

content analysis of the reviews in the 250 jackets we examined showed that there is, in general, a strong correlation between adjectival ratings and the substantive comments. It is rare to find strongly negative substantive comments and "excellent" or "very good" adjectival ratings, or strongly positive comments and "fair" or "poor" adjectival ratings. Program directors are aware, however, that one reviewer's "excellent" may be equivalent to another's "very good."

Occasionally, there are notable discrepancies between adjectival ratings and substantive comments. One program director observed:

Frequently, people who are "soft" raters will give you a very long critical review, really pointing out what they think is wrong. Then they give you an adjectival rating that is not at all commensurate with the review.

It appears that interpretation of reviews often involves considerable subtlety.

Frequently, a review is short, extremely affirmative, and rates a proposal "excellent." Some program directors maintain that this kind of review has little or no value. As one man commented:

We get reviews where somebody will say, "This is a great proposal; should be supported, excellent." That is of no use to us.

But another program director said that such reviews can be meaningful in light of the reviewer's past performance. He said:

I sent off a proposal to a _____ who is also a very good friend of mine and I got back a hastily scrawled one line comment—"I like all of these proposals." I know what he is telling me and if there were any questions asked, he would have written me a long, carefully detailed letter because he has done this in the past.

In these two cases, the reviews were "objectively" similar—both short and affirmative. However, they were responded to differently because of the program directors' knowledge of the reviewers in question.

A particularly thorny problem for program directors is evaluating negative comments by reviewers. They must determine whether such comments are based on relevant or irrelevant criteria. The following quote illustrates the difficulty:

If I get back a review that says, "On page 4, he says such and such," I have to make up my mind if he is really nitpicking because he doesn't like the man, or because he really doesn't feel this is the right way to go.

SECTION 4 | Interpretations of Reviews by Program Directors

INTERPRETATIVE PROCESS

The completed review forms that come back to program directors include adjectival ratings (ranging from poor to excellent) and written comments by reviewers. Evaluation of the reviews consists not only of taking an average of the ratings but also of careful assessment of both the ratings and the comments. In other words, the program director must interpret what reviewers have presented to him.

Program directors maintain that the adjectival ratings can be evaluated only in context. Reviewers tend to develop patterns in the rating of proposals. Some use the entire rating scale; others restrict their range. Some never rate above "very good," and some never below "good." Knowledge of a reviewer's range can affect how a rating is interpreted:

We don't just look at the box that is checked. That's another thing we build up over the years, we know what people's standards are. Most of the proposals that I recommend for granting have reviews above very good. On the other hand, I have a proposal that I recommended for a grant last year that had two reviews and they both were just good. I am absolutely certain that the proposal with those two investigators was the strongest I supported. Why? I know that the people reviewing the proposal will infrequently rate a proposal higher than good.

Knowledge of a reviewer's range is useful in understanding ratings, and further interpretation is based on the written comments. Our

88 | PEER REVIEW IN THE NATIONAL SCIENCE FOUNDATION

When asked how he could tell whether a negative review was a result of a personal antagonism between the applicant and reviewer, one program director responded:

It shows through by the choice of words of the reviewer. When you read it, there are tip-offs in the kinds of language a reviewer will use. They move outside of scientific language to more colorful language to describe what's wrong with the proposal.

Or, he may use his own expertise in the area in evaluating the quality of the work being proposed:

There are indications in the proposal to permit you to make judgments. There are peripheral things like a man is working in one area, but he's using a kind of instrument that I've used, an analytical instrument I'm very familiar with. This gives me a clue as to how valid the criticisms of the work are.

Or, he may call in a "neutral" third party who is knowledgeable in the area and may be able to "review the review."

If I'm not technically competent, I'll telephone someone who is neutral on the issue. You sort of divide the scientific community into two camps. There will always be fringe areas and various people that won't be in either camp, so you then try to find neutral observers to talk to.

A variation on the strategy just mentioned is contacting the applicant and asking him to respond to the criticisms that have been made. One program director described this strategy:

Very often if I am not able to come to a definitive conclusion about something on someone's review, I will call the investigator and say, "I wasn't able to understand the following question, can you help me?" Very often the man will come through with a delightfully detailed answer to the question and then follow it up with a letter. There have been times when this sort of discussion has tipped the balance in favor of a person.

These illustrations suggest that most program directors spend a great deal of time on interpretation of reviews. They view this process as a basic component of their job—as essential to a fair appraisal of the scientific merits of proposals and thus to decisions concerning them.

The interpretation and evaluation of reviews and the decision making based upon these activities are what make the activities of program directors highly demanding and professional—far beyond the routine clerking they may sometimes appear to be. This may be one reason why program directors insisted that the numerical ratings of proposals were of less use to them than were the substantive comments.

If it were possible to do the program director's job simply by

computing mean scores from adjectival ratings, and to fund solely on the basis of those ratings, the position would require less scientific expertise and probably would be less attractive. One program director put it simply:

We don't only look at the box designations, we look at the actual comments. If you were merely tabulating evaluations, you wouldn't need someone with a Ph.D. to handle the job.

To interpret peer reviews intelligently and to decide whether further reviews are needed, whether a reviewer is using unfair criteria, whether one reviewer is in a better position to appraise a proposal than another, requires considerable knowledge of the social and intellectual structures of scientific fields.

TYPES OF PEER REVIEW CASES

The problems faced by program directors in decision making are best understood by examination of representative cases, which we shall now do.¹ We will present representative examples, drawn from several programs, of six different types of cases that we have identified. These are:

Unproblematic Cases:

Reviewer agreement—positive evaluations

Reviewer agreement—negative evaluations

¹Before doing this, however, we make two general observations. The first concerns the criteria used by reviewers in evaluating applications. These differed significantly from one reviewer to another and, equally important, from one program to another. In algebra, the overwhelming majority of the comments concerned the reviewer's evaluation of the applicant's ability. In mathematics it is difficult to predict whether the applicant will actually be able to solve the problem he is proposing to work on. Reviewers' comments, therefore, are usually focused on the skill of the mathematician and the significance of the problem. But in chemical dynamics and solid-state physics, proposals are frequently to do experimental work. In these programs, reviewers are more likely to comment on the experimental design as well as the significance of the problem and the ability of the applicant. Second, reviews frequently combine two types of comments: technical substantive ones and evaluative ones. Although we did not have the time to do systematic counts of these two types of comment, it was the independent impression of all three authors that a significant majority of the reviews concentrated on evaluative statements. The majority of reviews were short, generally taking up no more than two-thirds of a page. Some, however, were several pages long. In the reviews excerpted below, we have concentrated on evaluative comments. The reader should be aware that many reviews contained detailed substantive comments.

- Reviewer agreement on quality of "borderline cases"
- Reviewer disagreement on quality of "borderline cases"
- Weighing different decision criteria
- Apparent discrepancies between reviews and final funding decision

UNPROBLEMATIC CASES: REVIEWER AGREEMENT—POSITIVE EVALUATIONS

Case 1

The first example is a case that was not problematical for the program director. There was complete intellectual agreement among the reviewers about the scientific merit of the proposal. The panel gave the proposal a mean rating of 1.2, and mail reviewers gave it an average rating of 1.7. (Recall that ratings range from 1 for "excellent" to 5 for "poor." In most fields, ratings tend to cluster more toward the "excellent" to "good" ratings.) After describing briefly the subject matter of the proposal, the first reviewer begins a terse evaluation:

The quality of work, the insight, and the general productivity of programs headed by _____ are, in my judgment, unsurpassed in the nation. The methodology which they have developed and the kinds of questions which their research addresses are models for the rest of the work to follow. . . . In summary, I give my highest recommendation for the continued support on this pioneering research group. (rating: 1.0)

A second reviewer of this proposal sums up his reaction with the following comments, quoted here in their entirety, except for identifying information:

I strongly recommend that this proposal be funded. The objectives of the proposal are important to both basic and applied _____ . The experimental procedures have been designed with care and appear to be sound. The previous research of _____ indicates that he could conduct this research program in an excellent manner. Based upon my personal observations, I rate _____ as being possibly the most competent experimental scientist in this research area. The equipment and facilities that he has developed are excellent for the proposed research. I am a highly critical person, but I found nothing of substance to criticize in this excellent proposal. (rating: 1.0)

A third reviewer, somewhat more critical, had this evaluation:

Although I do not find the proposal to be particularly well written—particularly in terms of a sufficient explanation of experimental design and techniques to be

employed, I would still strongly endorse funding of this proposal. This endorsement is based primarily upon the excellent work which has emanated from this group in the field of _____. I have carefully followed the progress of these people over the past several years, and I am rather well convinced that some of the most incisive and yet technically advanced work in the area of _____ in the nation has been performed by this group. . . . Had this proposal come from most any other group in the country I would not provide the same enthusiastic endorsement on what is specifically contained in the proposal. (rating: 2.0)

This last reviewer is clearly emphasizing one of the NSF's stated criteria, the "track record" of the principal investigator.

The comments of the review panel are not available, but from its rating, the comments must have been strongly favorable. Even a cursory examination of the comments made by reviewers suggests that this proposal should pose few problems for a program director, unless the program has no funds to distribute.

Case 2

Another case, in the algebra program, was equally easy for the program director, although the reviewers focused more on evaluations of the abilities of particular principal investigators than on the proposal. One reviewer had this to say:

"X" and "Y" are certainly on the first team in _____ with an excellent record for producing important work. There is no doubt that they should be supported. (rating: 1.0)

The next reviewer gives specific ratings to each of the four principal investigators:

I shall treat each of the four principal investigators separately. The easiest is "X." I think. No one would question the statement that he is the best person in the field, at least until recently. Whatever he does will be important. If anyone deserves support in this area, it is he. _____ is difficult for me to judge, since I neither know him nor his work very well. He is certainly very capable, but since he publishes very little, it would perhaps be good to get an opinion on him from someone more closely involved with _____. _____ as I have indicated before, is in a unique position, he is extremely bright, and what he does well . . . there is no one, except perhaps _____, who can touch him. His really brilliant work to appear in _____ on the problem of _____ is a good example of this. . . . He is very likely to produce first rate work, and should certainly be supported.

We have presented only a few out of scores of possible examples of positive evaluations that offer few problems for the program director. The forms of these evaluations vary: some concentrate on the quality of the proposed work; others on a combination of the quality of the proposal and the performance record of the principal investigators; and still others focus almost exclusively on the prior track record and prominence of the principal investigator in his field. But they all agree substantially that the grant should be made.

In sum, outstanding proposals that receive uniformly high ratings from reviewers almost invariably are funded. One program director put it this way:

There are some proposals that are very outstanding, that present new ideas, and are something that the NSF should be supporting because they have all the difficulty—very outstanding proposals get funded no matter what!

UNPROBLEMATIC CASES: REVIEWER AGREEMENT—NEGATIVE EVALUATIONS

Case 3

In this case a proposal received uniformly negative reviews. These are selected review comments; the project was given a final rating of 3.0. The first reviewer commented:

This proposal needs quite a bit of additional work prior to funding. . . . A small sample (one man working for one-half time for 3 years) will contribute little to the store of knowledge. No science is presented or planned. With these animals it is even more imperative than in some other studies that a clear appreciation of one's scientific objectives, hypotheses be clearly stated and that experimental designs be appropriate to test those hypotheses. (rating: 3.3)

A second reviewer also is skeptical:

The objectives of the proposed study are clearly stated but the work is not put in a strong conceptual framework. The investigator proposes to do in more detail the same sorts of things that have been done elsewhere on the [animal]. Some of the primary faults of past work are not overcome because the same approach is being taken in this study. (rating: 2.8)

A third reviewer damns the proposal with faint praise: Although this study probably will not in itself break new ground in terms of basic concepts, it should result in a solid contribution to knowledge of [animal] biology. (rating: 3.0)

Interpretations of Reviews by Program Directors | 93

In sum, there was no strong enthusiasm for this proposal from any of the reviewers, although none of them thought that it represented particularly bad science. Nonetheless, the lack of enthusiasm made it relatively easy for the program director to recommend that the request for support be declined.

