

# Insiders, Outsiders, and Efficiency in a National Science Foundation Panel

David Klahr *Carnegie-Mellon University*

**ABSTRACT:** *At three recent meetings of a National Science Foundation panel, nearly 200 proposals received almost 1,400 reviews from "insiders" (the panel members) and "outsiders" (the ad hoc reviewers). Although the ratio of ad hoc time to panelist time is about 2.1, the ad hoc reviewers' quantitative ratings have scant independent influence on the final proposal ratings. Nevertheless, it may be that the ad hoc reviewers provide crucial qualitative information, acting more as "expert witnesses" than as "judges." A second finding is that the outcome of about one third of the proposals (the very good and the very poor) can be reliably predicted by the independent panelist assessments that occur before the panel meeting.*

The time that scientists spend reviewing others' research proposals is time not spent on their own work. Nevertheless, the peer review system imposes a substantial burden of reviewing on the scientific community. This article assesses the efficiency of the peer review procedure of a National Science Foundation (NSF) panel, and it proposes some changes that would make the procedure more efficient without compromising its effectiveness.<sup>1</sup> The analysis is based on nearly 1,400 ratings of over 190 proposals considered at three meetings of the Sub-Panel on Memory and Cognitive Processes of the National Science Foundation.

## How the Panel Operates

### Pass 1

The panel consists of six members serving staggered 3-year terms. Prior to a panel meeting, each proposal is read by at least three panel members (two secondary reviewers and one primary reviewer).<sup>2</sup> Each panel member reads about half of the total set of proposals (one sixth as a primary reviewer and one third as a secondary reviewer). In addition, each proposal is sent to between 5 and 10 ad hoc reviewers ("ad hocs"), about half of whom complete the reviews.

The output from all of the reviewers, be they panelists or ad hoc reviewers, is the same: a narrative

evaluation of the merits of the proposal and a quantitative rating on a 5-point scale (1 = excellent, 2 = very good, 3 = good, 4 = fair, 5 = poor). For each proposal, the *Pass 1 mean* is the mean of the three panelist ratings; the *primary rating* is the rating of just the primary reviewer; the *secondary rating* is the mean rating for the two secondary reviewers; and the *ad hoc rating* is the mean over as many ad hoc ratings as are available. Note that the premeeting ratings are based on independent evaluations: Neither the panelists nor the ad hocs discuss proposals among themselves, and only rarely is an ad hoc rating read by a panelist prior to making a rating.

### Pass 2

The second phase of the review process occurs at the panel meeting in Washington, D.C. The discussion of each proposal starts with the primary reviewer presenting his or her review and rating of the proposal and summarizing the ad hoc reviews and ratings. Next, the secondary reviewers present their own reviews and ratings, and then there ensues a

---

I am indebted to Joseph Young, Program Officer for the Subpanel on Memory and Cognitive Processes of the National Science Foundation (NSF), for supplying the summary information from the panel meetings (I was a panel member from 1981 to 1983.) The study is neither funded nor endorsed by the NSF. Thanks go to William Chase, Sheldon Cohen, Robyn Dawes, Lynne Reder, Robert Siegler, and Catherine Sophian for their helpful comments on earlier versions of this article. This article is dedicated to the memory of the late William Chase—friend, colleague, scientist, and unselfish peer reviewer.

Requests for reprints should be sent to David Klahr, Department of Psychology, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pennsylvania 15213.

<sup>1</sup> Recent investigations of peer review in general (Cicchetti, 1980; Peters & Ceci, 1982), and proposal review at NSF in particular (Cole, Cole, & Simon, 1981; Hensler, 1976; Mitroff & Chubin, 1979), have devoted considerable attention to possible sources of nonoptimality, such as sex bias (Moore, 1978), institutional bias, scientific conservatism, and regional preferences, as well as issues of reliability (McReynolds, 1971; Scarr & Weber, 1978; Scott, 1974) and validity (Gottfredson, 1978). This study departs from previous work by focusing on the key components of the review process itself, rather than on substantive properties of proposals or demographic properties of reviewers.

<sup>2</sup> There are a few proposals, for things such as conferences, for which the Program Officer requests ratings from the entire panel.

general discussion of the proposals. When the discussion is completed, all the panelists make a final quantitative rating of the proposal<sup>3</sup> using the same 5-point scale described above, and the primary reviewer makes a final recommendation. I will refer to the mean (over all panelists) of this rating, made after the panels' discussion, as the *Pass 2 rating*.

