Harriet Zuckerman and Robert K. Merton

Sociology of refereeing

Is the fate of a paper submitted to Physical Review unfairly affected by the standing of its author or referee? Two sociologists examine the evidence.

Nearly all scholarly journals use referees to screen submitted manuscripts. Physical scientists recognize the significance of the referee system: Some defend the system, and others attack it. But refereeing itself has not been systematically assessed. By studying the archives of *The Physical Review* for the years 1948 to 1956 (before the separate publication of *Physical Review Letters*) we have been able to come to a few conclusions about the workings of the referee system. Although some of the results were expected, others are surprising. Younger physicists, for example, are more likely to have their papers accepted than are older physicists, and the physics "establishment" does not appear to have any bias toward publishing the papers of its own members. The referee system here apparently does what it is supposed to do: Sift out the good papers from the bad.

Two types of journals

A preliminary survey of disciplines ranging from history and philosophy through the physical sciences showed us a pattern for rejection: The more humanistically oriented a journal, the higher the percentage of manuscripts its editors reject; the more experimentally and observationally oriented a journal, the lower the percentage rejected. These variations probably reflect the varying extent of agreement on standards of scholarship in the different disciplines.

A high rejection rate, we found, does not imply harsh refereeing of marginal papers. Journals with high rejection rates apparently receive a large proportion of manuscripts that the editor and his referees consider not merely borderline cases but papers that fail to meet even minimal standards of scholarship. The editor of a journal that rejects nine out of ten papers notes that he turns down about four of these nine as hopelessly inept and unpublishable in any journal. The editor of another journal that rejects about 85% of all submitted papers reports that about 20% "were so clearly unacceptable that I didn't want to waste a referee's time with them . . . We still get a flow of articles of a thoroughly amateurish quality."

These differences are important because they influence the attitudes of journal editors toward borderline papers. Editors and referees, of course, want to avoid errors of judgment altogether. But recognizing that they are not infallible, editors of different kinds of journals prefer different kinds of mistakes. The editorial staff of high-rejection journals evidently prefer to risk rejecting manuscripts that the wider community of scholars (or posterity) would consider publishable rather than risk publishing papers that will be widely judged as substandard. The editorial staff of low-rejection journals apparently prefer to risk the occasional publication of papers that do not mea-

Harriet Zuckerman and Robert K. Merton are in the sociology department at Columbia University, where Zuckerman is an assistant professor and Merton is Giddings Professor.

sure up rather than to overlook work that may turn out to be significant.

The Physical Review is a low-rejection journal that physicists like to read[1] and to publish in.[2] How have its editors and referees made their selections? And how does the social stratification system within physics affect the process?

Modeling the system

During the nine years from 1948 to 1956 *The Physical Review* received 14 512 manuscripts. The editor-in-chief of The American Physical Society, Samuel A. Goudsmit, gave us access to the correspondence between authors, editors and referees, as well as to records of editorial decisions on allocating the manuscripts to referees, referees' evaluations and final actions on all the papers.

For our analysis we considered physicists (as all scientists) to form a "status hierarchy" (the term comes from the sociologist Max Weber) based on honor and esteem. Rank and authority are acquired through performance but, once acquired, tend to be ascribed for an indeterminate time. This combination of acquired and ascribed status brings strains into the process of judging scientific papers: Judgments *by* scientific authorities, whose status rests largely on their own past performance, may be given more weight than their intellectual cogency merits. And judgments *about* the work of ranking scientists may be affected by the judge's doubts of his competence to criticize a great man or by his fear of affronting an influential person. The hierarchy of excellence then can work in two ways against the ideal of an unbiased evaluation of scientific work.

Keeping this stratification system in mind, we divided the contributors to *The Physical Review* between 1948 and 1956 into three ranks. In the first rank are the 91 physicists who by 1956 had received at least one of the ten most respected awards in physics, such as the Nobel Prize, membership in the Royal Society or membership in the US National Academy of Sciences.[3] The second rank is formed by those 583 physicists important enough to be included in a biographical and bibliographical survey by the American Institute of Physics Center for the History and Philosophy of Physics. The remaining 8864 contributors form the third rank.

