Responsible Authorship and Peer Review

James R. Wilson

Department of Industrial Engineering, North Carolina State University, USA

Keywords: authorship principles, refereeing guidelines, peer review

ABSTRACT: *In this article the basic principles of responsible authorship and peer review are surveyed, with special emphasis on (a) guidelines for refereeing archival journal articles and proposals; and (b) how these guidelines should be taken into account at all stages of writing.*

> In the South Seas there is a cargo cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas—he's the controller—and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things cargo cult science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land.
>
> Now it behooves me, of course, to tell you what they're missing. … It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty—a kind of leaning over backwards. For example, if you're doing an experiment, you should report everything that you think might make it invalid—not only what you think is right about it: other causes that could possibly explain your results; and things you thought of that you've eliminated by some other experiment, and how they worked—to make sure the other fellow can tell they have been eliminated.
>
> … In summary, the idea is to try to give *all* of the information to help others to judge the value of your contribution; not just the information that leads to judgment in one particular direction or another.
>
> Richard P. Feynman, "*Surely You're Joking, Mr. Feynman!*" [1 (pp. 310–311)]

Address for correspondence: James R. Wilson, Professor and Head, Department of Industrial Engineering; 2401 Stinson Drive, Riddick Labs 328, North Carolina State University; Raleigh, NC 27695-7906, USA; jwilson@eos.ncsu.edu (email).

Paper received, 21 June 2001: revised, 4 March 2002: accepted, 5 March 2002.

1. Introduction

In the context of research in science and engineering, the essential principles of responsible authorship and peer review are captured in Feynman's memorable exhortation for "utter honesty" and "leaning over backwards" in all phases of performing and documenting such research; and clearly these precepts extend to refereeing (reviewing) archival journal articles and proposals submitted by other researchers. Instead of concentrating the discussion on the problems that can arise from self-deception or conscious, deliberate misconduct, the main focus of this paper is on exemplary practices in writing articles and proposals and in reviewing the articles and proposals of others.

As one of the instructional modules on research and professional ethics that were developed through the Research Ethics Initiative of the Graduate School of North Carolina State University, this paper was originally intended to serve as a primer for graduate students on authorship and peer review; and this revised and expanded version of the paper is intended to address a broader audience of interested readers. Two notorious episodes in the history of science illustrate the need for strict adherence to the principles of responsible authorship and peer review.

1.1 "*The Science of Things That Aren't So*"

In addition to performing Nobel Prize–winning research, the American physicist Irving Langmuir explored extensively a subject he called "pathological science," defining this as "the science of things that aren't so." Although he never published his investigations on this subject, he presented a colloquium on pathological science at General Electric's Knolls Atomic Power Laboratory on December 18, 1953. Subsequently Robert N. Hall, one of Langmuir's former colleagues at General Electric, transcribed and edited a recording of Langmuir's presentation so that it could be published in the October 1989 issue of *Physics Today*.[2] This article is a fascinating account of famous cases of self-deception by scientists working in a broad diversity of disciplines.

1.1.1 *The Discovery of N Rays*

Perhaps the most remarkable case documented by Langmuir and Hall concerns the discovery of N rays by the French physicist René Blondlot in 1903. This exotic form of radiation was claimed to penetrate inches of aluminum while being stopped by thin foils of iron. When N rays impinged on an object, Blondlot claimed a slight increase in the brightness of the object; but he admitted that great experimental skill was needed to detect the effect of these rays.

During the period from 1903 to 1906, over 300 papers were published on N rays by 100 scientists and medical doctors around the world.[3] When the American physicist Robert W. Wood learned about the discovery of N rays, he went to France to observe Blondlot's experimental procedure. At that time Blondlot was using a spectroscope fitted with an aluminum prism to measure the refractive indices of N rays. Although Blondlot's experiments were performed in a darkened room, a small red (darkroom) lantern enabled Blondlot to see a graduated scale for measuring to three significant figures the position of a vertical thread coated with luminous paint. The thread was supposed to brighten as it crossed the invisible lines of the N-ray spectrum. According to Langmuir and Hall, Wood asked Blondlot the following question:

> … from just the optics of the thing, with slits 2 mm wide, how can you get a beam so fine that you can detect its position to within a tenth of a millimeter? [2 (p. 43)]

Blondlot is reported to have given this reply.

> That's one of the fascinating things about N rays. They don't follow the ordinary laws of science … You have to consider these things by themselves. They are very interesting but you have to discover the laws that govern them. [2 (p.43)]

His suspicions aroused at this point, Wood used the cover of the darkened room to remove the prism and put it in his pocket. Wood then asked Blondlot to repeat some of his measurements. With the critical component of the experimental apparatus missing, Blondlot obtained exactly the same results. In a letter published in *Nature*, Wood[4] exposed Blondlot's experiments on N rays as a case of self-deception. Although Wood's letter killed research on N rays outside France, it is interesting to note that the French Academy of Sciences chose Blondlot to receive the 1904 Le Conte Prize—even though the other leading candidate was Pierre Curie, who together with Marie Curie and Henri Becquerel had shared the 1903 Nobel Prize in physics for pioneering work on radioactivity.

Langmuir and Hall also discuss a number of other anomalous phenomena, and they analyze the main symptoms of pathological science (or cargo cult science, to use Feynman's more colorful expression). These symptoms are summarized in Table 1. The case of N rays exhibits all of these symptoms. As a prelude to the detailed discussion of the principles of responsible authorship (Section 2), note that the foremost concern of all authors should be to avoid lapsing into self-deception and pathological science. Beyond the N ray episode, the "Cold Fusion Fiasco" reveals that there are important practical as well as philosophical grounds for following accepted principles of responsible authorship and peer review; and this point is elaborated in Sections 2 and 4 below.