Although the Program Officer has some flexibility in responding to the panel's recommendation, proposals with Pass 2 ratings above 1.5 (approximately the top 10%) almost always get funded, whereas those with Pass 2 scores below 2.5 (the bottom 60%) rarely do.

### The Cost of the Review Process

Peer review is an essential part of an open and vital scientific enterprise, and its benefits are unquestionable. However, it is not without costs. In addition to the creative effort of the principal investigator, there are various institutional costs, and at the review end of the process are the activities of the full-time NSF staff, as well as the cost of room, board, and travel for each panel meeting.

Although these costs are substantial, they are not considered here, because they are outside the scope of the decision variables I will be addressing. Instead, I focus on the opportunity cost of reviewers' time. An informal survey of about 20 colleagues and fellow panelists revealed that experienced panel members can review three or four proposals a day, whereas ad hoc reviewers spend from a few hours to a full day on a proposal, with most at the upper end of this distribution. Assume that a panelist devotes one third of a day to each proposal and that an ad hoc reviewer devotes at least half a day. Because each proposal is read by three panelists and an average of four ad hoc reviewers, we can conservatively estimate that during Pass 1, each proposal consumes approximately 3 person-days: 2 days of total ad hoc time and 1 day of total panel time.

The full panel discussion of each proposal averages 15 minutes, so that the additional time consumed by each proposal during Pass 2 is about 1.5 person-hours. The panel meeting itself, although a critical culminating step in the entire review process, has little impact on the 2:1 ratio of ad hoc time to panelist time.<sup>4</sup>

Estimation of the dollar costs requires the compounding of the time estimates with estimates of the

opportunity cost of reviewers' time. Using a conservative \$200 per day yields \$600 per proposal. The opportunity cost of reviewers' time for a typical 60-proposal meeting is approximately \$36,000, and two thirds of it is incurred by the ad hoc reviewers.

It is clear that this particular form of peer review imposes a substantial cost—a cost that is rarely estimated and that is diffused over the scientific community at large. Several questions about the efficiency of this process can be asked: (a) Is it necessary for all proposals to be discussed and rerated at the panel meeting? Can some be automatically accepted or rejected? (b) What is the relative influence of ad hoc reviewers and panelists on the fate of a proposal? Is it commensurate with the relative cost of ad hoc reviewers and panelists? Is it consistent with the expectations or norms of the scientific community? (c) Can the participants in the process—panelists, ad hoc reviewers, and proposal submitters—be better informed about the process?

### Distribution of Individual and Mean Ratings

The data base consists of all Pass 1, Pass 2, and ad hoc ratings at three meetings: spring 1982 (75 proposals), fall 1982 (46 proposals), and spring 1983 (70 proposals). For each successive meeting, there were two new panelists, so that by the third meeting, only two of the original six panelists remained. This was a more rapid turnover than the panel normally experiences, but it gives the analysis somewhat more generality than it would have if the same six panelists participated in each meeting.

### Relation Between Pass 1 and Pass 2 Ratings

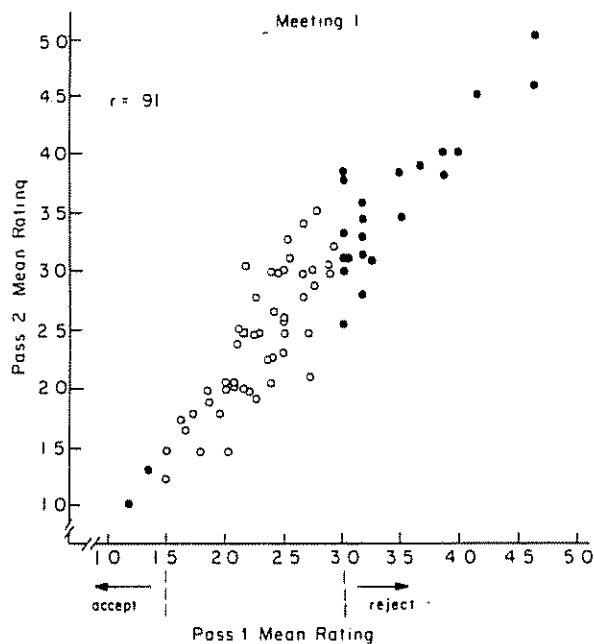
Figures 1 and 2 show that a proposal's Pass 1 mean is an excellent predictor of its Pass 2 mean: both for Meeting 1 (Figure 1) and for Meetings 2 and 3 combined (Figure 2),  $r = .91$ . The narrow band along which proposals are distributed suggests that the fate of some proposals is completely determined by the Pass 1 mean. In particular, at Meeting 1, every proposal with a Pass 1 mean better than 1.5 received a Pass 2 mean better than 2.0, and no proposal with a Pass 1 mean equal to or worse than 3.0 got a Pass 2 mean better than 2.5. Given the fact that proposals with Pass 2 means above 2.0 always get funded whereas those below 2.5 do not, the following decision rule is suggested: If the Pass 1 mean is 3.0 or below, then reject the proposal; if it is above 1.5, then accept it.

