

This two-part article reviews the current literature on journal peer review. Research on this subject has grown during the 1980s and 1990s, and has increased our awareness of both the myths and facts about peer review. Part 1 summarizes research findings on the participants in the system (the appointment mechanisms of editors and referees, and reviewer tasks and qualifications), and systemic problems of reliability, accuracy, and bias. Part 2 describes current research on how fraud, favoritism, and self-interest may affect the review system and on such policy issues as interference of particularistic criteria; connections among editors, authors, and referees; and double-blind reviewing. Although the literature indicates that peer review has many problems, the author concludes that it is difficult to imagine how science could advance without such a key quality control mechanism.

Peer Review for Journals as It Stands Today—Part 1

JUAN MIGUEL CAMPANARIO

University of Alcalá

Weight opinions, do not count them.

—Seneca

Printed academic journals are the major venue through which scientists communicate their results, vent their opinions, and exchange observations. Scientific journals are the means by which the scientific community certifies accumulations and additions to its body of accepted knowledge and the means by which scholars compete as if in a mental Olympiad for prestige and recognition (Hargens 1988; Merton 1957). Not surprisingly, publication in academic journals is judged as an indicator of performance of an individual

Author's Note: The author thanks Jerry Keller for his help in the writing of the article and the *Science Communication* referees for their comments. Address correspondence to Juan Miguel Campanario, Grupo de Investigación en Aprendizaje de las Ciencias, Departamento de Física, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain; phone: 34-1-8854926; fax: 34-1-8854942; e-mail: fscampanario@alcala.es.

Editor's Note: Part 2 of this review article will appear in the June 1998 issue of *Science Communication* (Volume 19, Number 4).

Science Communication, Vol. 19 No. 3, March 1998 181-211
© 1998 Sage Publications, Inc.

effort, and on many occasions it fosters an author's career advancement (Diamond 1986; Lindsey 1976; Peters and Ceci 1982a; Tuckman and Leahey 1975). This prestige—and the reason we trust journal quality—rests on the process by which manuscripts are evaluated before publication; that is, the peer review system. Typically, two or more reviewers assess the soundness of a manuscript's ideas and results, its methodological and conceptual viewpoint, its quality, and its potential impact on the world of science.

Ziman (1968, 111) said "the referee is the linchpin about which the whole business of Science is pivoted." Others have referred to editorial board members and external referees as the "gatekeepers of science" (Beyer 1978; Crane 1967; Glogoff 1988; Pipke 1984; Zsindely and Schubert 1989; Zsindely, Schubert, and Braun 1982). Editors similarly mark their personal characters on their journals, and may adopt one or more of various roles: gatekeeper, shifter, arbitrator, curmudgeon, or champion of scientific orthodoxy (Roediger 1987, 227) and, sometimes, a scout who scurries the scientific no-man's-land to identify and publish the papers with relatively substantive impact (Roediger 1987, 240). As former *American Sociological Review* editor Rita J. Simon (1994) noted, the editor of a scientific journal has the authority at whim to launch a new author or a new book into the world of competing journals, or to withhold the person's achievements and thus shutter the career of an aspiring scientist.

Because peer review constitutes such a key part of science, it would be natural to assume that it has long been a major object of study, and that study of peer review ought to be a strategic research goal analyzing the dynamics of science and the patterns by which new knowledge is transferred. However, research on peer review is relatively recent and scarce; many discussions of it are based on personal observations rather than systematic data gathering (Scharschmidt et al. 1994). It is an area that is understudied (Kassirer and Campion 1994), uncomprehended, and unchartered (Peters and Ceci 1982b, 252). LaFollette (1992, 121), for example, noted the scarcity of research on such questions as how referees apply standards, the specific criteria established for making a decision, and who makes the final decision on a manuscript. In 1986, the Institute for Scientific Information, in clustering the citations in their files, identified a research front titled "Objectivity of Reviewers in Peer Review," which consisted of only two core studies and twelve cited papers (Garfield 1986b). Armstrong (1997) recently identified fifty-six empirical studies on peer review, but only three of them had been published before 1975.

Among the first modern studies on peer review were those accomplished by Zuckerman and Merton (1971) and Crane's (1967) research on the patterns of scientific communication. In 1982, a controversial paper by Peters and

Ceci (1982a) published in *Behavioral and Brain Sciences* stimulated a heated debate on peer review in academia. Bailar and Patterson (1985) later noted the absence of studies on editorial peer review and issued an appeal for a research agenda. Among others, Garfield (1986a, 1986b) and Rennie (1993) have raised concerns about the necessity of research on peer review. As a result of this interest, a series of International Congresses on Peer Review in Biomedical Publication were initiated (the First International Congress on Peer Review in Biomedical Publication was held in Chicago, Illinois, 10-12 May 1989; the Second International Congress on Peer Review in Biomedical Publication was held in Chicago, Illinois, 9-11 September 1993; and the Third International Congress on Peer Review and Global Communications was held in Prague, 17-20 September 1997).

Many of the commentaries, editorials, and articles on peer review have been written by editors and referees themselves, and the debate about the effectiveness of refereeing can be found on many editorial pages and letters sections of journals (Garfield 1986a). A significant part of the research on peer review has examined differences in journal practices (Scanlan 1991), but editors and journals have also conducted research on peer review as part of journal management (Gidez 1991). Another popular method is to survey editorial boards or journal referees. Some journal editors have opened their files for research on peer review procedures, although this practice is controversial because of the risk of breaching confidentiality of the review system. New methodologies are being developed to study editorial processes. Citation analysis is especially promising as a method for analyzing bias and particularistic criteria in peer review.

According to Peters and Ceci (1982a), research on the peer review system requires considerable time, persistence, and a tolerance for lack of cooperation from journal editors and referees, but some observers have found the methods themselves questionable. The Peters and Ceci study was criticized because the authors used deception (Fleiss 1982; Mindick 1982). Similar criticism was made when Epstein (1990) submitted a fictitious manuscript to 140 social work journals in an attempt to demonstrate their bias toward positive results; Epstein's manuscript about this research was rejected by the leading journals in social work, and the National Association of Social Workers Disciplinary Board opened an inquiry. After Mahoney carried out his study of peer review, charges of ethical misconduct were filed, and there were attempts to have him fired (Mahoney 1990). Some authors (Feinstein 1991; Kemper 1991) have criticized such studies, arguing that bypassing informed consent and standard ethical guidelines for research could have important negative consequences for the subjects, the researcher, and the scientific community.

This article is Part 1 of a two-part summary of the growing body of research on peer review and referees' assessments and, as such, attempts to identify those areas in need of more work. The review is intended for a broad audience, including those authors, referees, and editors who would like to know more about peer review as it stands today. My approach is to bring together the results from a disparate body of research and to organize some of what we know about peer review for journals into a coherent discussion.

I have located the studies using such databases as Science Citation Index (SCI) and Social Sciences Citation Index (SSCI) and through references within papers located in such databases. Because empirical studies on peer review have been published in such divergent fields, the terminology often varies indiscriminately. Armstrong (1997) observed that the words "peer review," in fact, seldom appear in the titles of studies on that topic, and consequently some may have been inadvertently omitted. This article focuses on journal peer review, not on the review process for books or proposals. It also does not address electronic publishing, a new trend in scientific communication.

In Part 1, I describe current research about the various participants in the peer review system, focusing on the appointment mechanisms of editors and referees, and reviewer tasks and qualifications, and about some of the systemic problems of reliability, accuracy, and bias. In Part 2, I describe current research on how fraud, favoritism, and self-interest may affect the review system and the research relevant to such policy issues as interference of particularistic criteria; connections among editors, authors, and referees; and double-blind reviewing.

Participants in the System

Credentials of Referees, Editorial Board Members, and Editors

Although most journals send papers to two or more referees, other journals resort to the single initial referee system. In this system, manuscripts submitted for publication are sent to a first reviewer, and if he or she pronounces a positive verdict, the journal's editors almost always grant the green light for printing the article. Should the referee recommend that the article is not worth the trouble, the editor is compelled by virtue of professionalism to request a second opinion from an independent reviewer. This last system is used more in the natural sciences.

According to Michels (1995, 218), the first function of editorial peer review is to ensure that decisions regarding publication are made by those

most qualified to assess the standard and appropriateness of a submitted manuscript. Are the reviewers or editorial board members the most qualified to evaluate manuscripts? In other words, are they experts in the field or fields represented by the journal? Various studies shed light on this question.

Schulman, Sulmasy, and Roney (1994) surveyed the editors-in-chief at fifteen major medical journals about their refereeing procedures, but the journal editors responded in a manner that brought forth little understanding of their reviewers' formal training in the fields in which they were supposed to be experts. Referees of *Physiologia Plantarum*, surveyed by Murphy and Utts (1994), seemed very qualified; 83 percent of those answering a survey were in senior university positions, 57 percent had received their Ph.D. degrees at least fifteen years prior to the date of the survey, 60 percent had published six or more papers in the last three years, and 75 percent had been main authors of at least one paper. Yankauer (1991) studied reviewers for the *American Journal of Public Health* and found that 31 percent of them had not been listed as authors of a source publication in the 1987 Science Citation Index and 15 percent had not been cited at all. Lindsey (1978) studied the publication records of editorial board members for some psychology, social work, and sociology journals and discovered that the majority of the members of editorial boards in social work had never published an article abstracted by the major abstracting services, but members of sociology and psychology journal boards had published a number of articles. In addition, high citation counts for the published work of editorial board members in psychology and sociology suggested that their contributions have had a substantial impact on their fields. Similarly, Zsindely, Schubert, and Braun (1982) found a significant correlation between the mean number of citations to international chemistry journals editors and the impact factor of the journals in question. Parallel conclusions were drawn in the field of analytical chemistry (Braun and Bujdoso 1983). Hamermesh (1994) studied refereeing processes for seven economics journals and found that the number of years after Ph.D. averaged sixteen. Over half of the referees were cited at least ten times per year, and 25 percent were cited at least fifty times, but this result should be compared to data showing that only eighty-five American economists were cited at least fifty times per year during the period 1971-1985 (Medoff 1989). Zsindely and Schubert (1989) studied the citations received by 769 medical editors in twenty-eight subfields and discovered that in all but three of the subfields, the editors-in-chief were, on average, less cited than their own journals. They concluded that editing a scientific journal requires different qualities than those needed to be a prolific and highly cited author. Bakker and Ritger (1985) analyzed the representation of scientists from different countries in the editorial boards of the most influential and prestigious

journals within forty-eight fields of biomedical and clinical research and found that scientific performance and eminence may be important reasons for appointment as editor, but nationality biases played a significant role in the composition of editorial boards.