Table 1: Langmuir's Symptoms of Pathological Science

1. The maximum effect that is observed is produced by a causative agent of barely detectable intensity, and the magnitude of the effect is substantially independent of the intensity of the cause.
2. The effect is of a magnitude that remains close to the limit of detectability or, many measurements are necessary because of the very low statistical significance of the results.
3. There are claims of great accuracy.
4. Fantastic theories contrary to experience are suggested.
5. Criticisms are met by *ad hoc* excuses thought up on the spur of the moment.
6. The ratio of supporters to critics rises up somewhere near 50% and then falls gradually to oblivion

1.1.2 *The Cold Fusion Fiasco*

In recent years the most well-known example of pathological science can be found in the cold fusion experiments performed by Martin Fleischmann and B. Stanley Pons at the University of Utah. This episode is fully documented in the book *Cold Fusion: The Scientific Fiasco of the Century*.[5] On March 23, 1989, Fleischmann and Pons gained worldwide attention by announcing at a press conference that by means of a simple electrochemical process, they had successfully induced a sustained nuclear fusion reaction at room temperature in a small jar on a laboratory tabletop. They claimed to measure amounts of excess heat in their experiments that offered the promise of a virtually inexhaustible source of power for the planet, not to mention immense financial rewards for the University of Utah as well as the discoverers of this remarkable electrochemical process.

Less than two weeks prior to their press conference, Fleischmann and Pons submitted their initial paper on cold fusion to the *Journal of Electroanalytical Chemistry*; and that paper was published only four weeks after submission.[6] A long list of errata soon followed[7]—including the name of Marvin Hawkins, a coauthor who was somehow omitted from the original paper. Over the next six months, many labs around the world rushed to perform experiments designed to confirm the results claimed by Fleischmann and Pons. Hastily announced confirmations of cold fusion were followed by equally swift retractions when additional experiments failed to yield the claimed excess heat. In November 1989 a U.S. Department of Energy panel on cold fusion released its final report, concluding that "the experimental results of excess heat from calorimetric cells reported to date do not present convincing evidence that useful sources of energy will result from the phenomena attributed to cold fusion";[5 (p.89)] and thus cold fusion was completely discredited by the mainstream scientific community.

Huizenga summarized the lessons of the cold fusion fiasco as follows:

> The University of Utah's handling of cold fusion is a striking illustration of what happens when scientists circumvent the normal peer-review process, when scientists use the press as a conduit to disseminate information about a claimed discovery in an unrealistic and overly optimistic tone, when scientists require too many miracles to account for their results, when research is done in isolation by scientists who are outside their field of expertise, when data are published by private communication rather than by those responsible, when administrators use potential royalties to force premature publication and when university administrators lobby for large federal funds before the science is confirmed. Cold fusion is an example of bad science where the normal rules and procedures of the scientific process were violated. One can only be amazed by the number of scientists who reported confirmation of cold fusion by press conference, only to follow later with a retraction or at least a confession of irreproducibility. Reproducibility is the essence of science. It has taken upwards of some fifty to one hundred million dollars of research time and resources to show that there is no convincing evidence for room-temperature fusion. Much of this effort would not have been necessary had normal scientific procedures been followed. The idea of

> producing energy from room-temperature fusion is destined to join N rays … as another example of a scientific aberration.
> … The general scientific enterprise is vibrant and healthy and has weathered the cold fusion flurry with only minor bruises and scratches. The cold fusion fiasco illustrates once again, as N rays … did earlier, that the scientific process works by exposing and correcting its own errors.[5 (pp. 235–236)]

The cold fusion episode highlights the potentially grave consequences that result when the standard procedures of peer review are not followed (see Section 5.1), and when researchers fail to adhere to the principles of responsible authorship (see Section 2.1).

2. Principles of Responsible Authorship

In an essay on the ethical basis of science, Glass[8] succinctly summarized the main premise underlying our discussion of responsible authorship with a quotation from the book *Science and Human Values*:

> …we must be able to rely on other people; we must be able to trust their word.[9 (p.57)]

The need for such trust in science was elaborated in Chapter 1 of the fifth edition of the *CBE Style Manual*.

> Scientists build their concepts and theories with individual bricks of scientifically ascertained facts, found by themselves and their predecessors. Scientists can proceed with confidence only if they can assume that the previously reported facts on which their work is based are indeed correct. Thus all scientists have an unwritten contract with their contemporaries and those whose work will follow to provide observations honestly obtained, recorded, and published. This ethic is no more than science's application of the ancient Golden Rule: "Do unto others as you would have them do unto you." It is an ethic that should govern everyone in the community of scientists when they serve as authors, editors, or manuscript referees.[10 (p.1)]

(Regrettably, this initial chapter was omitted from *Scientific Style and Format*,[11] which is the sixth edition of the *CBE Style Manual.)*

2.1 *Principles of Authorship*

It follows from our main premise that the authors of a scientific work must have participated sufficiently in the work so as to take public responsibility for its content, and they must be willing and able to respond to questions about the work. Moreover, at a minimum an author should have made substantial contributions to all the following aspects of the project as explained in the key document "Uniform Requirements for Manuscripts Submitted to Biomedical Journals" promulgated by the International Committee of Medical Journal Editors (ICMJE):[12, 13]

- conception and design, or acquisition of data, or analysis and interpretation of data;
- drafting the article or revising it critically for important intellectual content; and
- final approval of the version to be published.

The following are not sufficient to justify authorship:
- participating solely in acquisition of funding;
- participating solely in collection of data; or
- supervising the overall activities of the research group.

It should be noted that although the ICMJE criteria for authorship have been widely publicized, they have not been universally adopted even within the biomedical sciences, in part because a substantial gap remains between the guidance provided by these criteria and the problematic aspects of multiple authorship that commonly arise in practice.[14]

Beyond the minimal authorship requirements listed above, Houk and Thacker[15] elaborate the following ways to earn the status of coauthorship:
- contribution of original ideas;
- design and writing of an approved protocol;
- responsibility for acquisition of data;
- responsibility for and leadership of the performance of the study;
- analysis and critical interpretation of data—including review and evaluation of previous studies;
- drafting, revising, and reviewing the manuscript;
- responsibility for the final manuscript; or
- willingness and ability to defend the publication.