Cronbach and Gleser (1957) considered a class of multistage decision processes, of which the proposed rule is a simple case. They also proposed a procedure that evaluates costs and benefits of alter-

<sup>3</sup> Occasionally during Pass 2, a panelist will be busy working on a final statement about some other proposal and will not attend sufficiently to the discussion to be able to cast a vote.

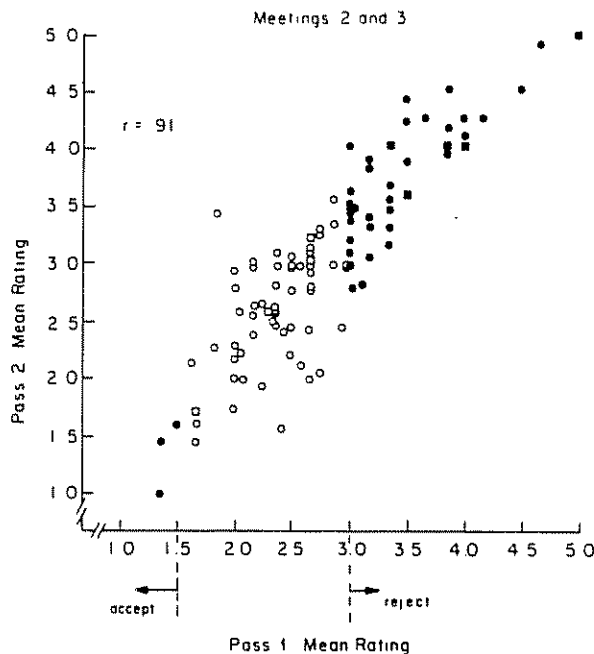
<sup>4</sup> Note that this ratio is for *aggregate* ad hoc time versus *aggregate* panel time. Because there are hundreds of ad hoc reviewers but only six panelists, most ad hoc reviewers see only one or two proposals a year, whereas the long-suffering panelists read over 50 in a typical year.

**Figure 1**  
Mean Pass 2 Rating Versus Mean Pass 1 Rating  
for Meeting 1 (75 Proposals)



Note: Closed circles correspond to proposals that would be eliminated from further discussion at the panel meeting by the suggested decision rule. Open circles correspond to mid-range proposals. Squares correspond to two proposals with identical values.

**Figure 2**  
Mean Pass 2 Rating Versus Mean Pass 1 Rating  
for Meetings 2 and 3 (116 Proposals)



Note: Closed circles correspond to proposals that would be eliminated from further discussion at the panel meeting by the suggested decision rule. Open circles correspond to mid-range proposals. Squares correspond to two proposals with identical values.

native rules. In the case of the rule proposed here, the evaluation is straightforward. When the rule is applied to Meetings 2 and 3, not a single proposal is misclassified. Adoption of this rule would reduce by about one third the number of proposals discussed at the meetings.

#### Individual Ratings by Panelists and Ad hocs

Cumulative distributions of individual ratings by panelists and ad hocs are shown in Table 1. There are two consistent patterns within and across meetings.

