 = 3.66$ ($P = 0.3$)
		Majority ethnicity	Western	8 % (6)	33.3 % (25)	37.3 % (28)	21.3 % (16)	75	
			Non-Western	4.4 % (3)	27.5 % (19)	33.3 % (23)	34.8 % (24)	69	
			Cramer's *V*	N/A	0.0629	0.0417	0.1500		0.1594
			χ^2 (*P* value)	N/A	$\chi^2 = 0.57$ ($P = 0.4503$)	$\chi^2 = 0.25$ ($P = 0.6170$)	$\chi^2 = 3.24$ ($P = 0.0719$)		$\chi^2 = 3.66$ ($P = 0.3$)
Reviewers	Reviewers - all papers	Reviewer ethnicity	Western	8.9 % (61)	29.8 % (210)	32.5 % (229)	29.1 % (205)	705	
			Non-Western	7.2 % (16)	26.6 % (59)	39.2 % (87)	27.0 % (60)	222	
			Cramer's V	0.0225	0.0301	0.0604	0.0194		0.0517
			χ^2 (*P* value)	$\chi^2 = 0.47$ ($P = 0.4976$)	$\chi^2 = 0.84$ ($P = 0.3594$)	$\chi^2 = 3.38$ ($P = 0.0660$)	$\chi^2 = 0.35$ ($P = 0.5541$)		$\chi^2 = 2.48$ ($P = 0.1156$)

Bolded/shaded numbers denote statistically significant results given $\alpha = 0.05$, and Bonferroni correction

with minor revision, (2) major revision, and (3) rejection. Finally, we compared the two more positive recommendations combined (accept and minor revision) to the two more negative recommendations combined (major revision and reject).

Single authors

Table 5 presents median Cramer's *V* values for single authored papers' recommendations as compared to the recommendations of reviewers belonging to the same gender or ethnicity. For both genders and ethnicity groupings, the 3X2 comparison of author-received and reviewer-given recommendations for "high" (accept and minor revision combined), major revision, and rejection yielded the highest Cramer's *V* values; notably, females had stronger associations between role status and recommendation level distribution than did males, and non-Westerners more so than Westerners, suggesting that each category commonly perceived as more powerful had less difference between proportions of recommendation levels given and received than its less powerful counterpart.

Considering 2X2 comparisons, in gender the strongest association between role status (author or reviewer) and recommendation proportion for females and males both is between "high" recommendations (accept + minor revision) and "low" (major revision + rejection), with a Cramer's *V* value of 0.1941 for females (reflecting fewer "high" recommendations received than given), and 0.1461 for males (reflecting more "high" recommendations given than received).

In the ethnicity comparisons, median Cramer's *V* values were notably lower for almost all categories than for either of the genders, suggesting closer alignment between proportions of recommendations given and received per ethnicity. As with gender, 2X2 comparisons with the highest Cramer's *V* values for Westerners is between "high" and "low" recommendations (0.1101), while for non-Westerners the highest Cramer's *V* value is for Major revision versus non-Major revision, reflecting the tendency of non-Western reviewers to assign the Major Revision recommendation level disproportionately compared to the amount received.

Multiple authors

Table 6 presents median Cramer's *V* values for multiple authored papers' recommendations with majorities of each gender and ethnicity group as compared to the recommendations of reviewers belonging to the same gender or ethnicity.

