Letters

The Peer Review Question

"Chance and consensus in peer review" by S. Cole, J. R. Cole, and G. A. Simon (20 Nov., p. 881), reports on a portion of a study initiated by the National Science Foundation (NSF) while I was director and carried out by the Committee on Science and Public Policy (COSPUP) of the National Academy of Sciences. Overall, peer review has served us well. But it is not perfect, and we should strive for new insights—such as those provided by COSPUP and the Coles—that lead to improvements. Indeed, there have been many changes in recent years to refine and improve peer review.

From an experiment based on 50 proposals submitted to each of three NSF programs, the authors conclude that about 25 percent of NSF funding decisions would be reversed if the proposals were rerated by other, equally qualified, reviewers. Their conclusion is based on comparisons of rating quintiles for each set of 50 proposals. The authors point out that the reversals were probably due to "real and legitimate differences of opinion among experts about what good science is or should be." They do not observe that, if the three programs are grouped together (to improve the statistical base), of the decisions in the top plus the bottom quintile, 87 percent were unchanged. For the second and fourth quintiles, the degree of concurrence drops from 9 out of 10 to 7 out of 10. We are entitled to wonder whether any system of rating could do better than that. As observed by Cole, Cole, and Simon, the fact that concurrence is lost in the middle quintile—at the margin—cannot be regarded with surprise.

Nor should we be surprised that there are differences of opinion about highly original research that departs from existing approaches and traditions. In addition to their numerical evaluations of proposals, reviewers submit written narratives containing detailed comments and criticisms. It is the role of NSF to review these comments, in addition to the ratings, and to reach funding decisions only after careful analysis and judgment; decisions are not based on numerical averages. Often, the decision to fund research depends upon additional reviews aimed at issues identified in the initial reviews and further discussion and negotiation of changes with the principal investigators. It would have been more appropriate to compare NSF decisions with COSPUP decisions reached after a knowledgeable scientist had analyzed, questioned, and integrated the various individual peer reviews.

One also should recognize that NSF procedures today are in many ways different from those used when the grants studied by the Coles were evaluated. For example:

- Verbatim anonymous reviews are now given to principal investigators.
- There is a formal appeal process for reconsideration of declined proposals.
- Specific guidelines for managing the peer review process are provided to program officers.
- External oversight by knowledgeable scientists of program officer decisions is a requirement.
- An Office of Audit and Oversight has been established.
- More consideration is given to recent scientific accomplishments of the principal investigator.

Particularly important are the opportunities to request reconsideration of declined proposals and to review anonymous copies of verbatim evaluations. These changes are designed to help the proposer understand better the basis on which decisions were made and to challenge decisions. Another highly significant change is the requirement that reviewers consider explicitly in their review the capability and creativity of the principal investigator as evidenced by recent accomplishments as well as the scientific quality and importance of the proposed research (other evidence of capability is examined in the case of younger scientists). As noted by Cole *et al.*, the importance of a research contribution is easier to judge in retrospect than in advance.

It is unfortunate that the article and a subsequent press conference elicited a press response that emphasized "luck" as an element in the decision process and that more cogent points were missed.

My reason for commenting on the article is not to diminish the conclusions of the study. Rather, I want to show that peer review itself is continually reviewed and modified. I regard this as a healthy process, consistent with the practice of science. I, and others at NSF, welcome studies such as this and look forward to a continuing debate over ways to improve the grant selection process.

RICHARD C. ATKINSON
Office of the Chancellor,
University of California,
San Diego, La Jolla 92093

As one of those responsible for the National Academy of Sciences' study on peer review (*1*), I was disappointed in the *Science* article's treatment of part of our work. I believe it misrepresented our results and is needlessly provocative.

I read the same data differently. Consider table 2 for example. Each quintile consists of only 10 proposals, a rather small sample. If we combine the data from the three fields, we find 4.7 reversals out of 30 proposals when we compare COSPUP's mean rating with NSF's decisions in the first quintile. That is, in the top quintile the reversal rate is 16 percent. In the next quintile the reversal rate is 31 percent; in the middle it is 44 percent; in the fourth quintile, 30 percent; and in the fifth quintile, 10 percent (*1*, p. 31, table 7).

These percentages will evoke varying degrees of surprise, but the phrase "luck of the reviewer draw" is inappropriate. The reversal rates do call for a closer look at causes of reviewer disagreement. We are dealing here with the uncertainties of predicting which among a group of high-quality proposals will lead to the very best science. Original ideas, breaking with past traditions, are bound to be evaluated differently by experts with divergent views of the most promising directions for future work. Furthermore, quite a number of reversals are inevitable when limited funding makes close calls necessary.

Perhaps because I prepared the final draft of the relevant sections, I believe a more balanced discussion of this issue can be found in our original report. Moreover, the concerns of Phase Two are not limited to funding reversals. We would particularly like to call attention to our observations and recommendations for peer review in section 4. The National Science Foundation, partly in-

dependently and partly in response to these recommendations, has instituted a number of procedural changes including extensive reconsideration of declined proposals.

I. M. SINGER*
Department of Mathematics, University of California, Berkeley 94720

References and Notes

1. *Peer Review in the National Science Foundation: Phase Two of a Study* (National Academy of Sciences, Washington, D.C., 1981). The authors of the study are J. R. Cole and S. Cole, with the Committee on Science and Public Policy (COSPUP), not J. R. Cole, S. Cole, as in reference 3 of the 20 November *Science* article.

* Past Chairman, COSPUP, 1975–1980.