1. Between Pass 1 and Pass 2, there is always a substantial reduction in the number of highly rated proposals. For example, the proportion of panelist ratings equal to or better than 2.0 (*very good*) goes from .41 to .28 at Meeting 1, from .35 to .26 at Meeting 2, and from .30 to .12 at Meeting 3. This reduction in the number of highly rated proposals might be attributed to the "problem finding" nature of group discussions. According to this explanation, the positive aspects of a proposed study are usually acknowledged by all of its reviewers, but different weaknesses are discovered by different panelists. During the discussion preceding the Pass 2

rating, panelists are more likely to be highly susceptible to new negative arguments, and they will, accordingly, lower their ratings. Another possible explanation for the shift is that the ratings at Pass 1 and Pass 2 serve different purposes. At Pass 1, the intent of a panelist giving a high rating might be to call a particular proposal to the attention of the other panelists and to suggest that it is worthy of serious consideration. At Pass 2, there is a shared expectation among the panelists about the operational consequences of the ratings. Panelists quickly learn that proposals with scores below 2.5 tend not to get funded. What appears to be a continuous scale, is, in effect, a binary variable. If the Pass 2 rating is much less than 2.0, it is operationally a "no" vote, whereas if it is 2.0 or better, it is essentially a "yes" vote.<sup>5</sup> Thus, at Pass 2, a rating functions as a

<sup>5</sup> Perhaps "no" is too strong. The official policy is that any proposal receiving a 3.0 or better from the primary reviewer on Pass 2 is recommended for funding, whereas anything lower corresponds to a recommendation to decline funding. In general, a rating between 2.0 and 2.5 means that if there is sufficient money to fund everything with a higher rating, then it would not be unreasonable to fund this one, too, whereas a rating of 2.0 or better means that the panelist expects the proposal to be funded.

**Table 1**  
Cumulative Frequency Distributions for Individual Ratings by Panelists and Ad Hoc for Meetings 1, 2 and 3

Review	Ratings								N
	1 0	1 5	2 0	2 5	3 0	3 5	4 0	4 5	
Meeting 1 (75 proposals)									
Pass 1	6	15	41	60	77	81	96	96	227
Ad hoc	18	22	50	55	76	79	94	95	294
Pass 2	2	10	28	45	65	85	97	97	392
Meeting 2 (46 proposals)									
Pass 1	7	11	35	49	76	80	94	94	150
Ad hoc	17	22	45	50	74	77	91	91	184
Pass 2	1	8	26	39	65	78	93	94	227
Meeting 3 (70 proposals)									
Pass 1	0	9	30	50	73	84	92	93	214
Ad hoc	0	22	50	54	76	79	94	94	326
Pass 2	0	6	12	27	56	79	92	94	326

Note. Entries show cumulative percentage of proposals with a rating equal to or better than the interval indicated. Last column shows total number of individual ratings. Most proposals are rated by only three panelists on Pass 1, but a few are rated by more than three on Pass 1, and most are rated by 5 or 6 panelists on Pass 2.

categorical judgment rather than as a signal to the other panelists.<sup>6</sup>

2. The distribution of ad hoc ratings shows a striking consistency across meetings. These ad hoc ratings are more "lenient" than the panel ratings, and they depart from the NSF suggestion that ratings of 1.0, 2.0, and 3.0 should correspond roughly to cumulative distributions of 10%, 33%, and 66%. The ad hoc reviewers' distributions for these ratings are approximately 18%, 48%, and 75% for Meetings 1 and 2. For Meeting 3, the ad hoc distribution remains almost unchanged from the previous two meetings, except for the absence of any 1.0s. These consistently high ratings by ad hoc reviewers make it impossible for the panel to interpret them categorically, for there are insufficient funds to support the nearly 50% of proposals that ad hoc reviewers rate 2.0 or better.

### Factors Contributing to the Pass 2 Rating: Ad Hoc and Panel Ratings

To what extent can the Pass 2 rating be predicted by the three components of the Pass 1 review process (primary, secondary, and ad hoc ratings)? One way to answer this question is to run multiple regressions, using the three components as predictor variables for the Pass 2 rating. The result of interest is the amount of variance accounted for by each of the

**Table 2**  
Amount of Variance in Pass 2 Ratings Accounted for by Pass 1 Ratings for All Meetings

Analysis	Primary M	Secondary M	Ad hoc M	R <sup>2</sup>
Full set (N = 191)	(1) 56 (2) 20 (2) 25	(2) 28 (1) 64 (3) 23	(3) 1 (3) 1 (1) 37	85 85 85
Mid-range subset (N = 123)	(1) 25 (2) 34 (2) 16	(2) 31 (1) 22 (3) 24	(3) 3 (3) 3 (1) 19	59 59 59

Note. The columns labeled "primary," "secondary mean," and "ad hoc mean" show the percentage of variance accounted for by each of those variables when entered in a step-wise linear regression in the order indicated, and the last column shows the total R<sup>2</sup> (%).

variables. The first analysis to be described uses the entire set of 191 proposals from all three meetings.