In recent years, there has been a rapid proliferation of large, interdisciplinary research teams writing articles with tens or even hundreds of potential authors; and it is not surprising that authorship disputes are occurring more frequently.[14,16] From a larger perspective, an even more serious problem in such situations is the lack of a clear-cut assignment of responsibility among all the potential authors for ensuring the integrity of the entire article and for answering any questions about the article that may arise after publication. These issues have led many scientists and scientific editors to propose alternatives to the traditional model for authorship outlined above. In particular, Rennie, Yank, and Emanuel[14] proposed replacing the concept of an author with the concept of a *contributor* whose role in the work is precisely specified in a footnote to the published paper; moreover in the same footnote at least one contributor (usually a coinvestigator) is designated as a *guarantor* for the integrity of the article as a whole. Davidoff[16] asserts that full disclosure of the contributions of each collaborator merely ensures that the same high standards for accurately reporting the scientific information in the article are also applied in the assignment of credit and accountability for the work.

The selection of authors (or contributors) for a paper should be jointly agreed by all the collaborators on a project as soon as the group has decided on the assignment of responsibilities and workload for all members of the group. Failure to achieve such consensus at the outset can have serious repercussions later. Perhaps the most extreme example of such a failure is the original paper on cold fusion,[6] which Fleischmann and Pons prepared in such haste that they inadvertently omitted their coauthor Marvin Hawkins—even though Fleischmann and Hawkins had done the bulk of the work in the months prior to the submission of the paper;[5] and there is some evidence that Hawkins heavily pressured his collaborators to be included as a coauthor in the list of errata[7] that soon followed.[17] Considerations of the division of labor naturally lead to the question of who shall be the primary author (or contributor).

2.2 *Principles of Primary Authorship*

The primary author (that is, the author listed first in the article's byline) must have demonstrated the ability and willingness to exert scientific leadership of the project so as to (a) assume responsibility for a major professional aspect of the work, and (b) ensure that all the project objectives are met. Thus the primary author of a paper is generally chosen based on an evaluation of that individual's contributions to the conception, planning, and execution of the study. Selection of the primary author often occurs after the experimental work has been performed but just before the paper is written. If two or more authors have contributed equally to the project, then the primary author should be the one who by mutual consent actually coordinates the overall writing of the paper. According to Houk and Thacker,[14] individuals who satisfy one or more of the following criteria should be considered as candidates for primary authorship:

- Originality of contribution—the primary author made an original theoretical or methodological contribution that proved to be a highly important basis for the paper.
- Major intellectual input—throughout the study, the primary author generated ideas on the study design and modifications, on ensuring availability and use of appropriate experimental subjects or material, on productively conducting the study, on solving measurement problems, on analyzing and interpreting data in a particular way, and on preparing reports.
- Major feature of the manuscript—the primary author originated and developed the feature of the paper that is of central importance.
- Greatest overall contribution—the primary author did the most work, made the study succeed, provided intellectual leadership, and analyzed and interpreted the data.

No one is entitled to primary authorship solely because of administrative position or expertise in a particular subject or discipline. Selection of the primary author should reflect a consensus of the paper's coauthors on the most deserving individual. For the alternative contributorship model of collaboration, Rennie, Yank, and Emanuel[14] and Davidoff[16] recommend that in arriving at mutually agreed-upon descriptions of their individual contributions to the paper, the collaborators should also arrive at a mutually agreed-upon ranking of the relative importance of those contributions; and this ranking should be used to determine the order of contributors in the byline and in the contributors list.

It should be noted that other designations such as *first author, senior author, corresponding author*, and *last author* are sometimes used to denote different divisions of responsibility than were described above for the primary author.[18] For example, the senior author may be a faculty adviser (mentor) who coordinated the overall writing of a paper on a student's doctoral dissertation research, with the student listed as the first author, followed by any other members of the student's supervisory committee who played a significant role in the work, and with the senior author listed last. The term *primary author* is used in this paper to simplify the discussion.

2.3 *Writing the Paper*

In writing the paper, the primary author should coordinate the contributions of all the coauthors, who are responsible for both the style and content of their respective sections of the paper. (For guidelines on writing style and "mechanical" issues not related to subject-matter content, see the comprehensive list of references on technical writing given in reference 19.) In addition, to facilitate the passage of a paper through the peer review process, Section 4.1 provides a checklist of key questions to be answered in a referee's report that authors should keep firmly in mind while writing a paper.

In all phases of documenting the work, each author should adhere strictly to Feynman's precepts of "utter honesty" and "leaning over backwards" to avoid the pitfall of self-deception as exemplified by the case of N rays described above. Moreover, authors should guard against selective underreporting of experimental results that are "disappointing" or "uninteresting"—that is, results for which no significant effects could be detected. Chalmers[20] asserts that underreporting research is scientific misconduct, pointing out that in the biomedical sciences this practice can result in the continued use of medical treatments that are unnecessarily risky, unpleasant, or costly because practicing physicians who prescribe those treatments are unaware that the expected benefits of the treatments have not been reconfirmed in more extensive follow-up experimentation. Moreover, underreporting research is linked to the perceived bias of the peer review system in favor of papers that report the detection of significant effects (see Section 4.4).

An auctorial pitfall complementary to underreporting of research results is overselling such results—that is, the increasingly common tendency for authors to indulge in rhetorical exaggeration of the merits of their work while eschewing an exhaustive discussion of the deficiencies of that work. The cold fusion fiasco provides an extreme example of overselling research. Often the rationale for such behavior is that for research of major importance, the task of publicizing the limitations and defects of the work may safely be left to rival researchers who have competing techniques to promote.[21] In contrast to this point of view, Peter Medawar, the winner of the 1960 Nobel Prize in medicine for his work on tissue transplantation, made the following statement in his book *Advice to a Young Scientist*:

> *I cannot give any scientist of any age better advice than this: the intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not.* The importance of the strength of our conviction is only to provide a proportionately strong incentive to find out if the hypothesis will stand up to critical evaluation.[22(p.39)]

(The emphasis in the quoted statement is Medawar's.)