To make our analysis most clear cut and useful we have included only those papers with a single author, about 80% of which were eventually accepted. (Multiple-author papers, often reports of experimental data, are accepted about 95% of the time, so that there is little variation of the kind that interests us). This restriction gives us 55 physicists from the first rank who submitted papers and 343 from the intermediate rank. The remaining 8864 are represented by a sample of 659 authors. The 354 referees who judged the single-author manuscripts have been stratified in the same way, with 12% in the first rank, 35% in the second and 53% in the third.

We looked for answers to four questions: Do contributors located in various parts of the stratification system submit papers at different rates? Are there patterns of allocating manuscripts to referees at particular levels in the hierarchy and are these allocations related to the status of the authors? Are differences in the fraction of papers accepted based on the professional standing of the physicists submitting the manuscripts? Finally, are any such differences linked to the relative standings of author and referee?

Submission rate and status

Eminent scientists publish not only better papers but also more papers than run-of-the-mill scientists. We are not surprised to find that they also submit more manuscripts for publication. During the period we studied, highest-rank physicists submitted an average of 4.1 single-author papers to *The Physical Review*, intermediate-rank physicists averaged 3.5 and third-rank physicists 2.0. (These differences are, of course, amplified if multiple-author papers are included; higher-rank physicists have greater opportunities for collaboration than lower-rank ones. And, in the general population of physicists, the differences would probably be greater than in this self-selected population of paper writers.)

How does the ratio of submitted to published papers vary for the different strata? If the ratios were all the same, we would conclude that all would-be contributors have the same standards and competence and that the refereeing process results in uniform rates of acceptance for scientists at all levels of the stratification system. We get our first intimation that this is not so when we note the rates of submission and acceptance for papers from the 17 foremost university physics departments[4] and from less distinguished departments. Although the submission rates are nearly equal, about 91% of the papers by physicists in the foremost departments were accepted, as against 72% from other universities.

One interpretation of the departmental differences is that the work of scientists in the upper strata is evaluated less severely and that these authors are given the benefit of the doubt by editors and referees. Another interpretation ascribes the different outcomes of the evaluation process to differences in the quality of the manuscripts. The same standards are, according to this interpretation, rather uniformly applied, but, on the average, scientists in the better departments tend to be more capable, have greater resources for investigation and more demanding internal standards, and are more likely to have their papers examined by competent colleagues before sending the papers in for publication. These opposing interpretations are not, of course, mutually exclusive; our job is to disentangle the components. Suppose, for example, that all papers submitted by Nobel laureates are accepted. Would some of these have been rejected if submitted by scientists of distinctly lower standing?

These difficulties of analysis would be largely avoided if authors were altogether anonymous to referees. But arrangements designed to assure anonymity work imperfectly. As Goudsmit has said:

> "Removing the name and affiliation of the author does not make a manuscript anonymous. A competent reviewer can tell at a glance where the work was done and by whom or under whose guidance . . . One must also remove all references to previous work by the same author, all descriptions of special equipment and other significant parts of the paper. Nothing worth judging or publishing would be left."[5]

We look to the archives of *The Physical Review* for evidence of particular evaluation patterns.

Who judges the papers?

We observed first that, quite clearly, higher-ranking authors have a greater proportion of their papers judged only by one or both of the two editors and not by outside referees. About 87% of the papers submitted by physicists of the highest rank were judged only by the editors, in contrast to 73% of the papers from the intermediate rank and 58% of the remaining papers. An immediate consequence is seen in Table 1: the higher the rank of the physicist, the more prompt the decision taken on his manuscript. As we shall see, the more problematic papers are the ones that are sent to referees. Who are these referees?

Manuscripts should be evaluated by experts in a particular subject. We are not surprised that the outside referees were drawn disproportionately from the high-ranking group. Nearly 12% of the 354 outside referees were from this group (compared with only 5% of the authors), and this 12% contributed one third of all referee judgments. We can classify the referees by other means as well, and their composite portrait is clear. Whether gauged by their own prestige, institutional affiliation or research accomplishments, they are largely drawn from the scientific "elite," as we expect from the principle of expertise.