Case 4

Consider a second example of clear agreement about the relatively poor quality of a proposal—this one from a social science discipline. Here are selected representative comments from the referees:

I'm not convinced that the research as now planned would be worthwhile. There has always been agreement that one should try to trace out all the costs and gains from a project. As long as the gain is judged to be worth the research cost, go ahead and fund. In this proposal, I am unenthusiastic about this prospect. (rating: 3.0)

The review just quoted was the most affirmative one received for this proposal. Another reviewer labeled the effort as a "crackpot proposal" and went on to explain why the substance of it was meaningless, rating it "poor." Another reviewer rated it "poor," saying:

I recommend that the project not be approved. Most importantly the proposal reflects no special insight concerning the three elements of the proposed study.

The final reviewer was also extremely negative. He said:

This proposal contains no evidence that the author has anything to say about this well-worn topic. There is some evidence that he does not even understand what progress has been made on these questions. (rating: 5.0)

The program director's summary reflected these reviewers' evaluations. He said:

We recommend that this disappointing proposal be turned down. The reviewers are unanimous in criticizing this proposal because of the absence of any original ideas or insights.

Two program directors summed up reactions to proposals that receive uniformly poor ratings:

There are some that are not competitive. They are dealing descriptively with phenomena and they are not addressing it in such a way that they are trying to understand and explain it. . . . It has no scientific merit.

You can read the proposal and see the sort of thing that they are doing and just know that it is just definition chasing, or generalization for the sake of generalization, and it's not exactly attacking any problems. And even if he solved the things he wanted to do, it wouldn't move [the field] forward at all. It's a weak guy and a weak problem.

PROBLEMATIC CASES: REVIEWER AGREEMENT ON QUALITY OF "BORDERLINE CASES"

In some cases there was substantial reviewer agreement that the proposals were neither excellent nor poor, and they were given relatively uniform ratings in the middle of the scale. They were judged to represent good but not outstanding science. For the program director these cases are "borderline"; they are not clearly either acceptable or rejects.

Borderline cases require that an accept-reject decision rule be applied within a restricted range of assessed quality. If program directors and panel members cannot distinguish between a set of proposals in terms of their estimated scientific worth, what criteria do they use in making decisions? This question requires further research.

Case 5

In the first borderline case we considered, there was substantial reviewer agreement that the proposal fell somewhere near the middle of the quality range—that the quality of work was likely to be mediocre. The proposal was given a summary rating of "good" by the mail reviewers and a "good" by the program director. It was declined.

One reviewer raised serious questions about the proposed methods:

The topic that _____ propose to investigate is an important one. I am not that confident, however, that the proposed research methods would bear fruition in the time allowed. (rating: 3.0)

A second also raised questions about the proposed method and rated the proposal "fair" to "good." A third reviewer said the following:

The proposed research, with its emphasis on _____ is a scientifically worthwhile project, and it should be attempted. I do have some questions about the proposed techniques of analysis.

He goes on to comment that the scientific competence of the investigators is sufficient to carry out the proposed research. His rating was "good."

The most critical reviewer said:

I have serious doubts concerning the contribution the proposed research will make toward improving our basic knowledge in this area. . . . Unfortunately, more than half of the author's listed references are not refereed, open-literature published articles. Furthermore, they are inaccessible to the reviewer, which makes it difficult to assess pertinent work which has been done in the past on this proposed problem. (rating: 4.0)

The program director made the following comment in his summary statement:

Very steady, if uninspired work. We just cannot afford to risk funds on a program that may be weak. The comments are excellent in scope and should be helpful to the investigator. I recommend declination.

Unlike the "clear" decline, this proposal was judged to have some merit. Had there been more resources available, it might have been funded. Also, the program director's comments imply that a resubmission responsive to reviewers' comments might be more successful the second time around.

This case is representative of many that fall near the cutting point in decisions. The principal investigator is not in competition for funds with the scientists whose proposals are given uniformly excellent reviews or with those who get uniformly poor reviews. He is competing with other principal investigators whose proposals also fall near the borderline.

Case 6

In our second borderline case, from a social science discipline, the proposal again received only intermediate ratings. Two mail reviewers rated it "good," one panel member judged it as "very good," another as "good," and another as "fair." The proposal requests somewhat more than \$75,000. Here is a sampling of the reviews. One, by a panel member, gives the proposal mixed reviews:

The questions posed by [the principal investigator] are potentially quite interesting. Somehow, by the time [he] gets down to the details of his proposal, he has made them seem much less interesting. . . . In spite of my lack of enthusiasm for this proposal, I do think that [the principal investigator] does deserve some support. Much of his work shares the quality of this proposal: The topics are interesting and often reasonably original, the questions are important, the analysis is competent, and yet the whole thing is a bit prosaic. Yet his earlier work with _____ is among the best recent work in

Another panel member was even less sure of the merits of this proposed work:

The bulk of the present proposal consists of a reprint of a paper which _____ wrote 9 months ago.

Another reviewer also expressed mixed views:

The subject is an important one. . . . The conceptual framework is straightforward, almost obvious; and accordingly, the main results obtained so far are obvious. Some important questions are overlooked. . . . Some questions are raised and answered incorrectly. . . . All in all, this would not be a bad piece of work, and he will undoubtedly get a publication out of it. I seriously question that it is worth _____ dollars.

The proposed project was funded. How was an affirmative decision made on this proposal and not on others that received roughly similar comments from peers? It appears that characteristics of the applicant might have influenced the decision, and indeed the principal investigator was an established social scientist in a major department. The program director summarizes these comments in justifying the decision:

The reviewers find that _____ is raising some interesting questions about an important problem. Most also find that the theoretical aspects of the project are the most satisfying part . . . even though they point out that there are significant omissions. . . . They foresee difficulties with the empirical testing because it is vague and not closely related to _____. [X] is an excellent _____ whose last several applications have been marginal. In this case, [A] and [B] feel he deserves support, and [C] expects he will get a publication out of the project. [D] was very supportive in the panel discussion. The program director believes that [A's] lukewarm support outweighs [E's] emphatic negative, that the project is a good one and merits approval.

Case 7

The decision in case 6 must be viewed in light of at least one other—one in which the ratings are good and the qualifications of the principal investigator are sound but in which the research proposal was declined. Here are some selected comments by reviewers in such a case:

Reviewer 1:

If [X] can deliver what he promises, it will surely be a major contribution to _____. I recommend and indeed urge support of this project.

Reviewer 2:

I am not in a position to comment on the difficulty of the proposed research. . . . I am . . . indifferent to this research.

Reviewer 3:

Another one of these narrow, highly technical proposals where some results of some use are highly likely but important results are ruled out almost a priori. . . . I would say that [X] is correct in claiming that the approaches he intends to follow are more likely to be productive than . . . [those] pursued by [Y].

Reviewer 4:

The competence and potential of the principal investigator are beyond question, and the conceptual framework of the project is sound. The chief question is whether the successful achievement of the . . . objectives of this proposed research will constitute a greater payoff to _____, if successfully carried through.

It is difficult to see what distinguished the first borderline case that received an award from the second that was declined. Although the reviews were mixed, but generally affirmative, the program director acknowledges that he comes out feeling much as did one of the reviewers who questioned the value of the project: "But the program director ends up concurring with the judgment expressed by [reviewer 3], mainly that this is another one of these narrow, highly technical proposals where some results are almost ruled out a priori."

PROBLEMATIC CASES: REVIEWER DISAGREEMENT ON QUALITY OF "BORDERLINE CASES"

Some cases fall near the borderline because the reviewers agree that they represent good but uninspired science. Others fall near the borderline as a result of significant reviewer disagreement, some reviewers being very affirmative and others very negative, giving the proposal a mean rating in the borderline area.

Case 8

This case involves one of the most difficult decisions faced by program directors. It produced a set of highly discrepant reviews. The mail reviews yielded an average rating of 3.0, which was high (low) enough in this program to preclude funding; the panel members gave it an average score of 2.3. The program director's final rating, however, was 1.9. Excerpts from the program director's final comments follow:

At first reading the proposal looks plausible and helpful in understanding _____. Upon reflection it seems that the proposal also describes a

lot of work that may not be useful. The implications and significance of this research are exaggerated. . . . Much of the field research conducted by this research group in the past is flawed due to faulty experimental design. We believe that this proposal is also weak in experimental design. There are too many questions on research concepts and methodology to warrant NSF support.

The project was funded. The combination of this comment by the program director and the mixed reviews by the referees called into question the final decision. We reinterviewed the program director about this case and asked for an explanation of the decision. Before looking at the program director's *post hoc* rationalizations of his decision, consider the disagreement among the referees about the merits of the proposal.

One referee's comments were essentially reproduced by the program director in his summary. This is frequently done by program directors, who go through referees' reports and extract verbatim or paraphrased sections of them for their own summaries. In this case the reviewer quoted by the program director, a panel member, gave the proposal a 4.0 ("fair") rating. But two other reviewers liked the proposal. One rated the proposal "excellent," 1.0. His one-page review, which contains little detail, says that the problem addressed is important and that the group is well equipped to conduct the research. A second affirmative review came from another program director in the same division, who rated it 2.0:

Although I am not qualified to view this proposal as an expert, I did find the proposal interesting and generally understandable. I feel that the investigators are well qualified and the background and facilities for this work are excellent. . . . Overall, based on past productivity and probable future contributions from the group, I rate the proposal as good to very good.

Another panel member, in a cursory review, rated the proposal at 1.8 and supported funding. But still another reviewer rated it as "poor," 5.0, and presented a five-page, detailed, negative critique of the proposal. However, yet another reviewer rated the proposal as "very good," noting:

The research area proposed is one of the most exciting and potentially valuable areas extant to both _____. The _____ team has been a consistent contributor to this field—some excellent, some bad; but on the whole, good to excellent.

Another reviewer thought the project was "excellent." The review, in its entirety, with deletions only for identifiable comments was:

This is a very good proposal—straightforward, workable and should contribute considerably to our knowledge of _____. [Professors A and B] have produced clean experimental results in the past and are proposing to do so in the future. As well as being good basic research, the outcome most likely will be of practical value in _____.

Also located in the files is an instructive letter to the program director from one of the principal investigators:

I know that you are aware of the unfortunate competitive bitterness that has emerged in the field of _____. We believe it is unlikely that objective reviews can be obtained from anyone involved in the _____ program, so I am attaching a list of independent researchers who are familiar with the field and might be expected to provide fair appraisals. Perhaps you could give this some consideration. Meanwhile, we are carefully avoiding the temptation to criticize those who have attacked us, in the hope that better relations can be developed. We deeply regret the necessity to touch on this subject.

The program director is faced with a difficult decision. He has before him totally conflicting reviews, widely disparate in their appraisal of the scientific merit of the proposed research. The reviewers, who come from reputable backgrounds and many of whom are distinguished scientists in their field, disagree sharply about the qualifications of the research group proposing the research. The program director also is aware of the history of conflict within the scientific specialty. The proposal is borderline not because of agreement among referees that the proposal is of medium quality, but because of the averaging of discrepant evaluations.

We discussed the decision with the program director. Here is the exchange between us:

I: This is a grant to _____. You had written up what seemed to be a negative summary of it. [Handed the jacket to the program director.]

S: You would be interested to know that I cited this to the guys yesterday as one of the most extremely difficult proposals that I had to decide on, and I've been here going on my seventh year. The field is so polarized, as you can see from the reviewers; they are either all good or all bad. They are either 1's or 4 or 5. In the whole field these two [principal investigators] approach the subject in a way that has antagonized many of the people. . . . They have some proponents, however, that think that they are right. So this is a very difficult thing. As I remember, I was going to decline it.

I: Yes, look at your own summary in there. That's what confused us most, otherwise we would have understood it. Your summary was taken from a

review by _____, who gave it a score of 4.0. He was a member of the panel.

S: He is our expert on these things. He is the closest thing we could come to a fairly unbiased person on this. . . . Let me tell you the whole history of this. We were all prepared to decline it and sent it down to the front office to decline it. [Mr. X], who was then deputy division director and a _____, knew of this difficulty. _____ seemed to be dealing with this sort of thing and felt that there were too many higher ratings along with the low ratings—it was so evenly divided. [Mr. X] said if you can't get a clear reading—if some of them just "damn it to hell" and some say "good"—it's better to play it cautious than to let it go because this . . . area of _____ is one of the areas that I'm least competent in, so I would trust my own judgment less than in any other area. He said to send it to someone else. So we sent it to another batch of reviewers. We got no review from one and a 1 and a 2. So we got good readings from both of those; and on the basis of that, I proposed that we make the grant.