Another perspective on overselling research results is the following:

> Over the past twenty years, I have accumulated considerable experience in mediating extremely acrimonious disputes between researchers acting as "severe Popperian critics" of each other's work. Much of this hard-won experience was gained during the nine years that I served as a departmental editor and former departmental editor of the journal *Management Science*. To avoid reopening wounds which have not had much time to heal, I will not go into the particulars of any of these cases; but I feel compelled to draw some general conclusions based on these cases.
>
> In every one of the disputes that I mediated, the trouble started with extensive claims about the general applicability of some simulation-based methodology; and then failing to validate these claims independently, reviewers and other researchers proceeded to write up and disseminate their conclusions. This in turn generated a heated counterreaction, usually involving claims of technical incompetence or theft of ideas or both. Early in my career I served as the "special prosecutor" in several of these cases. Later on I moved up to become the "judge," and in the end I was often forced to play the role of the "jury" as well. In every one of these cases, ultimately the truth emerged (as it must, of course)—but the process of sorting things out involved the expenditure of massive amounts of time and energy on the part of many dedicated individuals in the simulation community, not to mention the numerous professional and personal relationships that were severely damaged along the way.[23 (p.1410)]

In brief, when authors violate Feynman's precepts of "utter honesty" and "leaning over backwards" by overselling their work, the cost to the scientific enterprise of policing these individuals rapidly becomes exorbitant.

3. Role of the Peer Review System

The main purpose of the peer review system is to serve the community of researchers—and ultimately to benefit society—by providing *expert advice* to:

- editors of archival journals who must make decisions on acceptance, rejection, or revision of papers submitted to their journals; and
- program managers of funding agencies who must make funding decisions on the research proposals submitted to their programs.

At the most basic level, the peer review system performs a quality-control function that is essential to maintaining the self-correcting character of the scientific research enterprise.

The Royal Society of London is frequently given credit for introducing the concept of refereeing or reviewing scientific manuscripts when that organization took over official responsibility for publication of the *Philosophical Transactions* in 1752.[24] However, it was not until the first decade of the twentieth century that increasing specialization forced editors in virtually all scientific disciplines to seek systematically the advice of subject-area experts on the publishability of highly technical articles.[25] In recent years a parallel to this development can be found in the rise of open software standards—and the principal advantage claimed for this approach to software development is that superior software products result from "peer review" by the users and developers of such software.[26]

Beyond the "sifting and sorting" function described above, the peer review process can also substantially improve the quality of papers that are ultimately published. In particular, conscientious referees can provide invaluable feedback to the author on revisions that will substantially improve the clarity and readability of a paper.

However, there are some things peer review cannot do.

- Peer review cannot detect fraud.

As mentioned above, proper functioning of the peer review system depends critically on the honesty of authors. Relman elaborated on this point:

> When authors say what they did and what they observed, there must be a presumption of honesty, because reviewers and editors cannot know what occurred in the laboratory. Occasionally, internal inconsistencies or implausible results might raise suspicions of malfeasance, but it is usually difficult, if not impossible, to recognize fraud solely from perusal of a manuscript. Detection of fraud or other malfeasance is the responsibility of the author's co-workers or supervisors at the institution where the work is done, not of reviewers or editors.[27 (p. 273)]

The Darsee case[28] is perhaps the most well-known example of scientific fraud to occur in recent years. Dr. John R. Darsee was a young clinical investigator in the Cardiac Research Laboratory of the Brigham and Women's Hospital (a teaching affiliate of Harvard University) who was caught fabricating data by his associates and supervisors at Harvard in May 1981. A series of investigations ensued—first at Harvard (1981–1982); then at the National Heart, Blood, and Lung Institute (1981–1982); and finally at the National Institutes of Health (1983). Ultimately Darsee was found to have fabricated research publications starting when he was a biology student at Notre Dame, continuing through his medical residency and cardiology fellowship at Emory University, and ending at Harvard. Altogether seventeen primary journal articles and fifty-three abstracts had to be retracted as a result of these investigations. Although Darsee's coauthors on the faculties of Emory and Harvard were found to have had no part in the fraud, these coauthors were unaware of Darsee's fabrications because they had little or no contact with the work that was reported in the retracted publications. Relman[29] elaborates the lessons learned from this case concerning the imperatives of responsible authorship and the limitations of the peer review system.

- Peer review cannot certify the validity of a manuscript.

There is the widespread misconception, particularly among those outside the scientific community, that passing peer review guarantees the truth of the paper's results; and then highly publicized failures like the cold fusion episode lead to questions about the value of the peer review system. Relman put this issue in the proper perspective.

> Even at its best the system can guarantee the truth of a manuscript no more than it can the honesty of an author. Rather, its function is to hold a scientific report to the best current standards, to ensure that the design and method are acceptable by those standards, and to ensure that the data are properly analyzed and interpreted. As knowledge in the field develops, new developments will improve methods and modify older concepts. Even the best current research will probably be superseded by more sophisticated and insightful work, which might reveal unsuspected limitations or flaws in previous reports.[27 (p.276)]

Clearly conscientious peer review is essential to the continued advance of science. Without it, the literature would be flooded with papers of wildly varying quality; and readers would face the overwhelming task of sifting through all this material to identify research contributions of value.

4. Problems with the Peer Review System

In recent years the peer review system has begun to show increasing signs of distress.

4.1 *Nonperformance of Editors and Reviewers*

By far the most serious problem with the peer review system for archival journal articles is simple dereliction of duty—

- by editors who refuse to take responsibility for "hard" editorial decisions, preferring to operate by majority vote of the referees; and
- by referees who cannot be bothered to read and evaluate carefully the work of other researchers, leaving editors as referees of last resort.

There is substantial anecdotal evidence that in some areas of science and engineering, this problem is growing worse over time; and this trend may be related to the increasingly intense competition for publications and funding.[30] Perhaps the most egregious failure of the peer review system in recent years was the publication of the initial paper on cold fusion by Fleischmann and Pons[6] in the *Journal of Electroanalytical Chemistry* just four weeks after submission; and there is some evidence that this paper was reviewed only by the editor of the journal and not by independent referees.[5 (pp.219–220), 17]

Of even greater consequence is the problem of nonperformance by editors. For her pioneering work on radioimmunoassay, R. S. Yalow[31] received the 1977 Nobel Prize in physiology or medicine—and yet the key paper that led to this recognition was rejected by *Science* and by the *Journal of Clinical Investigation*. The editor of the latter journal finally agreed to publish the paper in question only after extensive negotiations. Yalow made the following incisive remarks about the problem of nonperformance by editors.

