



Estudios de Economía Aplicada

ISSN: 1133-3197

secretaria.tecnica@revista-eea.net

Asociación Internacional de Economía
Aplicada
España

Seidl, Christian; Schmidt, Ulrich; Grösche, Peter
The Performance of Peer Review and a Beauty Contest of Referee Processes of Economics Journals
Estudios de Economía Aplicada, vol. 23, núm. 3, diciembre, 2005, pp. 505-551
Asociación Internacional de Economía Aplicada
Valladolid, España

Available in: <http://www.redalyc.org/articulo.oa?id=30123301>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

The Performance of Peer Review and a Beauty Contest of Referee Processes of Economics Journals

CHRISTIAN SEIDL*, ULRICH SCHMIDT** and PETER GRÖSCHE***

Department of Economics, University of Kiel, Germany; **Department of Economics, University of Hannover, Germany, *Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Essen, Germany*

Email: seidl@economics.uni-kiel.de

Comments Editors' note

Estudios de Economía Aplicada applies the peer-review system as a quality-control device to the submitted manuscripts. The editors believe that, in spite of its weakness, peer review is able to correct their subjective judgments and contributes to improve manuscripts and to warrant a better final selection.

It falls into the responsibility of editors of scientific journals that apply peer review to evaluate its actual working and to discuss improvements for its better use.

The editors hold that it is, in particular, important to guarantee

- an adequate and objective selection of the referees,
- double-blind peer review,
- the absence of favoritism and ideological biases,
- competent and well-founded referee reports,
- short response time.

Recognizing deficiencies in peer reviewing is a pre-requisite of improving its efficiency. To openly admit that the editors of learned journals, as well as their advisors and referees, are not free from biased views and from pursuing their own interests, can be helpful to control discrimination and subjectivism.

The work of Christian Seidl, Ulrich Schmidt and Peter Grösche, seemed to the editors of *Estudios de Economía Aplicada* an interesting and important contribution to the necessary development and reform of peer-review procedures.

In publishing this paper as an article in this issue of *Estudios de Economía Aplicada*, the editors decided to overrule a previous negative peer-review assessment of one of the top ten learned economics journals.

ABSTRACT

Peer review influences decisively personal advancement, research opportunities, salaries, grant-funding, promotion, and tenure. It claims to exert quality control of manuscripts and improve them, to promote innovative research, foster dissemination of new research, and to serve as a means to rank researchers, journals, and institutions. Nowadays per review mainly serves the purpose of imprinting a signal of quality on a scholar's research. This requires a perfect performance of peer review. However,

JEcLit Number: B40,A11.

Artículo recibido en mayo de 2005 y aceptado para su publicación en octubre de 2005.

Artículo disponible en versión electrónica en la página www.revista-eea.net, ref.: E-23301.

empirical research has shown that peer review lacks validity, impartiality, and fairness, which makes its claim to imprint manuscript excellence dubious. This is demonstrated in the first part of this paper which surveys peer-review research across all disciplines.

In the second part of this paper, we report on an internet investigation conducted among economics authors. We found that there is a group of some eight top economics journals which fall at the bottom in most rankings. Moreover, we found that authors appreciate competence and carefulness of referee reports even more than manuscript acceptance.

Keywords: Peer review, Referee processes, Publishing, Economics journals.

This paper provides, in its first part, a very concise survey on the basic problems and the performance of peer review in general.¹ In its second part, we give an account on an internet survey which we conducted among economists for economics journals.

1. BASIC PROBLEMS AND THE PERFORMANCE OF PEER REVIEW

1.1 Publish or Perish: The Central Role of Peer Review

Scientific journals were originally launched in the seventeenth century to speed up the publication process because printed books, the most acceptable way until then to publish new research findings, was time-consuming (Szenberg 1994, p. 305; Kronick 1990; Burnham 1990). In our time, the role of publications has fundamentally changed. Beyond serving the dissemination of new scientific ideas and discoveries, they have become the main vehicles for personal advancement, research opportunities, salaries, grant-funding, promotion and tenure. Getting research published is decisive for the fate of both individual researchers and entire academic departments (Peters and Ceci 1982a, p. 187; Mahoney 1982, p. 220). Unsatisfactory publication records may cause the extinction of academic careers of individual scholars and, in terms of universities and their departments, damaged their reputation, impair a department's grant-funding abilities.

Whenever the stakes are so high, all kinds of manipulation, which concern the performance of peer review, are around the corner. The challenge of fair and impartial peer review has, therefore, been taken up by many disciplines (Cf., e.g., Juhasz et al. 1975, pp. 177-8; Speck 1993). The American Medical Association has so far sponsored

1. This part gives only a short account on the literature. A much more comprehensive study on *The Performance of Peer Review: An Interdisciplinary Report*, is presently under elaboration. For a more detailed discussion of the literature we refer to this forthcoming study. For a good survey on the main problems of peer review see also Armstrong (2002).

four international congresses on peer review in biomedical publication.² Concerning the analysis of peer review processes, economics lags behind. Of course, there is Shepherd's (1995) impressive collection and some other seminal research, but, as compared with other disciplines, much research must still be done if economics is to become at par with other disciplines.

In the next section, we shall state the aims of peer review as they are commonly stated in the literature, as well as a fairness requirement. Thereafter, we shall briefly confront the aims and the accomplishment of the fairness requirement with the empirical facts.

1.2 The August Aims of Peer Review and the Fairness Requirement

Peer review has been harnessed to the following ends (Horrobin 1990; Cicchetti 1991):

1. Quality control. It has three facets, to wit:
 - a. Minimization of Type I and Type II errors, i.e., defective papers should not be accepted and good papers should not be rejected (Laband 1990, p. 342).
 - b. Signal of manuscript excellence when it is published in a prestigious journal (Shepherd 1995, p. 134).
 - c. Prevention of plagiarism and fraud to protect intellectual property rights.
2. Quality improvement of manuscripts (Goudsmith, 1969; Laband 1990; Goodman et al. 1994; Hamermesh 1994, p. 156).
3. Impartial promotion of innovative research, in particular, new paradigms.
4. Selection of projects for grants funding. This field of peer review has gained importance in the post-world-war-II era, where the financing of research was shifted to funding.
5. Admission of papers for conference presentation.
6. Quality rankings of researchers, journals, and institutions (Garfield 1972; Hawkins et al. 1973; Malouin and Outreville 1987).

Moreover, when pursuing these goals, peer review should proceed in an impartial and fair way. On this requirement, we concur with Peters and Ceci (1982b, p. 252), who insist:

that science must uphold a fairness doctrine. This, to us, means that everyone should have fair access to journal space and federal funds. Fair is defined here as being judged on the merit of one's ideas, not on the basis of academic rank, sex, place of work,

2. Guarding the Guardians (1990); Second International Congress on Peer Review (1994); Third International Congress on Peer Review (1998), Fourth International Congress on Peer Review (2002).

publication record, and so on. Peer review in science is a tribunal of sorts. It is instrumental in the great ‘sorting’ process which ultimately is linked to the dissemination of professional rewards. It is necessary that we all insist on equal justice in this system. More important, peer review should be used in such fashion that it enhances rather than impedes the progress of science.³

Given the central significance of decent peer review procedures, it is not intelligible that some observers consider violations of fairness in matters of peer review as minor peccadilloes only. However, violations of fairness represent severe offences against proper scientific conduct. Recently, there has been a great stir in Germany, because Richard Hoyzer, a soccer umpire, had manipulated soccer games for the sake of own financial emoluments from bets. He was prosecuted and even arrested because of his betrayal. But the stakes in soccer games are modest as compared to the damage inflicted upon the scientists of whole countries due to unfair peer review processes. It is both a waste of intellectual potential and of millions of taxpayer money for the benefit of small gangs of good buddies.

1.3 Quality Control of Manuscripts

1.3.1 Minimization of Type I Errors and Type II Errors

The paramount aim of peer review is accuracy in identifying the quality of papers. The prevalent measure to assess accuracy of quality identification is *validity*. It is defined as the correlation of the true intrinsic quality of a paper and the quality assessment as it emerged from peer review (Kraemer 1991, p. 153).

Efficient quality control and promotion of innovations should winnow the “good” papers from the “bad” ones. Yet the “bad” papers do not bear an imprint like the mark of Cain; “badness” is an intrinsic characteristic whose revelation can be very tricky and laborious. Referees have nothing but signals to infer the quality of a paper (Ben-Yashar and Nitzan 2001). Here, the validity of peer review comes in. It harnesses expert knowledge to spot “good” and “bad” quality papers. However, referees’ judgments can be fraught with Type I and Type II errors. The former means acceptance of a “bad” paper, the latter rejection of a “good” paper.

Now, both types of error trade off, that is, as one value gets larger, the other gets smaller. If the editor(s) instruct(s) referees to minimize the probability of Type I errors (Beyer 1978, pp. 80-2), then this boosts the probability of Type II errors and, thereby, also the paper rejection rate. If the editor(s) want(s) referees to minimize the probability of Type II errors, then the probability of Type I errors is maximized and

3. This is not undisputed. Bailar (1991, p. 138), for instance, argued that “the fundamental objective of peer review, and of the manuscript selection process in general, is not ‘fairness’ to authors (though this may be a welcome byproduct). It is to improve decisions” (of the editors).

the paper rejection rate is at its minimum. Fixing the paper rejection rate elicits, in turn, the respective consequences.

Both kinds of errors bring editors into disgrace. Their behaviour is governed by the greater disgrace inflicted to them by Type I errors or Type II errors. When Type II errors are intelligible in a clear-cut way by the profession, editors will shy away from Type II errors. This happens in particular for the natural sciences such as physics, chemistry, and mathematics (Cole 1991, p. 140). Indeed, journals in these disciplines have the lowest rejection rates (Zuckerman and Merton 1971; Lazarus 1982; Adair 1982). However, when Type I errors can more easily be identified than Type II errors, editors shy away from Type I errors. This is typically the case for the humanities whose journals exhibit record rejection rates.

Greater aversion of committing Type I or Type II errors, respectively, seems to have given rise to the adoption of different refereeing modes: Low-rejection journals employ sequential refereeing (a second referee report is solicited only if the first is negative), whereas high-rejection journals reject many papers without refereeing at all, and employ simultaneous refereeing for the rest. The editors of simultaneously refereed journals tend to consider a single negative referee report - irrespective of its quality - as a sufficient reason to reject a paper (see also Zuckerman and Merton 1971, p. 78; Bakanic et al. 1990, p. 378; Hargens and Herting 1990b, p. 97).

Referee reports can also be considered in terms of *reliability*, which is simply the correlation between the quality assessments of a paper by different referees (Gulliksen 1950, p. 78; Kraemer 1991, p. 153). Note that high reliability may coexist with low validity (when referees express similar evaluations which suffer from a high error; cf. Gilmore 1991), and low reliability may coexist along with high validity (when different referees realize different true characteristics of a paper).

The large majority of empirical studies evidenced a rather weak reliability among referees.⁴ Some scholars argue that heterogeneous referees are chosen because they provide different background of expertise for the respective manuscript.⁵ However, such improvements in information require prudent editors who are able to digest and process the plentiful information (Stricker 1991; Crandall 1991; Gordon 1982, p. 213). However, referee reliability gains momentum for high-rejection-journals whose editors are panicked at the sheer idea of ever committing a Type I error, and, therefore, tend to give credence to the referee report which is most hurtful to the author. This attitude prevents publication of many deserving papers (Cole 1991). Griffith (1982, p. 215) acknowledged “that low interjudge reliability probably cannot be changed;

4. Ingelfinger (1974, p. 690); Lazarus (1982); Cicchetti (1991, pp. 121-7); Marsh and Ball (1989, p. 153); Mahoney (1977, pp. 170-1); Munley et al. (1988, p. 200); Fiske and Fogg (1990, p. 595); Blank (1991, p. 1059); Bakanic et al. (1987, p. 632).

5. Wilson (1978, pp. 1698-9); Hargens (1988, p. 148); Hargens and Herting (1990a, p.2; 1990b, p.94); Bailar (1991, p. 138); Kiesler (1991, p. 151); Kraemer (1991, pp. 152-3); Roediger (1991, p. 159).

the system becomes a strange one when such low reliability is coupled with very high levels of rejection.”

1.3.2 Peer Review as a Signal of Manuscript Excellence

Curiously enough, journal articles are not being read widely; they serve another purpose. There is more than a grain of truth in Takashi Negishi’s sarcastic remark: “There is a rumor that on the average an academic paper has less than two readers including a referee.”⁶ No economist can ever regularly read even a small subset of the 700 economics journals of the Coupé (2000b) list. The ubiquity of economics journals has caused a shift in their central focus. Whereas half a century ago their main service was the dissemination of research, their main purpose is nowadays to imprint a signal of quality on a scholar’s research. The typical journal subscriber notes that N.N.’s paper appears in the table of contents, and infers that N.N.’s work must be important. After looking for papers concerned with the reader’s own specialty, (s)he casts a glimpse at some abstracts, and subsequently entombs the journal on the shelf, never to be opened again (Shepherd 1995, pp. 134-5).

Towards the end of the 1950’s, the American Psychological Association became interested in the problem of efficient and effective communication of scientific information. In 1960, the National Science Foundation granted funds to the American Psychological Association to carry out research on scientific information exchange in psychology. W.D. Garvey and B.C. Griffith were entrusted with this research, which they pursued over the decade 1960-70. A very condensed report of their findings was published in the *American Psychologist* in 1971 (Garvey and Griffith 1971). We feel that their findings are even today of utmost relevance.

The central message of their findings is that journals perform only a minor role in the dissemination of findings, let alone initiating new research efforts. For instance, for research efforts in psychology, ideas for less than one out of seven originated from sources such as journal articles, presentations at national meetings, etc. (Garvey and Griffith 1971, p. 354). The bulk of communication and dissemination of current research runs over informal outlets such as personal communications, technical reports, discussion papers, and preprints. When it comes to manuscript submission to a journal, the author “is likely to have disseminated his findings on several occasions prior to submitting it to a journal” (Garvey and Griffith 1971, p. 358). As research is an ongoing process for most scholars, they typically consider the research cycle as complete when a manuscript has been submitted to a journal. Meanwhile, new research had already been started by some 70% of researchers (Garvey and Griffith 1971, p. 358).

6. Shepherd (1995, p. 51). Merton (1968, p. 59) argued that in medicine and psychology only 0.5% of published articles are read by one member of the profession.

When the manuscript has toiled through the referee process and is published, the “article, now part of the primary literature of science, becomes a permanent record of completed research; that is, it becomes part of the ‘archives’ of science.” (Garvey and Griffith 1971, p. 359.) As its content has already been communicated to the scholar’s closer scientific community, it is no wonder that at most 1% of all articles are actually being read (Garvey and Griffith 1971, p. 358).