### Full Proposal Set

For the full set of proposals, ad hoc ratings are completely redundant with Pass 1 ratings as predictors of Pass 2 ratings. The first line of Table 2 shows that the primary ratings account for 56% of the variance in the Pass 2 mean, the secondary ratings for 28%, and the ad hoc ratings for 1% when entered in that order. If the order of the secondary and primary ratings in the regression is reversed, then the amount of variance accounted for is similarly reversed, as shown in the second line. If the ad hoc ratings are entered first, then they account for 37% of the variance, but each of the other two variables still accounts for a substantial amount (line 3). When the secondary and primary ratings are entered first, they collectively account for almost 85% of the variance in Pass 2 ratings, whereas the additional amount contributed by ad hocs is essentially zero.<sup>7</sup>

The variance accounted for by the first variable in a step-wise regression is equal to the amount accounted for by that variable alone. If only one of the three variables were to be used, the amount of variance accounted for by the primary, secondary, and ad hoc ratings would be 56%, 64%, and 37%, respectively. Although ad hoc ratings are highly correlated with the Pass 2 outcomes ( $r = .61$ ), they are completely redundant with the panelists' ratings. Thus, they provide a modest confirmation of panelists' judgments, but they have little independent influence on Pass 2 ratings.

<sup>6</sup> A more detailed analysis of criterion levels and Pass 1 to Pass 2 shifts by primary and secondary reviewers is presented in Klahr (1983).

<sup>7</sup> The variance accounted for by ad hoc ratings drops from 37% when entered first to 1% when entered last because ad hoc ratings are positively correlated with the two other predictors ( $r[\text{ad hoc-primary}] = .55$ ,  $r[\text{ad hoc-secondary}] = .47$ ).

### *The Mid-Range Subset*

Recall that for about one third of the proposals, there was little change from Pass 1 to Pass 2, and this enabled us to formulate an accept/reject rule for those proposals. It is possible that the ad hoc influence might be evidenced for those proposals that are not automatically accepted or rejected by the decision rule. In order to test this hypothesis, I constructed a *mid-range set* consisting of the 123 proposals with Pass 1 means less than or equal to 1.5 and above 3.0.

Overall, substantially less of the total variance is accounted for by the three predictors on this subset (59%) than on the full set (85%). This is to be expected, given the range reduction caused by excluding the highest and lowest values of the Pass 1 mean. Correspondingly, each of the predictors alone accounts for less variance than on the full set (25%, 22%, and 19% for primary, secondary, and ad hoc ratings, respectively). But here, too, the contribution of ad hoc ratings is completely redundant with the panelists' ratings.<sup>8</sup>

In searching for some independent effect of ad hoc ratings, one could consider many special subsets of proposals. For example, perhaps the ad hoc reviewers are important when there is high variance on the panel's Pass 1 ratings and low variance among the ad hocs. That is, panelists might turn to the ad hoc reviewers' judgments when they disagree among themselves.<sup>9</sup> Several such subsets were constructed, but to no avail: The general pattern of ad hoc redundancy remained.

### *Summary of the Regression Analysis*

We can consider two different ways to interpret these analyses. We can ask how much the total  $R^2$  would be reduced if one of the components was not used. If the ad hoc ratings were not used at all, we would get the following reductions in amount of variance accounted for: full set, no effect; mid-range subset, from 59% to 56%. Deletion of either secondary or primary ratings would substantially reduce the  $R^2$  for the full and mid-range sets. Examination of other plausible subsets failed to uncover any situation in which elimination of ad hoc ratings would distort the final outcome of the panel review process.

Alternatively, we can ask which component would be the best predictor if it was the *only* one available. In this situation we would choose the

<sup>8</sup> Another consequence of selecting proposals with mid-range means is the slight negative correlation between primary and secondary panelist ratings ( $r = -.15$ ). The effect on the regression (Table 2) is to have the second, rather than the first, of the two panel components account for more of the overall variance.

<sup>9</sup> In general, ad hocs and panelists disagree about when to disagree: The correlation between the panel's Pass 1 variances and the ad hoc variances is very low ( $r = .13$ ).

secondary mean for the full set and the primary rating for the mid-range subset. There was no case for which the ad hoc ratings were the most important.