I: The only thing that I don't understand about it is in your summary you selected Mr. A's summary as probably the closest to your own.

S: Well, this is the place for the panel summary.

I: But that is essentially negative the way it was written. Then it was sent downstairs to _____ recommending to decline and they questioned it down there and then sent it back and you sent it out for additional reviews. Is there any indication that some external pressure was brought to bear on the Foundation to fund this proposal?

S: No, [Mr. X] felt it wasn't getting a fair shake—I felt, for gosh sakes! I've agonized with this thing for so long.

Indeed, it was clear that the program director had agonized over the decision, and it was not entirely clear to him that he had made a correct one. Later on, in discussing other problematic proposals, the same program director came squarely to the difficulty in dealing with borderline proposals:

You get advice, and if it all says the same thing, there is no problem. If either it says turn it off or turn it on, O.K. But often they don't. You get advice that's different. On the same proposal people will say it's sloppy, and others will say it's nicely put together. It's amazing when you go through this. In general, there is a certain amount of consensus. But just the ones that I've gone through lately somebody says "it's shoddily put together" and the other person says "it's the best prepared proposal I've seen in a long time." So what do I do? For my money I take my choice. So it's up to me to evaluate in part. I've read the

proposal, and I've read the evaluations, and I know something about some of the guys. There are antagonisms, feuds and all sorts of things in the field.

PROBLEMATIC CASES: WEIGHING DIFFERENT DECISION CRITERIA

Case 9

In another type of problematic decision the program director must weigh different criteria of evaluation. Suppose that a proposal is considered good potential science, but the budget is very large (but necessary), or that the research facilities are poor at the applicant's institution. How does the program director handle such complications? The case that we use for illustration received strong affirmative evaluations of the competence of the principal investigator (with one notable exception), but the additional factor—the size of the budget—attracts the attention of the reviewers. The first reviewer rates the project between excellent and very good:

This is a very good application from a solid worker. The problem is interesting. They have made good progress and will continue to do so. The budget is very large and it seems to me that this should be supported by NSF. I don't think the application is so outstanding as to deserve an unusual outlay on the part of NSF, but some support should be given if needed.

A second referee, who rates it as "very good," has more reservations:

First, the investigator, Dr. X, is a very active researcher, from a good department, and has published many papers. He manages to attract good postdoctoral students and must be regarded as experienced and a well established man in his field. . . . Critique: This application is terribly verbose, disorganized and confusing to someone outside this particular area of research. Nevertheless, it is obvious that this research plan contains a lot of good ideas, and that the applicant has both the expertise and the manpower to do such experiments. On its scientific merits, this is clearly a very good application. However, I feel very strongly that Dr. X is already very generously supported. At a time when many bright young scientists cannot get funded, it seems almost unethical that others receive enormous sums. This may be understandable to top people, but Dr. X does not appear to be one of them. I would not fund this application for the reasons given above.

Finally, a third reviewer rates the project as "excellent," but voices some qualifications. One can sense some ambivalence in his comments:

_____ is a brilliant _____ . He has been extremely productive in many aspects of _____ and has been, in recent years,

one of two or three leading investigators in the hot field of _____ We toil in the same vineyard and he scoops us three times out of four. His record is truly excellent and promises to remain so even after he assumes some administrative responsibilities at [University C]. There is no point in my analyzing his research. I know it is truly outstanding. He deserves, without question, continuing NSF support. The only real question is how much the NSF should support his move to [University C]. Having pushed him for this and other jobs, I feel somewhat responsible and desperately want him to succeed. He should, and will, do a beautiful job. I also know that he is entering a research vacuum and needs all the equipment and staff he asks for. However, I cannot see how the NSF can afford to build up the whole show. He deserves everything he asks for, but he simply ought to get most of it elsewhere. . . . Therefore, my recommendation is to give him everything he wants, but since that is impossible, I would fund him at a level higher than he gets now.

The program director's final decision was to fund, commenting: "An excellent project from a seasoned _____."

The program director in this case is asked to make a decision about the appropriate weight to give to two factors—the quality of the proposed work and the size and type of NSF commitment to a research program. There is little disagreement on the track record or quality of the proposed research, but there is a general question about whether the NSF should be in the business of building a research program for the applying scientist at his new location. The program director in this instance chose to ignore the monetary considerations in his summary, although he did limit the requested funding to 1 year rather than the requested 3 years of support.

PROBLEMATIC CASES: APPARENT DISCREPANCIES BETWEEN REVIEWS AND THE FINAL FUNDING DECISION

Let us consider even more complex cases—those in which the decision seems to be at odds with the reviewers' evaluations. It is important to note that the case described below is a statistical anomaly. Moreover, it is drawn from a program in which the program director believes in taking a far more active role in the decision-making process than do most of the other directors that we interviewed. The case is presented because it illustrates several important problems in the peer review process.

Case 10

The proposal, submitted by a single principal investigator, was reviewed by four reviewers: one rated it "excellent," two "very good,"

Interpretations of Reviews by Program Directors | 103
and the fourth "good plus." Here are some selected, but entirely representative, comments by the reviewers:

Reviewer 1:

This is an excellent proposal which should lead to development of new techniques for _____. In this area Professor [B] has succeeded in providing _____. In the past, he has had remarkable success and should continue. The proposal is totally deserving of support.

Reviewer 2:

The determination of the structure of _____ is a difficult and important problem, and at present time there is little information in this area. The proposer hopes to deal with the question of the _____ and this would have many applications to _____ as well as shedding light on _____. I feel that these questions are important and that the proposer brings strong qualifications and techniques to those problems.

Reviewer 3:

The problem of funding _____ is indeed difficult. Consequently, I believe the proposal is worthwhile. In view of the proposer's past accomplishments, I believe that he is likely to accomplish his primary objectives. I do question the proposed duration of the grant.

Reviewer 4:

My only reservation about this proposal is that it is not clear why one should study _____. It is of course important to find units in _____, but I do not see any reason why units of the form stated in the proposal are useful. However, it does appear that the proposer will probably find some useful results. . . . In conclusion, I only question the value of the proposed research. However, I should add that I am not really an expert in this [specialty]. I am sure that those persons who work in a field more closely related to the proposal can give a more accurate judgment of its worth. I rate the proposal between "good" and "very good," but closer to "good."

This case is particularly interesting only because the project was declined for funding. In view of the material in the jacket, this decision seemed without explanation. While we could not understand the highly technical comments about the science, the quality evaluations were explicit and clear enough. The ratings seemed very high, and the one negative comment on the proposal was by a referee who, by his own admission, was not a specialist in the research area covered by the proposal. In a second interview with the program director, we asked

S: I remember that one well. We play funny games here. Once in a while you get a proposal and you are absolutely certain about how the proposal is going to turn out. I find that a lot of the problem involved in sending proposals out is knowing the reviewers and interpreting them. This is the hardest part. It's always a problem being able to calibrate the extent to which a reviewer reflects the value of a proposal. That's something of a fine-tuning process that I think just comes after a while. What I will frequently do with a proposal that I know very well is send it out to people that I have never used before. That's what I did with this one. . . . I have never before this used either _____, or _____, or _____ [the three reviewers who gave the proposal strong, positive reviews]. This is three of the four reviewers. Reviewer "X" I knew. I know [the principal investigator] and I know his work. [Professor X], let me say in advance, on the basis of his track record, is a very, very polite gentleman who has never done anything substantial, who works on problems that are really quite routine, comes up with reasonable results. In the best of all possible worlds, or even in the very late sixties, when our funding level [was higher], we would have supported X. It's reasonable work, it's of some interest, it's the sort of thing that will be publishable, somebody will read it and grunt over it, but certainly nothing outstanding. . . . What I did with this one was send it out to four people, and I discovered of those four people, three of them were reviewers who, for one reason or another, were overly generous. So, what I'm saying is that this isn't really a fair application of the review process because in some sense I had a good idea of what the reviews were going to say in advance, or *should have said in advance*. I was using this proposal not to test the reviews of the proposal but to test the reviewers themselves. [Emphasis added.]

I: We have some notes here of the review of [the fourth reviewer who rated the proposal between "good" and "very good"] where he says, "I should add that I'm not really an expert in _____ theory. I am sure that the persons who work in the field more closely related to the proposal can give a more accurate judgment of its worth."

S: Excuse me for interrupting, but that's a gentlemanly way of saying that this [research area] isn't really all that interesting. [Reviewer 4] is a very, very brilliant young man, and he is right there in the middle of the mainstream of [the field] and his statement of "I am not an expert. . . ." can be interpreted as that this area of [the field] isn't really all that interesting.

If an observer from outside the scientific community—for example, a GAO auditor—examined this case without detailed knowledge of the

field, it might appear to be a perfect illustration of inequity or bias—a case in which the program director disregarded the peer reviews—and it may indeed be such a case. A list of reasons that could lead a program director to decline this proposal with no sound basis for the declination could easily be developed. Here we seem to have an example of the excessive exercise of power, and this could be ammunition for the critics of the present structure of peer review. When the program director states that he knew beforehand what treatment the proposal should receive in peer reviews, he suggests that his judgment precedes the review process. Moreover, he implies that if the reviews differ from his preconception, he will alter his evaluation of the reviewers before he alters his preconception.

It may be, in fact, that this program director has for some functionally irrelevant reason declined a perfectly acceptable, even high-quality proposal. But before we can conclude that, we must consider closely some of the points made by him that may at first seem like rationalizations for a bad decision. In effect, he raises the question of weighing reviews from peers based on his own knowledge of their scientific capabilities and of their past records as reviewers. This seems like an acceptable mode of operation. Some reviewers tend to be very tough on proposals; others lenient. A casual perusal of comments, without knowledge of a reviewer's own history, could lead an outsider to incorrect inferences about the decision process. Further, some reviewers may be relatively unknown, both as scientists and as peer reviewers, while others are "world class." A terse comment by the best man or woman in a field, or one that damns a proposal with faint praise, might in fact provide a program director with more useful information than he could possibly get from five or six less-qualified reviewers. What is the program director to do, therefore, when he is faced with conflicting reviews in which the reviewer he believes to be most qualified is in a minority?

The interview excerpt quoted above also presents other problems. In his *post hoc* interpretation, the program director tells us that he can read between the lines of reviewer 4, that he can tell that the reviewer is actually quite negative about the substance of the proposal, that the reviewer really is unenthusiastic about it, and that the reviewer believes that it should be declined. Clearly, on the surface this position is not very convincing, and it may be that even below the surface, the reviewer was not trying to communicate a negative evaluation to the program director. In some cases—particularly in those involving a long-standing relationship between the reviewer and the program director—subtle judgments can be conveyed through the use of language. Most scientists who serve on review panels, or who have had

occasion to read many referee reports for journal articles, are fully aware that there are code words and modes of negative judgments in such reports.

The program director says he uses certain proposals not only for evaluating the scientific merits of their content but also as a device for evaluating new reviewers. The initiation of new reviewers into the peer review system is not done systematically at the NSF.

Finally, this example underscores how difficult it is in the end to demonstrate conclusively that particularistic criteria have influenced a specific decision. To be sure, this case represents a statistical anomaly, but even here it is difficult indeed to "prove" that an unfair judgment has been reached or, correlatively, to "disprove" the claim that it was a fair judgment.

A PARADOX ABOUT INFLUENCES ON REVIEWER RATINGS

Even the casual reader of this report must have noticed an apparent inconsistency in the data reported. On the one hand, the quantitative data on the influences on ratings suggest that the personal and social characteristics of principal investigators do not affect in a major way the evaluations of quality by peers. On the other hand, we have presented qualitative data—drawn from interviews with program directors and panel members as well as from NSF proposal jackets—that suggest that the track records of particular principal investigators explicitly enter the peer evaluations and have an important effect on funding decisions. How is it possible to find frequent references to the track records, eminence, and social statuses of scientists in actual reviewer comments and yet find low correlations between these same variables and proposal ratings?