What had been noticed by Garvey and Griffith some four decades ago has since become much more pronounced due to the advent of the electronic age. The web enables researchers to become quickly aware on the most recent research in their field, and hardly any one will wait for two or even more years until important research has eventually been published. Rather than being a means of rapid information, publication in paper journals has become a more or less pompous funeral for a manuscript. It is, taken by itself, questionable whether the publication ceremony is, thus, worth its effort. If, on top of that, peer review lacks impartiality and fairness, the imprint of manuscript excellence becomes dubious, too.

1.3.3 Prevention of Plagiarism and Fraud

In Section 1.4 we will demonstrate that referees are “astonishingly ignorant of what had recently been published in their own journals” (Horrobin 1982, p. 217). Thus, it is no wonder that peer review is largely impotent to prevent plagiarism and fraud. Moreover, retractions and corrections of and comments on published articles demonstrate referee and editor inattentiveness in manuscript review.⁷ Therefore, some editors try to stifle all respective moves.⁸ Thus, the instances of plagiarism and fraud which have, in spite of all that, come to the fore, are but the tip of an iceberg. Even those few cases suffice to demonstrate the impotence of peer review to effectively prevent plagiarism and fraud.⁹

1.4 Quality Improvement of Manuscripts

Peer review can bring about quality improvement of manuscripts if referees spend sufficient time to evaluate a manuscript and if they possess skills that are not inferior as compared with the authors’ expertise. Yet referees do not seem to spend too much

7. This applies to authors as well; cf. Pfeifer and Snodgrass (1990); Horton (2002). Moreover, retractions do not seem to have much effect: Budd et al. (1998) found that retracted articles continue to be cited as valid work in the biomedical literature after publication of the retraction in 94% of the observed cases.

8. Cf., e.g., Broad (1980a,b); Pfeifer and Snodgrass (1990); Friedman (1990); McCutchen (1991, pp. 33-5).

9. Cf., e.g., Shepherd (1995, p. 23); Broad (1980a,b; 1981a,b); Broad and Wade (1982); Steward and Feder (1987); Friedman (1990); Sayre (1975); Manwell and Baker (1979); von Rauchhaupt (2002); Wade (1976).

time on refereeing a paper. Empirical research showed average refereeing times between 37 minutes and 5.4 hours expended on a manuscript.¹⁰

Referee replication studies of published articles have revealed major errors in the manuscripts, which were obviously overlooked by the referees. Although the studies vary, they show that less than one third of all published manuscripts would have passed a careful referee process.¹¹

However, the most spectacular failures of peer review were disclosed by using bogus manuscripts to test referee quality. This method was pioneered by Ross (1979; 1982) for the field of belles-lettres. He typed up Jerzy Kosinski's 1968 novel *Steps* and submitted it, untitled, under the pseudonym Erik Demos to 14 major publishing houses and 13 literary agents. Ross (1982, p. 236) reports: "The highly acclaimed novel, which had won the prestigious National Book Award for fiction in 1969, was rejected by all (including Random House, its original publisher). No one recognized the work, and no one thought it deserved to be published."

Impressed by Ross' study, Peters and Ceci (1980, 1982a) selected 12 top-ranked psychology journals and resubmitted a randomly selected (and slightly camouflaged) article, which had been published between 18 and 32 months ago, as a manuscript to the same journal which had published this article. Referees and editors, taken together 38 persons, were involved in manuscript evaluation. Only three of them (in different journals) actually detected that the respective paper had already been published in the very same journal. This amounts to a rate of 92%. Out of the remaining nine manuscripts, one was accepted and eight were rejected. Peters and Ceci spotted two possible explanations for their results, to wit, (i) less competence of the initial referees, and (ii) systematic bias in favor of authors' status and institutional affiliation. The first explanation can be ruled out on purely statistical grounds, leaving only the latter explanation.

A similar experiment was conducted by Epstein (1990) for 108 social work journals and for 38 journals belonging to allied disciplines. Epstein modified a widely cited article by Purcell et al. (1969), which has had more than 50 citations in the Social Science Citation Index since its publication. The Epstein results were similar to Peters and Ceci's in their devastating findings, but space does not allow dealing with them in greater detail.¹² Meyer (1999, pp. 546-9) conducted a similar experiment of re-

10. Cf., e.g., Jauch and Wall (1989, p. 163); Lock and Smith (1990, p. 1343); Yankauer (1990); McNutt et al. (1990, p. 1373); Nylenna et al. (1994, p. 150). Analyzing the quality of referee reports of 420 manuscripts submitted to the *British Medical Journal*, Black et al. (1998) found only one major characteristic which determined the quality of referee reports, viz. the time spent on a review. Quality improvement was observed for times up to three hours, but not beyond.

11. Cf., e.g., Dewald et al. (1986); Garfunkel et al. (1990a); Hubbard and Vetter (1996); Schor and Karten (1966); Murray (1988); Eichhorn and Yankauer (1987); Evans et al. (1990); Altman (2002).

12. In a follow-up study, Epstein (2004) observed 73.5% inadequate referee reports.

submitting previously published and camouflaged articles to the very same economics journals. Unfortunately, he had to discontinue his experiment after five weeks only because of prosecution by a furious editor of an economic journal who seemed to have reasons to hush up referee incompetence of this journal.¹³

A comprehensive study was carried out by Baxt et al. (1998). In fall 1994, they sent a bogus manuscript to 262 referees of the *Annals of Emergency Medicine*, a 25 years-old journal with a circulation of more than 25,000. "The work was stated to have been a double-blind, placebo control study of the effect of intravenous propranolol on pain in 100 consecutive patients who presented to the emergency department with acute migraine headaches. The work purported to demonstrate that propranolol significantly reduced the pain of migraine compared with placebo." (Baxt et al. 1998, p. 311.) There were ten major and thirteen minor errors placed in the manuscript. 203 reviews were received (response rate 78%), of which 15 recommended acceptance, 67 revision, 117 rejection, and 4 had no clear-cut recommendation. The referees recommending rejection had the highest hit rate of the errors. Yet even this group identified only 39.1% of the major errors and 25.2% of the minor errors. Only 31% of referees identified even a single statistical error, 14.8% of the referees misspelled propranolol throughout their reviews, only 9 referees identified the two existing previous articles on the use of propranolol for migraine headaches,¹⁴ and only 3 referees identified the two references which were fabricated and did not exist in the literature. 68% of the referees failed to realize that the conclusions were not supported by the data.

Direct manuscript improvement due to peer review was studied by Goodman et al. (1994). They asked 44 experts (physicians and epidemiologists) to evaluate the quality of 111 manuscripts accepted by the *Annals of Internal Medicine* from March 1992 to March 1993.¹⁵ Two versions of each manuscript were evaluated, the version originally submitted and the version sent to the printer for publication after all modifications based on peer review, editors' comments, and copyediting. All manuscript versions were reformatted to show identical appearance. Authors' names and affiliations were removed. The two manuscript versions were randomly assigned to different experts. Goodman et al. (1994) supplied their experts with a questionnaire consisting of 34 items on various criteria of manuscript evaluation.

Goodman et al. (1994, pp. 13-4) observed 33 improvements of the 34 items after peer review and editing. Alas, although they bravely championed manuscript improvement due to peer review, only four items showed statistically significant

13. Mahoney (1982, p. 221) reported that in one of his studies of the peer-review system "the emotional intensity and resistance of several participants were expressed in the form of charges of ethical misconduct and attempts to have me fired."

14. The bogus manuscript had stated that this study represented the first use of propranolol for the treatment of acute migraine headaches, in spite of previous published work by Fuller and Guiloff (1990) and Banerjee and Findley (1991).

improvements. Moreover, with respect to the 32 multi-refereed manuscripts, interreferee reliability was low. The intraclass correlation coefficients were 0.12 for the average score and 0.02 for the percentage scores (Goodman et al. 1994, p. 14). These results show that the case for manuscript improvements due to peer review can enjoy but feeble support. If it is present at all, its effect is very modest.

In spite of the long spells it takes for authors to receive reactions to their manuscripts (at least in the humanities), referees do not expend much time to review a paper, seem to be markedly less, rather than more, skilled than the authors of the manuscripts, and, on top of that, indulge in considerable carelessness. Notice that we do not contend that this picture holds for any particular referee, but that it characterizes the average referee pattern.¹⁶

1.5 Impartial Promotion of Innovative Research

1.5.1 Avoidance of Type I Errors Impedes Innovation

When referees and editors try to minimize Type I errors, they inevitably maximize Type II errors, which strongly discriminates against innovative papers whose novelty is recognized by a minority of referees only. When this meets with editorial attitudes to reject manuscripts in the presence of a single negative referee report, it becomes obvious that innovative papers encounter immense difficulties getting published. This is, in particular, the case for the humanities,¹⁷ although the literature is also glutted with Type II errors in the natural sciences.¹⁸

This implies, conversely, a simple recipe to get one's papers published: The paper "should be written in one day, with one idea or less, it should fit well in the literature, and in particular in the literature with the editor or the potential referees in it." [Chichilnisky in Shepherd (1995, p. 57).] In response to this environment, authors concentrate mainly on methodological aspects of manuscripts on the traditional topics, rather than directing their effort on innovative research. Drawing on his experience

15. They had to rely on the judgment of experts different from the original referees to investigate whether manuscripts had actually improved due to revision initiated by the referee reports.

16. Horrobin (1982, p. 217), past editor of two journals, *Prostaglandines and Medicine and Medical Hypotheses*, remarked that, from his own experience, about one-third of referee reports were accurate as they report on important issues and are fair in their recommendations. About one-third are accurate but obsessed with the trivial and recommend revision or rejection on inadequate grounds, and about one-third are objectively inaccurate. He adds: "What constantly astonishes me is the intemperate language in which many reports in the last two categories are couched. The lack of sound judgment among people who have the fate of science and the lives of others in their hands is appalling."

17. Cf., e.g. Chichilnisky in Shepherd (1995, p. 57): "in my experience, the more innovative and interesting the paper, the more likely it is to be rejected."

18. Cf., McCutchen (1991, p. 31); Ingelfinger (1974, p. 691); Yalow (1978; 1982, p. 244).

as an editor of a major economics journal, Ellison has presented a formal model to explain the declining importance of a paper's main contribution in favor of its technical perfection, due to extensive reviewing. "Researchers react by spending less and less time developing new insights and more and more time padding and polishing papers." (Ellison 2002b, p. 1024.)

However, both referees and editors would strongly deny that they are inimical to innovative research – and we may well assume that they do their best not to impede it. However, being strongly averse to Type I errors, they look for obvious and cheap signals of paper excellence. Several signals stand out, viz. personal knowledge of the author, the prestige of the author's institution, and affirmative results of the manuscript. The first two signals give rise to favoritism, the last one represents a confirmatory bias.

1.5.2 Favoritism

Editors are glutted with manuscripts (Goodstein 1982, p. 213). They neither have enough time to screen them thoroughly, nor engage a sufficient number of referees to do it for them. Having thus little idea of a paper's quality, they employ personal knowledge of the author¹⁹ and the prestige of the author's institution as signals. Manuscripts which bear the imprints of such signals are passed on to referees; others are rejected without any further attention. Both signals work out as favoritism.

Moreover, note that an "editor not only has a commitment to quality control, but a vested interest in the prestige of the journal. Just as authors attract kudos in terms of where they publish, so journals undoubtedly gain prestige in terms of whom as well as what they publish."²⁰ A nice anecdote was contributed by Rosenthal (1982, p. 235) in his comment on the Peters and Ceci (1982a) paper: "It reminded me of the 15 to 20 articles I had written at UND [University of North Dakota] that I was not able to publish in mainstream psychological journals. After I had been at Harvard a few years, most of those same articles *were* published in mainstream journals." Phenomena like this led Scarr (1982, p. 53) to the insight: "A rose grown in major universities seems to smell sweeter than the same variety submitted to journals from less halcyon fields."

Due to the flood of papers, editors often encourage referees to stiff reviewing. Moreover, they have to grant their referees absolute anonymity to prevent the pool of referees from drying up. Referees who economize on time in the preparation of their reports are tempted to use the same signals (Honig 1982, p. 217), and, thereby, reinforce the bias of editorial screening. When the author is unable to show these credentials, they resort to some trivial statements and suggest rejection of the paper.

19. Cf. Wilson (1982); for empirical evidence cf. Fisher et al. (1994, p. 145).

20. Over (1982, p. 230). Cf. also Perlman (1992), Fisher et al. (1994, pp. 145-6).

There are scores of attempts at showing empirical evidence of favoritism,²¹ alas, most of them had to rely on data of accepted or published papers rather than submitted papers (Hodgson and Rothman 1999, p. F180) because such data are hardly ever released. Yet favoritism can only be evidenced with respect to data on submitted papers broken down both for regions and institutions. A rare exception is the work by Link (1998). She investigated all original submissions received by the journal *Gastroenterology* in 1995 and 1996 and found a statistically significant bias of all referees in favor of papers submitted by US authors; this tendency was still more pronounced for US referees.

This leads us to a more palpable objective reason of favoritism, viz. the language problem. English has by now become the *lingua franca* of science. This puts non-native speakers with their different cultural backgrounds, their imperfect command of English, and their sometimes clumsy English expression at a severe drawback. Now, a language which had become the *lingua franca* of science bears not only decisive advantages for the native speakers, it has also responsibility for science as a whole and should not discriminate against earnest scholars just because of their imperfect command of the ruling language. Peer reviewers should face the challenge to provide stylistic help rather than suggest rejection of a manuscript because they eschew the effort to try to understand it properly.

1.5.3 Confirmatory Bias

Scores of empirical studies have shown that manuscripts with positive results have much greater chances of being published.²² It seems that referees as well as editors are predisposed towards positive results and dislike papers reporting negative results. The same applies to significant versus nonsignificant results; papers reporting significant results have greater chances of being published than papers reporting nonsignificant results.²³ This applies even for large randomized trials.²⁴ In order to remedy this pitfall of traditional peer review processes, a specialized journal, viz. the *Journal of Negative Results in Biomedicine*, has been established in 2002 by Harvard University, but it seems that only few authors did submit respective manuscripts to this journal. Obviously, authors, too, are possessed by the confirmatory bias.

21. Fusfeld (1956); Cleary and Edwards (1960); Yotopoulos (1961); Siegfried (1972); Bairam (1994); Elliott et al. (1998); Hodgson and Rothman (1999); Kalaitzidakis et al. (1999; 2001); Kocher and Sutter (2000); Coupé (2000a,b).

22. Cf., e.g., Smart (1964); Goodstein and Brazis (1970); Greenwald (1975); Mahoney and Kimper (1976); Kerr et al. (1977); Mahoney (1977); Rowney and Zenisek (1980); Chalmers (1985); Dickersin (1990); Callaham et al. (1998). In contrast to that, Olson et al. (2002) did not evidence a confirmatory bias.