> It is the editor's responsibility to determine whether the reviewer's reports or the rebuttals by the authors have more merit. In the cases cited the problem was not in the review process per se, but in the lack of editorial competence. On occasion I have written to remind editors that their office is not simply a letter drop, that theirs is the responsibility to act as judges between the reviewer's report and the author's response. In fact in my Nobel lecture (Yalow 1978), I published the initial letter of rejection from the *Journal of Clinical Investigation* of work that was to prove to be of fundamental importance to the development of radioimmunoassay. Eventually we reached a compromise with the editor, and the paper was published. I have since had the opportunity of writing to other editors who rejected our papers saying, "You may not become as famous as [the editor] in being identified in a Nobel lecture, but you are on the right track."[32 (p.244)]

4.2 *Conflicts of Interest of Reviewers*

With increasing specialization and fragmentation of disciplines, competitors within narrow subdisciplines are more frequently called upon to evaluate each other's papers and grant proposals, creating the potential for serious conflicts of interest. In particular, such conflicts of interest may tempt referees to engage in the following types of misconduct:

- misappropriation of ideas—that is, stealing ideas from the papers and grant proposals that a referee is asked to evaluate; and
- misappropriation of priority—that is, delaying or obstructing the publication or funding of a referee's rivals so that the referee can be the first to publish a result or to receive funding for work in a particular area.

Misappropriation of ideas is particularly hard to guard against since it may occur unintentionally or subconsciously. In situations where the solution to one of the referee's own research problems is found in a paper or grant proposal sent to the referee for confidential review, permission to use the ideas in question must ultimately be sought from the author and acknowledged in papers that exploit those ideas.

Goodstein gave the following analysis of the problem of conflicts of interest of reviewers:

> Peer review is not at all suited, however, to adjudicate an intense competition for scarce resources such as research funds or pages in prestigious journals. The reason is obvious enough. The referee, who is always among the few genuine experts in the field, has an obvious conflict of interest. It would take impossibly high ethical standards for referees to fail to use their privileged anonymity to their own advantage. Most scientists do hold themselves to high standards of integrity, but as time goes on, more and more referees have their ethical standards eroded by the unfair reviews they receive when they are authors. Thus the whole system is in peril.

> … Recently, as part of a talk to a large audience of mostly young researchers at an extremely prestigious university, I outlined this analysis of the crisis of peer review. The moderator, a famous senior scientist, was incredulous. He asked the audience how many disagreed with my heresy. No one responded. Then he asked how many agreed. Every hand in the house went up.[30 (p.402)]

4.3 *Inadequate Recognition and Encouragement of Innovation*

Examples abound of groundbreaking ideas that were rejected for publication or not funded because of the inherent conservatism (or lack of imagination?) of reviewers; in particular see McCutchen[33] and Horrobin.[34] Yalow summarized the gist of the problem with the following memorable remark.

> There are many problems with the peer review system. Perhaps the most significant is that the truly imaginative are not being judged by their peers. They have none![32 (p.244)]

Editors must be constantly on the lookout for highly innovative submissions, ensuring that referees of the highest quality are engaged in evaluating such work; and ultimately editors must make an informed judgment on the publishability of the work based on a careful reading of the work itself as well as the referees' reports on the work.

4.4 *Publication Bias*

In the fields of education, medicine, and psychology, there is clear evidence of publication bias reflecting the direction and strength of study results.[35] Such bias is the main cause of underreporting research discussed above. Other causes of bias in referees' reports include: jealousy; revenge; and prejudice against certain topics, individuals, or institutions.[33] It is the job of the editor to detect such referee bias and compensate for it in making editorial decisions.

4.5 *Variability of Reviewers: Assassins, Demoters, Pushovers, and Zealots*

Siegelman[36] coined the term *demoter* to describe a referee who recommends rejection of submitted papers much more frequently than the norm for "mainstream" referees, whereas a *pushover* recommends acceptance much more frequently than the norm for "mainstream" referees. Similarly, he coined the term *assassin* to denote an extreme demoter; and a *zealot* is an extreme pushover. Demoters and assassins contend that the standards for publication should be elevated since too many papers of marginal quality are currently being published; and thus demoters and assassins often provide lengthy reviews of a submitted paper that detail irreparable inadequacies in the paper's theoretical and experimental results or in the documentation of those results. By contrast, pushovers and zealots often seek to increase the visibility of their academic subdisciplines by making a concerted effort to increase the number of high-quality publications in those subdisciplines; and thus pushovers and zealots often provide lengthy reviews of a submitted paper that include helpful suggestions for revision. To avoid the potential for unfair treatment of authors, editors should carefully monitor the performance of reviewers, ensuring some balance in the types of referees assigned to each submitted paper.

5. Guidelines for Peer Review

5.1 *Archival Journal Articles*

Two of the main reasons for breakdowns in the operation of the refereeing system are
- misconceptions by referees about the job they are supposed to do; and
- misperceptions by referees about the incentives for doing a good job of refereeing, and the consequences of doing a poor job.