This apparent discrepancy might be explained in several ways. One explanation is that referees first read and evaluate the contents of scientific proposals on their merits. They decide whether they should be funded or not and then construct arguments to support their decisions and justify their ratings. They comment on the technical science and then, because of explicit instructions from the NSF, they refer to the track records of the principal investigators and their ability to carry out the proposed research. This explanation assumes that the basic adjectival evaluations are made on the contents; the proposals are given tentative ratings; and then the qualitative comments are constructed to conform to NSF criteria and to justify the evaluations of

the proposals. In justifying ratings it is certainly easier and less time consuming to make evaluative comments on applicants' qualifications than to make substantive comments on the contents of proposals.

At first this explanation seems almost naive, because we can instantly think of cases in which the evaluation was influenced by knowledge about principal investigators. The first two things that most reviewers look at are the titles and names of the principal investigators. As soon as the name is observed, a set of "evaluations" has already been made. But this may be true only for a very small fraction of the population in a field—those who are widely known.²

Each of us can think of a few scientists to whom we would give excellent ratings even if their proposals were disappointing. In some such instances, reviewers may actually doubt their ability to understand correctly the contents of the proposals. As one program director said:

There are certain [scientists] who have been doing such good work for a while that unless something drastic happens to them mentally, you know that as long as there is federal support in [the field], these guys should get support.

These few scientists notwithstanding, it seems plausible that the ratings of proposals and the evaluations of the characteristics of principal investigators could be quite independent of each other. Further, even if a reviewer knows an applicant's work, in most cases he can keep that knowledge from influencing his evaluation of a current proposal. We would conjecture that only in a small number of cases does the high quality of past performance override the judgment that a proposal is weaker than expected.

Another explanation for the seeming discrepancy between the quantitative and qualitative data relates to self-selection of applicants for NSF grants. It is possible that a substantial majority of applicants have good track records and that differentiation among them is usually based upon the perceived quality of the proposal. Clearly, one possible reason for the low correlation between ratings and applicant characteristics is that most applicants to the NSF are already among the top scientists in their fields. Thus, if many scientists are going to be denied funding, a good share of them are apt to have substantial track records

²Data from an unpublished study (Cole, 1971) indicate that random samples of scientists located at ACE-ranked institutions generally have low visibility to their colleagues. The average scientist's name is known to only approximately 30 percent of the people working at these same ACE-ranked institutions.

108 | PEER REVIEW IN THE NATIONAL SCIENCE FOUNDATION
or to be located at major centers of scientific activities. The presence of substantial numbers of relatively eminent scientists among those denied funding, of course, reduces the association between ratings and the characteristics of the applicants. At this point we have only limited and fragmentary data on the extent to which NSF applicants differ in eminence from a random sample of scientists in their fields.

SECTION

5

Influence on Decision

Thus far, in tracing the course of a proposal from the time that it is submitted until a decision is made, we have investigated the ways in which reviewers are selected, the possible interactions between reviewers and applicant characteristics, and the influence of principal investigators' characteristics on ratings given by peer reviewers. After the peer reviewer makes his evaluation of the proposal, he sends his rating back to the program director. When the program director has received the reports of mail reviewers and when the panel (for those programs with panels) has met, the program director must make a decision to fund or not to fund. In the previous section, we discussed the ways in which the program director interprets reviews of proposals.

We would expect that the ratings given proposals by peer reviewers (both *ad hoc* mail reviewers and panels) should be the most important influence on program directors' decisions. We also would expect that program directors would exercise some discretion in making grants and that there would not be perfect correlations between ratings given by peer reviewers and funding decisions. In this section we attempt to estimate the effect of peer ratings and some of the characteristics of applicants on funding decisions.

Identifying the factors affecting the granting decisions presents a difficult methodological program—how to deal with a dichotomous dependent variable. Although conceptually the proposals could be ranked from the “most meritorious” to the “least meritorious,” in fact they must be divided into two groups: those to be funded and those to

be denied funding. To deal with this problem, we exhibit and analyze tables as well as use special analytic techniques that rely upon additional assumptions. Each method has advantages and disadvantages. For the purpose of assessing the effect of one independent variable when the others are fixed, the most satisfactory approach is probably that based on contingency tables, which does not require the assumptions of normality and linearity. However, for each of the 10 programs we have approximately 120 cases, too few for an analysis of this type to yield a reasonably precise quantitative measure of the effect of one variable when the others are fixed. Therefore, we decided to use correlation and multiple regression techniques. The particular form we use is probit analysis, which was developed by biologists studying drug dosage levels and which has been used extensively by economists and political scientists for analyzing data involving dichotomous dependent variables. (See Appendix B for a discussion of this method.) These techniques assume that the probability of funding is a normal distribution function evaluated at a linear function of the independent variables. We do not have enough data to test whether these assumptions are valid throughout the study. In fact, some of the tables show that the linearity assumption does not hold. In these cases the tables give more information than probit (see Table 54, for example, and the discussion concerning it on pages 145-147).

We are aware of the deficiencies of probit analysis described above, and also of the desirability of discovering a reasonably accurate structural model rather than simply imposing a model (probit) from which to infer the degree of effect of certain variables. Nevertheless, we have chosen this statistical tool for two reasons: (1) It is standard methodology for social science that yields a measure of association R^2 , which is at least qualitatively meaningful, even if the assumptions are not satisfied exactly. (2) When we compare qualitative results inferred from the tabular data with the probit results and with the standard regression analysis (which also assumes linearity) we have found almost complete agreement.

Table 35 presents the proportions of variance in decisions explained by the mean ratings given proposals by mail reviewers and the proportions of variance explained by the mean ratings given by panels (for programs that use panels). The last column of Table 35 shows the total amounts of variance in decisions explained by both the ratings given by mail reviewers and the ratings given by panel members. (These are estimates from probit multiple regression equations.) With two exceptions, the ratings given by peer reviewers (both the *ad hoc* mail reviews and the panel reviews) explain the bulk of the variance in decisions

TABLE 35 Probit Estimated R^2 for Rating of Reviewers (and Panels Where They Exist): 10 Programs

Program	Dependent Variable = Decision		
	1 Mean Rating Given by Mail Reviewers	2 Mean Rating Given by Panel	3 Total (Mail + Panel)
Algebra	0.76	—	0.76
Anthropology	0.63	0.52	0.83
Biochemistry	0.74	0.52	0.86
Chemical Dynamics	0.92	—	0.92
Ecology	0.44	0.70	0.77
Economics	0.50	0.56	0.78
Fluid Mechanics	0.57	— ^a	0.57
Geophysics	0.49	—	0.49
Meteorology	0.92	—	0.92
Solid-State Physics	0.70	—	0.70

^aPanel ratings for this program were unavailable.

made. In fact, in two of the programs, chemical dynamics and meteorology, almost all the variance in decisions is explained by the ratings given by mail reviewers.

The two programs in which less variance in decisions is explained by ratings are geophysics and fluid mechanics. The geophysics program uses both *ad hoc* mail review and a panel. Unfortunately, however, no data are kept on panel meetings or the ratings given by panel members. Thus, the ratings and the influence that they presumably exerted on the program director cannot be included in this quantitative analysis. However, if we compare geophysics with other programs that use panels, we can estimate that if we had the data on the panel ratings, the total amount of variance explained by the peer review evaluations would be significantly greater for geophysics. The mail review ratings alone explain 49 percent of the variance in that program. This is comparable to the amount explained by the mail ratings in other programs using panels. For example, in economics, 50 percent of the variance was explained by mail review ratings alone, and in ecology, 44 percent was explained by mail review ratings alone. In biochemistry and anthropology, slightly more variance was explained by the mail reviews alone. If geophysics were run the same as, let us say, ecology and economics, we can presume that the total amount of variance explained by both panel ratings and mail review ratings would be approximately 75 percent. Without data on the panel ratings this must remain a plausible speculation.

Thus, upon closer inspection, we tentatively conclude that there is probably no contradiction between the accounts of the program directors and the data reported in Table 35.

ACCUMULATIVE ADVANTAGE

Although the ratings given by the peer reviewers are clearly the most significant influence on program directors' decisions, the correlation between these ratings and decisions, as we have said, is not perfect. We wanted to observe the extent to which the program director was influenced by some of the personal characteristics of applicants.

Past work in the sociology of science has led us to expect that eminent scientists would be at an advantage in the competition for NSF funds. A social process called "accumulative advantage" (Merton, 1942; Cole and Cole, 1973; Allison and Stewart, 1974; Zuckerman, 1977) is related to this expectation. In this process, people who have been rewarded at time 1 have a better chance of being rewarded at time 2, independent of the quality of their role performance at time 2.

We have enough data to make a rough test of the extent to which accumulative advantage operates in the NSF peer review process. Except for the effect of past funding history, the data show little evidence that it significantly affects the funding of proposals by the NSF. Since, however, there is a growing literature in the sociology of science developing this concept, it is important that we explain its meaning and expand on our findings. The idea is a simple one: it is predicated upon the notion that young scientists who are most talented and who have opportunities for obtaining resources and facilities can, in fact, make use of those resources and facilities to become more productive scientists than others within their age cohort who do not have the same access to resources and facilities.

Critics of the peer review process as it currently operates assert in so many words that accumulative advantage, not based upon demonstrated ability, operates in a self-reinforcing way to exclude able scientists from the funds needed for research. An "in group" of scientists, predominantly located at the most prestigious and powerful universities, becomes established, whose members give funds to their own colleagues, thereby enabling them to publish more scientific papers and to lay claim to future resources on the basis of their established track records.

Few scientists would argue that past performance should not count in evaluating potential to make scientific discoveries. But how much do

The reason for the relatively small amount of variance explained in decisions by peer review ratings in the fluid mechanics program is difficult to determine. It is possibly a result of the management style of the program director. It is possible for a program director to make decisions that do not completely correspond with the mean ratings of proposals. It is also possible that the fluid mechanics program director did not depend as much upon the adjectival ratings as upon the verbal comments made by the peer reviewers and that the correlation between the adjectival ratings and the substance of the comments was less in this program than in others. Nonetheless, a substantial amount of variance in decisions is explained in fluid mechanics, the program that we believe shows the least influence of the peer reviewer ratings.

The general conclusion to be drawn from Table 35 is that the crucial variable in determining whether or not a grant is made is the mean of the adjectival ratings given by the mail reviewers and the panel reviewers. This seems to contradict the accounts of the program directors that we reported in the previous section. Most of the program directors told us that they paid more attention to the substantive comments than to the numerical or adjectival ratings. There are several plausible explanations for this seeming contradiction.

1. The program directors may have exaggerated the necessity of being able to interpret the substantive comments in order to enhance the importance of their role.
2. As noted above, there is a high degree of correspondence between the substantive comments and the numerical ratings. Therefore, we would expect to find a high correlation between mean ratings and decisions even if the program directors were, in fact, basing their decisions primarily on the substantive comments.
3. It is in the problematic cases, in which the program director's decision is difficult, that careful examination of the substantive comments is the most important. These difficult cases take up most of the program director's time and therefore stand out in his view. Thus, the program directors may have been more likely to report their procedures on these cases than on the numerically more frequent ones that were less problematic.
4. The correlation between peer review ratings and decisions was not perfect. This may reflect the lack of perfect correspondence between the substantive comments and the numerical ratings. For cases in which the two parts of the peer review evaluation do not correspond, the program director may be giving greater weight to the substantive comments than to the numerical ratings.

track record and other indications of status count in the peer review process, and what evidence exists to substantiate or repudiate claims that accumulative advantage operates to reward some scientists at the expense of others equally qualified? These questions were raised in our interviews with program directors.

Program directors differ on how much weight should be given to track records in evaluating proposals. Some believe the track record is most important:

I think that track record counts very heavily. I would rather fund a somewhat average proposal from an outstanding investigator than I would fund an outstanding proposal from an average investigator, simply because it's not so difficult to sit down and say, "This is an obvious problem; it should be solved." Everyone can pick out important problems, but not everyone can solve them. This has to be taken into account very heavily.