23. Cf., e.g., Sterling (1959); Bozarth and Roberts (1972); Atkinson et al. (1982).

24. Krzyzanowska et al. (2003).

This tendency is rather alarming because it is falsification rather than verification which is the logically appropriate method of advancing science. The confirmatory bias among authors, referees, and editors puts Popper (1959) upside down. Even the 1974 edition of the *Publications Manual* of the American Psychological Association raised outright hostility against negative results (p. 21). Mahoney's (1977, p. 174) comment still holds good after more than a quarter of a century:

The ironic feature of confirmatory bias is the fact that it is fundamentally illogical. Positive results and negative results experiments are not equivalent in their logical implications. In fact, while they have unquestionable bearing on the subjective aspects of belief, *successful experiments have no necessary logical bearing on the truth status of their source* (i.e., a theory or hypothesis). ... *It is only negative results (contrary-to-prediction) experiments which carry logical implications.* ... Despite this clear mandate from logic, however, our research programs and publications policies continue in their dogmatically confirmatory tradition.

A bias related to the confirmatory bias has been noticed for the source of the funding of the research. Whereas experimental drugs were recommended as treatments of choice in only 16% of trials funded by nonprofit organizations, this percentage rose to 51% of trials funded by profit organizations (Als-Nielsen et al. 2003).

1.6 Are Nonprofit Journal Editors Good Economists?

Let us start with an example: The number of members both in the American Physical Society and the American Sociological Association increased about two and a half times between 1950 and 1965. But, whereas the number of pages published annually by *The Physical Review* increased 4.4 times during this time from 3,920 pages to 17,060 pages, the number of pages published by the official organ of the American Sociological Association remained about the same (Zuckerman and Merton 1971, p. 79).

As to ten top nonprofit economic journals, only three of them increased the number of pages by more than 10% between 1985 and 2001 (Bergstrom 2001, p. 189). As the number of pages per article has also considerably increased (Ellison 2002a), these joint effects sky-rocketed the manuscript rejection rates of the top nonprofit economics journals.

In contrast to that, the average number of pages published annually by ten top commercial economics journals increased between 1985 and 2001 from 913 to 1700, that is, by 86.2%. Whereas the average price per printed page rose from 10 Cents to 15 Cents between 1985 and 2001 for the nonprofit journals, the respective average price rose from 30 Cents to 82 Cents for the commercial journals (Bergstrom 2001, p. 189). This policy enabled the publishers of commercial journals to reap enormous profits (Bergstrom 2001, pp. 187 and 192). They accomplished that by and large by

publishing only specialized journals. This demonstrates the inability of the editors of nonprofit economics journals to have filled in this growth segment of the publication market. Other than journals such as the *Physical Review*, *Nuclear Physics*, *Nuevo Cimento*, the *Journal of the Chemical Society-London*, the *Journal of the American Medical Association* and the *Journal of Experimental Psychology*, to name but the most important ones, the nonprofit economics journals did not establish specialized subjournals, but increased the rejection rates instead²⁵ and referred excellent papers to the more specialized (commercial) journals. Neither did they have the imagination to introduce page charges, following Sam Goudsmit's lead of the *Physical Review* journals in the 1940's. This device allows, first, low manuscript rejection rates, second, reasonable subscription rates, and it works, third, in the direction of shorter papers. While economists have pioneered in research on economic incentives, the editors of their learned nonprofit journals seem to be incapable of their application.

Thus, the shortcomings of peer review have given rise to widespread uneasiness. Colman remarked that "... authors ... often find themselves in a Kafkaesque situation analogous to that of a person prosecuted and condemned in a court of law without any right of defense."²⁶

The dichotomous views on peer review and the sharp contrast between different voices induced us to carry out an own empirical investigation. It is presented in the second part of this article.

2. A BEAUTY CONTEST OF REFEREE PROCESSES OF ECONOMICS JOURNALS

2.1 The Survey

At the end of 2001 and at the beginning of 2002 we addressed some twenty thousands persons twice by e-mail asking them for their online responses to seven

25. Unfortunately we cannot adduce widespread statistical evidence that rejection rates in economics journals have indeed increased, as the respective figures are released by the editors of only very few economics journals. There is, however, broad indirect evidence for this fact: (1) economics papers are roughly twice as long as they were 25 years ago [Ellison (2002b, p. 998)], (2) The average number of pages published by nonprofit economics journals has increased by but 18% between 1985 and 2001 [Bergstrom (2001, pp. 189 and 192)], (3) the number of academic economists has considerably increased [just look at the number of attendants at economics congresses and at the increased number of such congresses], (4) the productivity of academic economists has not decreased [look at the number of papers presented at economics congresses]. 26 Colman 1991, p. 142. Cf. also Ingelfinger (1974, p. 687); Cole (1991); Crandall (1982; 1991); Eckberg (1982); Mindick (1982, p. 227); Horrobin (1982, p. 217); Hogan (1982, pp. 216 and 218); Armstrong (1982, p. 197); Glenn (1982); Griffith (1982, p. 215); Howe (1982, p. 218); Nelson (1982); Ziman (1982, p. 245):

questions concerning their experience with referee processes of economics journals. Some 6,000 persons in total had a look at our questionnaire, but only a bit more than 10% of these individuals started to respond to our questionnaire.

As professional institutions in the Anglo-Saxon world did not even respond to our inquiry, let alone did offer their cooperation we used other sources of mail addresses in the hope of capturing many economists. Many addresses were those of professional people who had never published. Response rate of 3% may seem to be small, but all senders of spam mails would consider such a response as a great success. Moreover, election forecasts and public opinion polls rest on much lower percentages and yet provide reliable results. Thus, although we would have liked to have more and better data, our results provide nevertheless interesting and important insights.

2.2 The Journals

In our survey we used two groups of journals. We call them *invited* and *contributed* journals, respectively. The invited journals centred on the famous Diamond (1989) list presented in Table 1, which comprises 27 economics journals. We enlarged this list by 49 journals taken from the A and B categories of the VSNU (Vereniging van Samenwerkende Nederlandse Universiteiten - Association of Universities in the Netherlands) economics journals ranking list, which we considered as important enough to be included. This makes 76 invited journals. Furthermore, we allowed our respondents to contribute journals to list at their own discretion.

Table 1.

American Economic Review	Journal of Financial Econ.
Brooking Papers on Ec. Act.	Journal of International Econ.
Canadian Journal of Economics	Journal of Labor Economics
Econometrica	Journal of Law and Economics
Economic Inquiry	Journal of Mathematical Econ.
The Economic Journal	Journal of Monetary Econ.
Economica	Journal of Political Economy
Economics Letters	Journal of Public Economics
European Economic Review	Oxford Economic Papers
International Economic Review	Quarterly Journal of Econ.
Journal of Development Econ.	Rand Journal of Economics
Journal of Econometrics	Review of Economic Studies
Journal of Economic Literature	Review of Econ. and Statistics
Journal of Economic Theory	

This procedure yielded a total of 359 journals. Space does not permit to include a list of all journals in this article.²⁷ However, the structure of responses shows that our choice of the 76 invited journals provided a good match with respondents' experience: Among the 73 journals showing at least 10 responses to Question 1, only four were not listed among the invited journals. All ten journals which attracted one hundred or more responses are also members of the Diamond list. All journals of the Diamond list, with the exception of the *Brooking Papers on Economic Activity*, elicited at least 11 responses to the first question. Moreover, the responses show that the Diamond list ignores some renowned (mostly non-American) journals, which existed well before 1989.

For the presentation of the results of our study, the data were broken down as follows: For the analysis of relationships between respondents' attitudes, we employed *all* data irrespective of how many responses per journal we had. For descriptive documentations with respects to particular journals we arbitrarily settled on at least *five valid responses* for the respective journals. To report on journals with less than five valid responses would probably convey a distorted picture. As some respondents had chosen to drop out during the survey, the set of journals decreases somewhat for later posed questions. For Questions 1 and 2, 110 journals had at least five valid responses, for Question 3 we had 107 journals, and for Questions 4-7 we had 106 journals.

For the purpose of this paper we prepared, moreover, a concise summary documentation of subjects' responses. We narrowed down the set of papers to a cut-off benchmark of at least twenty valid responses to Question 7 (the last question).²⁸ 51 journals satisfy this condition. Table 3 shows them ordered according to the ranking of the journals with respect to Question 7 asking for subjects' general satisfaction with the journals' referee processes. All rank numbers are taken from the more comprehensive tables.

2.3 Respondents

A survey of researchers' experience with referee processes can follow several routes. One possibility is to address successful authors who managed to get their

27. The list of all 359 journals can be downloaded from our homepage <http://www.wiso.uni-kiel.de/ifs/chair/peerreview.php> as Table 1*. It contains all journals for which the first question was answered (respondents could answer subsequent questions only by passing Question 1 first). In this table, the invited journals are marked with an *asterisk*, the journals of the Diamond list among them with a *diamond*, and the contributed journals are *unmarked*.

28. The documentation of the results based on a benchmark of at least five valid responses can be downloaded from our homepage <http://www.wiso.uni-kiel.de/ifs/chair/peerreview.php> as Tables 3*-8*.

papers published in a journal. Alas, this approach is in danger of biasing responses in favor of the respective journal because one asks the lucky ones whose work had been accepted. The other way is to address economists at large and thus collect also the experience of the less lucky ones which is, however, crucial for a valid picture of the performance of referee processes. Although this approach might be in danger of attracting mainly frustrated authors who wish to take revenge on allegedly unfair refereeing, we shall demonstrate below that our data do not suffer from such biases. Rather they are biased in the opposite direction.

We took the second route and asked several institutions for e-mail addresses of economists. Fortunately for our survey, many of them lent us their help. In particular, we received help from the *European Economic Association*, the *Verein für Socialpolitik*, the editorial board of the *Economics Bulletin*, from IZA [*Forschungsinstitut zur Zukunft der Arbeit – Institute for the Study of Labor*] and from *Inomics*. Our faculty colleagues Sönke Albers and Joachim Wolf also provided good advice. Some e-mail addresses of economists were collected by us. Several institutions, mostly from the Anglo-Saxon world, did not even reply to our inquiries, let alone offer us their cooperation. These were, in particular, The American Economic Association, The Econometric Society, and The Royal Economic Society, as well as some less well-known Asian economic associations. This refusal of cooperation implies that American, Asian and Pacific economists are unfortunately underrepresented in our survey (cf. Table 2). We had only the choice to work with the data available or dispensing with our endeavour at all. As we had to cope with biased data anyway, we felt that the results would justify data shortcomings.

Table 2.

Rank	Country	#	Rank	Country	#
1	Germany	167	11	Denmark	16
2	USA	113	11	Sweden	16
3	United Kingdom	44	12	Norway	13
4	Spain	38	13	Canada	10
5	Italy	33	14	Israel	9
6	Austria	31	14	Portugal	9
7	France	28	15	Turkey	5
8	Switzerland	27		Rest	31
9	Netherlands	21			
10	Belgium	19		Sum	630

The data of all respondents underwent a plausibility test. This led to the elimination of the data of 9 respondents, representing their joint responses to 22 journals, for various reasons.²⁹ As these were typing errors, jokes, or attempts at manipulation,³⁰ the data of these subjects had to be eliminated. Concerning the descriptive results for the particular journals (at least five usable responses), we were left with 630 respondents to the first question, of which 551 participated in the survey through to the seventh question. In the aggregate³¹, we could dispose of between 4538 (for Question 1) and 3791 data per question (for the entries to Questions 2-7 cf. the number of entries in Table 8).

2.4 Reactions

In addition to responses to our questions, many subjects sent us comments and suggestions. The tenor of their reactions was helpful, sympathetic, or critical.

Many helpful comments drew our attention to related work, some of which entered into the first part of this paper. Most of the general work, had, however, to be deferred to a comprehensive study on *The Performance of Peer Review: An Interdisciplinary Report*, which is presently under elaboration.

Numerous sympathetic, some of them even enthusiastic, reactions came from all strata of respondents. Many commentators argued that we should have posed more and more detailed questions. Yet, it is true that we started originally with a far more comprehensive list of questions, but decided to confine the questionnaire to but seven questions for fear of too many drop-outs. Our experience with this survey showed ample evidence that we were right in doing so: Only about eight per cent out of all persons originally interested in our survey embarked on responding to all questions. However, we conjecture that some drop-outs might also have been caused by inefficient web services in some countries. Other sympathetic scholars took the occasion of our survey to broach their own uneasiness with the current referee situation.

Critical comments were received from only a few prominent economists. For instance, a renowned economist urged us: “please stop sending me reminders about this research. Such research projects are dangerous and misleading!” Comments like this suggest that research directed at referee processes of learned journals seems not to be favored among some of the profession’s most prestigious scholars.

29. The elimination criteria were: Response time exceeding 250 weeks (3 respondents), having received more than 10 referee reports (5 respondents), and having received referee reports without having submitted a paper (1 respondent).

30. We checked also computer IP addresses for similar evaluations, but did not observe suspicious similarities of responses in the cases of multiple uses of the same computers.

31. Counting all journals irrespective of how many responses we had per journal.

2.5 The Questionnaire

When a respondent connected to our server, (s)he was first presented with a general plea to participate in the survey. Then the respondent was shown a list of our 76 invited journals and asked to select those journals with which (s)he had experience as an author. Furthermore, the respondent was prompted to add further economics journals of his or her choice. Both sets together formed the particular respondent's journal set.

Then the respondent was asked the first question and asked to respond to subsequent items for the selected journals. For the Questions 1, 2, 3, and 7, (s)he was urged to respond to the respective question for all journals in his or her set. As to the first three questions (s)he could proceed only after the respective question had been answered for all journals in the set.³² Questions 4-6 could only be answered if the respondent had actually received a referee report. As some journals reject manuscripts without having solicited referee reports, a respondent may not have experience with referee reports of all journals to which (s)he had submitted manuscripts. Therefore, for Questions 4-6 we allowed for passing to the next question without having responded to the respective question for all journals. Moreover, during the response process, a respondent could also opt to eliminate some journals altogether from his or her journal set. While the questions answered up to this point were kept in our data, these journals were then dropped for the subsequent questions. This device was intended to encourage respondents to complete the questionnaire even if s(he) realized that (s)he had initially proposed a larger set of journals than s(he) was able or willing to evaluate.³³

The questionnaire consisted of the following seven questions per journal.³⁴

Question 1: "After submission of your paper, how long did it take on average to get a reply other than just a confirmation that your paper had been received?"³⁵

Question 2: "How many referee reports did you receive on average?"

Question 3: "How many papers did you submit to this journal and how many papers were accepted?"

Question 4: "Were the referee reports competent?"

Question 5: "Did the decision of the editor match the referee report?"

Question 6: "Were the referee reports carefully done?"

Question 7: "How was your overall satisfaction with the procedure of paper submission to the respective journal?"

32. This method warrants that subjects could concentrate on meaningful comparisons among the journals of their set for the same aspect of evaluation.

33. The elimination of journals was easily accomplished. The respondent had only to erase a little hook after the journal.

34. The screenshots of all seven questions can be downloaded from our homepage <http://www.wiso.uni-kiel.de/ifs/chair/peerreview.php>.