As Gleser[37] points out, many referees think that a manuscript must be checked line by line for errors; and seeing that this will be extremely time-consuming, they continually put off the task. On the contrary, the referee's main responsibility is to serve the editor as an "expert witness" in answering certain key questions about the manuscript—*and most of these questions can be answered under the assumption that the manuscript is error-free*. The eminent mathematician G. H. Hardy got to the nub of the matter by saying that a referee must answer three questions about a piece of research offered for publication:[38 (p.119)]

3. Is it true?
2. Is it new?
1. Is it interesting?

Although Hardy prescribed the order of these questions, they are numbered in reverse to suggest that in many cases the work required to produce a competent review is not nearly so great as it initially appears to be, since a negative answer to Question 1 (Is it interesting?) eliminates the need for investing the time and energy necessary to give correct answers to the other two questions. Truesdell provides a cogent example of a review that concisely answers all three of Hardy's questions:

> This paper, whose intent is stated in its title, gives wrong solutions to trivial problems. The basic error, however, is not new… .[39 (p.561)]

Forscher,[40] Gleser,[37] and Macrina[18 (pp.62–66)] provide comprehensive guidelines for refereeing; see also Table 2 below. Although some research suggests that little improvement in referees' reports results from providing referees with a checklist, training, or editorial feedback,[41,42] the generalizability of these conclusions is unclear.[43,44] On the other hand, there is some evidence that improvement in referees' reports may result from eliciting specific evaluations of separate, concrete aspects of the paper under review rather than a general global evaluation of the entire paper.[45] The questions in Table 2 are designed to elicit the desired specificity in referees' reports.

If a paper passes the initial screening that consists of answering Questions 1–8 in Table 2, then it is necessary to undertake the verification of technical correctness required to answer Questions 9 and 10. If competent referees had scrutinized the initial paper on cold fusion by Fleischmann and Pons[6] with the objective of answering Questions 9 and 10 in Table 2, then the serious flaws in this work would have been uncovered immediately.[5 (pp.219–220)] Finally we note that Question 11 of Table 2 requires the referee to verify that the paper includes an explicit discussion of the limitations of the study. Essentially Question 11 of Table 2 asks if the authors have adhered to Feynman's ideal of "utter honesty" and "leaning over backwards" in reporting their results. In an analysis of changes made by authors to papers that were ultimately published in the *Annals of Internal Medicine*, Purcell et al.[46] found that the most common required revision was the inclusion of information on study limitations.

Additional tips on effective refereeing are given by Waser, Price, and Grosberg.[47] A set of questions similar to those given in Table 2 can be found on the home page of the *ACM Transactions on Modeling and Computer Simulation* by visiting

<http://www.acm.org/pubs/tomacs/review/review.html>.

Table 2: Key Questions to be Answered in a Referee's Report

1. Are the problems discussed in the paper of substantial interest? Would solutions of these problems materially advance knowledge of theory, methods, or applications?
2. Does the author either solve these problems or else make a contribution toward a solution that improves substantially upon previous work?
3. Are the methods of solution new? Can the proposed solution methods be used to solve other problems of interest?
4. Does the exposition of the paper help to clarify our understanding of this area of research or application? Does the paper hold our interest and make us want to give the paper the careful reading that we give to important papers in our area of specialization?
5. Are the topic and nature of this paper appropriate for this journal? Are the abstract and introduction accessible to a general reader of this journal? Is the rest of the paper accessible to a readily identified group of readers of this journal?
6. Are the clarity and readability of the manuscript acceptable? Is the writing grammatically correct?
7. Does the manuscript contain an adequate set of references? Is adequate credit given to prior work in the field upon which the present paper is built?
8. Is the material appropriately organized into an effective mix of text, figures and tables? Are data given in tables better presented in figures or in the text?
9. Is the work technically correct? Are the main conclusions justified by the experimental data and by logically valid arguments? Are the theorems stated and proved correctly given the assumptions? In practical applications of the theoretical results, do the authors check the validity of the underlying assumptions?
10. Are there gaps in the discussion of the experimental methods or results? If there are such gaps, can the closing of these gaps be considered (i) essential, (ii) desirable, or (iii) interesting? Are the experimental methods described in sufficient detail so that other investigators can reproduce the experiments?
11. Have the authors explicitly addressed the limitations of their study?

5.2 *Grant Proposals*

Rosenzweig, Davis, and Brown[48] argue that it is much more damaging to a discipline to suppress important contributions than to fund or publish questionable research since new ideas cannot have any effect unless they are developed and publicized, while serious errors are usually detected and corrected. Thus they argue that reviewers of grant proposals should adopt the attitude that a proposal does not deserve funding unless it is daring, novel, or interesting; and in such cases, it is not sufficient simply to make brief, mildly positive statements of approval of the proposed research. Instead it is necessary to include in the review detailed answers to the following questions.

1. Why is the proposed research important?
2. What contribution will the proposed research make?
3. Why are the investigators qualified to do the work?

Of course similar specificity is required in discussing the weaknesses of a proposal; but it is highly desirable to guard against the well-known tendency for reviewers to provide much more detailed comments on the negative aspects of a proposal than on the positive aspects since such an imbalance tends to introduce an excessively conservative bias into the deliberations of review panels. Rosenzweig, Davis, and Brown summarize their recommendations thus:

> In order to convey a more accurate impression of our collective labors, we all need to make a conscious effort to tolerate diverse ideas and unconventional approaches, and to promote independence and originality. Robert Reich [former U.S. Secretary of Labor in the Clinton Administration] has written that "Technological innovation is largely a process of imagining radical alternatives to what is currently accepted." Thus it can thrive only if dissent is tolerated. In our reviews, we must encourage that dissent and emphasize the advances it will make possible.
> … The scientific enterprise … is in its very nature mutualistic and collaborative. … It is up to us … to use the peer review system carefully and wisely. Only then can it serve the goal that we all share: the rapid advancement of our discipline.[48(pp.154–155)]

Spier[49] contends that when peer review of grant applications results in sponsorship of truly innovative research, this fortunate outcome probably occurs in spite of the review process rather than because of it. He suggests that some proportion of research funding (say, 10%–20%) should be devoted to long-term support for key researchers who are identified by a radically new selection process, possibly involving randomized project selection or a lay panel that will reflect pressing social priorities. On the other hand, Chubin asserts that

> Wise stewards and editors make defensible decisions based on peer advice most of the time. If they did not, we would not still pay homage—in word and deed—to peer review.[50 (pp.111)]

To address the problems of the peer review system, Chubin calls for a more "scientific" approach to public accountability of the operation of this system. As in the case of peer review of archival journal articles, significant improvements in peer review of grant proposals will require formulation of effective guidelines for reviewers—especially in written evaluations of potentially innovative proposals.