How Editors and Editorial Board Members Are Appointed

The appointment criteria and the mechanisms used to appoint editors and referees is another neglected research area. Mystery appears to shroud the process of selecting and assigning referees (Hamermesh 1994, 153). For instance, 50 percent of the journals listed in *Index Medicus* do not publish clear information concerning their peer review practices and selection of referees (Colaianne 1994). Sometimes, editors use sophisticated computer databases to select referees, but the criteria by which referees are included in these databases are far from public. Sometimes authors are allowed (or asked) to give names of potential referees for the papers they submit, and often journals give authors the option to suggest referees who should be excluded. Appointment mechanisms should thus be a key research topic in the sociology of science but, regrettably, few studies have addressed this topic.

Journal editors of the American Psychological Association (APA) are nominated by search committees that dedicate considerable effort to secure names of well-qualified individuals who match the prerequisites (Eichorn and Van den Boss 1985). Openings are at times advertised in APA journals because an editor's mandate confers not only power but also a heavy burden of work; it is not unusual for between 50 percent and 80 percent of nominees to decline appointment as editors. APA search committees review résumés, letters of recommendation, and records of previous publications in order to manufacture a rank-ordered list of candidates. The Publication and Communication Board discusses in executive session the capabilities of candidates, and the head of the Publication and Communication Board invites the highest ranked candidate to accept the position. Should the candidate decline, the board proceeds down the list. Editorial boards are usually appointed by the editor. Their composition depends on a number of factors, but can range from twelve to fifty persons.

In a study on the appointment of editors and referees, Lindsey (1977) used path analysis to study the influence of editorial board members in the screening and publication process in psychology. His findings suggest the existence of two kinds of editors: the high producers and the eminent scholars. The first type reviews many manuscripts and substantially influences the selection process. As Lindsey (1977, 582) noted, "They are the workhorses of the journal operation enterprise." In contrast, the eminent scholar is a

distinguished person who is characterized by a high citation number, independent of volume or productivity, and who occupies a seat on the editorial board even though he participates minimally in the reviewing process. Lindsey concluded that psychology journals purposely appoint members who serve as figureheads; the data suggest that in psychology, the most eminent authors were also not the true gatekeepers.

There is further evidence that appointment to an editorial board may be based on partiality and on particularistic or ascriptive criteria. Yoels's (1971) study of the doctoral origins and appointment patterns of editors at the *American Sociological Review* from 1948 to 1968 demonstrates that graduates of Chicago, Columbia, and Harvard tend to favor colleagues from their own alma maters. In a later study, Yoels (1974) found that the top five U.S. universities account for over 70 percent of the total board members in the leading sociology journals. Beyer (1978), on the other hand, observed that sociologists rank institutional affiliation as the most important factor in deciding editorial board appointments. Scientists from other disciplines placed less emphasis on institutional affiliation (Willis and McNamee 1990).

How Referees Are Chosen

If the above is true, that is to say, if particularistic criteria influence appointment of editorial board members, one might assume that the same occurs in referee selection. Hamermesh (1994) studied the appointment of referees for seven economics journals and observed that, in the whole sample, almost 12 percent of referees were from the same university department as their colleagues, the editors (in one journal, the fraction surpassed 30 percent). Hamermesh (1994, 156) concluded that the extent to which editors rely on colleagues from their own departments attests to the role of propinquity in the choice of referees.

There may be other circumstances that influence the choice of referees, and sometimes the most eminent individuals in a determined field are not selected as the gatekeepers. Busy schedules and the deadlines of the reviewing processes may make a referee refuse to review a paper. Study of refereeing practices in the *Journal of Clinical Investigation* found that a disproportionate number of refusals to review a paper originated from the high-status group of reviewers (Stossel 1985). Consequently, editors are obliged to search for less eminent, younger, and more inexperienced colleagues, who may agree to referee sometimes just to promote their own careers.

A referee in dire need might also turn to a colleague for help. In a study of the refereeing practices of journals in forty-three disciplines, Juhasz et al. (1975) discovered that referees occasionally subcontract junior associates to

do the “dirty work” (such as rederiving equations). In studying refereeing practices in library journals, Glogoff (1988) found that referees admitted that some of the manuscripts they were supposed to review were passed on to subreferees to evaluate the article in part or entirely. Other research shows that a small fraction of medical referees pass their manuscripts on to a colleague for partial or total review (Lock and Smith 1991).

To make matters more confusing, some editors deliberately choose young scholars as referees (Fyfe 1994; Guest 1994). Doctoral students with exceptional writing and communicative skills may replace older colleagues burdened with deadline schedules (Glenn 1976; Romanelli 1995). This practice is not necessarily harmful; some editors report that the quality of reviews varies inversely with the referee seniority and status (Finke 1990; Judson 1994). Sometimes the qualified reviewers are overburdened and have no time for reviews. A study of 226 reviews of 131 consecutively submitted manuscripts to the *Journal of General Internal Medicine* showed that when a reviewer was around forty years old or younger, from a top academic institution, and well-known to the editor, the probability that he or she would produce a fairly elaborate review reached 87 percent. Similarly, in a study of 1,600 reviews for the *Journal of Clinical Investigation*, it was found that the fraction of poorly graded reviews was the largest in the high-status reviewer group, whereas the low-status reviewers submitted the highest proportion of well-documented reviews (Stossel 1985).

Reviewer Incentives and Tasks

The first thing that attracts our attention when investigating the role of editors and referees in the process of scientific communication is that, in times of rampant commercialism, this important task is carried out almost entirely by volunteers. In most cases, these highly qualified experts are not compensated in any way for their work. Some years ago, the *New England Journal of Medicine* used to pay \$5 per review, and some reviewers rightfully complained that, if that was the price that this eminent journal placed on their opinion, the *New England Journal of Medicine* should seek referees elsewhere (Ingelfinger 1974, 690). Editors for whom reviewing is a part-time activity within an active academic career are generally paid only a token honorarium. APA editors are paid modest salaries (Eichorn and Van den Boss 1985).

This corps of editors and referees invests many hours. In a survey of the *American Journal of Public Health* reviewers, Yankauer (1991) found that participants invested more than 3,300 hours of uncompensated labor to the journal. More than 3,000 individuals reviewed manuscripts for *The Journal*

of the American Medical Association during 1983 (cited by LaFollette 1992, 115). Referees for *Academy of Management Review* and *Academy of Management Journal* declared that they spend a mean of 5.4 hours reviewing a paper (Jauch and Wall 1989). Biology referees surveyed by *The FASEB Journal*, the official publication of the Federation of American Societies for Experimental Biology, certified that they spend a mean of 3.5 hours per manuscript reviewed. A survey by the *Journal of General Internal Medicine* found that the mean review time is three hours, with 20 percent of reviewed material requiring six or more hours (McNutt et al. 1990). In the 1970s, Relman (1978) estimated that the review process of the *New England Journal of Medicine* required six to seven person-years of peer review, two person-years of in-house work, and \$100,000 of office expenses. In 1985, *The Journal of the American Medical Association* received 3,446 manuscripts, all of which were read by in-house editors and 1,413 of which were sent to outside referees (Biggs 1990). Costs of the peer review system have remained quite high. Biomedical journals spend an average of \$75-\$150 per manuscript (Scanlan 1991).

Some commentators suggest that referees only carry out this work to gain privileged access to the most recent advances in the discipline (Glogoff 1988; Laband 1990), or for the satisfaction of promoting an important intellectual product or out of an altruistic desire that the author's idea will be given due credit and deserved attention (Simon, Bakanic, and McPhail 1986). The prestige and recognition that one earns within one's field is also an important factor (Endres and Wearden 1990; Yankauer 1991). Most journals acknowledge publicly the effort of their external referees, and all the journals publish a list of their editorial board members in each printed issue. However, in a survey of referees for journalism and mass communication journals, 52 percent of the participants reported that refereeing for journals did not count toward promotion or secured tenure, and as a result the authors concluded that "refereeing is an act of love" (Endres and Wearden 1990, 49). This key role playing can be considered a service to the community as well (Guest 1994, 98) or among the most important professional services that gifted scientists provide (Zentall 1991). According to Glenn (1976, 180), article refereeing is a "charity work" in the sense that there are only minimal formal recognition and rewards for performing it, and no penalties for refusing it. One reviewer has complained that refereeing work is "a demanding, difficult, time-consuming and very ungrateful task" (Jauch and Wall 1989, 171). Hence, the journal peer review system depends and relies on the referees' sense of professional obligation, generosity, understanding, and good humor (Biggs 1990). But there is also a strong emotional component involved in the refereeing task. This aspect was made clear by Siegelman (1991), who

observed that, on occasion, reviewers protested the final decision concerning acceptance or rejection of a paper if made against their advice.