35. In the respective cell subjects were asked to indicate the response time in weeks.

Note that the responses to Questions 1-3 are numbers such as the spell to get a first reply, numbers of referee reports received, and numbers of papers submitted and accepted. In contrast to that, the responses to Questions 4-7 result from mouse click to one out of seven fields on Likert scales. In our results, the worst value is coded with a 0, and the best with a 6, so that 3 forms the mean coded value of each Likert scale if all values were clicked with equal frequency. Notice, of course, that data from Likert scales are necessarily subjective data. Were interested in the opinions of our respondents Likert scales is the proper way to capture perceptions.

2.6 Results

To have some kind of measuring rod for comparisons, we often calibrate our results against those journals which are commonly regarded as being the core economics journals. Although there exist several categorizations (e.g., Burton and Phimister, 1995), we decided to stick to the well-known Diamond (1989) list.

2.6.1 Descriptive Results

Recall that, for the particular journals, we confined our attention to those journals which commanded at least five valid responses. This reduced the set of journals given attention to a domain of 106 to 110 journals. The journals were ranked according to the given responses, where equal responses led to equal ranks. We used the following ranking criteria:

- Question 1: Shorter response times.
- Question 2: Greater number of referee reports.
- Question 3: Higher individual acceptance rates (papers accepted / papers submitted).
- Question 4: Higher competence of referee reports.
- Question 5: Higher matching of editorial decision and referee reports.
- Question 6: Higher carefulness of referee reports.
- Question 7: Higher overall satisfaction with referee process.

Complete data for the present study are presented in tables which can be downloaded as Tables 3*-9* from <http://www.wiso.uni-kiel.de/ifs/chair/peerreview.php>. For the concise summary documentation of subjects' responses presented in Table 3 of this paper, we employed a cut-off benchmark of at least twenty valid responses to Question 7. 51 journals satisfied this condition. In order to provide background information, we indicated in Table 3 the ranks R of the more comprehensive tables. As to Table 3, we had to settle on one ordering criterion; we used overall satisfaction (Question 7). As overall satisfaction is the most important characteristic, we arranged the columns of Table 3 in reverse order of the presentation of questions. The findings of our paper rest, however, on the more comprehensive results.

Table 4 provides a concise summary of descriptive results. In the remainder of this subsection the descriptive results are discussed in greater detail.

With respect to *response time*, the *Quarterly Journal of Economics* stands out as the speediest one with a mean turn around time of 0.613 weeks, that is, 4.29 days. However, given its high subjective rejection rate of 93%, this means that the managing editor(s) of the *Quarterly Journal of Economics* reject(s) many of the submitted manuscripts without ever having consulted a single referee as to rejection or acceptance of a paper. Indeed we registered $n = 91, 89, 88$ responses to Questions 4, 5, 6, but $n = 106$ responses to Question 7 which means that several subjects did not receive a referee report at all. These values resemble those obtained for *Economics Letters*, for which we registered $n = 129, 125, 123$ responses to Questions 4, 5, 6, but $n = 154$ responses to Question 7. Given a mean response time of 14.77 weeks and a meager mean of 0.86 referee reports, this leads us to conjecture that the decision to reject a paper without having sent it to a referee takes the editor of *Economics Letters* considerable time. Moreover, we cannot exclude that some authors counted a letter from the editor only as a true referee report.

As compared to other disciplines, e.g., the natural sciences, economics journals seem to take a particularly long time to reach a decision. Hardly any journals decide in fewer than 10 weeks, and more than half of them need 20 weeks and more to take a decision. 20 journals of the Diamond list (out of the 26 remaining ones) need more than 20 weeks to make a decision.

A mean of more than two *referee reports* is the exception rather than the rule. Among the Diamond journals, only *Econometrica* and *Economic Inquiry* reach a mean number of referee reports above two. These are the only Diamond journals which rank among the first twenty ranks with respect to the mean number of referee reports. *Economics Letters* ranks last among the journals from the Diamond list (0.86 referee reports per respondent). Recall that respondents might have considered the managing editors' rejection as a valid referee report.

We used subjects' reports on total numbers accepted by and submitted to a respective journal to compute the journal's individual *acceptance rates*. The data show that the more reputed journals have lower acceptance rates, which was to be expected. Indeed, 17 journals out of the Diamond list figure among the last 25 ranks.³⁶ Cross-disciplinary comparisons show that manuscript rejection rates are much higher in the humanities than in the natural sciences.³⁷ By and large a rejection rate of more than 80% in the humanities contrasts with an acceptance rate of some 80% in the natural sciences.

36. However, there are some exceptions to this regularity. For instance, four journals of the Diamond list rank below 50 (out of 94 ranks).

37. Cf., e.g., Zuckerman and Merton (1971), Lazarus (1982), Adair (1982), Hargens (1988; 1990).

Table 3

Journal	n	Satisfaction (7)			Carefulness (6)			Match (5)			R
		R	Ø	STD	R	Ø	STD	R	Ø	STD	
Finanzarchiv	38	6	4.53	1.43	20	4.16	1.39	28	4.97	1.20	29
Math. Soc. Sc.	30	8	4.47	1.81	12	4.35	1.52	25	5.03	1.55	9
J.Economics/ZfN	70	12	4.33	1.67	25	4.09	1.41	21	5.12	1.38	32
J. Math. Economics ♦	24	13	4.25	1.76	11	4.36	1.53	7	5.39	1.31	15
Scott. J. Polit. Ec.	20	15	4.20	1.91	51	3.62	1.91	34	4.91	1.64	67
Soc. Choice Welfare	32	16	4.19	1.71	14	4.29	1.58	16	5.22	1.18	10
J. Inst. Theor. Ec./ZfgStW	49	17	4.18	1.50	19	4.16	1.46	33	4.94	1.28	30
Economic Inquiry ♦	24	19	4.13	1.90	30	4.00	1.83	29	4.96	1.75	11
J. Ec. Dynamics Control	33	20	4.09	2.07	31	3.97	1.82	31	4.94	1.61	24
J. Econometrics ♦	23	21	4.09	2.04	23	4.09	1.66	54	4.64	1.40	14
J. Public Ec. Theory	29	23	4.03	1.57	13	4.31	1.29	17	5.20	0.85	18
J. Population Economics	24	25	3.92	1.74	25	4.09	1.44	20	5.13	1.10	41
Ec. Theory	51	27	3.84	1.96	28	4.04	1.70	18	5.17	1.42	23
Eur. J. Political Ec.	76	28	3.84	1.67	45	3.68	1.61	46	4.73	1.54	47
Econometrica ♦	98	29	3.82	1.91	17	4.23	1.88	37	4.88	1.36	12
Weltwirtsch. Archiv	62	30	3.81	1.77	24	4.09	1.23	24	5.05	1.37	37
Canad. J. Economics ♦	37	33	3.73	1.84	36	3.81	1.47	27	5.00	1.33	50
Games Ec. Behav.	66	34	3.73	1.76	35	3.81	1.83	10	5.35	0.97	21
Management Science	31	36	3.68	2.04	40	3.71	1.70	44	4.81	1.58	43
Scand. J. Economics	74	37	3.66	1.93	49	3.64	1.73	53	4.64	1.51	69
Economics Letters ♦	154	38	3.63	1.99	71	3.27	1.88	9	5.36	1.13	49
J. Public Economics ♦	112	39	3.59	1.81	34	3.83	1.68	68	4.45	1.52	42
J. Economic Theory ♦	94	41	3.56	1.93	39	3.73	1.91	23	5.09	1.16	34
J. Development Ec. ♦	42	42	3.52	1.94	37	3.78	1.72	51	4.67	1.43	73
Empirical Economics	44	43	3.50	1.85	32	3.91	1.61	42	4.83	1.41	44
Public Choice	57	47	3.39	2.10	73	3.16	1.87	26	5.02	1.23	64
American Ec. Review ♦	171	48	3.37	1.83	55	3.55	1.86	65	4.49	1.61	27
J. Ec. Beh. Org.	67	49	3.37	1.91	61	3.51	1.86	58	4.61	1.60	60
Review Ec. Studies ♦	84	50	3.37	1.73	29	4.03	1.53	47	4.72	1.58	20
J. Industr. Ec.	46	51	3.35	1.69	27	4.04	1.38	43	4.83	1.14	40
J. Mon. Credit Banking	29	52	3.35	1.72	44	3.69	1.97	27	5.00	1.23	79
Rev. Ec. Statistics ♦	71	53	3.34	1.93	54	3.55	1.82	67	4.46	1.60	31
International Ec. Rev. ♦	76	54	3.33	1.87	41	3.70	1.59	38	4.87	1.41	39
J. International Ec. ♦	60	55	3.32	1.73	57	3.54	1.58	61	4.56	1.38	52
Theory and Decision	26	56	3.31	1.87	48	3.64	1.60	36	4.88	1.64	55
Int. J. Game Theory	30	60	3.17	2.04	43	3.96	1.83	14	5.24	1.12	36
Oxford Ec. Papers ♦	51	62	3.10	1.96	70	3.28	1.93	74	4.31	1.80	76

Table 3

i	h (5) STD	Competence (4)			Acceptance Rate (3)			Number (2)			Spell (1)		
		R	Ø	STD	R	Ø	STD	R	Ø	STD	R	Ø	STD
97	1.20	29	3.82	1.43	16	0.73	0.38	57	1.63	0.49	14	13.26	7.26
03	1.55	9	4.39	1.50	33	0.62	0.45	50	1.69	0.73	39	17.56	11.18
12	1.38	32	3.76	1.47	27	0.66	0.43	30	1.88	0.64	20	14.58	11.58
39	1.31	15	4.26	1.36	26	0.87	0.33	61	1.59	0.83	97	29.07	29.70
91	1.64	67	3.23	1.80	53	0.47	0.50	67	1.48	0.65	32	16.33	11.49
22	1.18	10	4.36	1.32	17	0.72	0.41	34	1.82	0.71	80	23.71	16.50
94	1.28	30	3.78	1.39	28	0.66	0.44	37	1.81	0.71	36	16.76	10.20
96	1.75	11	4.30	1.55	24	0.68	0.39	17	2.04	0.43	45	18.57	13.37
94	1.61	24	3.94	1.59	39	0.58	0.46	16	2.08	0.88	69	22.07	18.25
64	1.40	14	4.29	1.43	31	0.64	0.44	20	1.94	0.74	82	24.44	16.79
20	0.85	18	4.13	1.14	50	0.48	0.49	38	1.81	0.48	74	23.14	12.37
13	1.10	41	3.65	1.53	14	0.74	0.40	24	1.89	0.63	70	22.10	12.90
17	1.42	23	3.98	1.52	44	0.54	0.45	71	1.42	0.62	58	21.22	19.65
73	1.54	47	3.56	1.62	35	0.61	0.47	25	1.89	0.60	55	20.65	17.26
88	1.36	12	4.30	1.91	88	0.20	0.31	13	2.12	0.80	68	22.04	14.57
05	1.37	37	3.69	1.50	47	0.51	0.41	80	1.25	0.56	21	14.71	13.73
00	1.33	50	3.51	1.43	54	0.45	0.48	44	1.74	0.55	54	20.58	11.89
35	0.97	21	4.06	1.69	58	0.43	0.43	47	1.71	0.56	90	26.22	16.23
81	1.58	43	3.64	1.73	56	0.44	0.46	7	2.36	0.88	51	20.03	17.14
64	1.51	69	3.21	1.68	60	0.43	0.46	19	1.95	0.51	50	19.98	12.41
36	1.13	49	3.55	1.81	49	0.50	0.44	86	0.86	0.55	22	14.77	10.70
45	1.52	42	3.64	1.57	74	0.34	0.40	27	1.88	0.54	67	21.66	11.91
09	1.16	34	3.76	1.77	76	0.32	0.37	48	1.70	0.72	89	25.86	17.00
67	1.43	73	3.16	1.73	67	0.38	0.44	35	1.82	0.56	100	32.47	24.29
83	1.41	44	3.61	1.62	20	0.69	0.44	39	1.80	0.53	76	23.22	17.82
02	1.23	64	3.28	1.68	45	0.54	0.46	69	1.42	0.61	31	16.19	14.33
49	1.61	27	3.86	1.71	91	0.17	0.33	33	1.85	0.75	66	21.64	15.83
61	1.60	60	3.39	1.64	40	0.56	0.47	60	1.60	0.73	85	25.11	19.30
72	1.58	20	4.08	1.42	87	0.22	0.38	29	1.88	0.66	95	28.33	17.72
83	1.14	40	3.66	1.40	80	0.30	0.39	32	1.86	0.59	63	21.50	16.10
00	1.23	79	2.97	1.70	72	0.34	0.46	49	1.69	0.60	47	18.67	11.88
46	1.60	31	3.78	1.56	69	0.36	0.44	45	1.73	0.70	60	21.36	13.99
87	1.41	39	3.67	1.64	75	0.32	0.44	26	1.89	0.88	93	27.78	15.20
56	1.38	52	3.48	1.55	84	0.26	0.38	21	1.91	0.48	86	25.14	13.20
88	1.64	55	3.44	1.58	25	0.67	0.43	72	1.41	0.73	77	23.25	19.43
24	1.12	36	3.73	1.70	61	0.43	0.44	56	1.64	0.64	101	33.76	30.35
31	1.80	76	3.06	1.82	78	0.31	0.43	43	1.75	0.66	62	21.48	14.00

.../...

Table 3 (cont.)

Journal	n	Satisfaction (7)			Carefulness (6)			Match (5)			C
		R	\bar{O}	STD	R	\bar{O}	STD	R	\bar{O}	STD	
Kyklos	53	63	3.09	2.11	56	3.55	1.74	19	5.14	1.22	54
J. Applied Economics	23	66	3.04	2.23	51	3.62	1.69	27	5.00	1.05	56
Rand J. Economics ♦	60	67	3.03	1.97	47	3.66	1.92	60	4.59	1.37	37
Cambridge J. Ec.	20	69	2.95	1.73	50	3.63	1.80	56	4.63	1.89	77
European Ec. Review ♦	187	70	2.95	1.83	69	3.32	1.73	50	4.68	1.40	65
Economica ♦	52	71	2.94	1.92	76	3.08	1.91	57	4.63	1.48	75
Southern Economic J.	22	72	2.91	2.11	77	3.05	1.96	39	4.86	1.36	78
Quarterly J. Economics ♦	106	73	2.90	2.05	75	3.08	2.03	79	4.86	2.02	57
J. Labor Economics ♦	32	74	2.88	1.90	63	3.43	1.81	55	4.63	1.65	33
J. Human Resources	30	75	2.87	2.11	68	3.36	1.87	78	4.08	2.20	38
Oxf. Bull. Ec. Statistics	29	76	2.86	2.18	53	3.57	1.83	30	4.96	1.19	68
Economic J. ♦	115	77	2.86	2.01	62	3.50	1.86	71	4.38	1.72	59
J. Monetary Economics ♦	34	80	2.79	2.03	66	3.39	2.12	32	4.94	1.41	63
J. Financ. Ec. ♦	28	81	2.79	1.85	58	3.54	1.62	63	4.54	1.67	66
J. Political Economy ♦	88	85	2.13	1.92	81	2.61	1.96	66	4.47	1.72	80

Table 4: Concise Summary of Descriptive Results

Item	Remarks
Response time	Median: 20.5 weeks; QJE: 0.613 weeks; 20 journals of the 26 Diamond journals need more than 20 weeks.
Number of referee reports	Median: 1.75; only 24 out of 110 journals provided at least 2 referee reports.
Acceptance rates	Median: 0.5; 23 Diamond journals below 0.5.
Competence	Median: 3.641; even distribution of Diamond journals; only 12 below 3.
Matching	Median: 4.826; only 7 below 4.
Carefulness	Median: 3.69; even distribution of Diamond journals; only 10 below 3.
Satisfaction	Median: 3.524; 36 score 4 or better, 23 worse than 3.