Majorities

Table 6 presents median Cramer's *V* values for multiple authored papers' recommendations with majorities of each gender and ethnicity group as compared to the recommendations of reviewers belonging to the same gender or ethnicity. As with single-authored papers, 3X2 comparisons contain the largest Cramer's *V* for each of the categories. Unlike the single-author-based comparisons however, both pairs of categories were found to have relatively similar Cramer's *V* values. In considering 2X2 comparisons for gender, the highest Cramer's *V* value for males was for Major Revision versus non-Major Revision (0.1307—reflecting more major revision recommendations received than given), while for females the largest Cramer's *V* value was for Rejection versus non-Rejection (0.1061—reflecting more rejections given than received). For ethnicity, the strongest differences in

Table 5 Median Cramer's *V* values for single author recommendations versus reviewers of matching gender or ethnicity

	Males	Females	Westerners	Non-Westerners
Accept + Minor, Major, Reject	0.2139	0.2535	0.1223	0.1357
Accept + Minor, Major + Reject	0.1461	0.1941	0.1101	0.1003
Minor, non-Minor	0.1107	0.1099	0.0984	0.0623
Major, non-Major	0.1243	0.1049	0.1049	0.1222
Reject, non-Reject	0.1138	0.1509	0.0998	0.1017

Table 6 Median Cramer's V values for majority author recommendations versus reviewers of matching gender or ethnicity

	Males	Females	Westerners	Non-Westerners
Accept + Minor, Major, Reject	0.1923	0.1712	0.1919	0.2053
Accept + Minor, Major + Reject	0.1177	0.1041	0.1028	0.0825
Minor, non-Minor	0.1061	0.1006	0.1026	0.0827
Major, non-Major	0.1307	0.0991	0.1023	0.1184
Reject, non-Reject	0.1119	0.1061	0.13	0.1646

2X2 comparisons were "high" versus "low" for Westerners (0.1026—reflecting more "high" recommendations received than given) and rejection versus non-rejection for non-Westerners (0.1646—reflecting substantially more rejections received than given).

Presences/Absences

Table 7 presents Cramer's *V* values for multiple authored papers' recommendations for papers with at least one author from a given gender or ethnicity group as compared to the recommendations of reviewers belonging to the same gender or ethnicity. As with other analytic frames, 3X2 comparisons had the highest Cramer's *V* values for each category; as with single author-based comparisons, those in Table 7 based on female or non-Western author groups were found to have markedly higher Cramer's *V* values than their male or Western-based counterparts. Of the 2X2 comparisons for gender, rejection versus non-rejection was found to have the highest Cramer's *V* for papers authored by no females as compared to male reviewers' recommendations (0.0784—with slightly higher proportions given than received), while for papers with one or more female authors as compared to female reviewers' recommendations, major revision versus non-major revision had the highest Cramer's *V* (0.1022—also with more given than received). Considering 2X2 comparisons for ethnicity, papers with no Western authors as compared to non-Western reviewers were found to have the highest of all Cramer's *V* values for 2X2 comparisons, with (0.1938 for rejection vs. non-rejection, reflecting more received than given). Finally, for papers with one or more Western authors as compared to Western reviewers, rejection vs. non-rejection also had the highest Cramer's *V* (0.0848—reflecting more rejections given by Western reviewers than received by papers with one or more Western authors).

Table 7 Median Cramer's *V* values for author recommendations based on presence/absence of a category vs. reviewers of paired gender or ethnicity

	Female (Absent)/ Male reviewers	Female (Present)/ Female reviewers	Westerner (Absent)/ Non-W reviewers	Westerner (Present)/ W reviewer
Accept + Minor, Major, Reject	0.1278	0.1633	0.2213	0.1175
Accept + Minor, Major + Reject	0.0732	0.0902	0.103	0.0607
Minor, non-Minor	0.077	0.085	0.0873	0.061
Major, non-Major	0.0775	0.1022	0.0831	0.0612
Reject, non-Reject	0.0784	0.095	0.1938	0.0848

Discussion

We begin by addressing each of our research questions and hypotheses in light of our results, before stepping out into the wider theoretical implications.

Author gender

Our first research question investigated whether female-authored manuscripts received significantly different recommendations from male-authored papers, hypothesizing that female-authored papers would be assigned a higher proportion of rejection and major revisions compared to male-authored manuscripts. Data from singly-authored papers provide extremely little support for this hypothesis (with female-authored papers receiving greater proportions of acceptance and major revision, and male-authored papers receiving more minor revisions and rejections). However, multiply-authored papers provide some support: multiply-authored papers *without* female authors are recommended to be accepted as-is or with minor revisions in greater proportion than papers *with* one or more female authors and male-majority authored papers were recommended for acceptance and minor revision at a greater rate than papers with a majority of female authors (although neither of these differences among multiple-authored papers was found to be statistically significant).