The symbolic rewards are probably a real incentive. David L. Ransel, editor of *American Historical Review*, once observed that he received several letters a week from historians offering their services as reviewers for the journal (Coughlin 1989). Others have also spoken about the scholars' willingness to participate in the reviewing tasks (Zuckerman and Merton 1971). Yet, in other areas, it is not so easy to recruit external referees. Psychology editor Ronald A. Finke complained that the job of finding people willing to review submitted manuscripts is very often difficult, if not impossible (Finke 1990).

Editors and reviewers often serve on several journals simultaneously. Twelve percent of editorial board members of three school psychology journals, for example, served on two or three journals (Kawano et al. 1993, 413). Fifty-four of the 575 editorial advisers to the top twenty-five economics journals served on the editorial board for two journals, and 8 of these people served on three different journals (Gibbons and Fish 1991). Referees for the *American Journal of Public Health* surveyed by Yankauer (1991) reviewed manuscripts for a mean of 3.6 journals; *British Medical Journal* referees reviewed manuscripts for a mean of 5.0 journals (Lock and Smith 1991); *Academy of Management Journal* and *Academy of Management Review* referees reviewed manuscripts for an average of 3.1 theory journals and 2.6 empirical journals (Jauch and Wall 1989); and referees from different biology journals surveyed by *The FASEB Journal* reviewed manuscripts for a mean of 3.4 journals, with 5 percent of them reviewing between 10 and 14 journals. Almost one third of surveyed reviewers stated that they had reviewed six to ten papers during the last year (Gidez 1991). Medical referees were found to review manuscripts for an average of 4.3 journals, and the mean number of reviews during the last twelve months was 12.9 for men and 6.8 for women (Nylenna, Riss, and Karlsson 1994). Referees may be evaluating as many as five or six manuscripts from different journals at the same time (Garfunkel et al. 1990); one sociology referee stated that he reviewed between twelve and eighteen papers per year (Pressman 1994), and a prestigious psychologist declared that he reviewed about fifteen manuscripts per year (Hartley 1993). In a survey of referees for thirty-one library journals, Glogoff (1988) found that 40 percent of the referees had reviewed one to three manuscripts in the previous twelve months, and 8 percent stated that they had reviewed more than fifteen manuscripts in the same twelve-month period. As LaFollette (1992, 122) noted, exceptional reviewing performance most likely attracts more requests for a service of a given reviewer.

Systemic Problems of Reliability, Accuracy, and Bias

Reliability of Review

Reliability is usually judged by the consistency of scores granted by a given judge concerning the same item in successive trials, or it may also be the consistency of scores given by different judges. The classical interpretation underlines the fact that without reliability, there cannot be validity, and that even where reliability is found, there could be low validity (Gilmore 1991, 148).

To date, the most complete and critical analysis of the reliability of the peer review system has been published by Cicchetti (1991), and it originated a heated debate in *Behavioral and Brain Sciences*. Cicchetti surveyed a wide spectrum of research designs in many areas from physics to behavioral sciences and, as a first conclusion, identified some errors in the handling of statistics from previous studies on referees' reliability. The main, but not surprising, result of his inquiry was a very low reliability level detected in most of the studies. Many studies reviewed by Cicchetti demonstrated an agreement among referees even less than would have occurred by chance. In the behavioral sciences, for example, the reliability for nearly 3,000 papers averaged a disappointing 0.21. Referees can agree on acceptance, resubmission, or rejection, but for different and sometimes quite conflicting reasons. Cicchetti (1991) also found that the converse was true; that is, reviewers often agreed on the evaluation of a given manuscript but drew different conclusions about its publishability. Fiske and Fogg (1990) found similar evidence of this phenomenon in referee reports from nine APA journals. They compared pairs of referees who agreed on the same point to pairs who did not agree at any common point, and concluded that the overall recommendations contained similar discrepancies. These results seemingly contradict Polanyi's suggestion that "two scientists strange to each other, but acting as referees in the publication of the same paper, usually reach the consensus concerning its approximate value" (quoted in Simon, Bakanic, and McPhail 1986, 261). D. B. Ranking, editor of the *Journal of the American Statistical Association*, asserted that "all who routinely submit articles for publication realize the Monte Carlo nature of review process" (quoted in Eysenck and Eysenck 1992, 393).

The low reliability of the peer review system is well documented; however, there exists disagreement over how this fact should be interpreted. In the debate that followed the Cicchetti (1991) study, some authors expressed concern. Bornstein (1991b, 139), for example, complained that "if one

attempted to publish research involving an assessment tool whose reliability and validity data were as weak as that of the peer review process, there is no question that studies involving this psychometrically flawed instrument would be deemed unacceptable for publication.” Others argue that the goal of peer review is not to obtain high reliability, but to help editors in their decision making (Bailar 1991; Stricker 1991). For some authors, the lack of reliability even indicates solid science (Eckberg 1991), and some authors argue that high reliability is associated with low inference (McAleese 1993). Schönemann (1991) raised doubts about the benefit of incremental improvement in reliability; he estimates that the validity of peer review is negative overall. Because the existing peer review system does not really select the best papers for publication, any improvement in reliability would lead to a disappearance of error variance, and this would obviously make things worse.

The origin of the low interreferee consent is complex. According to Kiesler (1991), editors do not select referees at random; consequently, a high interreferee accord is less than probable. The choice is made in view of conflicting interests or simply because they match the areas with which they are best acquainted. Editors wish to know what weaknesses each reviewer saw (Fiske and Fogg 1990); a manuscript in medicine can be reviewed, for example, by a statistician, a clinician, and a biochemist, each examining it from his or her particular viewpoint. The study by Fiske and Fogg (1990) confirms the diversity and uniqueness of reviewers’ comments—referees often did not overtly disagree on particular points but instead focused on different topics existing in the papers, and, based on their own conclusions, made different recommendations to the editors. Similarly, a study of referees’ comments on manuscripts submitted to the *American Sociological Review* from 1977 to 1981 revealed that the comments more often complemented than contradicted one another (Bakanic, McPhail, and Simon 1989). About 11 percent of the manuscripts sampled received contradictory comments from the referees, but generally on theoretic and stylistic issues. However, most of the comments were neither contradictory nor similar. Instead, referees simply made observations and commented on different aspects of the presented manuscript. Authors who submit papers to professional journals seem to be aware of these phenomena; a survey of authors of accepted and rejected papers prompted a number of complaints about manuscript review patterns practiced by the APA (Cofer 1985). Only 33 percent of participants stated that referees had focused criticism on the same aspects of the papers.

Another factor that contributes to low reliability is the lack of objective criteria that define good scientific work. No uniform standard exists for evaluating articles, nor are there any universal guiding procedures for reviewers (Sternberg 1985). Hence, a flaw in experimental design may render a

paper worthless in the eyes of one reviewer, whereas this error may be of no significance to another. A survey by Kerr, Tolliver, and Petree (1977) of 429 reviewers from nineteen leading management and social science journals revealed considerable dissimilarities in response to the same items, as stated by different reviewers for the same journal. The authors concluded that journal norms and guidelines are not as relevant to acceptance as are personal factors.

Siegelman (1991) went one step further and tried to identify refereeing patterns that might explain the lack of comprehensive interreferee agreement. He analyzed the reports of 660 referees from *Radiology*, which requires its referees to grade manuscripts on a scale from one to nine. Siegelman was thus able to identify categories of reviewers who consistently scored with great deviation from the mean. Interestingly enough, referees in various categories differed less in the quality of the reviews, and more in their perceptions of what a good manuscript is. Referees in the most critical and in the least critical categories were among the highest ranking with respect to their prestige and experience with the journals. Both the most critical and the least critical reviewers provided detailed reports containing valuable commentaries on the originality, consistency, and validity of the examined material. Similar results were found in studies by Saksena et al. (1995) and Cummings, Frost, and Vakil (1985). In the latter study, researchers divided *Academy of Management Journal* referees into two categories according to their comments, labeling them either "coaches" or "critics." Coaches were referees who "reinforced well-done papers"; they were "helpful and suggestive . . . promoting improvements . . . and providing reasons for recommendations" (Cummings, Frost, and Vakil 1985, 479). Critics were those referees who were "almost exclusively critical . . . skeptical and relatively insensitive to the form and tone with which their criticism was communicated" (p. 479). These results indicate that, in addition to intrinsic scientific merit, the future of a submitted manuscript depends on the kind of referees assigned for review.

Accuracy of Review

Perceptions of the goals and merit of peer review vary widely. Garfield (1986b, 9) maintained that "refereeing is probably the most efficient and effective method for distinguishing the promising from the meritorious, at least to an extent when it is proven otherwise." Lederberg (1978) also spoke positively about the merit of peer review, calling it the "glue that holds the scientific establishment together" (p. 1315). In their study on refereeing practices, Jauch and Wall (1989) concluded that referees were professionals

who allocate substantial time and effort to their discipline and a lot of attention to their colleagues, making sincere attempts to improve research and serve as writing guides. Ingelfinger (1974, 692) asserted that peer review brings "law and order" to biomedical investigation, analysis, and communication, whereas Kassirer, one of his successors as editor of *The New England Journal of Medicine*, called it "indispensable" (Kassirer and Campion 1994). Bailar and Patterson (1985, 654) agreed that "although individual decisions based on peer review may be criticized with vigor, the system as a whole seems to be accepted as working reasonably well," and Legendre (1995, 36) pointed out that "even though the peer review process is far from perfect, it is the best system to ensure quality and reliability." However, even editors of scientific journals complain about the system (e.g., see the experts who expressed negative opinions about peer review in Eysenck and Eysenck 1992, 393).