The median values are the medians of the mean values for the individual journals

Table 3 (cont.)

i	h (5)	Competence (4)			Acceptance Rate (3)			Number (2)			Spell (1)		
		STD	R	Ø	STD	R	Ø	STD	R	Ø	STD	R	Ø
14	1.22	54	3.44	1.55	70	0.36	0.46	85	1.11	0.76	4	10.07	8.32
00	1.05	56	3.43	1.69	22	0.69	0.45	66	1.48	0.80	104	38.48	28.35
59	1.37	37	3.69	1.67	85	0.23	0.38	51	1.67	0.70	83	24.49	16.14
63	1.89	77	3.05	1.57	51	0.47	0.48	15	2.10	0.88	79	23.50	16.58
68	1.40	65	3.26	1.56	77	0.31	0.41	28	1.88	0.61	84	25.08	16.66
63	1.48	75	3.11	1.76	63	0.42	0.47	42	1.75	0.60	75	23.21	14.21
86	1.36	78	3.00	1.95	62	0.42	0.49	64	1.56	0.64	46	18.65	14.53
86	2.02	57	3.42	1.90	94	0.07	0.21	79	1.25	0.82	1	0.61	3.70
63	1.65	33	3.77	1.38	90	0.19	0.35	59	1.61	0.60	88	25.79	17.66
08	2.20	38	3.68	1.60	64	0.42	0.49	52	1.66	0.87	48	19.29	14.06
96	1.19	68	3.22	1.76	73	0.34	0.45	83	1.19	0.86	52	20.11	11.08
38	1.72	59	3.40	1.71	82	0.27	0.41	23	1.90	0.78	65	21.58	13.54
94	1.41	63	3.29	1.92	81	0.29	0.41	68	1.46	0.72	102	36.65	30.81
54	1.67	66	3.24	1.70	93	0.10	0.31	73	1.40	0.55	10	12.44	8.04
47	1.72	80	2.89	1.91	92	0.15	0.33	70	1.42	0.59	91	26.77	21.28

Competence of the referee reports is, on the whole, judged rather favorably. Only twelve out of 106 journals were rated below 3.0 (out of a maximum 6.0), among them only two Diamond journals, viz. the *Journal of Political Economy* and the *Journal of Financial Economics*. Seven out of the journals of the Diamond list (recall that the *Brooking Papers* dropped out) score at 4.0 or better. Competence of referee reports seems to be not positively correlated with the reputation of a journal. Neither the journals of the Diamond list nor of the invited journals bunch at the upper or at the lower end; they appear to be rather evenly distributed among the ranks. For instance, 14 journals of the Diamond list rank ahead of, and 12 rank behind the mean rank of 44.

Our results show that most journals score rather well with respect to *matching* of the managing editors' decisions with the recommendations of the referee reports. However, this signals a good performance of peer review if and only if referees' judgements are valid. If they are just reliable, and the managing editor decides blindly in accordance with them, this need not be a proxy for good refereeing because referee hostility or incompetence may be but insufficiently monitored by the editor.³⁸ Attentive editors should interfere in the latter case, which would be reflected in lower matching scores. Companying letters to the editors may also be harsher than the referee reports for the authors. Editorial deviation from referee recommendations may also be prompted by high backlogs of manuscripts which may goad editors' zeal to curb the growth of the queues of papers agreed to be published. Such independent decisions by editors may help explain the lower rankings of some prestigious journals, such as the *American Economic Review*, the *Journal of Political Economy*, the *Review of Economics and Statistics*, the *Journal of Public Economics*, the *Economic Journal*, and the *Quarterly Journal of Economics*. When an editor, because of space limits, is forced to reject manuscripts furnished with good referee reports, (s)he may well give in to favoritism, which will be discussed below.

38. With respect to the editorial decision to accept or reject a manuscript, voices have been aired which encourage editors to use their discretionary powers wisely and - if necessary - should not shy at overriding referees' recommendations (Bailar 1991, p. 138). Stricker (1991, p. 164) disputes that good editors should behave like psychometric clerks who simply add up the scores that a manuscript gets from the referees. He argues that "good editors are not clerks. They read the manuscript, appraise the reasons reviewers give for their recommendations, and weigh all the information about it ..." He is paralleled in this view by Glenn (1982, p. 212), Rodman (1970, pp. 355-356), and Goodstein (1982, p. 213). Crandall (1991) pleads along the same lines that editors should be super referees. He deplores that too many editors do not behave in this way. He suspects "that many editors do not even read the papers for which they are supposed to have editorial responsibility." Scarr (1982, p. 54), editor of *Developmental Psychology* and the *American Psychologist*, has made a case for editorial responsibility. She refers editors, who shirk their duties, to one of Harry Truman's wise insights: "If you can't stand the heat, get out of the kitchen." Yalow (1982, p. 244) blamed reviewer and editorial incompetence for instances as revealed by the Peters and Ceci (1982) experiment. Simon et al. (1986, p. 270) report that only between 13 and 19% of authors' complaints against referee reports were successful.

Concerning *carefulness* of the referee reports, journals scores were similar to their scores on competence of the referee reports. Only 10 out of 106 journals scored less than 3, among them again the two notorious Diamond Journals, the *Journal of Political Economy* and the *Journal of Financial Economics*. Only 6 out of the 26 journals of the Diamond list (after dropping the *Brooking Papers*) scored at 4 or better than 4. As compared to competence, we observe a minor shift of the reputed journals to lower ranks: 11 journals of the Diamond list were rated above and 15 below the mean rank. Some reputed journals rank among the bottom 20 of carefully done referee reports, viz. the *European Economic Review*, the *Oxford Economic Papers*, *Economics Letters*, the *Quarterly Journal of Economics*, *Economica*, the *Journal of Political Economy*, and the *Journal of Financial Economics*.

Overall satisfaction with the whole procedure of paper submission proved as disappointing for the prestigious journals. Out of 106 journals, 36 scored at 4 or better; among them only five journals were from the Diamond list. Out of the 106 journals, 23 scored worse than 3; among them eight journals were from the Diamond list, to wit, the *European Economic Review*, *Economica*, the *Quarterly Journal of Economics*, the *Journal of Labor Economics*, the *Economic Journal*, the *Journal of Monetary Economics*, the *Journal of Financial Economics*, and the *Journal of Political Economy*.

2.6.2 Statistical Results

2.6.2.1 Response Biases

All data obtained from the subjects are, of course, subjective data. However, the data collected for Questions 1-3 have objective counterparts. In an attempt to correct for subjective biases, we sent a mail to the editors of all 110 journals for which we had received at least five valid responses to Question 1 and asked them for editorial data on the average response time to authors, the average number of referee reports solicited, and the average acceptance rate of manuscripts. Replies to these questions seemed to be easy, as editors of most journals are wont to keep regular statistics on these figures. Indeed, a few journals do even publish them (see, for instance, The Economic Journal Managing Editors' Report, 2005). Editors who did not respond to our mail were sent a reminder. We received responses from the editors of 52 journals (response rate: 47.3%), among them 7 responses from the 26 Diamond journals (response rate: 26.9%). Note that we could not check whether we received the true objective data, biased data or mere conjectures of the editors, but they represent an independent alternative data source which allowed inferences on possible biases. For the sake of a shorthand expression we address them as the *objective* data in this paper.

Table 5 gives a concise summary of our results. Their entries are the means (taken over all journals for which we had data) of the ratios of the mean responses of the

subjects and the responses of the editorial board. Table 5 shows us that the subjective response time exceeds the objective one by some 50%, that the subjective number of referee reports is slightly lower than the objective number of referee reports, and that the subjective acceptance rate exceeds the objective one by some 150%.

Table 5: Response Biases: Statistics

Subjective value divided by objective value	N	Min.	Max.	Mean	StD
Response Time	52	0.61*	4.73	1.49	0.69
No. of Reports	52	0.53	1.40	0.90	0.16
Accept. Rate	50	0.42**	8.89	2.50	1.34

All means significant at the 1% level (two-sided).

* Only 8 values smaller than 1.

** Only 4 values smaller than 1.

The most spectacular upward bias is noticed for the subjective acceptance rates. We may offer several explanations for that (all of which may have contributed to produce this result):

1. *Self-selection effect*: It seems that the more successful (and, perhaps, also the more active) scholars felt more attracted by our survey than the less successful ones.³⁹
2. *Survey-selection effect*: As our survey was directed to investigate authors' experience with referee processes, we had asked subjects to respond only for those journals with which they had experienced at least one referee process. This rules out manuscript submissions which were rejected immediately by the editors without soliciting referee reports. Of course, manuscripts which were infused into the referee process have positive chances of being accepted, whereas the crude (objective) rejection rate includes also purely editorial rejections.
3. *Cognitive-disonance effect*: Successful events are memorized, failures are mentally suppressed.
4. *Trend effect*: True acceptance rates fell with the lapse of time. Respondents who remember their submission history of papers amalgamate past with present experience, which, due to the influence of higher past acceptance rates, biased their perception of acceptance rates upwards. In other words, their samples with the respective journals comprise different situations of whose trend they were not or were only insufficiently aware.

39. Similar effects were observed by Sweitzer and Cullen (1994). They polled 209 authors for their satisfaction with peer review processes. 67% of the AR (accept with revision) authors, 43% of the RR (reject but may resubmit) authors, and only 30% of the RO (reject outright) authors responded to their questionnaires sent to unsolicited authors of the *Journal of Clinical Anesthesia*. Higher response rates of authors whose papers were accepted were also observed by Garfunkel et al. (1990b) for the *Journal of Pediatrics*.

The upward bias of the response time is most probably associated with the upward bias of the acceptance rates. A well-established result says that it takes journals shorter times to reject a paper than to offer a revision of the paper.⁴⁰ This is reinforced when many papers are rejected immediately by the editors without having been infused into referee processes. Given that the more successful authors were over-represented in our survey, this implies longer spells of response time.

Finally, although the subjective number of referee reports is slightly lower than the objective number of referee reports, this deviation is not dramatic. The respective numbers are not too far off reasonable correspondence.

Note, therefore, that our data are biased in favor of the more successful authors. However, as we did not pre-select our respondents a priori from the set of the successful ones, this upward bias is certainly less than it would have been, had we addressed only people whose papers were actually accepted for publication (e.g., by way of taking only the e-mail addresses of respondents whose manuscripts were already published in journals). On the other hand, a survey such as ours is endangered of attracting frustrated respondents who wish to deal a blow to those journals which they consider to have treated them unfairly. The entries in Table 5 show that this was certainly not the case.

2.6.2.2 Favoritism

Favoritism can manifest itself in three ways, to wit, personal, institutional, and regional favoritism. Personal favoritism means that certain persons enjoy preferential treatment with respect to refereeing and/or editorial decisions. The literature abounds with gossip about personal favoritism, yet, in order to demonstrate its presence, one needs inside data on referee processes and editorial decisions, to which we had no access. Hence we could not study personal favoritism.

We also had difficulties investigating institutional favoritism because, for reasons of respondents' anonymity, we have only a regional, not an institutional breakdown of data. The detailed data of the Coupé list⁴¹ of economics journals contain an institutional and regional data breakdown, alas, only for *published* manuscripts. Hodgson and Rothman (1999) have rightly objected that, in order to show institutional or regional favoritism, one needs data on regional and institutional distributions of manuscript submissions and acceptances rather than just data of the distribution of

40. Cf., e.g., Ellison (2002a, p. 955, Table 2); Omerod (2002) remonstrated the long response time of decision processes of the *Economic Journal*: In the year 2000, it took this journal 18 weeks to reject a paper and 28 weeks to offer a revision of a paper. See also The Economic Journal Managing Editors' Report (2005, p. 6, Table 4).

41. Both the regional and the institutional breakdown of the Coupé (2000b) list can be downloaded from homepages.ulb.ac.be/~tcoupe/update/journreg.html, and from homepages.ulb.ac.be/~tcoupe/update/journuni.html. Cf. also Coupé (2000a).

published manuscripts. Yet these data are not released by editors. Therefore, we shall, as a substitute, offer Table 6 based on the Coupé data, which shows the institutional publication concentration ratios (the ratio of published pages from authors of the three main contributing institutions and the number of published pages). We have confined the entries in this table, first, to our 110 journals and, second, to those journals for which the institutional publication concentration ratios account for at least 15% of the pages.

Table 6: Institutional Publication Concentration Ratios (above 15%)

Journal	%	Journal	%
Quart. J. Economics	32.86	National Tax Journal	17.77
Economic Record*	30.75	J. Ec. Man. Strategy	17.15
Investigaciones Econ.*	24.80	Cambridge J. Ec.	16.74
Australian Ec. Papers	24.47	J. Risk and Uncertainty	16.31
J. Applied Ec.	23.69	J. Publ. Econ. Theory	16.10
Konjunkturpolitik*	22.18	Canadian J. Ec.	15.97
J. Law and Economics	20.75	J. Financial Ec.	15.63
Finanzarchiv*	19.27	Pacific Economic Rev.	15.42
J. Law, Econ. Organi.	19.03	Rand Journal of Econ.	15.07
J. Political Economy	18.95	Research Policy	15.01
Econometrica	17.77		

* Journals with mainly national supply. Source: Coupé

Table 6 reveals the impression of institutional favoritism practiced by the *Quarterly Journal of Economics*. The Coupé data reveal that the three institutions represented on 32.86% of the QJE's pages are all American institutions (Harvard, MIT, and Chicago). However, although American institutions in total account for 84.5% of all published pages of the *Quarterly Journal of Economics* according to the Coupé data, our own data do not allow addressing these facts as regional or institutional favoritism. They do not allow inference on institutional nor on regional favoritism because authors institutions or regions outside the United States of America might exhibit a particularly low submission quota. Moreover, American authors may be similarly affected by institutional favoritism of the *Quarterly Journal of Economics* (if it really obtains) as authors from other regions. Table 6 serves the purpose of providing an illustration of the problems involved.