Regarding peer review, I question the interpretation given to the study (*1*) of which I was a coauthor with Cole and Cole. I reproduce one of three figures from the main report (Fig. 1); the overall impression is one of substantial agreement of the sets of reviews, disregarding two undependable data points. The corresponding figure of Cole *et al.* (p. 883) suggests greater disagreement because they chose to plot ranks and not to circle the most doubtful points. When ranks are plotted, many near-ties are displayed as sizable separations.

The statistics on variation are best evaluated by assuming an average of four reviewers per proposal, which is close to the NSF norm. On a scale that assigns 10 points to the difference between *Very good* and *Good*, the estimated error standard deviations are 3.8, 3.5, and 4.8 for the three areas of research, the largest being for economics. These bottom-line estimates are far less dramatic than the intermediate calculations stressed by Cole *et al.* Their last analysis compares 2 standard deviations whose ratio would be 1.00 if chance alone determined ratings. By their method I reach a ratio of 1.6 for an average of four reviewers, but the article indicates that the ratio is only 1.16, because a single review is unrealistically assumed and an error is made ($\sqrt{23.67} \neq 4.36$).

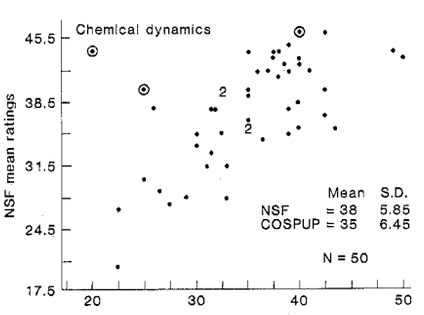

Fig. 1. Mean ratings given by NSF and COSPUP reviewers to proposals in chemical dynamics. Each circled point is based on data from only two COSPUP reviewers. [Reproduced from (*1*), p. 28]

The National Science Foundation deliberately chooses reviewers who bring different kinds of expertise to a proposal. Their properly diverse remarks inform the program director, and disagreement in their rating is not to be disparaged as random or nonrational.

LEE J. CRONBACH
Boys Town Center, Stanford University, Stanford, California 94305

References and Notes

1. J. R. Cole, S. Cole, with the Committee on Science and Public Policy (COSPUP), *Peer Review in the National Science Foundation: Phase Two of a Study* (National Academy of Sciences, Washington, D.C., 1981). COSPUP, as principal investigator, launched the study and engaged the Coles and asked the late Jack Kiefer and me to join the Steering Committee to provide technical guidance and review.

Other letters about the article by Cole et al. *and a reply from the authors will appear in a later issue.*—EDITORS

Globin Genes

Roger Lewin's article on globin genes (Research News, 23 Oct., p. 426) is highly informative. However, one point should be clarified. Mitiko Gō, who successfully predicted the exon/intron structure of primitive globin genes, is a "she," not a "he." We hope her work will inspire other women in science.

ANNE MCCAMMON
MacGregor Medical Clinics Association, Houston, Texas 77054
ANDREW MCCAMMON
Department of Chemistry, University of Houston, Houston, Texas 77004

According to Lewin, Alec Jeffreys has speculated that legumes might have obtained their globin-like gene through a horizontal transfer via an insect-borne plant pathogenic virus. This speculation was taken to imply that insect hemoglobin has three introns, similar to the plant leghemoglobins, rather than the two introns from vertebrate hemoglobins.

Most insects do not have hemoglobin (*1*). The three genera known to have hemoglobin are internal parasites of horses, aquatic flies, and aquatic bugs. All are highly unlikely to transmit plant pathogens. How, then, could insects transmit genes responsible for leghemoglobin in plants?

RICHARD T. ROUSH
Department of Entomology, Mississippi State University, Mississippi State 39759

References

1. R. F. Chapman, *The Insects: Structure and Function* (American Elsevier, New York, 1969).

I would not wish to saddle Jeffreys with the suggestion of insects alone as the only possible source of a horizontally transferred gene. The important points are these. Leghemoglobin is very similar to the gene in the animal kingdom. No close relative of legumes has a gene that is remotely like leghemoglobin. Horizontal transfer therefore looks likely. The question of the exact source is somewhat secondary. There are, for instance, many invertebrates with globin genes so far uncharacterized. If insects are poor candidates, then there are many more waiting in the wings.—ROGER LEWIN

Scientific Productivity

A few years ago I attended a committee meeting during which the granting of funds for certain research projects was discussed. As is usual in such cases, the grant request forms included an item about the recent publications of the scientists involved. Since the number of publications was assumed by some to correlate with the productivity and effectiveness of the author, a rip-roaring argument ensued about what could reasonably be termed a publication and what not. The merits of including institutional reports together with articles in prestigious journals, short notes versus letters to the editor, longer versus shorter contributions, and publications in local versus international journals were all bandied about, in much the same way as they recently have been in the columns of *Science* (News and Comment, 13 Mar., p. 1137; Letters, 24 Apr., p. 396; 1 May, p. 494; 29 May, p. 986; 25 Sept., p. 1450). On this occasion the chairman voiced the liberating opinion: Surely we are all capable of weighing as well as adding. This promptly put an end to the argument. Subsequently, grant applications have included references to all contributions which the author himself thinks are of sufficient merit to favorably influence the committee's deliberations. In practice the ability of the scientist to judge his own work in this fashion has been an important factor in helping committee members establish the competence of the scientist in question.

JRE LUTJEHARMS
National Research Institute for Oceanology, Council for Scientific and Industrial Research, Stellenbosch, 7600 South Africa

---

Erratum. In the Research News article, "Drug found to help heart attack survivors" (13 Nov., p. 774) two percentages were transposed. The article should have stated that 9.5 percent of the placebo group died, but that only 7.0 percent of the propranolol group died after 2 years of follow-up.