6. Carrots and Sticks in the Peer Review System

At all levels of the scientific enterprise, it is recognized that reviewing manuscripts and grant proposals is one of the most important ways in which individual researchers can contribute to the development of their discipline. McCutchen observed that

> Reviewing of journal articles and grant applications gives reviewers the intellectual pleasure of interacting with authors and proposers, as well as education that, I suspect, has led to more advances than generally realized. These rewards are legitimate.[33 (p.158)]

Since the best referees generally receive the best papers and proposals to review, those individuals enjoy the benefits of continual professional enrichment and renewal. Moreover, high-visibility editorial positions are usually filled from the ranks of prompt and insightful reviewers; and most universities and many other research organizations regard appointment to such positions as grounds for promotion and other forms of professional advancement.

The most serious consequence of bad refereeing is the long-lasting damage to an individual's reputation in the eyes of editors and program managers, who increasingly maintain computerized records on the performance of reviewers.[51] Some editors even go so far as to maintain two lists of referees, say the "A" list of good referees and the "B" list of bad referees; and when authors from either list submit a paper for review, the editor selects referees for that paper from the list to which the author belongs.

To further enhance the incentives for good reviewing, editors should provide timely feedback to referees on (a) the strengths and weaknesses of their reviews, and (b) the issues identified in other referees' reports on the same paper. As a professional courtesy, editors should include such feedback with their letters of appreciation to referees. For major journals with high rates of submission, some selectivity may be required to make this suggestion practical; in particular, editors might provide detailed editorial feedback to referees only in cases for which (i) the paper is judged to be a major contribution to the literature, or (ii) the editorial decision-making process is particularly difficult. Moreover, editors should strive to ensure that individuals who provide prompt and thorough refereeing will receive comparable service when those individuals submit their own papers for review.

7. Conclusions

In essence this article's central thesis is simply this: the proper functioning and continued advancement of the scientific enterprise depends critically on individual scientists living up to the standards of ethical conduct so memorably articulated by Feynman—not only in the design, execution, and documentation of their research projects, but also in their response to the challenges of responsible, professional peer review.

Acknowledgments: This work was supported by NSF Grant SES-9818359. Special thanks go to Tom Regan, Rebeca Rufty, Margaret King, and Nell Kriesberg (North Carolina State University) for many enlightening conversations on research ethics. Thanks also go to Stephanie Bird (MIT), Frank Davidoff (Executive Editor, Institute for Healthcare Improvement; Editor Emeritus, *Annals of Internal Medicine*), and the anonymous referees for numerous suggestions that substantially improved this paper.

REFERENCES

1. Feynman, Richard P. (1985) *"Surely You're Joking, Mr. Feynman!": Adventures of a Curious Character*, W. W. Norton & Co., New York.
2. Langmuir, Irving, and Hall, Robert N. (1989) Pathological science, *Physics Today* **42**: 36–48.
3. Nye, Mary Jo (1980) N-rays: An episode in the history and psychology of science, *Historical Studies in the Physical Sciences* **11**: 127–156.
4. Wood, Robert W. (1904) The *n*-rays, *Nature* **70**: 530–531.
5. Huizenga, John R. (1993) *Cold Fusion: The Scientific Fiasco of the Century*, Oxford University Press, New York.
6. Fleischmann, Martin, and Pons, Stanley (1989a) Electrochemically induced nuclear fusion of deuterium, *Journal of Electroanalytical Chemistry* **261**: 301–308.
7. Fleischmann, Martin, and Pons, Stanley (1989b) Errata, *Journal of Electroanalytical Chemistry* **263**: 187–188.
8. Glass, B. (1965) The ethical basis of science, *Science* **150**: 1254–1261.
9. Bronowski, J. (1965) *Science and Human Values*, rev. ed., Harper & Row, New York.
10. CBE Style Manual Committee (1983) *CBE Style Manual: A Guide for Authors, Editors, and Publishers in the Biological Sciences*, 5th ed. rev. and expanded, Council of Biology Editors, Bethesda, Md.
11. CBE Style Manual Committee (1994) *Scientific Style and Format: The CBE Manual for Authors, Editors, and Publishers*, 6th ed., Cambridge University Press, Cambridge.
12. International Committee of Medical Journal Editors (1997) Uniform requirements for manuscripts submitted to biomedical journals, *Journal of the American Medical Association* **277**: 927–934.
13. International Committee of Medical Journal Editors (2001) Uniform requirements for manuscripts submitted to biomedical journals [online], available online via <www.icmje.org> [accessed February 10, 2002].