On the other hand, some reviewers and program directors believe that funding decisions should be based solely on the merits of current proposals. One program director articulated this position:

We put a lot of emphasis on the document. Sometimes they are not as good as they should be. We faced this issue specifically in one area where some obviously competent people turned in proposals that said, "Send me money, I'm so and so." If we say, "All right, we know what he's doing; we ought to give him some money," it would be unfair to young investigators. So we have decided to take a deliberate stance and go with the document. It's got to be a good document, and if it's a famous investigator, then he better "pull his socks up."

Critics of peer review have expressed the extreme view that it is virtually impossible for eminent scientists to have proposals rejected, regardless of their quality. We talked to people at the Foundation about the treatment of eminent scientists. Dealing with proposals from eminent scientists occasionally does present problems for program directors. When the proposals from eminent scientists are imaginative and well prepared, they present few problems. Problems arise for program directors, however, when an eminent scientist submits a proposal perceived by the reviewers as being of lower than expected quality.

Most program directors believe that a relatively poor proposal from an eminent person generally means one of two things: (1) the person has peaked and is no longer doing high-quality work or (2) the person has not taken the time to write a good proposal because he has the attitude that if he asks for money, he should "automatically" get it. As one program director put it:

If we get a lousy proposal from an eminent person, we have to decide whether he is on his last legs or whether he made an assumption that everybody knew what he could do.

The assessment that an investigator is no longer capable of doing such good work is a difficult one to make, of course. This is largely dependent on the program director's ability to get an evaluation of the investigator's most recent work. As one person said:

When one of these people gets reviews that say, "All this person has done is dick around with definitions for the last 3 years," it won't go.

Another program director spoke of the difficulty in making such a judgment about a prominent scientist. He said:

This is tremendously difficult. Here is somebody that's really done something outstanding in the past, he's set up a whole subfield of science, and we're sitting here playing God, saying that his research productivity is coming to an end.

In cases in which the prevailing opinion is that the principal investigator has not taken the time to write a good proposal, different positions are taken by program directors. Some report that they are apt to give a person the benefit of the doubt and fund the proposal.

A variation on this position is to fund, but at a lower level than requested. A program director described this in the following way:

I got a so-so proposal from an "eminent." Since he was eminent and I would expect him to continue to do good work, I telephoned him and told him exactly what was going on. I told him, I'm not going to put \$50,000 into this, but I will give you \$20,000 with the expectation that good science would be done and that a better renewal would come in.

There are other ways in which reduced funding is employed in these cases. Some program directors give people terminal support, hoping that this will be taken as a threat and "light a fire" under the investigator and that he will resume working at the level he had worked at in the past.

Other program directors handled this discrepancy by declining to support the principal investigator but strongly suggesting that the proposal be resubmitted in a better, more complete form:

If they're still productive and through complacency just wrote a thin proposal, we won't banish them to Siberia. We'll say, "Give us some more details; what is it you're really up to?" And they'll probably make it.

There are those, though, who take a harder line and decline proposals from “eminent” with no implicit or explicit promise of another look. Some are just put off by the attitude that all one needs to do is submit a blank check because one is eminent. A program director described such a case:

There was one principal investigator who said he was too busy to take the time to write a good proposal. He got turned down. He said, “Do you mean to tell me that if Newton came to you, you would refuse him?” I told him, “Unfortunately, we do not deal with Newtons in our daily lives here.”

Another said:

There were a number of cases where we received proposals from eminent people in the field and the proposal was three pages. It said something like, “I’m me, give me the money.” There was no way that person was going to get something out of me or the panel. We were insulted.

What can we conclude from these qualitative interviews about the effect of accumulative advantage on peer review decisions? In general, program directors testify that the major factor in decisions is the comments and ratings given to proposals by mail reviewers and panel members. To the extent that these reviewers include in their evaluations comments about the track records or prior performance of principal investigators, such factors may enter into the decision-making process. But few program directors suggested that the status of a scientist in the social-stratification system of science was a major factor in their decisions. They asserted, with a number of exceptions, that their decisions were, for the most part, not particularly influenced by knowledge of the scientist’s academic affiliation, his rank, prior history of funding, and the like.

Questions about sex and race bias only elicited the response that few women and minority group members applied for funds, and that, when they did, these functionally irrelevant statuses had no influence on decisions. Program directors tended to stress a particular concern with younger scientists beginning their careers. Special efforts were made to get people going on research, not to “turn them off” if they had any potential—indeed, to favor the young in cases in which their proposals were similar in quality to those of older scientists.

But these assertions and comments by program directors are, we would expect, somewhat normative responses. We cannot expect program directors, section heads, and other NSF officials to state openly that they favor one type of scientist over another when they cannot distinguish them by the quality of their proposals. Actually,

some were very candid about the difficulty in arriving at decisions in closely competitive situations. To ascertain the extent to which the characteristics of applicants influence program directors’ decisions we turn to an analysis of the 1,200 applicants in 10 programs.

INFLUENCE OF PAST RESEARCH OUTPUT

To what extent do eminent investigators actually have a better chance of receiving grants, independent of the reviewer ratings of their proposals? We have no direct measure of “eminence.” Citations to past work and numbers of papers published, however, are good indirect indicators of scientists’ positions in the social-stratification system of science.

In order to analyze the relationship between scientific output as measured by citations and number of papers published and granting decision separately for each program, we shall use probit analysis.¹ The results are presented in Table 36. To be perfectly clear about the content of Table 36, note that it contains six probit regression equations for each field. The first three probit equations presented in columns 1, 2, and 3 represent two-variable relationships, one independent and one dependent variable. The data presented in column 4 represent probits involving three independent variables, that is, the three productivity measures: citations to recent work, citations to old work, and number of papers published in the last 10 years. Column 5 presents a probit regression that includes only the mean rating and the panel ratings. Column 6 contains probit regressions that include five independent variables: the mean rating; the panel rating; and the three productivity measures above. Column 7 simply represents the differences in the estimated R^2 that are obtained by subtracting the figures in column 5 from those in column 6.

The first three columns of Table 36 show the amounts of variance in each of the 10 programs accounted for by each of the three research-productivity measures separately. The data in these three columns are less significant for what they tell us about the influence on program directors’ decisions than they are concerning the types of scientists who are more or less likely to receive NSF grants. We would have the same expectations for these data that we had when we analyzed the influence on reviewer ratings. We would expect that scientists who had done the most important work in the past, as measured by the numbers

¹See the caveat at the beginning of this section.

of their publications and citations of them, would be more likely to receive NSF grants than those whose work had been judged to be less important. The data in column 1 of Table 36 indicate great variation in the extent to which citations to recent work (work published between 1965 and 1974) are correlated with decisions. The estimated R^2 goes from a low of 0.03 for anthropology to a high of 0.37 for biochemistry. However, in five of the fields—biochemistry, chemical dynamics, ecology, economics, and solid-state physics—this variable has at least a moderate correlation with decisions.² There is also a moderate correlation between citations to recent work and decisions in geophysics and meteorology. In fluid mechanics and algebra the correlations are lower.

What can we conclude from these data? Although the strength of the correlation varies significantly among the 10 programs, in most of the programs, scientists who have done work often cited in the past are considerably more likely to receive grants from the NSF. The correlation, however, is far from perfect. This indicates that the grant proposals of some scientists who are often cited are being turned down, while at the same time, the proposals of other scientists, with fewer citations, are being approved.

These findings suggest several interesting questions. Why are the proposals of scientists whose recent work is very highly cited turned down? Are these scientists becoming less productive (a fact known to reviewers although not evident from the longer-term citation index) or do they continue to do important work but do not write detailed, good proposals? Also, why are scientists funded whose work has not been cited at all? Are these predominantly young scientists without track records, or are grants being given to some established scientists whose work has not been widely used by the scientific community? We searched for answers to these questions, looking in depth at the "deviant" cases in 3 of the 10 fields: chemical dynamics, biochemistry, and solid-state physics. Deviant cases are here defined as scientists in the highest citation quintile who are *not* funded and scientists in the lowest citation quintile who *are* funded. We pulled the jackets on all these cases in three programs (biochemistry, chemical dynamics, and solid-state physics), and considering the vitae of the scientists, reviewers' comments, and comments of program directors, we assessed the reasons for funding decisions.

The general conclusion of this analysis was that the decision-making procedure followed by the NSF is, indeed, open. The proposals of some

²By social science standards, some of these correlations would be considered large.

TABLE 36 Influence of Productivity Measures on Decision (Probit Estimated R^2)

Program	Citations to 1965-1974		Citations to Work before 1965		Publications between 1965-1974		Three Productivity Measures		Panel Rating, Plus Three Productivity Measures		Difference
	1	2	3	4	5	6	7	8			
Algebra	0.08	0.04	0.00	0.14	0.76	0.82	0.06	0.82	0.06	0.06	
Anthropology	0.03	0.02	0.00	0.04	0.83	0.86	0.03	0.86	0.03	0.03	
Biochemistry	0.37	0.10	0.18	0.38	0.86	0.86	0.00	0.86	0.00	0.00	
Chemical Dynamics	0.31	0.10	0.24	0.32	0.92	0.93	0.01	0.93	0.01	0.01	
Ecology	0.20	0.07	0.03	0.22	0.77	0.80	0.03	0.80	0.03	0.03	
Economics	0.19	0.03	- ^a	0.26 ^a	0.78	0.78 ^a	0.01	0.78 ^a	0.01	0.01	
Fluid Mechanics	0.08	0.20	0.05	0.21	0.57	0.66	0.09	0.66	0.09	0.09	
Geophysics	0.15	0.08	0.20	0.21	0.49	0.59	0.10	0.59	0.10	0.10	
Meteorology	0.12	0.01	0.05	0.13	0.92	0.93	0.01	0.93	0.01	0.01	
Solid-State Physics	0.24	0.18	0.13	0.32	0.70	0.77	0.07	0.77	0.07	0.07	

^aFor this field, publications was excluded, since it had a slight negative association with the dependent variable.

120 | PEER REVIEW IN THE NATIONAL SCIENCE FOUNDATION
quite eminent scientists have been turned down, and many young and relatively unestablished scientists are funded.

There was no evidence for the conclusion that the highly cited scientists who are not funded are "over the hill." All highly cited scientists who were not funded show evidence of continuing to be highly productive. Proposals of highly cited scientists are turned down for several reasons. In some cases the proposals are very weak and receive unfavorable evaluations. For example, a declined proposal in chemical dynamics had two principal investigators, one listing 75 publications and the other 45. The reviewers were virtually unanimous in their conclusion that the proposal contained too little detail and that it was impossible to assess the merits of what would be done on the basis of the proposal. The program director concluded: "The work proposed here is not adequately described. Details of indication of importance are missing." One reviewer commented: "It [the proposal] was written in a casual, almost half-hearted way." Thus, the proposals of experienced, highly published chemists can be turned down if the reviewers agree that they give no indication of important work that will be done with the grant funds.

In other cases, eminent investigators are turned down because they already have other sources of funding. These may be other NSF funds or funds from other agencies. For example, one eminent principal investigator who was turned down in the chemical dynamics program reports that he concurrently held a large grant in the NSF RANN program.³ Another applicant to the chemical dynamics program, a member of the National Academy of Sciences, who was turned down, concurrently held an NSF basic research grant in another chemistry program. He also concurrently held a grant from the Petroleum Research Fund of the American Chemical Society. A rejected applicant in the biochemistry program reports that he is currently funded by the NIH with a \$50,000 grant and has applied for an additional \$100,000 from the NIH. In solid-state physics a rejected eminent applicant reports holding a \$22,000 grant in another NSF program and additional support from the Air Force.

³The unit of analysis for this study has been the individual applicant to the NSF. We have no data, nor does the NSF to our knowledge have data, on whether or not applicants to the NSF have received grants from other agencies in the past. Although eminent scientists may have an only slightly higher probability of receiving a grant from the NSF on a particular application, their overall probability of being funded either by the NSF or in general may be considerably higher. This could happen if they made multiple applications to the NSF and/or other funding agencies. Further research is needed on the varying probabilities of different types of scientists receiving funding from any source.

In some cases that we examined, we found that rejected proposals from eminent investigators had not received uniformly bad reviews but either mixed or good reviews. Also there was no evidence that the principal investigator had funds from another source. It is possible, of course, that in some cases decisions can be explained as mistakes of program directors.