Reviewer gender

Our second research question investigated differences from the perspective of the reviewer's gender. We hypothesized that female reviewers would provide lower recommendations (i.e., rejections and major revisions) compared to their male counterparts. Moderate support was found for this hypothesis; while female reviewers were found to assign major revision and rejection in greater proportion (and acceptance and minor revision in lesser proportion) than male reviewers, no differences were found to be statistically significant.

Author ethnicity

In parallel with the research questions on gender, we examined differences in recommendations received by perceived ethnicity. We anticipated that Western-authored

manuscripts would receive more favorable reviews than non-Western authored manuscripts. Data from both single-authored and multiple-authored papers provide some support for this hypothesis. Westerners are assigned higher proportions of acceptance and minor revision and lower proportions of rejection for singly-authored papers, as are multiple-authored papers with Western authors present (vs. those with no Western authors) and majorities of Western authors (vs. those with majorities of non-Western authors). While no comparisons of Western authors present versus absent were found to be statistically significant given $\alpha = 0.05/40$ (after Bonferroni correction), the overall trend among the different frames of analysis is suggestive.

Reviewer ethnicity

We also sought to identify differences in the practices of reviewers of varying perceived ethnicities. We surmised that Western reviewers would assign more acceptances than non-Western reviewers. Reviewer data provide some support for this supposition (although no differences were found to be statistically significant): Western reviewers were found to give acceptance and minor revision in higher proportion than non-Western reviewers, and major revision in lower proportion. However, Westerners were also found to give rejection recommendations in higher proportion than non-Westerners as well, confounding one of the main conjectures of the hypothesis.

Reciprocity

These findings suggest that reciprocity is present in the peer review process. Regarding gender as a social frame, it can be seen that papers with no female authors receive a higher proportion of higher recommendation levels than do papers with female authors, and the same discrepancy is true between papers with a majority of male authors and a majority of female authors. In turn, female reviewers are more likely than their male counterparts to assign lower recommendation levels (rejection or major revision), and less likely to assign higher recommendation levels (acceptance or minor revision). That said, this trend in gender is not terribly strong (and results for single-authored papers appear to confound this trend). In considering degrees of reciprocity, findings suggest that female reviewers are likely to give recommendations in the same proportion they receive on multiple-authored papers that have no male authors, or at least a majority of female authors (as opposed to the proportions of recommendations given to single-authored papers).

Turning to ethnicity as a social frame, the pattern of reciprocity is suggested more consistently across all three analytical frames (although not via a direct mapping of reviews received to reviews given); more simply put, Westerners and non-Westerners alike are more reciprocal as groups than are men and women. A higher proportion of rejections and lower proportions of acceptances and minor revisions were recommended for papers authored singly by non-Westerners than by Westerners, for multiply-authored papers without Western authors than with Western authors, and for multiply-authored papers with a majority of non-Western authors than with a majority of Western authors. In two of these three analytic frames, the more non-Western of the two options also received a lower proportion of major revisions (single-authored non-Western and non-Western majority). Non-Western reviewers however assigned higher proportions of major revision than did Western reviewers (and lower proportions of all other recommendations, including rejection). This suggests a reciprocal tendency toward a more negative baseline evaluation—a tendency tempered, one might theorize, by uncertainty, politeness, or an openness

to the benefits of editing based on personal experience (one may be more aware of the salvageability of a text if one has re-tailored a rejected paper into an accepted one).