An increasing body of research demonstrates that referees often commit mistakes while evaluating manuscripts. These errors can be classified as type I errors (recommending to publish papers of low quality) and type II errors (recommending not to publish papers that should have been published) (Laband and Piette 1994).

There exists an assumption that review by referees and journal editors ensures some degree of reliability. Once a paper has cleared peer review, we all want to believe (and take it for granted) that the ideas expressed in the manuscript are valid, and that the technical aspects of the methodology are satisfactory. Readers and users of journals seek some form of virtue and sincerity without the need to check up on their colleagues to assure that work is authentic (LaFollette 1992, 121). It is now common practice for peer-reviewed articles to be more readily accepted by courts and legislative committees (Bailar and Patterson 1985; Ingelfinger 1974; Legendre 1995). Well-performed reviewing saves scientists time and effort and allows them to build functional models based on reliable results.

Ingelfinger once observed that the editors of *The New England Journal of Medicine* tend to accept about 10 percent of manuscripts they should have rejected, and rejected about 10 percent of manuscripts that should have been accepted (cited in LaFollette 1992, 114). Complaints made by readers and editors who are dissatisfied with journal quality, sterile journal style, and irrelevant articles can be found in the pertinent literature (Frost and Taylor 1995, 14-15; Nelson 1982, 229). Citation data show that most papers are rarely cited. Although a typical paper is cited an average of fifteen times (anonymous foreword to *Current Contents*), the distribution of citations is quite skewed; a few papers are highly cited, and many articles are cited only a few times or not cited at all (Seglen 1992). According to Garfield (1989,

7), more than 56 percent of articles indexed by the Institute for Scientific Information from 1955 to 1987 were not cited or even self-cited. Only 10 percent of articles cited between 1961 and 1982 received ten or more citations, whereas only 3.4 percent of such articles were cited at least twenty-five times (Garfield 1985, Table 2). Only 2 percent of over 32 million papers that were cited at least once between 1945 and 1988 were cited more than fifty times (Garfield 1990). Confronted with data on the low citing record of so many papers, some authors complain that these papers should have never been published in the first place (e.g., see opinions collected by Hamilton 1990). Actually, it seems that many journals serve authors, not readers. As medical editor Allen Bard complained, "In many ways, publication no longer represents a way of communicating with your scientific peers, but a way to enhance your status and accumulate points for promotion and grants" (quoted in Hamilton 1990, 1331).

The editorial process is supposed to improve the quality of submitted manuscripts (Sternberg 1985). Some authors of highly cited papers have said that referees' suggestions were enough to improve the quality of their papers (Campanario 1993a, 1995a; Gans and Shepherd 1994). Readability of articles published in *Annals of Internal Medicine*, as measured by Gunning indexes, gained slightly through peer review and editorial processes (Roberts, Fletcher, and Fletcher 1994). Metoyer-Duran (1993) found that the writing style of rejected papers from *College and Research Libraries*, however, was superior to that seen in accepted papers, even when the editorial staff of that journal copyedited all accepted manuscripts. Using multiple regression, Laband (1990) showed that the length of referees' comments had a positive impact on subsequent citation of a sample of ninety-eight papers published in seven major journals; however, the representativeness of these results should be viewed with caution, because Laband obtained responses concerning only ninety-eight papers after a request to 731 authors.

Such data reveal one side of the coin, for there exists some evidence that papers reviewed by editors and referees are not error free, and that acceptance may well be influenced by other factors. Koren (1986) studied the acceptance rate of abstracts submitted to the 1986 annual meeting of the American Pediatric Society and found that the acceptance rate of abstracts written with modern text processors was higher than that for typed papers. This was especially true in the case of authors who submitted papers in both formats. Armstrong (1980) found a correlation coefficient of 0.7 between the prestige of ten management journals and their reading difficulty index, as measured by the fog indexes. In addition, he surveyed thirty-two faculty members and asked them to rate the competence of the research reported in manipulated texts from management journals. The content of passages was held constant

while readability was varied (e.g., eliminating unnecessary words, substituting easy for difficult words, breaking long sentences into two sentences, etc.). Passages that were more difficult to read were rated higher in research competence (Armstrong 1980). When physicist Alan Sokal (1996) submitted a parody manuscript to *Social Text*, he carefully included citations to reputable physics books and journal articles alongside appropriate jargon in postmodernist literary and political theory for the benefit of the journal's reviewers. Referees and editors seem to prefer sophistication over simplicity, as was demonstrated by a survey of 265 editorial board members organized by Lindsey and Lindsey (1978). These authors found that over 65 percent of the board members rated the sophistication of the statistical methods used to be very important or of the highest importance. In a survey of 723 university professors, Bradley (1981) found that 76 percent complained about the pressure that strictly subjective reviewers' preferences imposes on them; 73 percent of those surveyed suspected false criticism, and 8 percent executed changes in the articles following reviewers' suggestions, even though they felt the changes were inappropriate.

More serious research has been done on the quality of peer review as a filter. Garfunkel et al. (1990) sent twenty-five manuscripts that had been revised and accepted for publication in the *Journal of Pediatrics* to a team of two referees for new review. The outcome of the new revision was striking. Most manuscripts were thought to have sufficient drawbacks to warrant their further revision. Some current research results raise doubts on the predictive validity of referees' decisions. Gottfredson (1978) found only low to moderate correlations ($R = .28$) among reviewers' rating of psychological research papers and the number of citations received by those papers.

Some indirect hints about accuracy come from research on previously rejected manuscripts published elsewhere and their impact. Medical editor Jean Wilson (1978) analyzed the citations received by papers published by the *Journal of Clinical Investigation* and those rejected by that journal and published by other journals. She found that papers published in the *Journal of Clinical Investigation* were cited twice as often during the first four years after publication as those that were rejected and published elsewhere. Daniel (1993) studied the peer review process in *Angewandte Chemie*, tracing the fate of rejected manuscripts. He divided manuscripts into two categories: those that were posteriorly published in another journal and those whose fate was unknown. Of the 115 manuscripts previously rejected by *Angewandte Chemie*, or withdrawn voluntarily by their authors, a total of 88 (71 percent) were later published in other journals. Daniel also discovered that 83 percent of the communications that were rejected but subsequently published else-

where appeared in their original form or in a shorter version; that is, communications, letters, notes, and so on. These “rejects” were published in a total of thirty-nine journals; however, none of the rejected manuscripts appeared in a journal whose impact factor was greater than *Angewandte Chemie*. In addition, the frequency of citations of papers accepted by *Angewandte Chemie* were higher than those rejected by the journal yet published in another journal of lesser prestige. Nevertheless, Daniel does not know whether this might be simply due to the fact that papers in *Angewandte Chemie* are cited more often. The data on impact factors favor this hypothesis. Abby et al. (1994) searched for publications that were identical or similar to those rejected by the *American Journal of Surgery* during 1989 and found that 62 percent of those were not subsequently published in any other journal indexed by MEDLINE over the three-year period following rejection. Using the MEDLINE database, Chew (1991) located 245 manuscripts previously rejected by the *American Journal of Roentgenology* and discovered that 69 percent of major papers, 62 percent of case reports, and 33 percent of technical notes were published elsewhere, but in journals with lower impact factors than the *American Journal of Roentgenology*.

The results of some research raise doubts about the accuracy of reviewing. Murray (1988) evaluated statistical procedures in twenty-eight papers that had been published in the *British Journal of Surgery* and concluded that four of these should have been rejected, seven needed major revisions, and eleven required minor changes. Gardner and Bond (1990) analyzed forty-five articles published in the *British Medical Journal* and found that seven of the articles “were considered not to be of an acceptable statistical standard” (p. 1356). Brown and Baca (1986) found statistical irregularities in 54 percent and 50 percent of articles published in 1983 and 1984, respectively, in the *American Journal of Children Diseases*. Many articles on scientometrics report the use of statistical methods that require a symmetric cross-citation matrix, even though it is a well-known fact that these matrices are not symmetric (Campanario 1995b).

The use of statistical significance testing and its worldwide acceptance as the most common and proper statistical procedure have been under attack for the past fifty years (Bakan 1966; Carver 1978; Mehel 1967; Shaver 1993; Skinner 1956; Thompson 1996). Some authors claim that use of the statistical significance test is responsible for the lack of productivity in educational and psychological research (Shaver 1993). Thompson (1996) reviewed the etiology of the propensity to conduct statistical significance tests and identified many misconceptions on this topic. For example, many researchers equate an unlikely result with an inherently interesting result. Some forty-five years

ago, Yates (1951) warned that many scientists have often regarded the execution of a test of statistical significance on an experiment as the ultimate objective.

A naive acceptance of the normality of samples in statistical tests persists. Micceri (1989) investigated the distributional characteristics of 440 large-sample readings and psychometric measurements, 265 of which came from articles published in journals and the others from national or state tests administered in the United States. Approximately 90 percent of the distributions included 460 or more cases, and almost 70 percent included 1,000 or more. All these samples were found to be significantly nonnormal at the .01 alpha level. Of course, statisticians have developed methods to check the robustness of statistical tests under nonnormal distributions. Typically, these methods are based on the study of usual statistical test reliability under some particularly nonnormal distributions. However, the results obtained by Micceri (1989) also failed to support the types of distributions calculated in most of the previously researched reliability schemes. Consequently, many research papers on educational psychology with cardinal flaws are being accepted for publication by editors and referees; furthermore, these somehow were published in reputed journals, and were used in decision or policy making, and might serve as the starting point for subsequent research.