Physicists on the rise

Our curiosity led us to identify a mobile subgroup of physicists: the 49 contributors who were in the intermediate rank during the time of the study but who later moved into the highest rank. They were observed in the course of their ascent, and we found a striking prognostic result in the pattern of submission rates. The 49 mobile physicists were the most prolific in the sample. A whopping 47% of them submitted as many as 15 papers (single and multiple authors) as compared with the 18% of the highest rank, 11% of all the intermediates and 1.5% of the third rank. Plainly these were physicists at the *peak of their productivity*. Six of them have since become Nobel laureates, and we catch here, as with a camera, a phase in the process through which early productivity is converted into a later recognition by the social system of science.

Table 1. Duration of Editorial and Refereeing Process

Duration	Rank of author: Highest	Intermediate	Third
Less than 2 months	42%	35%	29%
2–4 months	47	45	41
More than 5 months	11	20	30
Total number of papers	(202)	(1027)	(972)

We can describe the allocation of referees to authors in terms of four simplified models—"oligarchical," "populist" and "egalitarian" models and the model of expertise. In the oligarchical model, the established elite alone can judge the work of those beneath it in the status hierarchy. The second model assigns power of judgment to "the people," in this case to lower-ranking physicists only. The strictly egalitarian model calls for assessment of papers only by juries of "status peers." And the model of expertise calls for the allocation of manuscripts to referees who, regardless of rank, are particularly competent to judge them.

As we look at Table 2 we see that the actual patterns of allocation are inconsistent with all models other than the model of expertise. If we assume that demonstrated expertise is substantially (if imperfectly) correlated with rank, we would expect a preponderance, but no monopoly, of referees who outrank authors. Authors would occasionally outrank referees in prestige (if not in competence), and judgment by peers would be more frequent for the higher ranks of authors. In the table we see that judgments by peers account for half of the papers by top-ranking physicists, 41% of the papers by intermediates and 20% of the papers of the rank-and-file.

The results suggest that competence and expertise were the principal factors in matching papers with referees. But whatever the criteria, we now know that the more highly placed physicists had power disproportionate to their number. How did they act in their positions of power?

Status and acceptance rates

Ethics and practicality both rule out the draconian experiment in which matched samples of referees, all unknowing, would independently judge the same manuscripts, which would be ascribed to physicists of varying ranks. At best we can bring together data that cumulatively intimate the extent to which judgments by editors and referees are affected by the author's status. The flow chart on page 33 summarizes the refereeing process. We see that 90% of the manuscripts submitted by top-ranking physicists have been accepted, compared with 86% for the intermediates and 73% for the rank-and-file.

These stratified rates result from a continuing process of evaluation, and in each phase of the process higher-ranking physicists fare better than lower-ranking ones. Looking at the flow chart we see that a larger proportion of their papers are accepted immediately, a smaller proportion rejected immediately and a smaller proportion considered problematic. And problematic papers by these higher-ranking physicists are more likely to get into print than are problematic papers by the lower-ranking physicists.

Before we look at other evidence to help us interpret these data, we recall that *The Physical Review* is a "high-acceptance" journal, and that high-acceptance journals apparently follow the rule "when in doubt, accept." There are several expressions of this rule in the behavior of *The Physical Review's* editors. The ratio of immediate acceptances to later (more problematic) ones was over four to one (65% to 15%) compared with a 1.5-to-one ratio (12% to 8%) for rejections. Among the papers needing further evaluation, the ratio of acceptances to rejections is still 1.7 to one. More judges were used, on the average, for rejected papers than for those ultimately published, indicating that the journal mobilized more institutional machinery to reject papers than to accept them. For high-accep-

Table 2. Rank of Referees Assigned to Authors of Different Ranks

	Rank of referees			
Rank of authors	Highest	Intermediate	Third	Total judgments by referees
Highest	50%	31	19	(36)
Intermediate	38%	41	21	(394)
Third	27%	46	26	(653)
Total				(1083)

tance journals such as *The Physical Review*, the potentially unacceptable papers are apparently the most troublesome ones.