Limitations

This study sought to serve as a proof-of-concept piece: to propose a new concept in studying bias in peer review and a methodological framework for such studies. However, there are a number of limitations to the present study. The first is the small-scale of the study: this serves as an exploratory study from which the results should not be generalized. Furthermore, there are always concerns with the use of proxy measures—in this case, gender and ethnicity. While validation tests in other studies have shown fairly reliable rates for the use of similar methods for assessing gender (e.g., Lariviere et al. 2013), the proxies for ethnicities are complicated by the nuances of this variable. For example, large countries with a long history of immigration may contain authors with surnames that reflect certain ethnicities, but carry little cultural, racial, or historic relation to the individual.

Conclusion and future research

This study has put forth a novel method for studying an important but overlooked facet of academic peer review—i.e., reciprocity as it exhibits via gender and ethnicity. Via this method, patterns of greater negativity in reviews of female and non-Western authored papers are mirrored in reviews conducted by these same groups (although non-Westerners are more likely to suggest major revision than outright rejection). However, it is also important to note what this study does not interpret. Our first four research questions seek to establish certain patterns of difference between recommendations given and received by genders and ethnicities; however we intentionally do not assess the quality of submissions and reviews per se, and so remain agnostic to whether or not undue bias toward or by certain social groups is at work. We also do not examine the positive effects of reciprocity: that is, the ways in which this conditioning leads to the development of standards in a community of practice. By limiting our emphasis to the alignment and deviations between patterns of reviews received by social groups and patterns of reviews given by those same groups, it is our hope that the methods presented here are shown to facilitate a deeper understanding of trends within academic peer review. This in turn can pave the way for more nuanced studies not only of what differences in reviewing tendencies exist among social groups (for authors and reviewers alike), but how those differences reverberate through disciplinary cultures through time.

Several areas of future research based on the current study suggest themselves. At first blush, the greatest limitation of the current study is the sample size; while the use of Monte Carlo sampling helps to expand the utility of the data available, a data set composed of multiple journals and more authors and reviewers distributed over a greater time period (or even more ambitiously, more data that connects the history of reviews received and reviews given for the same people) would greatly help in fleshing out the current study's findings. An expanded sample size could also allow for the investigation of cross-effects between gender and ethnicity as social factors, both in the intersection of gender and ethnicity in perceived identities of authors and reviewers and in the interaction between author's and reviewer's social identities.

More specific to the dataset of the current study, another future line of research could seek to account for the interdisciplinary nature of information science; it is possible that

trends in the number of authors per paper may differ by subject area, reflecting authors' previous academic traditions or the disciplinary context of the subfield (e.g., writing that draws from philosophy or critical studies may be more likely to be produced by a single author as opposed to an experimental study based in psychology). Such a line of research could, one theorizes, eventually yield a framework that could look to connect the currently scattered and isolated studies of peer review in various disciplines into a constellation through the framework of disciplinary cultures.

Acknowledgments The authors would like to thank JASIST Editor-in-Chief Blaise Cronin and Meghann Knowles (JASIST Editorial Office) for generously providing access to the data used in this study.

References

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), 1041–1068.

Blau, P. (1964). *Exchange and power in social life*. New York: Wiley.

Bornmann, L. (2011). Peer review and bibliometrics: Potentials and problems. In J. C. Shin, R. K. Toutkoushian, & U. Teichler (Eds.), *University rankings: Theoretical basis, methodology and impacts on global higher education* (pp. 145–164). Berlin: Springer.

Bornmann, L., & Daniel, H.-D. (2010). Do author-suggested reviewers rate submissions more favorably than editor-suggested reviewers? A study on atmospheric chemistry and physics. *PLoS One, 5*(10), e13345. doi:10.1371/journal.pone.0013345.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2007). Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics, 1*, 226–238.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). How to detect indications of potential sources of bias in peer review: A generalized latent variable modeling approach exemplified by a gender study. *Journal of Informetrics, 2*, 280–287.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2009). The influence of the applicants' gender on the modeling of a peer review process by using latent Markov models. *Scientometrics, 81*(2), 407–411.