Some applicants have received NSF grants who had either zero or very few citations to work published within the past 10 years. Some of these had published many articles but had not often been cited. As has frequently been pointed out, number of citations is only a rough indicator of the quality of work and should not be considered a decisive indicator in particular cases. Some scientists who have made significant contributions are not frequently cited, and others whose contributions are of less importance have been frequently cited. Thus, the finding that some scientists who appear to be ranked low in the stratification system (as indicated by the number of citations to their recent work) may in part be explained as a result of measurement error. In other cases we found examples of scientists who were relatively young and had little track record who submitted proposals that were highly rated and funded.

Returning to the question with which we began the analysis of the influence of past scientific productivity on NSF decision making, to what extent does past research productivity influence decisions independent of the ratings received? Since the ratings given by peer reviewers are so clearly the most important determinant of decisions, analysis of the influence of any of the principal investigators' characteristics studied in this report on the decisions made required determining how much variance was explained by those characteristics after controlling, so to speak, for the mean ratings of reviewers and panels.

Column 4 of Table 36 presents the estimated proportion of variance in decisions explained by combination of the three research-output measures (citations to recent work, citations to old work, and number of recently published papers).⁴ Column 5 gives the estimated proportion of variance explained by reviewer ratings. (These are the same figures that appeared in column 3 of Table 35.) Column 6 shows the total amount of variance in decisions explained by the three research-output variables and the reviewer ratings.

Subtracting the number in column 5 from that in column 6, we obtain

⁴Note that in some fields (biochemistry and chemical dynamics) there is a good deal of multicollinearity among the three output measures. Also note that in algebra the three variables together explain more variance than the sum of the amount explained separately. This is not unusual or indicative of error in probit equations.

an estimate of the influence of past research output on decisions, independent of reviewer ratings.⁵ This figure varies from a low of 0.00 for biochemistry to a high of 0.10 for geophysics.⁶ It is important to note that we are aware that all the covariance between ratings and past productivity is being assigned to the ratings. For many substantive problems such a stepwise procedure would be misleading, but here we are interested essentially in how much weight in decision making the program director gives to status characteristics beyond those status elements that are "built into" the rating system. In general, we conclude that there is very little evidence that scientists whose past work has been frequently cited and who have published many papers benefit more than slightly from accumulative advantage. Given equal proposal ratings, highly published and cited scientists are not much more likely than those with fewer papers and citations to receive NSF grants. Since we have already shown in section 3 that the ratings are not heavily influenced by applicant characteristics, we can tentatively conclude that accumulative advantage does not play a significant role in the 10 programs we have studied.

Given the very high correlations of mean ratings by reviewers and panels with decisions, it is not surprising that past research performance would add only a small amount to the explained variance.⁷ Program directors are using peer reviews to make their decisions in those cases in which reviews clearly differentiate among proposals.

⁵To be statistically more precise, we have an estimate of the proportion of variance explained by the research-output measures not already explained by reviewer ratings. ⁶Recall that we do not have panel data for geophysics, the addition of which could reduce this figure. Although one could list (for example) the amount of variation explained by ratings not already accounted for by citation index, we have chosen in column 7 to emphasize the amount of variation already explained by the reviewer rating. This is because we are primarily interested in how much weight in decision making the program director gives to status characteristics above and beyond those status elements that are "built into" the rating system.

The reader will also note that here as in section 3 we present R^2 statistics rather than regression coefficients that are part of the structural models. We do this for two reasons: the first is the substantive point made above, and the second is the difficulty in determining the relative weights to be given to various independent variables.

⁷Because the peer ratings explain so much variance there is not much left to be explained by other variables. However, even if we examine the proportion of remaining variance (after ratings have been entered in the equation) explained by the output measures, we find no program in which these variables explain more than 25 percent of the remaining variance.

GRANTING HISTORY

We now study how past funding by the NSF influences current funding decisions. Table 37 shows the proportion of applicants in each program who have not applied for NSF funds within the previous 5 years. In 7 of the 10 programs, 70 percent or more of the sample of applicants have applied for NSF funds between 1970 and 1974. Most of the nonapplicants probably have never applied for NSF funds. The proportion of first-time applicants in algebra, anthropology, and economics is significantly higher than in the other programs. This is probably because these 3 programs have been growing rapidly in the last 5 years.

We also have data on the number of years between 1970 and 1974 in which applicants have received NSF funds. The notion of accumulative advantage would lead us to expect that past grant recipients would have an advantage over nongrant recipients independently of the ratings their proposals received.

Most program directors claimed that the granting history of applicants had little effect on current decisions. Some said that the NSF records of granting history were in such disarray that it was difficult for program directors to obtain such information if they wanted it. Perhaps the program directors were answering normatively, but most claimed that whether or not an applicant was a recent or current grant recipient had little independent effect on decisions. They also pointed out that there was, in general, a fairly high correlation between the quality of past work supported by NSF grants and the quality of current proposals.

TABLE 37 Applicants Not Having Applied for NSF Funds between 1970 and 1974

Program	Percent Not Applying
Algebra	37
Anthropology	48
Biochemistry	30
Chemical Dynamics	24
Ecology	30
Economics	54
Fluid Mechanics	28
Geophysics	23
Meteorology	25
Solid-State Physics	27

Some programs are more concerned with funding history than others. These differences tend to be related to the structure of the scientific discipline. Some projects in physics, for instance, are not really conceived as projects of limited duration, even if program directors receive new proposals every few years from the principal investigators or heads of laboratories. When the NSF invests large sums of money in physical hardware required for specific lines of scientific inquiry, they often think of the investment in relation to a 6- to 10-year period—or longer. Grants that are made with this expectation are carefully reviewed by site visitors and program directors, but given available funds, there is an expected continuity of funding. In such cases, granting history clearly influences the probability of further support. One program director in physics put it this way:

Clearly, if you go into something like that (an \$8,000,000 to \$17,000,000 investment in equipment), you don't casually drop out of it. You are more likely to go in and say, "Look, we've got to beef this program up a little, but with new people, better people, so that we can really get this instrument to perform according to its capabilities."

The data on granting history are presented in Table 38. The data in the first column show that in most of the programs (with the exception of anthropology) past grant recipients do have a higher probability of receiving grants than do those who have received little or

TABLE 38 Influence of Granting History on Decision (Probit Estimated R^2)

Program	Dependent Variable = Decision			
	1 Years Funded, 1970-1974	2 Mean Rating Plus Panel Rating	3 Mean Rating, Panel Rating, Plus Years Funded	4 Difference
Algebra	0.12	0.76	0.76	0.00
Anthropology	0.01	0.83	0.85	0.02
Biochemistry	0.19	0.86	0.86	0.00
Chemical Dynamics	0.14	0.91	0.92	0.01
Ecology	0.17	0.77	0.78	0.01
Economics	0.08	0.78	0.78	0.00
Fluid Mechanics	0.15	0.57	0.64	0.07
Geophysics	0.16	0.49	0.56	0.07
Meteorology	0.07	0.92	0.92	0.00
Solid-State Physics	0.38	0.70	0.77	0.07

TABLE 39 Effect of Mean Rating of Reviewers and Granting History on Decision: Two Fields

Mean Rating of Reviewers	Solid-State Physics, % Receiving Grants			
	Former Grant Recipients	Former Nonrecipients	Totals	
2.00 or less	100 (37)	82 (28)	92 (65)	
2.01 or more	50 (12)	5 (44)	14 (56)	
TOTAL	88 (49)	35 (77)		

Mean Rating of Reviewers	Geophysics, % Receiving Grants			
	Former Grant Recipients	Former Nonrecipients	Totals	
2.00 or less	100 (27)	79 (14)	93 (41)	
2.01 or more	57 (37)	28 (32)	43 (69)	
TOTAL	75 (64)	44 (46)		

no support from NSF in the past 5 years. This is notably true in solid-state physics.

An especially significant question is the extent to which past grant recipients have an advantage over nonrecipients of grants when their proposals receive roughly the same ratings. As the figures in the last column of Table 38 indicate, accumulative advantage has an effect in 3 of the 10 programs: fluid mechanics, geophysics, and solid-state physics.⁸ In these three disciplines, among proposals receiving about the same peer review ratings, the program directors are more likely to make awards to current or recent grant recipients than to scientists who have received little or no NSF support in the past.

To illustrate the effect of this factor in two of the programs, we present the data in Table 39. In Table 38 we showed that granting history had an independent effect on decisions in solid-state physics. In Table 39, we use tabular analysis to show the independent effect of granting history on decisions when controlling for mean ratings by reviewers. We point out that in general, the mean ratings of reviewers determine whether or not grants are given. Ninety-two percent of applicants receiving mean scores of 2.0 or lower (a low number is a

⁸It is necessary to point out again that the finding in geophysics could be a result of not including the panel ratings.

high rating) received grants. Only 8 percent of those receiving such low scores did not receive grants. And only 14 percent of those who received scores above 2.0 received grants. We also point out that 88 percent of past grant recipients received grants, as compared with only 35 percent of nonrecipients. Moreover, the former grant recipients were considerably more likely to receive high mean scores than were the nonrecipients. About three-fourths of the former grant recipients received mean scores of 2.0 or lower, whereas only slightly more than a third of the nonrecipients received scores of 2.0 or lower.

Among applicants who received mean scores of 2.0 or less, 100 percent of former grant recipients received current grants. Among those who had not received NSF grants before but who received essentially the same scores on their current applications, 82 percent received grants. Even more interesting is what happened to applicants who got poorer scores, those higher than 2.0. Among those, fully 50 percent who were former grant recipients received current grants, whereas among people who had not been given grants in the last 5 years who received scores higher than 2.0, only 5 percent received current grants.

This is an empirical illustration of the effect of accumulative advantage. Those who have succeeded in the past have a better chance of succeeding currently, independent of the measure of role performance, that is, the mean ratings of the reviewers. This is a conservative test of the influence of the accumulative advantage factor because the indicator we are using as a measure of role performance, mean rating, itself may have been influenced by the past track records of the scientists.⁹

The second part of Table 39 shows the same relationship in geophysics. Among applicants who received mean ratings of 2.0 or lower, 100 percent of former grant recipients received current grants, and 79 percent of applicants who were not former grant recipients received current grants. Among those who received mean scores above 2.0, 57 percent of former grant recipients and 28 percent of nonrecipients received grants. We have found effects of accumulative advantage from granting history in solid-state physics and, to less extent, in geophysics. It is noteworthy that there was no evidence of accumulative advantage in 7 of the 10 programs we studied.

⁹It should also be pointed out, however, that there may be significant differences in the mean ratings within each of the two categories of this variable. These differences may account for some of the observed effect of granting history.

RANK OF CURRENT DEPARTMENT

A preferred position in the social structure of science is a reward for past achievements and a resource with which to do better research. Many of the advantages of place are self-evident: to be located at a superior university provides opportunities to use superior local equipment and facilities. Facilities needed for some kinds of research are often lacking at less prestigious institutions. Thus a scientist's current location has a probable influence on his ability to obtain additional resources. Further, location in a superior research setting allows for substantial useful interaction with other scientists of quality. How much does rank of current department influence decisions at the National Science Foundation?

We asked program directors whether scientists at the top universities had an edge in the decision-making process. Critics of peer review, including some prior program directors, have claimed that some social characteristics of scientists, particularly their university affiliations, influence decisions. One former program director whom we interviewed commented on this issue:

We really felt pressure from the major institutions to fund all their people. There are schools that for 3 years we never turned down a single individual. Someone turns in a weak proposal from one of these places and gets funded and someone from [low prestige institutions] sends in a strong one and doesn't get funded. We come up against very strong pressure. . . . They expect to be funded and they let you know about it in no uncertain terms. And we're at their mercy. We've used them. They're our top reviewers and frequently they're our top _____.

However, when another person from the same research area was asked about the preceding statement on the influence of institutional location, he said the following:

I don't have the sense that there is that pressure. Bring _____ in and ask him if he feels that there's pressure on him when he wants to decline someone from MIT or Berkeley or a place like that. I think he'll say, "Ugh!" I don't say we wouldn't get calls, but we don't mind declining people from those schools. The fact of the matter is that we don't very often because they are high-quality schools and their people are generally very good. There are cases where we do decline people in those institutions.