In contrast to the Coupé data, our survey produced data on manuscript acceptance rates. However, for reasons of respondents' anonymity, we have only a regional, not an institutional breakdown of data. Thus, although our data are incapable of identifying institutional and personal favoritism, they may be put to good use to investigate regional favoritism. Table 7 shows the results. Figures are used only if they concern more than five submitted papers for a region. We focused our analysis only on acceptance rates which are different according to a χ^2 test at least at the 10%

significance level for at least two regions. The regions considered are: USA, ENGL (UK, Canada, Australia, New Zealand, Israel), and EUR (non-English speaking European countries). The rest had to be neglected due to data problems. Among all arrangements of our 110 journals for the three regions, the hypothesis of inequality between them had to be sometimes rejected at the 10% significance level for only 13 of them. These journals are presented in Table 7. A ~ sign means that equality of acceptance rates cannot be rejected at the 10% significance level, and a > sign means that the hypothesis of equality is rejected at the 10% significance level. This allows us to assume inequality.

Table 7: Regional Favoritism

Journal	USA	ENGL	EUR	Preference
Econometrica	43.94	17.39	22.48	USA>EUR~ENGL
Economic J.	33.33	44.26	27.68	ENGL>USA~EUR
Ec. Theory	72.00	-	51.85	USA>EUR
J.Ec.Theory	53.06	17.86	44.62	USA~EUR>ENGL
J. Hum. Res.	46.15	66.67	23.08	ENGL~USA~EUR; ENGL>EUR
J. Indus. Ec.	42.86	0.00	37.29	USA~EUR>ENGL
J. Internat.Ec.	50.00	22.22	27.78	USA>EUR~ENGL
J.Mo.Cr.Bank.	54.55	80.00	21.74	ENGL~USA>EUR
J.Pol.Ec.	26.32	47.83	9.72	ENGL>USA>EUR
Man. Science	60.00	-	23.81	USA>EUR
Rev. Ec. Stud.	43.75	34.48	20.00	USA~ENGL~EUR; USA>EUR
Rev. Ec. Stat.	56.00	40.00	31.37	USA~ENGL~EUR; USA>EUR
Weltw. Arch.	76.47	56.25	50.65	USA~ENGL~EUR; USA>EUR

Table 7 shows us that, quite generally, we encounter regional favoritism against non-English speaking countries.⁴² Interestingly enough, the intersection of journals figuring both in Tables 6 and 7 encompasses only two journals, viz. *Econometrica* and the *Journal of Political Economy*. When comparing the journals in Table 6 with the respective *regional* concentration indices of the Coupé data (not given here), we find that all but two (*Journal of Economic Theory*, *Journal of Industrial Economics*) have a regional quota of more than 60% in favor of one region. However, the reverse does not hold; having a regional quota of published manuscripts of more than 60% according to the Coupé data does not imply that the respective journal will also be included in Table 7. In other words, regional favoritism, according to our survey data, is reflected in higher regional manuscript publication, but not vice versa. Higher regional manuscript publication data may also result from different regional supply, rather than from incidences of favoritism.

42. Recall that this may not be a proxy for chauvinism, but may simply reflect poor mastery of the English language.

2.6.2.3 Correlation Analyses

To identify relationships among the responses to our questions, we pooled the data for *all* journals (irrespective of the number of responses per journal), and combined them into a correlation matrix, Table 8.

Table 8: Correlation Matrix of Responses to Questions

	Response Time	Number	Accept. Rate	Competence	Editorial Match	Carefulness
Number	-.021 4333					
Accept. Rate	-.029 4049	.147** 4049				
Competence	-.112** 3974	.279** 3974	.322** 3974			
Editorial Match	-.056** 3858	.032* 3858	.177** 3858	.265** 3858		
Carefulness	-.099** 3817	.276** 3817	.314** 3817	.722** 3817	.281** 3817	
Satisfaction	-.256** 3791	.199** 3791	.467** 3791	.682** 3791	.312** 3791	.684** 3791

** Significance of correlation at the 1% level (two-sided).

* Significance of correlation at the 5% level (two-sided)

The lower lines in the cells denote the number of cases.

When considering response time, we find that it is negatively correlated with all responses. Although longer response times may also be caused by more and better referee reports, the negative correlation with all responses suggests that longer response times seem to be more associated with editorial inefficiency than with more or better referee reports.

When considering the number of referee reports, we observe a moderate, but positive correlation with the acceptance rate and with the qualitative responses. The positive correlation of the number of referee reports with the acceptance rates seems to be influenced by the occurrence of manuscript rejection without referee reports. These manuscripts have no chance of being accepted. Thus, whenever referee reports are solicited, the chance of acceptance becomes greater than nought. The small correlation between the number of referee reports and the editorial match shows that referee reliability becomes a problem in case of multiple referee reports. As editors of economics journals are wont to reject a paper whenever a single one among several

referee reports is somewhat critical, irrespective of how positive the other reviews are,⁴³ authors perceive an editorial mismatch with referee suggestions. This perception seems to have caused the small correlation. Concerning the rest, more referee reports are associated with the perception of higher competence and higher carefulness, and, by that way, with higher overall satisfaction.

Prima facie one might have expected that the manuscript acceptance rate exhibits the paramount correlation with overall satisfaction. However, while that correlation *is* substantial, it is much lower than the correlation between competence and overall satisfaction, and between carefulness and overall satisfaction. The perceptions of higher carefulness and higher competence of the referee reports are associated with higher acceptance rates. Concerning the correlation of editorial match with referees' recommendations and satisfaction, one would, however, have expected a higher correlation.

The highest correlation reported in Table 8 is the one between competence and carefulness of the referee reports. Obviously, our respondents hold that a referee who does competent work also does it carefully and vice versa.⁴⁴ Both qualitative responses have at the same time the paramount positive correlations with overall satisfaction with the referee process. Thus, competence and carefulness emerge as the most important positive features of referee processes in authors' perceptions. They are even more meaningful for overall satisfaction than the acceptance rate itself.⁴⁵ This gives rise to the conjecture that authors accept rejection of their manuscripts more easily when it is based on competent and careful referee reports. And conversely, they seem to be but moderately happy with the acceptance of their paper when it was based on incompetent and sloppy referee reports.

Finally, to make use of similarities among responses, we applied a factor analysis. We employed a principal-component analysis using a varimax rotation with Kaiser Normalization.⁴⁶ This produced a factor composed of the two components: carefulness and competence, each with a factor weight of 0.539, which means that their marginal

43. Cf., e.g., Zuckerman and Merton (1971, p. 78), Bakanic et al. (1990, p. 378), Hargens and Herting (1990b, p. 97), Kupfersmid and Wonderly (1994, p. 56). In a similar sense cf. also Ingelfinger (1974, p. 687), Crandall (1982; 1991), Cole (1991), Colman (1991, p. 142), and Eckberg (1982).

44. In our instructions for response to Question 6 we used the following remark to alert respondents that competence and carefulness need not coincide: "Concerning question 4, please note that competence and carefulness may be independent." The full set of questions inclusive of instructions can be downloaded from our homepage <http://www.wiso.uni-kiel.de/ifs/chair/peerreview.php>.

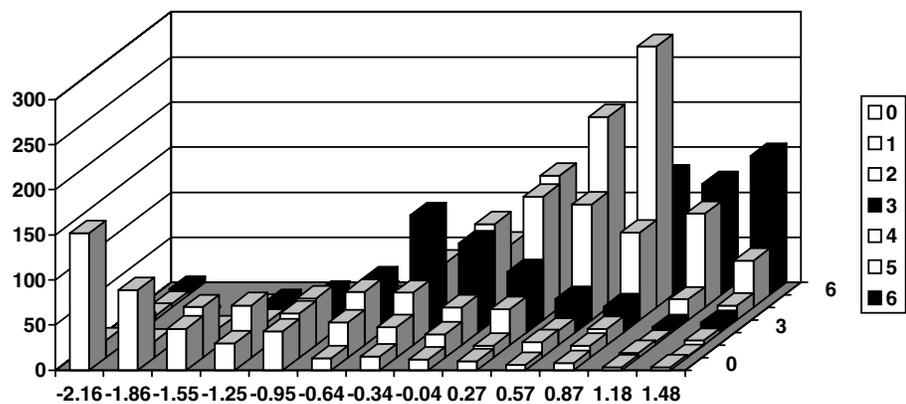
45. This result accords with the results of Garfunkel et al. (1990b) who did not find major differences in review evaluation among authors whose papers were accepted or rejected by the *Journal of Pediatrics*. However, our result for economics authors stands in remarkable contrast to the findings of Weber et al. (2002) who observed for authors of the *Annals of Emergency Medicine* that author satisfaction is associated with acceptance but not with review quality.

46. For the ease of calculation we shifted the Likert scales of questions 4-7 by 1 to Likert scales from 1 to 7. For the presentation in the figures, we stick to the scale range from 0 to 6.

rate of substitution is equal to -1. We call this factor *quality*. It explains 86.114% of the variance among the two characteristics competence and carefulness. Competence and carefulness are, therefore, good proxies for the factor “quality”. Our analysis yielded 48 factor levels, which we found to be arranged in terms of 13 distinct groups. Representing each group by its median allowed us to focus on 13 representative factor levels. A negative (positive) factor value means that a subject exhibits a less (better)-than-average evaluation of the respective journal. A factor value of zero corresponds to the average evaluation.

Associating these 13 factor levels with the seven levels of overall satisfaction shows a characteristic distributional pattern: For low quality we observe a positively skewed distribution of satisfaction. As quality increases, the distribution of satisfaction becomes symmetrical, and gradually becomes negatively skewed as quality approaches its peak. Note that, although this pattern is in a way due to bunching effects inherent in categorical measurement, it is, nevertheless, rather distinctive in this case. Figure 1 shows the respective graph, which arranges normalized quality at the abscissa, satisfaction at the ordinate. The vertical axis indicates the absolute frequency of our 3817 data points.

Figure 1: Distribution of Satisfaction for Factor Levels



Low levels of satisfaction are caused by a positively skewed distribution of factor values having their peak at the lowest factor level for the lowest level of satisfaction. For higher satisfaction levels the distribution of factor levels converges first to a symmetric distribution which is reached at the medium satisfaction level. For still higher satisfaction levels the distribution of factor levels assumes the shape of negatively skewed distributions. The factor distribution for the highest satisfaction level has its peak at the highest factor level.

Figure 1 depicts a mountain extending across the figure from the (-2.16, 0) coordinate point to the (1.48, 6) coordinate point. The steepness of this mountain on both sides of its ridge is captured by the correlation coefficient between the quality factor and overall satisfaction. Its value is 0.735. It is significant (two-sided) at the 1% level. This illustrates a good explanation of overall satisfaction with the referee process by competence and carefulness of the referee reports.

Finally, we have a look at the joint distribution of overall satisfaction and subjective acceptance rates. Figure 2 shows the respective graph. We observe negatively skewed distributions for all intervals of acceptance rates except the lowest acceptance rates (virtually rejections). Subjects whose papers are often rejected are not distinctly dissatisfied. Although they are not extremely enthusiastic about rejection, we encounter in this pattern a reflection of the appreciation of careful and competent referee reports. Good-quality referee reports may, thus, indeed cause authors to understand rejection of their manuscripts.

Figure 2: Distribution of Satisfaction with Referee Reports in Terms of Acceptance Rates

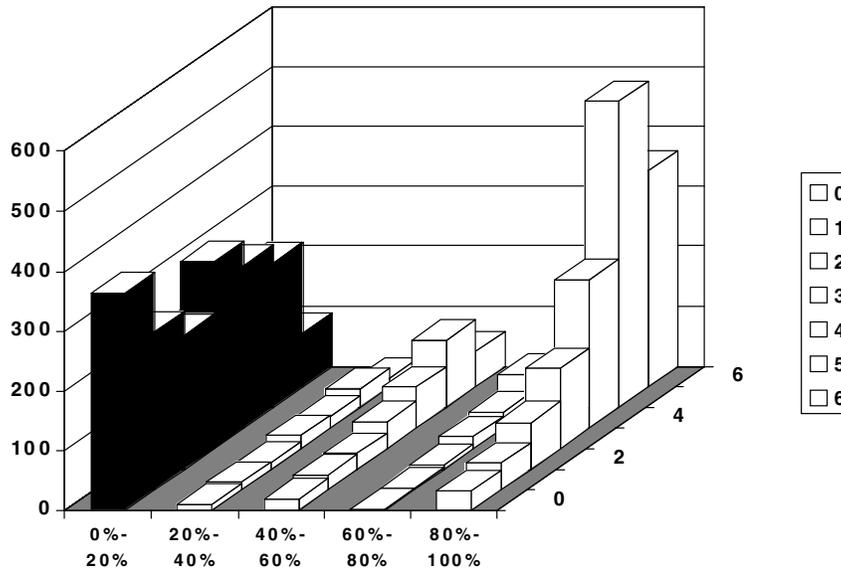
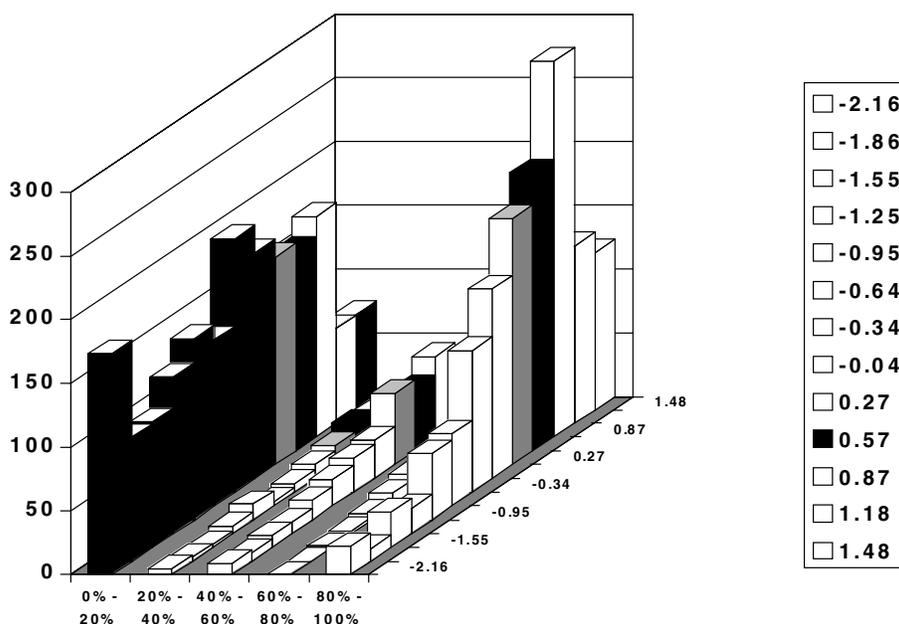


Figure 3 repeats this exercise for the 13 representative factor levels. For higher acceptance rates we observe negatively skewed distributions, while for the lowest acceptance rates satisfaction is rather evenly distributed with the exception of extreme happiness.