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

Borsuk, R. M., Aarssen, L. W., Budden, A. E., Koricheva, J., Leimu, R., Tregenza, T., et al. (2009). To name or not to name: The effect of changing author gender on peer review. *BioScience, 59*(11), 985–989.

Budden, A. E., Tregenza, T., Aarssen, L. W., Koricheva, J., & Leimu, R. (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution, 23*, 4–6.

Casement, P. J. (1991). *Learning from the patient*. New York: Guilford Press.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences of the United States of America, 108*(8), 3157–3162.

Cronin, B., & McKenzie, G. (1992). The trajectory of rejection. *Journal of Documentation, 48*(3), 310–317.

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

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*(2), 293–315.

Family name history. Retrieved from http://genealogy.familyeducation.com/family-names-surnames/meaning-origin

Genealogy data: Frequently occurring surnames from Census 1990 – names files. (1990). Retrieved from http://www.census.gov/genealogy/www/data/1990surnames/names_files.html

Genealogy main. Retrieved from http://www.census.gov/genealogy/www/index.html

Gilbert, J. R., Williams, E. S., & Lundberg, G. D. (1994). Is there gender bias in JAMA's peer review process? *JAMA: The Journal of the American Medical Association, 272*(2), 139–142.

Goldstein, W. N. (1991). Clarification of projective identification. *American Journal of Psychiatry, 148*(2), 153–161.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25*, 161–178.

Hamilton, N. G. (1990). Splitting and projective identification among healthier individuals. *American Journal of Psychotherapy, 44*(3), 414–422.

Homans, G. C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt Brace.
Klein, M. (1946). Notes on some schizoid mechanisms. *International Journal of Psycho-Analysis, 27*, 99–110.
Lariviere, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Global gender disparities in science. *Nature, 504*(7479), 211–213.
Last name meaning and origins. Retrieved from http://www.ancestry.com/learn/facts
Lee, C., 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.
Link, A. M. (1998). US and non-US submissions. *Journal of the American Medical Association, 280*(3), 246–247.
Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies, 60*(3), 531–542.
Manzari, L. (2013). Library and information science journal prestige as assessed by library and information science faculty. *The Library Quarterly, 83*(1), 42–60. doi:10.1086/668574.
Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O'Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Educational Research, 79*, 1290–1326.
Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist, 63*(3), 160–168.
Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.
Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies, 78*(1), 356–393.
Neuroscience, Nature. (2006). Women in neuroscience: A numbers game. *Nature Neuroscience, 9*(7), 853.
Nisonger, T. E., & Davis, C. H. (2005). The perception of library and information science journals by LIS education deans and ARL library directors: A replication of the Kohl-Davis study. *College & Research Libraries, 66*(4), 341–377.
Ogden, T. H. (1997). *Reverie and interpretation: Sensing something human*. Northvale: Jason Aronson.
Oswald, A. J. (2008). Can we test for bias in scientific peer-review. *IZA discussion paper 3665*. Bonn: Institute for the Study of Labor.
Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2002). The personal norm of reciprocity. *European Journal of Personality, 17*(4), 251–283.
Peters, G. (2013). "Baby Name Guesser". Retrieved from http://www.gpeters.com/names/baby-names.php*Surname database*. Retrieved from http://www.surnamedb.com
Thibaut, J., & Kelley, H. H. (1959). *The social psychology of groups*. New York: Wiley.
Valkonen, L., & Brooks, J. (2011). Gender balance in Cortex acceptance rates. *Cortex, 47*, 763–770.
Weller, A. C. (2002). *Editorial peer review: its strengths and weaknesses*. Medford: Information Today Inc.
Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature, 387*(6631), 341–343.
Wing, D. A., Benner, R. S., Petersen, R., Newcomb, R., & Scott, J. R. (2010). Differences in editorial board reviewer behavior based on gender. *Journal of Women's Health, 19*(10), 1919–1923.