We interviewed two other program directors from this section. They reported that they too had frequent contact with scientists in prestigious departments. In fact, they said that they tend to rely heavily on

those scientists for advice and reviewing. Essentially, all these program directors were reporting the same matter—having contact with prestigious scientists. Where they differed, though, was in their assessment of the nature of the contact. One program director experienced it as pressure, while others viewed it as a source of useful counsel.

Data showing the influence of rank of current department on granting decisions are presented in Table 40. Column 1 shows that in 8 of the 10 programs, rank of academic department is correlated in some degree with whether or not a scientist received a grant. The two programs in which it shows no meaningful correlation are anthropology and geophysics. We conclude that in most programs, scientists currently employed at prestigious institutions do have a somewhat better chance of receiving grants (although it may not be because of their institutions) than do scientists in less prestigious departments.

Once again, an important analytic question is whether scientists located in prestigious departments have a better chance of receiving grants than do their colleagues in less prestigious departments *independently of the ratings given their proposals*.

The proportion of variance explained by the ratings of reviewers and panels, plus the rankings of scientists' current departments, is presented in column 3. For example, in algebra, the mean ratings of the reviewers explained 76 percent of the variance. When we add the rank

TABLE 40 Influence of Rank of Current Department on Decision (Probit Estimated R^2)

Program	Rank of Current Department	Mean Ratings Plus Panel Ratings	Reviewer Ratings Plus Rank of Current Department		Difference
			Panel Ratings	Department	
Algebra	0.20	0.76	0.78	0.02	0.02
Anthropology	0.00	0.83	0.84	0.01	0.01
Biochemistry	0.19	0.86	0.86	0.00	0.00
Chemical Dynamics	0.12	0.91	0.93	0.02	0.02
Ecology	0.10	0.77	0.79	0.02	0.02
Economics	0.17	0.78	0.78	0.00	0.00
Fluid Mechanics	0.08	0.57	0.58	0.01	0.01
Geophysics ^a	0.05	0.49	0.49	0.00	0.00
Meteorology ^a	0.07	0.92	0.92	0.00	0.00
Solid-State Physics	0.28	0.71	0.74	0.03	0.03

^aRank of department scores based upon survey of NAS members (see Appendix B).

of the mathematician's department to the probit regression equation, we increase the proportion of variance explained by only 2 percent. In short, rank of department has no meaningful effect above and beyond that of reviewer ratings. If rank of department has any effect at all on funding, it works through its influence on reviewers rather than through the NSF program director and his superiors. We have already demonstrated that institutional location of applicants has only a minor influence on reviewer ratings.

PROFESSIONAL AGE

Most program directors claim that young applicants who have not had the opportunity to build track records are given special consideration. One program director specifically noted attempts to aid young scientists:

Because we try to encourage younger men in the field to get going, to get independent of their Ph.D. advisors, we do add on a few "brownie points" for being a young, new scientist.

Another program director spoke of the advantage that younger people have in the competition for funds. He said:

With young principal investigators we tend to be somewhat gentle—that given the guy has made a conscientious effort and has written quite a nice proposal, we would recommend something at a slightly lower level [mean rating] than I would a senior investigator.

This last point is important. Since the young investigator has a limited track record, if any, there is greater reliance on the content and quality of the proposal. As one program director stated:

Where the name of the young man is not going to mean much to the reviewer, there is not much that he can do about that. So how he writes his proposal has to be the main criterion.

Program directors frequently use additional information to establish or to fortify their own judgments of young people. In one program the known teachers of the young scientist, as well as his departmental affiliation, were used in reaching decisions:

You see what they published, if anything at all, you find out what their thesis advisor says about them, you try to find someone who wasn't as intimate with them as their thesis advisor, and you try to support what you consider the best of the lot.

It has also been noted that decisions made about young people are risky. One program director said:

I feel that it takes a minimum of 13 gambles to get a good payoff. You have to fund 13 people whom it's yes-no business to get one really good person, who 3 or 4 years later you can say, "I gave him his first break."

Others talked of getting "flack" for giving funds to unproven investigators, but in most cases they claimed that these gambles paid off.

Program directors frequently said that if the proposal ratings of younger and older scientists were roughly equal, they would give the edge to the younger scientists, because the older ones should have been able to construct better proposals and because the NSF had an obligation to help young scientists get started. One program director said:

Given two proposals where technically the merit would be equivalent, the bias would go to the young investigator.

In some scientific fields (according to the program director in charge) young people have almost no chance of funding. In physics, for example, many specialties are organized into large research teams, and younger scientists have little chance for independent funding:

The younger people in particle physics don't get a grant. There is no way you could set up your own group as a going enterprise and try to do experiments in particle physics. You need too much equipment, too much money, too much organization, and too much collaboration. The way young people get into it and eventually become independent is by joining existing groups.

Although most program directors suggested that youngsters actually had some advantage, some told us that the social structure of the scientific discipline precluded the funding of younger experimentalists.

To what extent do the data support the contention of many program directors that young people are given special consideration in the decision-making process? In order to analyze the influence of professional age, we divided the sample into two groups: those who received their Ph.D. degrees in 1970 or after (younger scientists) and those who received them before 1970 (older scientists). We did this because most scientists have established some track record by the time they have been out of graduate school for 5 years, and we were not interested here in the influence of variation in professional age among scientists who have already had the opportunity to establish track records. Twenty-nine percent of the applicants were classified as "young" scientists. Forty-six percent of young applicants and 54 percent of

TABLE 41 Influence of Professional Age on Decision (Probit Estimated R^2)

Program	Professional Age	Mean Rating		Difference
		Panel Rating	Reviewer Rating Plus Age	
Algebra	0.02 (-) ^a	0.76	0.80	0.04
Anthropology	0.03	0.83	0.83	0.00
Biochemistry	0.00	0.86	0.86	0.00
Chemical Dynamics	0.05	0.91	0.92	0.01
Ecology	0.02 (-) ^a	0.77	0.78	0.01
Economics	0.00	0.78	0.78	0.00
Fluid Mechanics	0.06	0.57	0.60	0.03
Geophysics	0.00	0.49	0.49	0.00
Meteorology	0.00	0.92	0.92	0.00
Solid-State Physics	0.14	0.70	0.76	0.06

^aIn these fields, young scientists had a slightly better chance of receiving grants than their older colleagues.

older applicants received grants.¹⁰ Thus, young applicants have almost as good a chance of receiving NSF grants as do more established applicants.

Column 1 of Table 41 shows probit estimates for professional age in each of the 10 fields. A negative sign next to the estimate indicates that young people are more likely to receive grants. The only field in which young people are significantly less likely to get grants is solid-state physics. But, as the last column of Table 41 indicates, in none of the fields does age explain much additional variance after reviewer ratings are entered into the probit regression equation.

OTHER CHARACTERISTICS OF APPLICANTS

We have data for three other characteristics of principal investigators: whether they are employed at Ph.D.-granting institutions, the rankings of the departments at which they earned their doctorates, and, for

¹⁰Some of the younger scientists may have been coprincipal investigators with older scientists. To see whether this had any effect on the findings, we took all grants that had more than one principal investigator and, using citations to recent work as a measure of eminence, excluded all but the most eminent investigator from the analysis. We found the correlations between professional age and decision to be virtually identical to the correlations obtained when the coprincipals were included. Thus the results about age have not been significantly influenced by the inclusion of coprincipal investigators.

academic scientists, their academic ranks (full professor, associate professor, or assistant professor). In general, none of these variables had a strong influence on granting decisions. Exceptions are the influence of academic rank and the rank of the department in which the scientist earned his Ph.D. in the solid-state physics program and the influence of rank of doctoral department in the anthropology program. For descriptive purposes, we report the probit R^2 estimates for influence of these variables on decisions in Table 42.

COMBINED EFFECTS OF THE NINE CHARACTERISTICS

We have been examining the extent to which some characteristics of applicants for NSF funds influence the decisions made by program directors. In particular, we have focused on the extent to which eminent, well-established investigators located at prestigious universities have a better chance of getting grants independent of the ratings given their proposals. We shall now discuss the combined effects of the characteristics we have been studying on decision. The data are presented in Table 43.

The first column of Table 43 shows the total proportion of variance

TABLE 42 Influence of Professional Rank, Rank of Ph.D. Department, and Type of Current Institution on Decision (Probit Estimated R^2)

Program	Academic Rank	Rank of Ph.D. Department	Type of Current Institution (Ph.D.-Granting or Other)
Algebra	0.02	0.10	0.01
Anthropology	0.01	0.13	0.02
Biochemistry	0.02	0.02	0.04
Chemical Dynamics	0.06	0.08	0.07
Ecology	0.00	0.00	0.04
Economics	0.01	0.02	0.01
Fluid Mechanics	0.01	0.00	0.00
Geophysics	0.00	0.00	0.05
Meteorology	0.00	0.04 ^a	0.06
Solid-State Physics	0.20	0.12	0.02

^aThe sign on this relationship is negative.

TABLE 43 Estimated Variance on Funding Decisions Explained by Characteristics of Principal Investigators and Ratings of Reviewers (Probit Estimated R^2)

Program	Nine Individual Variables ^a	Mean Rating of Reviewers (and Panels Where They Exist)	Nine Individual Variables Plus Mean Rating of Reviewers and Panels	Increase in Variance Due to Nine Individual Variables
Algebra	0.34	0.76	0.84	0.08
Anthropology	0.17	0.83	0.86	0.03
Biochemistry	0.51	0.86	0.86	0.00
Chemical Dynamics	0.39	0.92	0.96	0.04
Ecology	0.40	0.77	0.86	0.09
Economics	0.39	0.78	0.83	0.05
Fluid Mechanics	0.37	0.58	0.71	0.13
Geophysics	0.34	0.49	0.65	0.16
Meteorology	0.24	0.92	0.94	0.02
Solid-State Physics	0.70	0.70	0.91	0.21

^aThe variables include rank of Ph.D. department, professional age, rank of current department, academic rank, log of citations to work published in last 10 years, log of citations to work published more than 10 years ago, log of papers published in last 10 years, whether or not employed in a Ph.D.-granting institution, number of years funded between 1970 and 1974. We excluded in each field those variables that had negative correlation with decision. These were professional age in algebra, ecology, and meteorology; productivity in economics; and rank of doctoral department of meteorology.

explained by the nine characteristics for which we have data.¹¹ Essentially these characteristics describe the statuses of applicants in the stratification system of science as measured by their past productivity and the current rankings of their departments. They also include their granting histories with the NSF, the rankings of their doctoral departments, professional ages, and academic ranks.

The nine characteristics explain different amounts of variance in grant decisions in the different programs, from a low of 17 percent in anthropology to a high of 70 percent in solid-state physics (column 1 of Table 43). The significant finding, however, is not how much total variance is explained by these variables but how much additional variance above that explained by peer review ratings alone is explained by these variables. That is, when we add the nine variables to a probit

¹¹For a comparison of these results with those obtained from ordinary least squares analysis, see Appendix B.

regression equation in which we have already entered the ratings of peer reviewers, how much does the estimated R^2 increase?

Column 3 of Table 43 shows the total proportions of variance explained by a combination of all the variables, the ratings of the peer reviewers, and the nine characteristics. In column 4, we have subtracted the number in column 2 from the number in column 3, that is, the proportion explained by peer review ratings alone from the proportion explained by the combination of peer review ratings and principal investigator characteristics.

Clearly, in most of the programs we have studied, the characteristics of principal investigators add relatively little to the proportion of explained variance in decisions with three exceptions. The first is geophysics. We have already explained why geophysics is difficult to deal with. Because we do not have data on the panel ratings, it is very difficult to determine the extent to which the individual characteristics of geophysicist applicants influence the program director's decision. However, following the same logic used above, it appears to us that if we did have data on the panel ratings, the total proportion of variance explained by the individual characteristics of the principal investigators would not add much to the proportion explained by the ratings of mail reviewers and panel reviewers. This, however, must remain speculative. The two programs that show the most influence of the characteristics of principal investigators are fluid mechanics and solid-state physics. The nine individual characteristics explained an additional 13 percent of the variance in fluid mechanics and an additional 21 percent of the variance in solid-state physics.

In general, program directors' funding decisions are not heavily influenced by the characteristics of applicants beyond the effects of those characteristics on ratings.