### **Discussion**

The policy implications of this analysis depend on a key assumption about the appropriate criterion for the review process. If we take the final disposition of proposals as the "correct" decision, then the Pass 2 mean becomes the criterion, and any suggested procedural changes would be aimed at achieving the same outcome more efficiently. On the other hand, we might assume that certain factors *should* be more important, because on the face of it they seem like the correct ones. In particular, we might want to be sure that the ad hoc influence is greater than it is at present.

The resolution of the question is beyond the scope of this article. However, it is possible to lay out some arguments for, and consequences of, both sides of the issue. First, let us assume that the present system is producing the appropriate outcomes. If this is the case, then the ad hoc influence as measured by this analysis could not be much higher than it already is, even though it may be essential to the overall review process. Furthermore, it is possible to make the panel meetings more efficient and effective.

### *Why the Ad Hoc Ratings Are Inherently Limited Predictors of Pass 2 Outcomes*

Two different kinds of expertise are represented in the review process. One has to do with scientific issues and the other with the review process itself. With respect to the former, the ad hoc reviewers are often at an advantage compared to the panelists. For any specific proposal, the ad hoc reviewers may have more technical proficiency, a better sense of what can realistically be accomplished in the area, and greater familiarity with the track record of the principal investigator. However, the ad hoc reviewers are at a disadvantage when it comes to making a quantitative rating of the proposal. First of all, they are generally unfamiliar with the way the ratings get translated into decisions. Second, they do not have the same sense of scarce resources that the panelists do. It is essentially costless for an ad hoc reviewer to give a very high rating to a single proposal: Nearly 50% of the ad hoc ratings are 2.0 or higher (Table 1). On the other hand, the panelists realize that there must be an overall consistency to their ratings, and that only a few proposals will ultimately be funded. Thus, although the ad hoc reviewers may provide important *qualitative* information that is used by the panelists in sorting out the mid-range proposals, their *quantitative* ratings are ignored.

The panelists are also chosen for their expertise

in several areas, but they are generalists with respect to many of the proposals they must review. However, they are experts in the review process itself: They are more sensitive to the ultimate meaning of ratings, and they usually have more experience in evaluating a much wider variety of proposals than do the ad hoc reviewers. Furthermore, the panelists have the benefit of face-to-face discussion of the proposals prior to making their Pass 2 ratings, whereas the ad hoc reviewers obviously do not.

Collectively, these factors imply that ad hoc reviewers may play an important role that cannot be measured by the present analysis. The clinical judgment literature (Einhorn & Hogarth, 1981) makes a distinction between expert observations and expert decision rules. For example, detection of certain warning signs in a chest x-ray requires expert observational skills, but incorporating that observation into an entire medical profile and producing a diagnosis requires a different kind of skill. In the panel review process, the ad hoc reviewer's principal role is to encode key features of the proposal, whereas the panelist's expertise lies in the ability to combine these features.

#### *How to Clarify the Ad Hoc Reviewer's Job*

The ad hoc reviewer's role should be construed more as expert testimony and less as judgment. Ad hoc reviewers should continue to be instructed to comment on the strengths and weaknesses of proposals, but they should be informed that their final overall rating is not as important as their substantive comments.

#### *How to Increase the Panel's Efficiency*

The decision rule proposed earlier would have resulted in a 36% reduction in the number of proposals considered at the panel meetings. Having fewer proposals to discuss would either reduce the duration of the meeting or result in more extensive discussion of the proposals that really warrant it. Because there is a large fixed cost to panel meetings regardless of their duration, the latter alternative seems most efficient.

#### *Understanding the Feedback*

After all funding decisions are made, the full set of reviews from panelists and ad hoc reviewers is forwarded to the proposal initiator, accompanied by a brief summary of the panel discussion. Consider the puzzlement of an investigator attempting to reconcile a set of favorable ratings of his or her proposal with a negative funding action. The standard disclaimer of Program Officers is that "there were many excellent proposals, and insufficient funds to support them all." But the issue is neither that simple nor that ambiguous (for the implicit next

question is: How was the choice among the excellent ones made?).