2.6.2.4 Other Determinants of Satisfaction with Peer Review

When explaining satisfaction with peer review, other determinants come to mind. First, we could have made a regression analysis to estimate overall satisfaction. This would have required personalized data on the respondents which we do not have (with the exception of the country of residence). Had we required more personalized data, we would have experienced much more dropouts.⁴⁷

Figure 3: Distribution of Satisfaction with Referee Reports in Terms of Acceptance Rates for Factor Levels



Second, we could have used some objective journal data to estimate aggregate overall satisfaction of the respective journals. We could draw on the impact factor, or on several of the many attempts at journal rankings. Indeed, we tried hard to explain overall satisfaction by impact factors, but we failed. Impact factors are highly volatile. The results are highly dependent on the set of impact factors (with respect to the periods or aggregates of them) chosen. Moreover, impact factors form important inputs for most methods to compute journal rankings. Given the manifold of journal rankings,

47. Some colleagues told us privately that they were tempted to participate in our survey, but refrained from it for fear of being identified by some inferential techniques!

we would be at a loss to decide which is the most appropriate.⁴⁸ Various journal rankings can, of course, be compared with the journal ranking found in this study for overall satisfaction.⁴⁹ We doubt, however, that the ranking which conforms best with the ranking of our study would be the best ranking method. Rather it comes up to be a tautology.

3. CONCLUSIONS

Peer review in science is a tribunal of sorts. It influences decisively personal advancement, research opportunities, salaries, grant-funding, promotion, and tenure. Peer review claims to exert quality control of manuscripts, to improve manuscripts, to promote innovative research, to foster dissemination of new research, to select projects for grant funding, to screen papers for conference presentation, and to serve as a means to rank researchers, journals, and institutions.

Yet journals no longer serve the function of disseminating new research. As early as in 1971, Garvey and Griffith had already demonstrated that the bulk of communication and dissemination of current research runs over informal outlets such as personal communications, technical reports, discussion papers, and preprints. Eventual publication of a paper means that it had entered the archives of science, while its author had long ago started new research. This applies even more so in the electronic age.

The main purpose of journal publication nowadays is to imprint a signal of quality on a scholar's research. However, this requires an excellent performance of peer review. When peer review lacks validity, impartiality, and fairness, the imprint of manuscript excellence becomes dubious. Empirical research has shown that peer review lacks precisely these attributes. Instead of a sound balance of Type I and Type II errors, the humanities are extremely averse to Type I errors, which has sky-rocketed the journal rejection rates, while the natural sciences are perhaps too much keen to avoid Type II errors, which has resulted in fantastically high acceptance rates.

Referee replication studies and experiments to test referee quality have demonstrated poor general quality of peer review. Evidence of manuscript improvements due to peer review is very weak and statistically insignificant. Because of extreme aversion to commit Type I errors, the dissemination of innovations is impeded by peer review. Editors' and referees' search for easy signals of quality gave rise to institutional, personal, and regional favoritism and to a confirmatory bias in paper evaluations.

48. We mainly referred to the Diamond list because it is the classical attempt at defining journal excellence. Moreover, the journals of the Diamond list are contained in most journal rankings.

49. See Table 9* on our homepage <http://www.wiso.uni-kiel.de/ifs/chair/peerreview.php>.

Refusal of publishers and editors of nonprofit economics journals to expand the number of pages in their journals has boosted the rejection rates of economics journals. Other than journals in the natural sciences, they failed to establish either specialized spin-off journals or page fees. Thus, the publishers of nonprofit economics journals ceded the field to the commercial journals.

These limitations induced us to conduct an internet questionnaire investigation of authors of economics journals. We found much longer response times than what is customary in the natural sciences. The top journals had on average high rejection rates. While the top journals did not show particular differences from other journals with respects to the distribution of competence and carefulness of referee reports, they perform somewhat worse for overall satisfaction. Moreover, it is always the same group of some eight top economics journals which populate the bottom rungs in the respective rankings.

We observed response biases among our respondents: the subjective response time exceeds the objective one by some 50% and the subjective acceptance rate exceeds the objective one by some 150%. This may be explained by several effects (self-selection, survey-selection, cognitive dissonance, and trend). Only in rare cases could we isolate favoritism. When it occurred, it discriminated by and large against non-English-speaking countries, perhaps due to poor command of the English language.

A correlation analysis showed that competence and carefulness are highly correlated, and showed the paramount correlation with overall satisfaction, while the acceptance rate exhibited a smaller correlation with overall satisfaction. This suggests that the authors of economics journals have a higher esteem for good referee reports in comparison to a mere focus on the acceptance rate. In other words, they will understand a rejection of their paper if it is backed by well-founded reports.

Combining competence and carefulness into a factor “quality” showed that, as quality increases, the distribution of satisfaction follows, first, a positively skewed distribution higher levels of quality, a symmetrical distribution and approaches a negatively skewed distribution for the highest level of quality. When juxtaposing overall satisfaction [quality] and acceptance rates, we found negatively skewed distributions for all acceptance rates with the exception of the very lowest acceptance rates, for which the distribution is largely uniform. This confirms that manuscript rejection is tolerated provided that the referee reports are competent and carefully done.

Acknowledgements

This paper was presented at the ESA/Public Choice Conference 2003 in Nashville and at the PET 04 Conference 2004 in Beijing. For helpful comments we are indebted to Söhnke Albers, Ted Bergstrom, John Conley, Leigh Hobson, Alan Kirman, Stefan Traub, and Joachim Wolf. The usual disclaimer applies.

REFERENCES

- Adair, R.K. (1982), A Physics Editor Comments on Peters and Ceci's Peer-Review Study, *Behavioral and Brain Sciences* 5, 196.
- Als-Nielsen, B., Chen, W., Gluud, C., Kjaergard, L.L. (2003), Association of Funding and Conclusions in Randomized Drug Trials, A Reflection of Treatment Effect or Adverse Effects?, *Journal of the American Medical Association* 290, 921-8.
- Altman, D.G. (2002), Poor-Quality Medical Research. What Can Journals Do? *Journal of the American Medical Association* 287, 2765-7.
- Armstrong, J.S. (1982), Barriers to Scientific Contributions: The Author's Formula, *Behavioral and Brain Sciences* 5, 197-9.
- Armstrong, J.S. (2002), Discovery and Communication of Important Marketing Findings: Evidence and Proposals, *Journal of Business Research* 56, 69-84.
- Atkinson, D.R., Furlong, M.J., Wampold, B.E. (1982), Statistical Significance, Reviewer Evaluations, and the Scientific Process: Is there a Statistically Significant Relationship?, *Journal of Counseling Psychology* 29, 189-94.
- Bailar, J.C. (1991), Reliability, Fairness, Objectivity and other Inappropriate Goals of Peer Review, *Behavioral and Brain Sciences* 14, 137-8.
- Bairam, E.I. (1994), Institutional Affiliation of Contributors to Top Economic Journals, 1985-1990, *Journal of Economic Literature* 32, 674-9.
- Bakanic, V., McPhail, C., Simon, R.J. (1987), The Manuscript Review Process, *American Sociological Review* 52, 631-42.
- Bakanic, V., McPhail, C., Simon, R.J. (1990), If at First You Don't Succeed: Review Procedures for Revised and Resubmitted Manuscripts, *The American Sociologist* 21, 373-91.
- Banerjee, M., and Findley, L. (1991), Propanolol in the Treatment of Acute Migraine Attacks, *Cephalalgia* 11, 193-6.
- Baxt, W.G., Waeckerle, J.F., Berlin, J.A., Callahan, M.L. (1998), Who Reviews the Reviewers? Feasibility of Using Fictitious Manuscript to Evaluate Peer Reviewer Performance, *Annals of Emergency Medicine* 32, 310-7.
- Ben-Yashar, R., Nitzan, S. (2001), Are Referees Sufficiently Informed About the Editor's Practise?, *Theory and Decision* 51, 1-11.
- Bergstrom, T.C. (2001), Free Labor for Costly Journals?, *Journal of Economic Perspectives* 15, 183-98.
- Black, N., van Rooyen, S., Godlee, F., Smith, R., Evans, S. (1998), What Makes a Good Reviewer and a Good Review for a General Medical Journal? *Journal of the American Medical Association* 280, 231-3.

- Blank, R.M. (1991), The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from *The American Economic Review*, *American Economic Review* 81, 1041-67.
- Bozarth, J.D., Roberts, R.R. (1972), Signifying Significant Significance, *American Psychologist* 27, 774-5.
- Broad, W.J. (1980a), Imbroglio at Yale (I): Emergence of a Fraud, *Science* 210, 38-41.
- Broad, W.J. (1980b), Would-Be Academician Pirates Papers, *Science* 208, 1438-40.
- Broad, W.J. (1981a), Fraud and the Structure of Science, *Science* 212, 137-41.
- Broad, W.J. (1981b), Congress Told Fraud "Exaggerated", *Science* 212, 421.
- Broad, W.J., Wade, N. (1982), *Betrayers of the Truth*, Simon and Schuster: New York.
- Budd, J.M., Sievert, M.E., Schultz, T.R. (1998), Phenomena of Retraction. Reasons for Retraction and Citations to the Publications, *Journal of the American Medical Association* 280, 296-7.
- Burnham, J.C. (1990), The Evolution of Editorial Peer Review, *Journal of the American Medical Association* 263, 1323-9.
- Burton, M.P., Phimister, E. (1995), Core Journals: A Reappraisal of the Diamond List, *The Economic Journal* 105, 361-73.
- Callahan, M.L., Wears, R.L., Weber, E.J., Barton, C., Young, G. (1998), Positive-Outcome Bias and Other Limitations in the Outcome of Research Abstracts Submitted to a Scientific Meeting, *Journal of the American Medical Association* 280, 254-7.
- Chalmers, I. (1985), Proposal to Outlaw the Term 'Negative Trial', *British Medical Journal* 290, 1002.
- Cicchetti, D.V. (1991a), The Reliability of Peer Review for Manuscript and Grant Submissions: A Cross-Disciplinary Investigation, *Behavioral and Brain Sciences* 14, 119-35.
- Cleary, F.R., Edwards, D.J. (1960), The Origins of the Contributors to the A.E.R. During the 'Fifties', *American Economic Review* 50, 1011-4.
- Cole, S. (1991), Consensus and the Reliability of Peer-Review Evaluations, *Behavioral and Brain Sciences* 14, 140-1.
- Colman, A.M. (1991), Unreliable Peer Review: Causes and Cures of Human Misery, *Behavioral and Brain Sciences* 14, 141-2.
- Coupé, T. (2000a), Revealed Performances: World Wide Rankings of Economists and Economic Departments, unpublished manuscript, Free University of Brussels.
- Coupé, T. (2000b), ECARES/Université Libre de Bruxelles. Data on the main regions of journal contributors: homepages.ulb.ac.be/~tcoupe/update/journreg.html. Data on the main institutions of journal contributors: homepages.ulb.ac.be/~tcoupe/update/journuni.html.

- Crandall, R. (1982), Editorial Responsibilities in Manuscript Review, *Behavioral and Brain Sciences* 5, 207-8.
- Crandall, R. (1991), What Should Be Done to Improve Reviewing?, *Behavioral and Brain Sciences* 14, 143.
- Dewald, W.G., Thursby, J.G., Anderson, R.G. (1986), Replication in Empirical Economics: The Journal of Money, Credit and Banking Project, *American Economic Review* 76, 587-603.
- Diamond, A.M. Jr. (1989), The Core Journals of Economics, *Current Contents* 1, 4-11.
- Dickersin, K. (1990), The Existence of Publication Bias and Risk Factors for Its Occurrence, *Journal of the American Medical Association* 263, 1385-9.
- Eckberg, D.L. (1982), Theoretical Implications of Failure to Detect Prepublished Submissions, *Behavioral and Brain Sciences* 5, 209-10.
- Eichhorn, P., Yankauer, A. (1987), Do Authors Check Their References? A Survey of Accuracy of References in Three Public Health Journals, *American Journal of Public Health* 77, 1011-2.
- Elliott, C., Greenaway, D., Sapsford, D. (1998), Who's Publishing Who? The National Composition of Contributors to Some Core US and European Journals, *European Economic Review* 42, 201-6.
- Ellison G. (2002 a), The Slowdown of the Economics Publishing Process, *Journal of Political Economy* 110, 947-93.
- Ellison, G. (2002 b), Evolving Standards for Academic Publishing: A q-r Theory, *Journal of Political Economy* 110, 994-1034.
- Epstein, W.M. (1990), Confirmational Response Bias Among Social Work Journals, *Science, Technology, and Human Values* 15, 9-38.
- Epstein, W.M. (2004), Confirmational Response Bias and the Quality of the Editorial Processes among American Social Work Journals, *Research on Social Work Practice* [in press].
- Evans, J.T., Nadjari, H.I., Burchell, S.A. (1990), Quotational and Reference Accuracy in Surgical Journals: A Continuing Peer Review Problem, *Journal of the American Medical Association* 263, 1353-4.
- Fisher, M. Friedman, S.B., Strauss, B. (1994), The Effects of Blinding on Acceptance of Research Papers by Peer Review, *Journal of the American Economic Association* 272, 143-6.
- Fiske, D.W. and Fogg, L. (1990), But the Reviewers are Making Different Criticisms of my Paper! Diversity and Uniqueness in Reviewer Comments, *American Psychologist* 45, 591-8.
- Fourth International Congress on Peer Review in Biomedical Publication (2002), *Journal of the American Medical Association* 287, 2758-871.
- Friedman, P.J. (1990), Correcting the Literature Following Fraudulent Publication, *Journal of the American Medical Association* 263, 1416-9.