We take another step toward gauging the influence of the author's rank on the fate of his paper when we remember that scientific eminence and authority derive largely from past and not necessarily continuing accomplishments. In science then, as in other institutions, older men tend to have the power and authority. (This gerontocracy, some have said, may even be a good thing for science; it leaves the young productive scientists free to get on with their work and helps to occupy the time of those who are no longer creative.) If the sheer power and eminence of authors greatly affect refereeing decisions, then the older eminent scientists should have the highest rates of acceptance.

But in physics, the young man's science, we find the opposite. The younger rather than the older authors are more likely to have their papers accepted, and these age-graded acceptance rates hold true within each level of the hierarchy. In Table 3 we see that both eminence and youth contribute to a manuscript's chance for acceptance. Youth in fact is so important that the youngest group of third-rank physicists has as high an acceptance rate as the oldest group of high-rank physicists whose work, we suppose, is no longer as good as it once was. There comes a time in the life cycle of even the most distinguished physicist when he can no longer count on having his papers accepted in a major refereed journal such as *The Physical Review*.

Effect of relative ranks

The relative ranks of author and referee may be the important influence in the referee's decision. Self-explanatory names describe the possible forms of such bias: If referees and author are status peers, the referee could feel either *status solidarity* or *status competition*. If authors outrank referees, either *status deference* or *status envy* could be important. If referees outrank authors, their bias might take the form of *status patronage* or *status subordination*. All these forms of bias would result in rates of acceptance for each stratum of authors that differ according to the rank of the referee.

We see, however, in Table 4 that the data do not agree with any of the hypotheses. Referees of each rank accept the same proportion of papers from every stratum. The highest-ranking referees accept somewhat smaller proportions of papers, but they do this for every stratum. There is, in short, no preferential pattern. We come to a similar conclusion when we compare the decisions of referees about papers from major and minor university departments. Finally the rate at which referees accept papers by authors from varying academic ranks is unaffected by the academic rank of the referee.

The accumulated data suggest that referees apply the same standards to papers, whatever the source of the paper. But, agreeing that the refereeing process is impartial, we can still ask: Is refereeing needed?

How useful is refereeing?

The Physical Review, as the prime journal in its field, presumably applies exacting standards. All the same, the refereeing process results in about four out of five submitted manuscripts being accepted (some after revision) for publication. Does this mean that referees are superfluous? Like other observers of the referee system,[6] we think not. Referees, collectively engaged in sorting out good physics from bad, help editors, authors, other referees and the community of physicists.

Referees are most helpful to editors for papers that are difficult to assess; not all manuscripts exhibit neatly correlated arrays of intellectual virtues or vices. Apart from their expert judgment, the typically anonymous corps of referees can, more or less incidentally, protect the highly visible editor from the wrath of disappointed authors.[7]

The interests of authors and referees are, of course, not inherently opposed. Conscientious referees can and do suggest ways to improve papers. They sometimes link the paper with other work, they protect the author from unknowingly publishing duplications of earlier work, and as presumable experts in a subject they certify the paper as a contribution by recommending its publication. Not all referees, of course, are uniformly conscientious, but the problem of who judges the judges is not unique here.

Their role as referee helps the referees themselves, and sometimes creates problems for them. Particularly in fields without efficient networks of informal communication, or in rapidly growing fields, referees occasionally get a head start in learning about significant new work. Some referees report that, because they must scrutinize the papers so closely (contrasted with the often perfunctory scanning of articles already in print), they perceive potentialities for new lines of inquiry that were neither stated by the author nor previously considered by the referee. This unplanned evocative function of the paper often puts both referee and author under stress: What the referee defines as legitimate and appreciative borrowing or learning, the author may define as pilfering or downright plundering.[8] This sort of concern with intellectual property[9] is not new; Thomas Huxley, for example, expressed the same fears during the last century.[10]

The referee system becomes more important, Michael Polanyi has noted,[11] as science differentiates into extensive networks of specialities. The more specialized the paper, the more difficult it is for many to assess its worth. But although only a few may be competent to assess, many more may find the paper relevant to their work, and here the referee as deputy is particularly important. Scientists who use published results in fields tangential to their own rely on the referee system.