The apparent contradiction between high ratings and no funding is a consequence of several features of the review process described earlier. Panelist reviews are not distinguished from ad hoc reviews when they are forwarded to the investigator, and yet the ad hoc ratings—which tend to be the highest—have the least influence. Even if panel and ad hoc ratings were distinguished, the investigator would still be unaware of two critical facts: that the Pass 1 ratings (which are released to investigators) are only preliminary to the critical Pass 2 ratings (which are not released), and that Pass 1 ratings are almost always higher than Pass 2 ratings. If neither the ad hoc nor the panel Pass 1 ratings have the meaning attributed to them by the investigator, it is difficult to justify the current policy of forwarding them, rather than the Pass 2 rating, to the proposal initiator.

#### *Thinking the Unthinkable*

Suppose we make the alternative assumption about the criterion variable: that the panelists are really "wrong" here and that the final disposition of proposals should be based on the ad hoc reviewers' ratings, rather than on the panelists'.<sup>10</sup> The problem with this assumption is the pattern of high ad hoc ratings discovered in this analysis. The Program Officer would have many more highly ranked proposals than could be funded and would have to make a decision on some grounds *other* than the ad hoc ratings.<sup>11</sup> Rather than decentralize the decision-making influence, reliance on ad hoc reviewers rather than panelists would actually centralize it even further.

#### *A Final Note*

This report is based on nearly 1,400 ratings of almost 200 proposals evaluated by 10 different panelists and several hundred ad hoc reviewers. Although the analysis has demonstrated some stability over different meetings, panelists, and ad hoc reviewers, few other NSF panels follow the exact procedure described here, and it would be inappropriate to generalize to fundamentally different review processes (such as those at the National Institutes of Health). Nevertheless, the analysis may contribute to our understanding of the peer review process in several ways. First, it introduces some estimates of the costs

<sup>10</sup> Some NSF panels function in precisely this way: There is no fixed panel, and all reviews are ad hoc. In such cases the Program Director has much more discretion in how the ad hoc reviews are used.

<sup>11</sup> Simply adjusting the criterion to compensate for inflated ad hoc ratings would not help much. The skewed distribution produces such small differences in mean ad hoc scores that they cannot be taken very seriously.

that this process imposes on the scientific community at large. Although the estimates are highly approximate, it is important to consider the price we pay for the extensive peer review procedures that are presently in use. Second, I have suggested a straightforward methodology for determining the relative influence of different components of the peer review process. Panels with different procedures can use similar analyses for their own evaluation. Third, there are a few clear policy recommendations that follow from the analysis. Whether or not these suggestions are actually taken is less important than providing a framework in which to consider their effects.

#### REFERENCES

- Cicchetti, D. V. (1980) Reliability of reviews for the *American Psychologist*. A biostatistical assessment of the data. *American Psychologist*, 35, 300-303.
- Cole, S., Cole, J. R., & Simon, G. A. (1981) Chance and consensus in peer review. *Science*, 214, 881-886.
- Cronbach, L. J., & Gleser, G. C. (1957) *Psychological tests and personnel decisions*. Urbana: University of Illinois Press.
- Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, 32, 53-88.
- Gottfredson, S. D. (1978) Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist*, 33, 920-934.
- Hensler, D. R. (1976) *Perceptions of the National Science Foundation peer review process: A report on a survey of NSF reviewers and applicants* (Tech. Rep.; NSF Publication No. 77-33). Washington, DC: National Science Foundation.
- Klahr, D. (1983) *Panel review at NSF: A structural analysis*. Unpublished manuscript, Department of Psychology, Carnegie-Mellon University.
- McReynolds, P. (1971) Reliability of ratings of research papers. *American Psychologist*, 26, 400-401.
- Mitroff, I. I., & Chubin, D. E. (1979). Peer review at NSF: A dialectical policy analysis. *Social Studies of Science*, 9, 199-232.
- Moore, M. (1978) Discrimination or favoritism? Sex bias in book reviews. *American Psychologist*, 33, 936-938.
- Peters, D. P., & Ceci, S. J. (1982) Peer-review practices of psychological journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences*, 5(2), 187-255.
- Scarr, S., & Weber, B. L. R. (1978) The reliability of reviews for the *American Psychologist*. *American Psychologist*, 33, 935.
- Scott, W. A. (1974) Interreferee agreement on some characteristics of manuscripts submitted to the *Journal of Personality and Social Psychology*. *American Psychologist*, 29, 698-702.