- Fuller, G.N., and Guiloff, R.J. (1990), Propanolol in Acute Migraine: A Controlled Study, *Cephalalgia* 10, 229-33.
- Fusfeld, R. (1956), The Program of the American Economic Association Meetings, *American Economic Review* 46, 642-4.
- Garfield, E. (1972), Citation Analysis as a Tool in Journal Evaluation, *Science* 178, 471-9.
- Garfunkel, J.M., Ulshen, M.H., Hamrick, H.J. Lawson, E.E. (1990a), Problems Identified by Secondary Review of Accepted Manuscripts, *Journal of the American Medical Association* 263, 1369-71.
- Garfunkel, J.M., Lawson, E.E., Hamrick, H.J., Ulshen, M.H. (1990b), Effect of Acceptance or Rejection on the Author's Evaluation of Peer Review of Medical Manuscripts, *Journal of the American Medical Association* 263, 1376-8.
- Garvey, W.D., Griffith, B.C. (1971), Scientific Communication: Its Role in the Conduct of Research and Creation of Knowledge, *American Psychologist* 26, 349-62.
- Gilmore, J.B. (1991), On Forecasting Validity and Finessing Reliability, *Behavioral and Brain Sciences* 14, 148-9.
- Glenn, N.D. (1982), The Journal Article Review Process as a Game of Chance, *Behavioral and Brain Sciences* 5, 211-2.
- Goodman, S.N., Berlin, J., Fletcher, S.W., Fletcher, R.H. (1994), Manuscript Quality Before and After Peer Review and Editing at *Annals of Internal Medicine*, *Annals of Internal Medicine* 121, 11-21.
- Goodstein, L.D. (1982), When Will the Editors Start to Edit? *Behavioral and Brain Sciences* 5, 212-3.
- Goodstein, L.D. Brazis, K.L. (1970), Credibility of Psychologists: An Empirical Study, *Psychological Reports* 27, 835-8.
- Gordon, R.A. (1982), Optional Published Refereeing, *Behavioral and Brain Sciences* 5, 213-4.
- Goudsmit, S.A. (1969), What Happened to My Paper?, *Physics Today* 22/5, 23-5.
- Greenwald, A.G. (1975), Consequences of Prejudice Against the Null Hypothesis, *Psychological Bulletin* 82, 1-20.
- Griffith, B.C. (1982), Judging Document Content Versus Social Functions of Refereeing: Possible and Impossible Tasks, *Behavioral and Brain Sciences* 5, 214-5.
- Guarding the Guardians: Research on Editorial Peer Review. Selected Proceedings from the First International Congress on Peer Review in Biomedical Publication (1990), *Journal of the American Medical Association* 263, 1309-456.
- Gulliksen, H.O. (1950), *Theory of Mental Tests*, Wiley, New York.

- Hamermesh, D.S. (1994), Facts and Myths about Refereeing, *Journal of Economic Perspectives* 8, 153-63.
- Hargens, L.L. (1988), Scholarly Consensus and Journal Rejection Rates, *American Sociological Review* 53, 139-51.
- Hargens, L.L. (1990), Variation in Journal Peer Review Systems: Possible Causes and Consequences, *Journal of the American Medical Association* 263, 1348-52.
- Hargens, L.L., Herting, J.R. (1990a), A New Approach to Referees Assessments of Manuscripts, *Social Science Research* 19, 1-16.
- Hargens, L.L., Herting, J.R. (1990b), Neglected Considerations in the Analysis of Agreement Among Journal Referees, *Scientometrics* 19, 91-106.
- Hawkins, R.G., L.S. Ritter, S. Lawrence, and I. Walter (1973), What Economists Think of their Journals, *Journal Political Economy* 81, 1017-32.
- Hodgson, G.M. Rothman, H. (1999), The Editors and Authors of Economics Journals: A Case of Institutional Oligopoly? *The Economic Journal* 109, F165-F186.
- Hogan, R. (1982), The Insufficiencies of Methodological Inadequacy, *Behavioral and Brain Sciences* 5, 216.
- Honig, W.M. (1982), Peer Review in the Physical Sciences: An Editor's View, *Behavioral and Brain Sciences* 5, 216-7.
- Horrobin, D.F. (1982), Peer Review: A Philosophically Faulty Concept Which is Proving Disastrous for Science, *The Behavioral and Brain Sciences* 5, 217-8.
- Horrobin, D.F. (1990), The Philosophical Basis of Peer Review and the Suppression of Innovation, *Journal of the American Medical Association* 263, 1438-41.
- Horton, R. (2002), Postpublication Criticism and the Shaping of Clinical Knowledge, *Journal of the American Medical Association* 287, 2843-7.
- Howe, M.J.A. (1982), Peer Reviewing: Improve or be Rejected, *Behavioral and Brain Sciences* 5, 218-9.
- Hubbard, R., Vetter, D.E. (1996), An Empirical Comparison of Published Replication Research in Accounting, Economics, Finance, Management and Marketing, *Journal of Business Research* 35, 153-64.
- Ingelfinger, F.J. (1974), Peer Review in Biomedical Publications, *The American Journal of Medicine* 56, 686-92.
- Jauch, L.R., Wall, J.L. (1989), What They Do When they Get Your Manuscript: A Survey of Academy of Management Reviewer Practices, *Academy of Management Journal* 32, 157-73.
- Juhasz, S., Calvert, E., Jackson, T., Kronick, D.A., Shipman, J. (1975), Acceptance and Rejection of Manuscripts, *IEEE Transactions on Professional Communication* PC-18, 177-85.

- Kalaitzidakis, P., Mamuneas, T.P., Stengos, T. (1999), European Economics: An Analysis Based on Publications in the Core Journals, *European Economic Review* 43, 1150-68.
- Kalaitzidakis, P., Mamuneas, T.P., Stengos, T. (2001), Ranking of Academic Journals and Institutions in Economics, Discussion Paper.
- Kerr, R., Tolliver, J., Petree, D. (1977), Manuscript Characteristics Which Influence Acceptance for Management and Social Science Journals, *Academy of Management Journal* 20, 132-41.
- Kiesler, C.H. (1991), Confusion Between Reviewer Reliability and Wise Editorial and Funding Decisions, *Behavioral and Brain Sciences* 14, 151-2.
- Kocher, M.G., Sutter, M. (2000), The Market for Publications in Top Economics Journals, Concentration and Export Ratios of Authors' Current and Ph.D.-Affiliations for the Last Two Decades and Power Laws of Scientific Productivity, Discussion Paper 2000/1, University of Innsbruck.
- Kosinski, J. (1968), *Steps*, Random House, New York.
- Kraemer, H.C. (1991), Do We Really Want More 'Reliable' Reviewers?, *Behavioral and Brain Sciences* 14, 152-4.
- Kroninck, D.A. (1990), Peer Review in the 18th-Century Scientific Journalism, *Journal of the American Medical Association* 263, 1321-2.
- Krzyzanowska, M.K., Pintilie, M., Tannock, I.F. (2003), Factors Associated With Failure to Publish Large Randomized Trials Presented at an Oncology Meeting, *Journal of the American Medical Association* 290, 495-501.
- Kupfersmid, J., Wonderly, D.M. (1994), *An Author's Guide to Publishing Better Articles in Better Journals in the Behavioral Sciences*, Clinical Psychology Publishing Co., Brandon VT.
- Laband, D.N. (1990), Is There Value-Added From the Review Process in Economics?: Preliminary Evidence From Authors, *The Quarterly Journal of Economics* 105, 341-52.
- Lazarus, D. (1982), Interreferee Agreement and Acceptance Rates in Physics, *Behavioral and Brain Sciences* 5, 219.
- Link, A.M. (1998), US and Non-US Submissions. An Analysis of Reviewer Bias, *Journal of the American Medical Association* 280, 246-7.
- Lock, S., and Smith, J. (1990), What do Peer Reviewers Do?, *Journal of the American Medical Association* 263, 1341-3.
- McCutchen, C.W. (1991), Peer Review: Treacherous Servant, Disastrous Master, *Technology Review* 94/7, 28-40.
- McNutt, R.A., Evans, A.T., Fletcher, R.H., Fletcher, S.W. (1990), The Effects of Blinding on the Quality of Peer Review, A Randomized Trial, *Journal of the American Medical Association* 263, 1371-6.
- Mahoney, M.J. (1977), Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System, *Cognitive Therapy and Research* 1, 161-75.

- Mahoney, M.J. (1982), Publication, Politics, and Scientific Progress, *Behavioral and Brain Sciences* 5, 220-1.
- Mahoney, M.J., Kimper, T.P. (1976), From Ethics to Logic: A Survey of Scientists, in: M.J. Mahoney (ed), *Scientist as Subject*, Ballinger, Cambridge MA, 187-93.
- Malouin, J.-L., Outreville, J.F. (1987), The Relative Impact of Economics Journals: A Cross-Country Survey and Comparison, *Journal of Economics and Business* 39, 267-77.
- Manwell, C., Baker, C.M.A. (1979), The Double Helix: Science and Myth in the Act of Creation, *Bioscience* 29, 742-6.
- Marsh, H.W., and Ball, S. (1989), The Peer Review Process Used to Evaluate Manuscripts Submitted to Academic Journals: Interjudgmental Reliability, *Journal of Experimental Education* 57, 151-69.
- Merton R.K. (1968), The Matthew of Effect in Science, *Science* 159, 56-63.
- Meyer, D. (1999), Publishing Articles – Findings and Open Questions of an Empirical Study Conducted Among Editors of Economic Journals, *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 119, 531-59.
- Mindick, B. (1982), When We Practice to Deceive: The Ethics of a Metascientific Inquiry, *Behavioral and Brain Sciences* 5, 226-7.
- Munley, P.H., Sharkin, B., and Gelso, C.J. (1988), Reviewer Ratings and Agreement on Manuscripts Reviewed for the Journal of Counseling Psychology, *Journal of Counseling Psychology* 35, 198-202.
- Murray, G.D. (1988), The Task of a Statistical Referee, *British Journal of Surgery* 75, 664-7.
- Nylenna, M., Riis, P., Karlsson, Y. (1994), Multiple Blinded Reviews of the Same Two Manuscripts: Effects of Referee Characteristics and Publication Language, *Journal of the American Medical Association* 272, 149-51.
- Olson, C.M., Rennie, D., Cook, D., Dickersin, K., Flanagan, A., Hogan, J.W., Zhu, Q., Reiling, J., Pace, B. (2002), Publication Bias in Editorial Decision Making, *Journal of the American Medical Association* 287, 2825-8.
- Omerod, P. (2002), Dealing With Journal Submissions, *Royal Economic Society Newsletter* No. 117, 18.
- Over, R. (1982), What is the Source of Bias in Peer Review?, *Behavioral and Brain Sciences* 5, 229-30.
- Perlman, D. (1982), Reviewer 'Bias': Do Peters and Ceci Protest Too Much?, *Behavioral and Brain Sciences* 5, 231-2.
- Peters, D.P., Ceci, S.J. (1980), A Manuscript Masquerade: How Well Does the Review Process Work?, *The Sciences* 20, 16-9, and 35.

- Peters, D.P., and Ceci, S.J. (1982a), Peer-Review Practices of Psychology Journals: The Fate of Published Articles, Submitted Again, *Behavioral and Brain Sciences* 5, 187-95.
- Peters, D.P., and Ceci, S.J. (1982b), Peer-Review Research: Objections and Obligations, *Behavioral and Brain Sciences* 5, 246-52.
- Pfeifer, M.P., Snodgrass, G.L. (1990), The Continued Use of Retracted, Invalid Scientific Literature, *Journal of the American Medical Association* 253, 1420-3.
- Popper, K.R. (1959), *The Logic of Scientific Discovery*, Basic Books, New York.
- Purcell, K., Brady, K., Chai, H. (1969), The Effect on Asthma in Children of Experimental Separation From the Family, *Psychosomatic Medicine* 31, 144-64.
- Roediger, H.L. (1991), Is Unreliability in Peer Review Harmful? *Behavioral and Brain Sciences* 14, 159-60.
- Rosenthal, R. (1982), Reliability and Bias in Peer-Review Practices, *Behavioral and Brain Sciences* 5, 235-6.
- Ross, C. (1979), Rejected, *New West* 4, 39-41.
- Ross, C. (1982), Rejecting Published Work: Similar Fate for Fiction, *Behavioral and Brain Sciences* 5, 236.
- Rowney, J.A., Zenisek, T.J. (1980), Manuscript Characteristics Influencing Reviewers' Decisions, *Canadian Psychology/Psychologie Canadienne* 21, 17-21.
- Scarr, S. (1982), Anosmic Peer Review: A Rose by Another Name is Evidently not a Rose, *Behavioral and Brain Sciences* 5, 237-8.
- Schor, S., Karten, I. (1966), Statistical Evaluation of Medical Journal Manuscripts, *Journal of the American Medical Association* 195, 1123-8.
- Second International Congress on Peer Review in Biomedical Publication (1994), *Journal of the American Medical Association* 272, 91-170.
- Shepherd, G.B. (ed.) (1995), *Rejected, Leading Economists Ponder the Publication Process*, Thomas Horton and Daughters, Sun Lakes, Arizona.
- Siegfried, J.J. (1972), The Publishing of Economic Papers and its Impact on Graduate Faculty Ratings, 1960-1969, *Journal of Economic Literature* 10, 31-49.
- Smart, R. (1964), The Importance of Negative Results in Psychological Research, *Canadian Psychologist* 5a, 225-32.
- Speck, B.W. (1993), *Publication Peer Review, An Annotated Bibliography*, Greenwood Press, Westpoint (Conn.) and London.
- Sterling, T.D. (1959), Publication Decisions and Their Possible Effects on Inferences Drawn From Tests of Significance - or Vice Versa, *Journal of the American Statistical Association* 54, 30-4.

- Steward, W.W., Feder, N. (1987), The Integrity of the Scientific Literature, *Nature* 325, 207-14.
- Stricker, L.J. (1991), Disagreement Among Journal Reviewers: No Cause for Undue Alarm, *Behavioral and Brain Sciences* 14, 163-4.
- Sweitzer, B.J., Cullen, D.J. (1994), How Well Does a Journal's Peer Review Process Function? A Survey of Authors' Opinions, *Journal of the American Medical Association* 272, 152-3.
- Szenberg, M. (1994), Disseminations Scholarly Output: The Case for Eliminating the Exclusivity of Journal Submissions, *American Journal of Economics and Sociology* 53, 303-15.
- The Economic Journal Managing Editors' Report (2005), *Royal Economic Society Newsletter* No. 128 (January 2005), 3-7. Third International Congress on Peer Review in Biomedical Publication (1998), *Journal of the American Medical Association* 280, 213-302.
- Von Rauchhaupt, U. (2002), Betrug. Tief gestürzt: Ein deutscher Physiker fälschte systematisch, *Frankfurter Allgemeine Zeitung*, September 27, 2002.
- Wade, N. (1976), IQ and Heredity: Suspicion of Fraud Beclouds Classic Experiment, *Science* 194, 916-9.
- Weber, E.J., Katz, P.P., Waeckerle, J.F., Callahan, M.L. (2002), Author Perception of Peer Review. Impact of Review Quality and Acceptance on Satisfaction, *Journal of the American Medical Association* 287, 2790-3.
- Wilson, J.D. (1978), Peer Review and Publication, *Journal of Clinical Investigation* 61, 1697-701.
- Wilson, W.A. (1982), Research on Peer-Review Practices: Problems of Interpretation, Application, and Propriety, *Behavioral and Brain Sciences* 5, 242-3.
- Yalow, R.S. (1982), Competency Testing for Reviewers and Editors, *Behavioral and Brain Sciences* 5, 244-5.
- Yankauer, A. (1990), Who are the Peer Reviewers and How Much do they Review?, *Journal of the American Medical Association* 263, 1338-40.
- Yotopoulos, P.A. (1961), Institutional Affiliation of the Contributors to Three Professional Journals, *American Economic Review* 51, 665-70.
- Ziman, J. (1982), Bias, Incompetence, or Bad Management?, *Behavioral and Brain Sciences* 5, 245-6.
- Zuckerman, H., Merton, R.K. (1971), Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System, *Minerva* 9, 66-100.