On the other hand, there are many well-documented instances of type II errors that are much more damaging and significant than type I errors. Horrobin (1990) furnished striking examples of important scientific discoveries or innovative papers that were rejected outright by the peer review system. Some instances of mistaken judgment and rejections found final rest in the journal *Medical Hypotheses*, and proved to be extremely influential. Descriptions of similar instances of wrong rejections can be found in Aronson (1986), Astin (1991), Barber (1961), and McCutchen (1991).

Accounts concerning peer resistance to scientific discovery are scattered throughout the literature, making a systematic evaluation difficult. A study of retrospective commentaries by authors of highly cited papers showed that some of the authors of the commentaries had encountered enormous difficulties in publishing their original papers; sixty-four of those highly cited and prized papers were initially rejected, seven would eventually be among the most cited papers in the journals that eventually published them, and the Institute for Scientific Information would identify four of them as the core documents of a research front (Campanario 1993a, 1995a, 1996). The worst type of "referee's blunder" had also taken place; some papers that later secured the Nobel Prize for their authors had been rejected by the first journal to which they were submitted (Campanario 1993b, 1995a, 1997). In addition, eighteen of the most cited articles of all time encountered some difficulty or

resistance from referees. Three of these papers with initial publication problems turned out to be the most frequently cited by their respective journals (Campanario 1996).

Other important discoveries have been resisted by referees. The first law of thermodynamics was rejected by the prestigious German journal *Annalen der Physik*, and J. R. Mayer eventually published this law in an obscure journal (Colman 1982); Berridge's (1988) account of the role of inositol triphosphate and diacylglycerol as second messengers was initially rejected by *Nature*; Maiman's report on the first operation of a laser was rejected by *Physical Review* (Bloembergen 1993); Eyring's classic 1935 paper on the theory of the activated complex in chemical reactions was rejected by *Journal of Chemical Physics* (Laidler and King 1983); Belusov's account of oscillating chemical reactions was rejected by two journals (Vidal 1990); and early papers on chaos theory by Mitchell Feigenbaum were returned unpublished (Redner 1987).

Similar examples have occurred in the field of economics. Gans and Shepherd (1994) asked 140 leading economists, including all living winners of the Nobel Prize and the John Bates Clark Medal, to describe instances in which journals had rejected their papers. Some indicated that no journal had ever rejected their work, but other well-known economists had had papers rejected on quite a few occasions. Only three of the twenty prize-winning economists said that they had never had a paper rejected. Some of rejected papers were considered by their authors as among the best ones, considerably influential and among their most extensively cited.

In view of the above examples, there is more than eloquence alone in former *British Medical Journal* editor Stephen Lock's statement that "peer review favors unadventurous nibbling at the margin of truth rather than quantum leaps" (Lock 1985, 1560). Albert Einstein was so outraged by the refereeing in *Physical Review* that he ceased to publish his papers in that journal as a sign of a protest (Azbel 1993). The number of important, innovative, Nobel category papers that have been initially rejected is so high that I have proposed elsewhere the establishment of a hypothetical publication, the *Journal of Previously Rejected Important Papers* (Campanario 1995a).

To some extent, these types of data confirm what we already know from the history and philosophy of science about the reception of innovation. It is well-known that the conflict between new ideas and scientific (or religious) orthodoxy invokes a certain kind of "*odium professionale*" (Toulmin 1972). It is perhaps understandable that creative and imaginative scientists have been ignored because of their rebellious stands or unorthodox theories; Polanyi (1958, 149) noted, "The history of science records only happy endings."

According to Sol Tax, founder of *Current Anthropology*, "A bias that prevents competent work from entering the arena of community scrutiny is much more damaging than a bias that lets mediocrity work its way through to a podium of academic achievements" (quoted in Eysenck and Eysenck 1992, 394). What, then, is the relative frequency of such occurrences? A plausible way to answer this question would be to review journal manuscript files and trace the fate of individual papers (Campanario 1995a). This research program undoubtedly would yield valuable data on the validity of peer review system.

Is the System Biased Toward Positive Results?

There is strong evidence that, especially in the biomedical sciences, journals tend to publish only papers in which statistically significant results are reported (Beyer, Chanove, and Fox 1995; Hubbard and Armstrong 1992; Newcombe 1987; Salsburg 1985). Attention to this bias was drawn as early as 1962 (Melton 1962). Instances of papers that report nonsignificant results appear to be minimal in scientific and academic journals. Thus, for example, Sterling (1970) found that only 3 percent of the articles published in four psychology journals during a one-year period and 5 percent of a random sample of *Psychological Abstracts* citations printed studies that reported nonsignificant results. Only 6 percent of a sample of 1,334 papers published in three psychology journals reported the same type of results (Bozarth and Roberts 1972). In the biomedical sciences, there is a similar tendency. Weisse (1986) surveyed all 408 original articles published in 1984 in *The New England Journal of Medicine*, *Annals of Internal Medicine*, and *Annals of Surgery*. For *The New England Journal of Medicine*, 10 percent of the results were negative, 10 percent were neutral, and 80 percent were positive; for *Annals of Internal Medicine*, 1 percent of the results were negative, 10 percent were neutral, and 89 percent were positive; and for *Annals of Surgery*, 3 percent of the results were negative, 6 percent were neutral, and 91 percent were positive. Additional evidence demonstrating the bias against nonsignificant results has been found by Atkinson, Furlong, and Wampold (1982), Cohen (1979), and Van Heerden and Hoogstraten (1978).

Some surveys and experimental studies show editors' and referees' preferences for manuscripts reporting positive results. For example, Kerr, Toliver, and Petree (1977) surveyed 429 reviewers associated with nineteen leading management and social science journals, asking about their rationales for manuscript acceptance and rejection. Only 1 percent of reviewers stated that they would react favorably to a paper that reported an author's new theory yet yielded data that were not statistically significant. In this hypothetical situation, 44 percent of the referees surveyed stated that they might probably

or almost surely reject such papers on that basis alone. Interestingly, extremely negative reactions were elicited from three psychology journals (Kerr, Tolliver, and Petree 1977, 140). Atkinson, Furlong, and Wampold (1982) asked fifty-two consulting editors from two APA journals (*Journal of Counseling Psychology* and *Journal of Consulting and Clinical Psychology*) to rate a manuscript whose results were reported as either statistically significant, marginally significant, or nonsignificant. They observed that ratings on methodological rigor were markedly less positive when a study reported nonsignificant or marginally significant results than when such study reported significant results, even though the methodologies were identical in all three examples. In addition, the referees were over three times more likely to recommend publication when a manuscript had reported statistically significant results than if its results were not statistically significant. These findings should not be surprising because, according to Loftus (1993, 252), "During graduate training, students are taught that accepting the null hypothesis is unacceptable."

In the past, a bias toward significant results was considered beneficial by some authors. Beveridge (1950, 35) wrote that "it is an usually commendable custom not to publish investigations which merely fail to substantiate the hypothesis they were designed to test." Parapsychologists were among the first to become aware and sensitized to the problem and, as early as 1975, the Parapsychological Association Council adopted a policy opposing the selective reporting of positive outcomes (Bem and Honorton 1994). However, as late as 1983, the American Psychological Association negatively considered those studies reporting nonsignificant results unless repeated studies contradicted a strong theoretical or empirical base (American Psychological Association 1983). In the same year, the *British Medical Journal* advised prospective authors "who seek rapid publication of a paper (especially negative results) to submit the manuscript to a paid journal . . . negative results have never made riveting reading" (Minerva 1983).

It is clear from the above that the origin of the bias against nonsignificant results appears to have some of its roots in authors' attitudes and choices. Authors seem to be aware that papers with nonsignificant results are difficult to publish. In a survey of forty-eight authors and forty-seven reviewers of the *Journal of Personality and Social Psychology*, Greenwald (1975) found that authors were eight times more likely to estimate that they would submit a paper reporting significant results for publication than to do so for a paper reporting nonsignificant results. A survey of authors of published and unpublished clinical trials found that nonsignificant results were the main reason given why studies were not submitted for publication (Easterbrook et al. 1991). A study of ophthalmology abstracts reveals that one half of all studies

initially presented in abstract form were subsequently published in their entirety, but full publication was associated with significant results (Scherer, Dickersin, and Langenberg 1994).

More convincing evidence can be derived from the use of meta-analysis. Glass (1982) demonstrated that in ten meta-analyses, the average experimental effect from studies published in journals was greater than the corresponding effect estimated from theses and dissertations. Findings reported in journals were, on average, one-third standard deviations more favorable to the initial hypothesis of the investigation than findings reported in dissertations. Dickersin (1990) showed that the pooled effect of methods or therapies is greater when only the published results are counted; when unpublished but previously registered results are also counted, the pooled effect of such methods or therapies is normally diminished. Hence, some studies that obtain nonsignificant results are never published. This is the well-known "file drawer" problem (Rosenthal 1979). According to Bornstein (1991a, 446), "We actually have no idea how many potentially valuable findings go unpublished or are published in obscure journals where they are never scrutinized." As Mahoney (1982, 221) noted, "The contents of any contemporary psychological journal are likely to be a very selective sample of the evidence and ideas offered for publication."