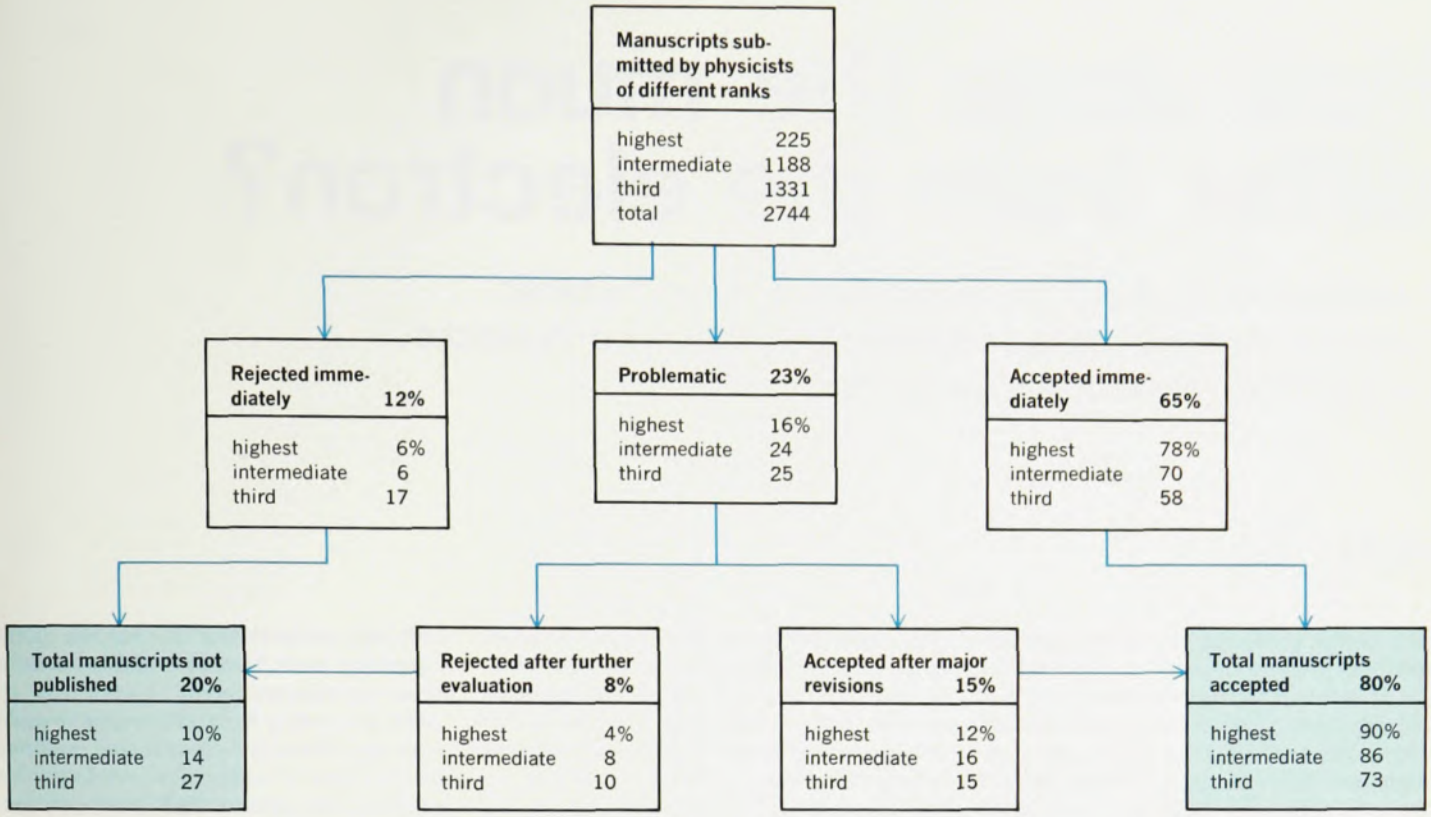
Journal editors are sometimes puzzled by scientists' willingness to serve in the anonymous and exacting role of referee. Participation in the referee system is widely diffused; almost 30% of a sample of high-energy theorists, for example, had acted as referees or done some editorial work for journals.[12] A sense of collective responsibility is probably one motive, and especially for young participants there is the symbolic reward of being thought expert enough to serve as a referee.