14. Rennie, Drummond, Yank, Veronica, and Emanuel, Linda (1997) When authorship fails: A proposal to make contributors accountable, *Journal of the American Medical Association* **278**: 579–585.
15. Houk, V. N., and Thacker, S. B. (1990) The responsibilities of authorship, in: CBE Editorial Policy Committee, eds. *Ethics and Policy in Scientific Publication,* Council of Biology Editors, Bethesda, Md., pp. 181–184.
16. Davidoff, Frank (2000) Who's the author: Problems with biomedical authorship, and some possible solutions, *Science Editor* **23**: 111–119.
17. Taylor, Craig (1999) The cold fusion debacle, presented at Research Ethics Institute, 13–16 June, at North Carolina State University, Raleigh, North Carolina.
18. Macrina, Francis L. (2000) *Scientific Integrity: An Introductory Text with Cases*, 2d ed., ASM Press, Washington, D.C.
19. Wilson, James R. (2001) Some guidelines on technical writing [online], Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina, available as <http://www.ie.ncsu.edu/jwilson/guide.html> [accessed February 26, 2002].
20. Chalmers, I. (1991) Underreporting research is scientific misconduct, in: *Peer Review in Scientific Publishing: Papers from the First International Congress on Peer Review in Biomedical Publication*, Council of Biology Editors, Chicago, pp. 169–177, also available as: Chalmers, I. (1990) Underreporting research is scientific misconduct, *Journal of the American Medical Association* **263**: 1405–1408.
21. Woodward, James, and Goodstein, David (1996) Conduct, misconduct and the structure of science, *American Scientist* **84**: 479–490.
22. Medawar, Peter B. (1979) *Advice to a Young Scientist,* BasicBooks, New York.
23. Wilson, James R. (1997) Doctoral colloquium keynote address: Conduct, misconduct, and cargo cult science, in: Andradóttir, S., Healy, K.J., Withers, D. H., and Nelson, B. L., eds. *Proceedings of the 1997 Winter Simulation Conference*, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, pp. 1405–1413.
    Available via <www.informs-cs.org/wsc97papers/1405.PDF> [accessed February 26, 2002].
24. Kronick, D. A. (1991) Peer review in 18th century scientific journalism, in: *Peer Review in Scientific Publishing: Papers from the First International Congress on Peer Review in Biomedical Publication*, Council of Biology Editors, Chicago, pp. 5–8, also available as: Kronick, D. A. (1990) Peer review in 18th century scientific journalism, *Journal of the American Medical Association* **263**: 1321–1322.
25. Burnham, J. C. (1991) The evolution of editorial peer review, in: *Peer Review in Scientific Publishing*: *Papers from the First International Congress on Peer Review in Biomedical Publication*, Council of Biology Editors, Chicago, pp. 9–26, also available as: Burnham, J. C. (1990) The evolution of editorial peer review, *Journal of the American Medical Association* **263**: 1323–1329.
26. Raymond, E. S. (2000) The cathedral and the bazaar [online], available on the web via <www.tuxedo.org/~esr/writings/cathedral-bazaar/cathedral-bazaar> [accessed February 26, 2002].
27. Relman, A. S. (1990) The value of peer review, in: CBE Editorial Policy Committee, eds., *Ethics and Policy in Scientific Publication*, Council of Biology Editors, Bethesda, Md., pp. 272–277.
28. Culliton, B. J. (1983) Coping with fraud: The Darsee case, *Science* **220**: 31–35.
29. Relman, A. S. (1983) Lessons from the Darsee affair. *The New England Journal of Medicine* **308**: 1415–1417.
30. Goodstein, D. (1995) Peer review after the big crunch, *American Scientist* **83**: 401–402.
31. Yalow, R. S. (1978) Radioimmunoassay: A probe for the fine structure of biologic systems, *Science* **200**: 1236–1245.
32. Yalow, R. S. (1982) Competency testing for reviewers and editors, *The Behavioral and Brain Sciences* **5**: 244–245.

33. McCutchen, C. W. (1997) Peer review: Treacherous servant, disastrous master, in: Elliott, D. E., and Stern, J. E., eds. *Research Ethics: A Reader*, University Press of New England for the Institute for the Study of Applied and Professional Ethics at Dartmouth College, Hanover N.H., pp. 151–164, also available as: McCutchen, C. W. (1991) Peer review: Treacherous servant, disastrous master, *Technology Review* **94**: 27–40.
34. Horrobin, D. F. (1991) The philosophical basis of peer review and the suppression of innovation, in: *Peer Review in Scientific Publishing: Papers from the First International Congress on Peer Review in Biomedical Publication*, Council of Biology Editors, Chicago, pp. 250–259, also available as: Horrobin, D. F. (1990) The philosophical basis of peer review and the suppression of innovation, *Journal of the American Medical Association* **263**: 1438–1441.
35. Dickersin, K. (1991) The existence of publication bias and risk factors for its occurrence, in: *Peer Review in Scientific Publishing: Papers from the First International Congress on Peer Review in Biomedical Publication*, Council of Biology Editors, Chicago, pp. 92–104, also available as: Dickersin, K. (1990) The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association* **263**: 1385–1389.
36. Siegelman, Stanley S. (1991) Assassins and zealots: Variations in peer review, *Radiology* **178**: 637–642.
37. Gleser, Leon J. (1986) Some notes on refereeing, *The American Statistician* **40**: 310–312.
38. Halmos, P. R. (1985) *I Want to Be a Mathematician*, Springer-Verlag, New York.
39. Truesdell, C. (1951) Review of "Equations of finite vibratory motions in isotropic elastic media. Surface force sufficient to maintain equilibrium," by García, G. (1950) *Actas Acad. Ci. Lima* **13**: 29–38, *Mathematical Reviews* **12**: 561.
40. Forscher, Bernard K. (1965) Rules for referees. *Science* **150**: 319–321.
41. Callaham, M. L., Wears, R. L., and Waeckerle, J. F. (1998) Effect of attendance at a training session on peer reviewer quality and performance, *Annals of Emergency Medicine* **32**: 318–322.
42. Callaham, M. L., Knopp, R. K., and Gallagher, E. J. (2002) Effect of written feedback by editors on quality of reviews: Two randomized trials, *Journal of the American Medical Association* to appear.
43. Jefferson, T., Alderson, P., Wager, E., and Davidoff, F. (2002) The effects of editorial peer review: A systematic review, *Journal of the American Medical Association* to appear.
44. Jefferson, T., Wager, E., and Davidoff, F. (2002) Measuring the quality of editorial peer review, *Journal of the American Medical Association* to appear.
45. Strayhorn, J., McDermott, J. F., and Tanguay, P. (1993) An intervention to improve the reliability of manuscript reviews for the *Journal of the American Academy of Child and Adolescent Psychiatry*, *American Journal of Psychiatry* **150**: 947–952.
46. Purcell, Gretchen P., Donovan, Shannon L., and Davidoff, Frank (1998) Changes to manuscripts during the editorial process, *Journal of the American Medical Association* **280:** 227–228.
47. Waser, Nickolas M., Price, Mary V., and Grosberg, Richard K. (1992) Writing an effective manuscript review, *BioScience* **42**: 621–623.
48. Rosenzweig, M. L., Davis, J. I., and Brown, J. H. (1988) How to write an influential review, *Bulletin of the Ecological Society of America* **69**: 152–155.
49. Spier, Raymond E. (2002) Peer review and innovation, *Science and Engineering Ethics* **8**: 99–108.
50. Chubin, Daryl E. (2002) Much ado about peer review, Part 2, *Science and Engineering Ethics* **8**: 109–112.
51. Abelson, P. H. (1992) Integrity of the research process, *Science* **256**: 1257.