The immediate consequence of the bias against nonsignificant results is that many researchers waste time, money, and effort in pursuit of investigative blind paths. The situation tends to self-perpetuate because a wrong hypothesis that bore negative results is rarely published. Other consequences stemming from the advancement of some disciplines are likewise dramatic and can even be dangerous. On occasion, the results reported in journals can overtly contradict the results reported in dissertations, as has occurred in the field of sex bias in psychotherapy. In this instance, the direction of the effect reversed when both sources were compared (Smith 1980). Similarly, a meta-analysis in the field of cancer research demonstrated a huge and significant survival advantage thanks to a combination chemotherapy, whose favorable results were possible by ignoring the nonsignificant results of unpublished trials (Easterbrook et al. 1991). In addition, as Dar (1987, 149) suggested, "When passing null hypothesis test becomes the criterion . . . for journal publications, there is no pressure on the researcher to build a solid, accurate theory; all he or she is required to do . . . is produce statistically significant results." Hersen and Miller (1992, 233) suggested that the bias toward positive results is one of the main risk factors leading to fraud or misconduct in science. Sometimes the old aphorism "publish or perish" should be updated to read "publish positive results or perish."

Is the System Biased Against Replication?

We can define replication as “the successive examination and reexamination of (someone’s) own findings and those of other scientists” (Forcese and Richer 1973, 10). As most scientists can attest, direct replications of existing studies are seldom rewarded. Granting agencies are reluctant to provide funds for replication alone, and there is considerable difference in status between being the first to publish a discovery and being the one who reaffirms it. Few scientists would like their names permanently identified as “replicators” (Weinstein 1979). In some fields, being second can mean the loss of a Nobel Prize. Journals prefer to publish new discoveries and tend to reject manuscripts that merely replicate previous findings. Most author guidelines for journals confirm the prevalence of this policy.

One study of 429 reviewers from nineteen leading management and social science journals found that referees do not regard replicative studies favorably, and 52 percent stated that a direct replication of an original study that added no new matter to the theory would count against the article or would probably (or almost surely) cause them to recommend rejection (Kerr, Tolliver, and Petree 1977). Beyer, Chanove, and Fox (1995) observed that when authors submitted manuscripts to the *Academy of Management Journal* claiming expressively novel content, their articles were more likely to be accepted; however, Chase (1970) found that editors of natural sciences journals ranked replicability of research techniques as a key criterion, in contrast to their peers in the social sciences.

Besides novelty, other factors could explain the bias against replication. Although formally recognized as a key process in science, replication is regarded as a waste of resources by many researchers, and its publication is not worth the effort. A common rationale, openly voiced by many journal editors, appears to be scarcity of printing space. This dilemma seems especially acute in the social sciences and humanities, where rejection rates can surpass 80 percent (Gordon 1978). Journals in psychology and other social sciences publish fewer replications of studies than do journals in the experimental sciences (Bornstein 1990; Bozarth and Roberts 1972; Greenwald 1975; Mahoney 1985). None of the 362 papers analyzed by Sterling (1970) reported a direct replication of previous studies. Bozarth and Roberts (1972) demonstrated that less than 1 percent of a sample of papers published in three psychology journals reported replication studies. Hubbard and Vetter (1996) reviewed 4,270 studies published in five business disciplines and found that only 266 (6.2 percent) involved replications and extensions. Some journals do attempt to allocate space to replication studies, however, such as the *Journal of Social Psychology* (Doob 1994).

References

- Abby, M., M. D. Massey, S. Galandiuk, and H. C. Polk. 1994. Peer review is an effective screening process to evaluate medical manuscripts. *Journal of the American Medical Association* 272:105-108.
- American Psychological Association. 1983. *Publication manual of the American Psychological Association*. 3d ed. Washington, DC: American Psychological Association.
- Armstrong, J. S. 1980. Unintelligible management research and academic prestige. *Interfaces* 10:80-86.
- . 1997. Peer review for journals: Evidence on quality control, fairness, and innovation. *Science and Engineering Ethics* 3:63-84.
- Aronson, N. 1986. The discovery of resistance: Historical accounts and scientific careers. *Isis* 77:630-46.
- Astin, H. S. 1991. *Citation Classics*: Women's and men's perceptions of their contributions to science. In *The outer circle: Women in the scientific community*, edited by Harriet Zuckerman, Jonathan R. Cole, and John T. Bruer, 57-70. New York: W. W. Norton.
- Atkinson, D. R., M. J. Furlong, and B. E. Wampold. 1982. Statistical significance, reviewer evaluations and the scientific process: Is there a (statistically) significant relationship? *Journal of Counseling Psychology* 29:189-94.
- Azbel, M. 1993. Could Columbus have passed peer review? *Physics Today* 46:13-14.
- Bailar, J. C. 1991. Reliability, fairness, objectivity and other inappropriate goals in peer review. *Behavioral and Brain Sciences* 14:137-38.
- Bailar, J. C., and K. Patterson. 1985. Journal peer review. The need for a research agenda. *The New England Journal of Medicine* 312:654-57.
- Bakan, D. 1966. The test of significance in psychological research. *Psychological Bulletin* 66:423-37.
- Bakanic, V., C. McPhail, and R. J. Simon. 1989. Mixed messages: Referees' comments on the manuscripts they review. *Sociological Quarterly* 30:639-54.
- Bakker, P., and H. Ritger. 1985. Editors of medical journals: Who and from where. *Scientometrics* 7:11-22.
- Barber, B. 1961. Resistance by scientists to scientific discovery. *Science* 134:596-602.
- Bem, D. J., and C. Honorton. 1994. Does Psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin* 115:4-18.
- Berridge, M. J. 1988. *Citation Classics* commentary on *Nature* 312 (1 August): 315-321. *Current Contents, Life Sciences* 31:16.
- Beveridge, W.I.B. 1950. *The art of scientific investigation*. New York: Vintage Books.
- Beyer, J., R. G. Chanove, and W. B. Fox. 1995. The review process and the fates of manuscripts submitted to *AMJ*. *Academy of Management Journal* 38:1219-60.
- Beyer, J. M. 1978. Editorial policies and practices among leading journals in four scientific fields. *Sociological Quarterly* 19:68-88.
- Biggs, M. 1990. The impact of peer review on intellectual freedom. *Library Trends* 39:145-67.
- Bloembergen, N. 1993. *Physical Review* records the birth of the laser era. *Physics Today* 46:28-31.
- Bornstein, R. F. 1990. Publication politics, experimenter bias and the replication process in social science research. *Journal of Social Behavior and Personality* 5:71-81.
- . 1991a. Manuscript review in psychology: Psychometrics, demand characteristics and an alternative model. *Journal of Mind and Behavior* 12:429-68.

- . 1991b. The predictive validity of peer review: A neglected issue. *Behavioral and Brain Sciences* 14:139-40.
- Bozarth, J. D., and R. R. Roberts. 1972. Signifying significant significance. *American Psychologist* 27:774-75.
- Bradley, J. V. 1981. Pernicious publication practices. *Bulletin of the Psychonomic Society* 18:31-34.
- Braun, T., and E. Bujdoso. 1983. Gatekeeping patterns in the publication of analytical chemistry research. *Talanta* 30:161-67.
- Brown, G. W., and G. M. Baca. 1986. A classification of original articles: *AJDC* 1983 and 1984. *American Journal of Diseases in Children* 140:641-45.
- Campanario, J. M. 1993a. Consolation for the scientist: Sometimes it is hard to publish papers that are later highly-cited. *Social Studies of Science* 23:342-62.
- . 1993b. Not in our nature. *Nature* 361:488.
- . 1995a. Commentary: On influential books and journal articles initially rejected because of negative referees' evaluations. *Science Communication* 16:304-25.
- . 1995b. Using neural networks to study networks of scientific journals. *Scientometrics* 33:23-40.
- . 1996. Have referees rejected some of the most-cited articles of all times? *Journal of the American Society for Information Science* 47:302-10.
- . 1997. The journal scout. *Scientist* 11:9.
- Carver, R. P. 1978. The case against statistical significance testing. *Harvard Educational Review* 48:378-99.
- Chase, J. M. 1970. Normative criteria for scientific publication. *American Sociologist* 5:262-65.
- Chew, F. S. 1991. Fate of manuscripts rejected for publication in the *AJR*. *American Journal of Roentgenology* 156:627-32.
- Cicchetti, D. 1991. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences* 14:119-35.
- Cofer, C. N. 1985. Some reactions to manuscript review from a questionnaire study. *Behavioral and Brain Sciences* 8:745-46.
- Cohen, L. H. 1979. Clinical psychologists' judgments of the scientific merit and clinical relevance of psychotherapy outcome research. *Journal of Consulting and Clinical Psychology* 47:421-23.
- Colaizzi, L. A. 1994. Peer review in journals indexed in *Index Medicus*. *Journal of the American Medical Association* 272:156-58.
- Colman, A. M. 1982. Manuscript evaluation by journal referees and editors: Randomness or bias? *Behavioral and Brain Sciences* 5:205-206.
- Coughlin, E. K. 1989. Concerns about fraud, editorial bias prompt scrutiny of journal practices. *Chronicle of Higher Education* 35:A4-A7.
- Crane, D. 1967. The gatekeepers of science: Some factors affecting the selection of articles for scientific journals. *American Sociologist* 32:195-201.
- Cummings, L., P. J. Frost, and T. F. Vakil. 1985. The manuscript review process: A view from the inside on coaches, critics and special cases. In *Publishing in the organizational sciences*, 2d ed., edited by Larry L. Cummings and Peter J. Frost, 469-508. Homewood, IL: Irwin.
- Daniel, H. D. 1993. An evaluation of the peer review process at *Angewandte Chemie*. *Angewandte Chemie International Edition* 32:234-38.
- Dar, R. 1987. Another look at Meehl, Lakatos and the scientific practices of psychologists. *American Psychologist* 42:145-51.
- Diamond, A. M. 1986. What is a citation worth? *Journal of Human Resources* 21:200-15.