Simon Pasternack, the present editor of *Physical Review*, has suggested that the very existence of the referee system serves as a control by anticipation. Knowing that their papers will be reviewed, authors take care in preparing them, and often the journal's high standards become their own. He also points out that even the "journals

Table 3. Acceptance of Manuscripts by Age and Rank of Author

	Rank of authors							
Age of authors	HIGHEST %	Number	INTERMEDIATE %	Number	THIRD %	Number	ALL RANKS %	Number
20–29			91	287	83	385	87	672
30–39	96	80	89	519	77	440	85	1039
40–49	95	58	83	236	73	79	83	373
More than 50	80	87	71	126	50	14	73	227
No information							61	423
All ages							80	2734

Evaluation of single-author manuscripts



that have little or no refereeing or editing . . . exist within a framework of the edited journals, which set the pattern . . ."[13]

Our remarks here on the functions of the referee system do not at all imply that it works perfectly. But although errors of judgment occur, the system of monitoring work before it enters the archives permits physicists to build upon the work of others with some degree of confidence.

* * *

This article was adapted and condensed by physics today *from an article that appeared in the January 1971 issue of* Minerva, *pages 66–100. The research reported here was supported by the National Science Foundation and made possible by the cooperation of Samuel Goudsmit and of Charles Weiner and his staff at the AIP Center for the History and Philosophy of Physics.*

References

1. S. Cole, J. R. Cole, "Visibility and the Structural Bases of Awareness of Scientific Research," Am. Sociological Rev. **33**, 412 (1968).
2. S. Kennan, F. G. Brickwedde, *Journal Literature Covered by Physics Abstracts in 1965*, AIP Report 68-1, New York (1968) Appendix 2; M. M. Kessler, *Technical Information Flow Patterns*, Lincoln Laboratories, MIT, Cambridge, Mass, page 247, 249 (1957); M. M. Kessler, *physics today*, March 1965, page 30.
3. S. Cole, J. R. Cole, "Scientific Output and Recognition: A Study in the Operation of the Reward System in Science," Am. Sociological Rev. **32**, 383 (1967).
4. H. Keniston, *Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania*, U. of Pa. Press, Philadelphia (1959).
5. S. A. Goudsmit, *physics today*, January 1967, page 12.
6. M. Polanyi, *Personal Knowledge*, Routledge & Kegan Paul, London (1958), chapter 6; J. M. Ziman, *Public Knowledge: The Social Dimension of Science*, Cambridge U. P., Cambridge, UK (1966) page 111; N. Storer, *The Social System of Science*, Holt, Rinehart and Winston, New York (1966) page 112; W. O. Hagstrom, *The Scientific Community*, Basic Books, New York (1965) page 18.
7. S. Raffel, *Abstracts* of the American Sociological Association (1968).
8. A. G. Prinz, *physics today*, August 1970, page 11.
9. R. K. Merton "Priorities in Scientific Discovery," Am. Sociological Rev. **22**, 635 (1957).
10. L. Huxley, *Life and Letters of Thomas Henry Huxley*, Volume 1, Macmillan, London (1900), page 97.
11. M. Polanyi, reference 7, page 163.
12. M. A. Libbey, G. Zaltman, *The Role and Distribution of Written Informal Communications in Theoretical High Energy Physics*, AIP/SDD-1, 49 (1967).
13. S. Pasternack, *physics today*, May 1966, page 40. □

Table 4. Referees' Decisions to Accept Manuscripts

	Rank of referees							
	HIGHEST		INTERMEDIATE		THIRD		TOTAL JUDGMENTS BY REFEREES	
Rank of authors	%	Number	%	Number	%	Number	%	Number
Highest	[too small to analyze]						50	36
Intermediate	55	150	62	160	62	84	59	394
Third	54	179	61	302	59	172	59	653
All ranks							59	1083