- Dickersin, K. 1990. The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association* 263:1385-89.
- Doob, L. W. 1994. Editing social psychology. In *Editors as gatekeepers: Getting published in the social sciences*, edited by Rita J. Simon and James J. Fyfe, 21-32. Boston: Rowman and Littlefield.
- Easterbrook, P. J., J. A. Berlin, R. Gopalan, and D. R. Matthews. 1991. Publication bias in clinical research. *Lancet* 337:867-72.
- Eckberg, D. L. 1991. When nonreliability of reviews indicates solid science. *Behavioral and Brain Sciences* 14:145-46.
- Eichorn, D. H., and G. R. Van den Boss. 1985. Dissemination of scientific and professional knowledge. *American Psychologist* 40:1309-16.
- Endres, F. F., and S. T. Wearden. 1990. Demographic portrait of journal reviewers. *Educator* 45:45-52.
- Epstein, W. M. 1990. Confirmational response bias among social work journals. *Science, Technology, & Human Values* 15:9-38.
- Eysenck, H. J., and S.B.G. Eysenck. 1992. Peer review: Advice to referees and contributors. *Personality and Individual Differences* 13:393-99.
- Feinstein, A. R. 1991. Construction, consent and condemnation in research on peer review. *Journal of Clinical Epidemiology* 44:339-41.
- Finke, R. A. 1990. Recommendations for contemporary editorial practices. *American Psychologist* 45:669-70.
- Fiske, D. W., and L. Fogg. 1990. But the reviewers are making different criticism of my paper! *American Psychologist* 45:591-98.
- Fleiss, J. L. 1982. Deception in the study of the peer-review process. *Behavioral and Brain Sciences* 5:210-11.
- Forcese, D. P., and S. Richer. 1973. *Social research methods*. Englewood Cliffs, NJ: Prentice Hall.
- Frost, P., and R. Taylor. 1995. Partisan perspective: A multiple-level interpretation of the manuscript review process in social science journals. In *Publishing in the organizational sciences*, 2d ed., edited by Larry L. Cummings and Peter J. Frost, 13-43. Thousand Oaks, CA: Sage.
- Fyfe, J. J. 1994. Cops and robbers in academe: Editing *Justice Quarterly*. In *Editors as gatekeepers: Getting published in the social sciences*, edited by Rita J. Simon and James J. Fyfe, 59-72. Boston: Rowman and Littlefield.
- Gans, J. S., and G. B. Shepherd. 1994. How are the mighty fallen: Rejected classic articles by leading economists. *Journal of Economic Perspectives* 8:165-79.
- Gardner, M. J., and J. Bond. 1990. An exploratory study of statistical assessment of papers published in the *British Medical Journal*. *Journal of the American Medical Association* 263:1355-57.
- Garfield, E. 1985. Uses and misuses of citation frequency. *Current Contents*, 10 October, 3-9.
- . 1986a. Refereeing and peer review. Part 1. Opinion and conjecture on the effectiveness of refereeing. *Current Contents*, 4 August, 3-11.
- . 1986b. Refereeing and peer review. Part 2. The research on refereeing and alternatives to the present system. *Current Contents*, 11 August, 3-12.
- . 1989. Citation behavior—An aid or an hindrance to information retrieval? *Current Contents*, 1 May, 3-8.
- . 1990. The most-cited papers of all time, SCI 1945-1988. Part 1A. The SCI top 100—Will the Lowry method ever be obliterated? *Current Contents*, 12 February, 3-15.

- Garfunkel, J. M., M. H. Ulshen, H. J. Hamrick, and E. E. Lawson. 1990. Problems identified by secondary review of accepted manuscripts. *Journal of the American Medical Association* 263:1369-71.
- Gibbons, J. D., and M. Fish. 1991. Ranking of economics faculties and representation on editorial boards of top journals. *Journal of Economics Education* 22:361-72.
- Gidez, L. I. 1991. The peer review process: Strengths and weaknesses, a survey of attitudes, perceptions and expectations. *Serials Librarian* 11:75-85.
- Gilmore, J. B. 1991. On forecasting validity and finessing reliability. *Behavioral and Brain Sciences* 14:148-49.
- Glass, G. V. 1982. Meta-analysis: An approach to the synthesis of research results. *Journal of Research in Science Teaching* 19:93-112.
- Glenn, N. D. 1976. The journal article review process: Some proposals for change. *American Sociologist* 11:179-85.
- Glogoff, S. 1988. Reviewing the gatekeepers: A survey of referees of library journals. *Journal of the American Society for Information Science* 39:400-407.
- Gordon, M. D. 1978. Disciplinary differences, editorial practice and the patterning of rejection rates for UK research journals. *Journal of Research Communication Studies* 2:139-59.
- Gottfredson, S. D. 1978. Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist* 33:920-34.
- Greenwald, A. G. 1975. Consequences of prejudice against the null hypothesis. *Psychological Bulletin* 82:1-20.
- Guest, A. M. 1994. Gatekeeping among the demographers. In *Editors as gatekeepers: Getting published in the social sciences*, edited by Rita J. Simon and James J. Fyfe, 85-106. Boston: Rowman and Littlefield.
- Hamermesh, D. S. 1994. Facts and myths about refereeing. *Journal of Economic Perspectives* 8:153-63.
- Hamilton, D. P. 1990. Publishing by—and for?—the numbers? *Science* 250:1331-32.
- Hargens, L. L. 1988. Scholarly consensus and journal rejection rates. *American Sociological Review* 53:139-51.
- Hartley, J. P. 1993. In defence of referees. *Psychologist* 6:200-201.
- Hersen, M., and D. J. Miller. 1992. Future directions: A modest proposal. In *Research fraud in the behavioral and biomedical sciences*, edited by David J. Miller and Michael Hersen, 225-44. New York: John Wiley.
- Horrobin, D. F. 1990. The philosophical basis of peer review and the suppression of innovation. *Journal of the American Medical Association* 263:1438-41.
- Hubbard, R., and J. S. Armstrong. 1992. Are null results becoming an endangered species in marketing? *Marketing Letters* 3:127-36.
- Hubbard, R., and D. E. Vetter. 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 publication. *American Journal of Medicine* 56:666-92.
- Jauch, L. R., and J. L. Wall. 1989. What they do when they get your manuscript: A survey of Academy of Management reviewer practices. *Academy of Management Journal* 32:157-73.
- Judson, H. F. 1994. Structural transformations of the sciences and the end of peer review. *Journal of the American Medical Association* 272:92-94.
- Juhasz, S., E. Calvert, T. Jackson, D. A. Kronick, and J. Shipman. 1975. Acceptance and rejection of manuscripts. *IEEE Transactions on Professional Communication* 18:177-85.

- Kassirer, J. P., and E. W. Campion. 1994. Peer review: Crude and understudied, but indispensable. *Journal of the American Medical Association* 272:96-97.
- Kawano, T., T. J. Kehle, E. Clark, and W. R. Jenson. 1993. School psychology journals: Relationships with related journals and external and internal quality indices. *Journal of School Psychology* 31:407-24.
- Kemper, K. J. 1991. Pride and prejudice in peer review. *Journal of Clinical Epidemiology* 44:343-45.
- Kerr, S., J. Tolliver, and D. Petree. 1977. Manuscript characteristics which influence acceptance for management and social science journals. *Academy of Management Journal* 20:132-41.
- Kiesler, C. A. 1991. Confusion between reviewer reliability and wise editorial and funding decisions. *Behavioral and Brain Sciences* 14:151-52.
- Koren, G. 1986. A simple way to improve the chances for acceptance of your scientific paper. *New England Journal of Medicine* 315:1298.
- Laband, D. N. 1990. Is there value-added from the review process in economics? Preliminary evidence from authors. *Quarterly Journal of Economics* 105:341-52.
- Laband, D. N., and M. J. Piette. 1994. Favoritism versus search of good papers: Empirical evidence regarding the behavior of journal editors. *Journal of Political Economy* 102:194-203.
- LaFollette, M. C. 1992. *Stealing into print: Fraud, plagiarism and misconduct in scientific publishing*. Berkeley: University of California Press.
- Laidler, K. J., and M. C. King. 1983. The development of transition-state theory. *Journal of Physical Chemistry* 87:2657-64.
- Lederberg, J. 1978. Digital communications and the conduct of science: The new literacy. *Proceedings of the IEEE* 66:1314-19.
- Legendre, A. M. 1995. Peer review of manuscripts for biomedical journals. *Journal of the American Veterinary Medical Association* 207:36-38.
- Lindsey, D. 1976. Distinction, achievement and editorial board membership. *American Psychologist* 31:799-804.
- . 1977. Participation and influence in publication review proceedings: A reply. *American Psychologist* 32:579-85.
- . 1978. *The scientific publication system in social science: A study of the operation of leading professional journals in psychology, sociology and social work*. San Francisco: Jossey-Bass.
- Lindsey, D., and T. Lindsey. 1978. The outlook of journal editors and referees on the normative criteria of scientific craftsmanship. *Quality and Quantity* 12:45-62.
- Lock, S. 1985. Letter to P.B.S. Fowler. 4 December 1984. *British Medical Journal* 290:1560.
- Lock, S., and J. Smith. 1991. What do peer reviewers do? *Journal of the American Medical Association* 263:1341-43.
- Loftus, G. R. 1993. A picture is worth a thousand *p* values: On the irrelevance of hypothesis testing in the microcomputer age. *Behavioral Research Methods, Instruments and Computers* 25:250-56.
- Mahoney, M. J. 1982. Publication, politics and scientific progress. *Behavioral and Brain Sciences* 5:220-21.
- . 1985. Open exchange and epistemic progress. *American Psychologist* 40:29-39.
- . 1990. Bias, controversy, and abuse in the study of the scientific publication system. *Science, Technology, & Human Values* 15:50-55.
- McAleese, R. 1993. Reform options for peer review. *Nature* 364:184.
- McCutchen, C. W. 1991. Peer review: Treacherous servant, disastrous master. *Technology Review* 94:28-40.

- McNutt, R. A., A. T. Evans, R. H. Fletcher, and S. W. Fletcher. 1990. The effects of blinding on the quality of peer review. *Journal of the American Medical Association* 263:1371-74.
- Medoff, M. 1989. The rankings of economists. *Journal of Economic Education* 20:405-15.
- Mehel, P. E. 1967. Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science* 34:103-15.
- Melton, A. W. 1962. Editorial. *Journal of Experimental Psychology* 64:553-57.
- Merton, R. K. 1957. Priorities in scientific discovery. *American Sociological Review* 22:635-59.
- Metoyer-Duran, C. 1993. The readability of published, accepted, and rejected papers appearing in *College and Research Libraries*. *College and Research Libraries* 54:517-26.
- Micceri, T. 1989. The unicorn, the normal curve and other improbable creatures. *Psychological Bulletin* 105:156-66.
- Michels, R. 1995. Peer review. *International Journal of Psycho-Analysis* 76:217-21.
- Mindick, B. 1982. When we practice to deceive: The ethics of a metascientific inquiry. *Behavioral and Brain Sciences* 5:226-27.
- Minerva, A. 1983. Views. *British Medical Journal* 287:1886.
- Murphy, T. M., and J. M. Utts. 1994. A retrospective analysis of peer review at *Physiologia Plantarum*. *Physiologia Plantarum* 92:535-42.
- Murray, G. D. 1988. The task of a statistical referee. *British Journal of Surgery* 75:664-67.
- Nelson, K. 1982. Reliability, bias, or quality: What is the issue? *Behavioral and Brain Sciences* 5:229.
- Newcombe, R. G. 1987. Towards a reduction in publication bias. *British Medical Journal* 295:656-59.
- Nylenna, M., P. Riss, and Y. Karlsson. 1994. Multiple blinded reviews of the same two manuscripts. *Journal of the American Medical Association* 272:149-51.
- Peters, D. P., and S. J. Ceci. 1982a. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences* 5:187-95.
- . 1982b. Peer-review research: Objections and obligations. *Behavioral and Brain Sciences* 5:246-52.
- Pipke, I. 1984. The gatekeepers: A multivariate study of accepted and rejected adult education research conference abstracts 1978-80. *Adult Education Quarterly* 34:71-84.
- Polanyi, M. 1958. *Personal knowledge*. London: Routledge and Kegan Paul.
- Pressman, S. 1994. Simultaneous multiple journal submissions: The case against. *American Journal of Economics and Sociology* 53:315-34.
- Redner, H. 1987. Pathologies of science. *Social Epistemology* 1:215-47.
- Relman, A. S. 1978. Are we a filter or a sponge? *New England Journal of Medicine* 299:197.
- Rennie, D. 1993. More peering into editorial peer review. *Journal of the American Medical Association* 270:2856-58.
- Roberts, J. C., R. H. Fletcher, and S. W. Fletcher. 1994. Effects of peer review and editing on the readability of articles published in *Annals of Internal Medicine*. *Journal of the American Medical Association* 272:119-21.
- Roediger, H. L. 1987. The role of journal editors in the scientific process. In *Scientific excellence: Origins and assessment*, edited by Douglas N. Jackson and J. Philippe Rushton, 222-52. Newbury Park, CA: Sage.
- Romanelli, E. 1995. Becoming a reviewer. Lessons somewhat painfully learned. In *Publishing in the organizational sciences*, 2d ed., edited by Larry L. Cummings and Peter J. Frost, 195-202. Thousand Oaks, CA: Sage.
- Rosenthal, R. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86:638-41.

- Saksena, S., J. D. Fisher, N. Madan, and B. Krause. 1995. Is justice blind? Blinded and unblinded abstract reviews. *PACE* 18:759-64.
- Salsburg, D. S. 1985. The religion of statistics as practiced in medical journals. *American Statistician* 39:220-23.
- Scanlan, B. P. 1991. Passengers, drivers and traffic cops on peer review road. *Serials Librarian* 19:71-73.
- Scharschmidt, B. F., A. DeAmicis, P. Bacchetti, and M. J. Held. 1994. Chance, concurrence and clustering: Analysis of reviewers' recommendations on 1000 submissions to the *Journal of Clinical Investigation*. *Journal of Clinical Investigation* 93:1877-80.
- Scherer, R. W., K. Dickensin, and P. Langenberg. 1994. Full publication of results initially presented in abstracts. *Journal of the American Medical Association* 272:158-62.
- Schönemann, P. H. 1991. In praise of randomness. *Behavioral and Brain Sciences* 14:162-63.
- Schulman, K., D. P. Sulmasy, and D. Roney. 1994. Ethics, economics and the publication policies of major medical journals. *Journal of the American Medical Association* 272:154-55.
- Seglen, P. O. 1992. The skewness of science. *Journal of the American Society for Information Science* 43:628-38.
- Shaver, J. P. 1993. What statistical significance testing is, and what it is not. *Journal of Experimental Education* 61:293-316.
- Siegelman, S. S. 1991. Assassins and zealots: Variations in peer review. *Radiology* 178:637-42.
- Simon, R. J. 1994. An effective journal editor: Insights gained from editing the *American Sociological Review*. In *Editors as gatekeepers: Getting published in the social sciences*, edited by Rita J. Simon and James J. Fyfe, 33-44. Boston: Rowman and Littlefield.
- Simon, R. J., V. Bakanic, and C. McPhail. 1986. Who complains to journal editors and what happens. *Sociological Inquiry* 56:259-71.
- Skinner, B. F. 1956. A case history in scientific method. *American Psychologist* 11:221-23.
- Smith, M. L. 1980. Sex bias in counseling and psychotherapy. *Psychological Bulletin* 87:392-407.
- Sokal, A. 1996. Transgressing the boundaries: Towards a transformative hermeneutics of quantum gravity. *Social Text* 46-47:217-52.
- Sterling, T. 1970. Publication decisions and their possible effects on inferences drawn from tests of significance or vice versa. In *The significance tests controversy: A reader*, edited by Denton E. Morrison and Ramon E. Henkel, 295-300. London: Butterworths.
- Sternberg, R. J. 1985. Tacit agreements between authors and editors. *Behavioral and Brain Sciences* 8:746-47.
- Stossel, T. P. 1985. Reviewer status and review quality: Experience of the *Journal of Clinical Investigation*. *New England Journal of Medicine* 312:658-59.
- Stricker, L. J. 1991. Disagreement among journal reviewers: No cause for undue alarm. *Behavioral and Brain Sciences* 14:163-64.
- Thompson, B. 1996. AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher* 25:26-30.
- Toulmin, S. 1972. *Human understanding*. Princeton, NJ: Princeton University Press.
- Tuckman, H. H., and J. Leahey. 1975. What is an article worth? *Journal of Political Economy* 83:951-67.
- Van Heerden, J. H., and J. Hoogstraten. 1978. Significance as a determinant of interest in scientific research. *European Journal of Social Psychology* 8:141-43.
- Vidal, C. 1990. Las ondas quimicas. *Mundo Cientifico* 10:184-92.
- Weinstein, D. 1979. Fraud in science. *Social Science Quarterly* 59:639-52.
- Weisse, A. B. 1986. Say it isn't no: Positive thinking and the publication of medical research. *Hospital Practice*, March, 23-25.

- Willis, C. L., and S. J. McNamee. 1990. Social networks of science and patterns of publication in leading sociology journals, 1960 to 1985. *Knowledge: Creation, Diffusion, Utilization* 11:363-81.
- Wilson, J. D. 1978. Peer review and publication: Presidential address before the seventieth annual meeting of the American Society for Clinical Investigation, San Francisco, California, 30 April 1978. *Journal of Clinical Investigation* 67:1697-701.
- Yankauer, A. 1991. Who are the peer reviewers and how much do they review? *Journal of the American Medical Association* 263:1338-40.
- Yates, F. 1951. The influence of "Statistical Methods for Research Workers" on the development of the science of statistics. *Journal of the American Statistical Association* 46:19-34.
- Yoels, W. C. 1971. Destiny or dynasty: Doctoral origins and appointment patterns of editors of the *American Sociological Review*, 1948-1968. *American Sociologist* 6:134-39.
- . 1974. The structure of scientific fields and the allocation of editorships on scientific journals: Some observations on the politics of knowledge. *Sociological Quarterly* 15:264-76.
- Zentall, T. R. 1991. What to do about peer review: Is the cure worse than the disease? *Behavioral and Brain Sciences* 14:166-67.
- Ziman, J. 1968. *Public knowledge: The social dimension of science*. Cambridge: Cambridge University Press.
- Zsindely, S., and A. Schubert. 1989. Editors-in-chief of medical journals: Are they experts, authorities, both or neither? *Communication Research* 16:695-700.
- Zsindely, S., A. Schubert, and T. Braun. 1982. Citation patterns of editorial gatekeepers in international chemistry journals. *Scientometrics* 4:69-76.
- Zuckerman, H., and R. K. Merton. 1971. Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva* 9:66-100.

JUAN MIGUEL CAMPANARIO is Assistant Professor, Department of Physics and Institute for Pedagogical Sciences, University of Alcalá, Madrid, Spain. He is a quantum chemist who has taught at the Universidad Nacional Autónoma de Nicaragua, and is currently working and publishing in the fields of cognitive sciences and science communication.