INFLUENCE OF GEOGRAPHIC LOCATION ON DECISIONS

The congressional act regulating the NSF states that the organization should "avoid undue concentration" of research funds. As noted in the introduction to this report, the NSF states that when the assessed quality of applications is roughly equal, geographic location is considered in making decisions. The NSF policy on geographic location is contained in the staff study entitled "Peer Review and Proposal Evaluation" (June 1975, Appendix I, pp. 7-8).

In the general competition for basic research grants, our policy has always been that, where a selection must be made from a group of projects of substantially equal merit, we will select those which contribute to the avoidance of undue concentration. This policy we attempt to apply uniformly in all programs. It should not be interpreted as meaning that projects must be strictly matched on an "all-other-things-beings-equal" basis. Rather, we should think in terms of a quality band or cut-off zone within which a considerable number of projects may be considered as having substantially equal merit. Within that cut-off zone selections should look both to geographical distribution and to distribution by type of academic institution.

The NSF presented a document to the House subcommittee investigating peer review in which they showed that with the exception of a few states (Massachusetts, New York, the District of Columbia, and California), the distribution of awards to the states matched closely the distribution of population, scientists, and eminent scientists.¹²

We have divided the 1,200 applicants we studied into East, South, Midwest, and West.¹³ We found that the proposals of 55 percent of eastern applicants, 35 percent of southern applicants, 51 percent of midwestern applicants, and 60 percent of western applicants were approved. The fact that the percentage of approvals of southern applicants was smaller than the others does not mean that they were discriminated against. The lower ratings could result from either the lower perceived quality of their proposals or the difference between track records and other characteristics of southern applicants and those of applicants from other areas of the country. In fact, as the data in Table 13 suggest, in 7 of the 10 programs, applicants from the South received poorer mean ratings than applicants from the other sections.

The more relevant question is how applicants from the different sections have fared when their proposals received roughly equal peer review ratings. Since we had too few cases in each program to analyze the effect of region for each program, we have combined the data from

¹²This appears in National Science Foundation, *An Analysis of the Geographical Distribution of NSF Awards as Compared with Other Selected Indicators*, 1975c. The figures in this report are based upon the distribution of NSF funds in all divisions.

¹³The groupings are: East—Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Delaware, Maryland, and District of Columbia. South—Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Alabama, Tennessee, Kentucky, Arkansas, Mississippi, Louisiana, and Texas. Midwest—Ohio, Michigan, Indiana, Illinois, Wisconsin, Minnesota, Iowa, Missouri, Oklahoma, Kansas, Nebraska, South Dakota, and North Dakota. West—Montana, Wyoming, Colorado, Idaho, Utah, New Mexico, Arizona, Nevada, California, Oregon, Washington, Alaska, and Hawaii.

Several cautions must be exercised in interpreting these data. First, we have combined the data from all 10 programs into 1 sample. It is possible that in some programs, geographic region is given more consideration than in others. It is also possible that the slight differences we have observed are a result of variables not measured in this study.

AMOUNT OF MONEY APPLIED FOR AND FUNDING DECISIONS

Many observers of the peer review system believe that the size of a project's budget influences its chances for funding. Principal investigators may try to shape their budgets with this belief in mind. Tailoring the budget size to the overall budget of an NSF program is surely required. Information about "ball park" budget figures is often obtained through informal communication between scientists and program officers. Data from the qualitative interviews attest to this fact. These contacts produce some restrictions on the range in budgets submitted to the Foundation. But how much does variation in budget size among submitted proposals affect final decisions? For the more than 1,000 proposals we have studied, we have data on the total amount of money requested.

Using the decision as the dependent variable, we computed probit estimates of the influence of amount of money applied for. Table 45 shows that in no field does the amount of requested funds noticeably

TABLE 45 Influence of Amount of Money Requested on Decision (Probit Estimated R^2): 10 Programs

Program	Estimated R^2
Algebra	0.12
Anthropology	0.01 ^a
Biochemistry	0.04
Chemical Dynamics	0.01 ^a
Ecology	0.00
Economics	0.00
Fluid Mechanics	0.00
Geophysics	0.04
Meteorology	0.10
Solid-State Physics	0.10

^aSign on relationship is negative.

hurt the chances of obtaining an NSF award. In only 2 of the 10 fields, anthropology and chemical dynamics, is there a negative correlation between the amount requested and the decision, and in these fields the associations are so weak that the estimated variance explained thereby is not significantly more than zero. The only estimates worth noting—those in algebra, meteorology, and solid-state physics—indicate that larger projects are more apt than smaller ones to be given awards.

We thought that, although the amount of money applied for had no negative influence on decisions among all the applicants to a program, it might have an influence on decisions for those who had not applied for NSF funds in the last 5 years. We therefore conducted the analysis separately for this group. The analysis showed that even among the applicants in this group, practically all of them new NSF applicants, those who applied for relatively large amounts of money had just as good chances of getting grants as those who applied for relatively little money.¹⁴

TABULAR ANALYSIS OF DATA ON DECISION

As we pointed out in section 3, regression analysis tells us the extent to which particular characteristics of the applicant enable us to predict ratings received from peer reviewers. Earlier in this section, we used probit analysis when decision was the dependent variable. This technique, however, does not allow us to make direct comparisons among people with distinctly different characteristics. In order to do this, we have used tabular analysis. The problem in using tabular analysis for decision is, as was pointed out above, that in most programs we have approximately 120 cases. This is too small a sample size to permit an accurate evaluation of the characteristics affecting decision. Therefore we treated applicants in all 10 programs in a combined sample. We standardized the data separately within each program and then analyzed the data using standardized scores instead of the absolute scores.¹⁵

In assessing the results to be presented below the reader must remember that we have combined the data from all 10 programs. As we have already pointed out, all programs do not operate in the same way. For example, in some programs younger investigators have an advan-

¹⁴Because of small numbers of cases in each program, this analysis treated the entire sample as a unit after standardizing amounts applied for separately for each program. The probit equation was then computed using the standard scores.

¹⁵For a description of the standardization procedure employed, see section 3, pages 53-54.

TABLE 46 Influence of Mail Ratings on Decision

Mean Rating of Mail Reviewers	Percent Receiving Grants	
	All Programs	Panel Programs Excluded
High	92 (382)	95 (239)
Medium	52 (384)	52 (259)
Low	10 (390)	9 (221)

Numbers of applicants are in parentheses.

tage while in other programs they are at some disadvantage. Likewise, in some programs the number of citations is fairly strongly associated with the granting decision, while in other programs there is little or no association between the two. These differences are blurred when we combine the data from all 10 programs into 1 sample. Nonetheless, the tabular analysis presented below does allow us to compare scientists with different characteristics.

Table 46 shows the relationship between the mean rating given a proposal by mail reviewers and the decision. Since the mean ratings were standardized separately within each field, we have roughly divided them into high, medium, and low groups. The first column of Table 46 shows that 92 percent of those proposals given high mean ratings by mail reviewers were funded. Among those given low mean ratings by mail reviewers only 10 percent were funded. Approximately half of the proposals given medium mean ratings were funded. The second column of Table 46 shows the same relationship; however, this time, data from the fields employing panels were excluded.¹⁶ In those fields the opinions of mail reviewers and panels may differ and program directors may be influenced by the panels. When we excluded panel fields, we found an even stronger relationship between the ratings of mail reviewers and decisions. Now 95 percent of proposals receiving high ratings are funded, and only 9 percent of those receiving low ratings are funded.

Among the fields not using panels, 53 percent of the proposals were funded. Therefore, if without any information we had to guess whether or not a proposal was funded or not funded, we would be correct in 53

¹⁶In this section of the analysis we have treated geophysicists as a nonpanel field, since we do not have any data on the panel ratings. When geophysicists were excluded from the data presented in column 2 of Table 46, we found that, among applicants receiving high mean ratings, 95 percent were awarded grants; among those receiving median mean ratings, 49 percent; and among those receiving low mean ratings, 5 percent.

TABLE 47 Influence of Panel Ratings on Decision (Only Fields with Panels)

Mean Rating of Panel Members	Percent Receiving Grants
High	84 (99)
Medium	52 (163)
Low	12 (144)

Numbers of applicants are in parentheses.

percent of the cases. Using knowledge of the trichotomized mean rating score we can guess correctly in 78 percent of the cases, an increase of 25 percentage points.¹⁷

Despite the fact that mean ratings by mail reviewers are so highly associated with decisions, we were still interested in those few deviant cases in which decisions did not correspond with the mean ratings. We listed all proposals in each field according to the mean ratings they received, with the "best" proposals at the top and the "worst" proposals at the bottom. We then looked at which were funded and which not funded. We were then able to identify cases in which decisions did not closely correspond to the mean ratings by mail reviewers. In half of these cases it turned out that there was a difference of opinion between the panel and the mail reviewers and the program director went along with the panel. We are still analyzing these other half to see if we can detect any systematic pattern among these cases in which the reviews did not determine the decision. Thus far, our analysis suggests that those who benefit from such discretionary decisions are slightly more apt to be eminent scientists, but once again the effect is small.

Table 47 presents the relationship between mean rating given a proposal by panel members (for those fields employing panels) and the decision. The association is strong, and the cases in which panel ratings are not in agreement with the decision could easily be cases in which

¹⁷This "modal" guessing maximizes the number of correct guesses both for prediction from the marginals and for prediction within the categories of the independent variable. "Proportional prediction" that would have reproduced the marginal totals on the dependent variable (Leo Goodman and William Kruskal, "Measures of Association in Cross-Classification II," *Journal of the American Statistical Association*, vol. 54, 1959, pp. 123-163) could also have been used. This would have lowered the proportion of correct guesses in both conditions, but their relative magnitudes (in this case 0.502 and 0.71) would have remained approximately the same. This more complex procedure would have been necessary had the mode of the dependent variable been in the same category for each category of the independent variable. Since this was not the case, we chose to use the simpler procedure.

mail reviewers and panel members disagreed and the program director went along with the mail reviewers.

In Table 48 we have divided the 1,200 applicants in the 10 different programs according to the standardized score on the number of citations to recent work. We divided them roughly into quintiles and then examined the proportion in each quintile that received grants. In the highest quintile, 78 percent received grants, and in the lowest quintile, 30 percent received grants. Since a total of 51 percent of the 1,200 proposals were funded, without any additional information we could guess the outcome correctly in 51 percent of the cases. With the citation information presented in Table 48 we could guess whether a project was funded or declined in 63 percent of the cases, or an increase of 12 percentage points.

As one would expect, investigators who have been successful in the past, as indicated by citations to their work, have a better chance of receiving an NSF grant than do those who have not been successful in the past. The effect is considerably larger at the extremes than it is in the middle of the distribution. But even at the extremes the number of citations to one's past work does not completely determine whether or not one receives a grant. Twenty-two percent of those whose work was most cited did not receive grants, and 30 percent of those whose work was least cited did receive grants. Thus, these data lead us to conclude that citations to one's work are moderately correlated with success in getting an NSF grant. We know that citations have some influence on ratings of proposals and that the ratings have a major influence on the decisions of program directors. Perhaps the most significant question is: To what extent are program directors influenced in their decision making by the characteristics of the applicants conditional on fixed values of the scores they receive from peer reviewers. In order to answer this question we must examine a three-variable table in which we look at the proportion receiving grants in different citation

TABLE 48 Influence of Citations in 1974 to Work Published between 1965-1974 on Decision (All Fields Combined)

Number of Citations Received	Percent Receiving Grants
High	78 (239)
	58 (239)
Medium	50 (238)
	41 (240)
Low	30 (244)

Numbers of applicants are in parentheses.

TABLE 49 Influence of Reviewers by Citations in 1974 to Work Published between 1965 and 1974 on Decision, Controlling for Mean Rating (All Fields Combined)

Citations to Work Published in 1965-1974	Percent Receiving Grants			Low Mean Rating of Reviewers
	High Mean Rating of Reviewers	Medium Mean Rating of Reviewers	Low Mean Rating of Reviewers	
High	99 (136)	60 (54)	32 (41)	
	92 (91)	58 (77)	9 (69)	
Medium	90 (61)	55 (92)	10 (78)	
	86 (50)	49 (85)	11 (97)	
Low	78 (44)	41 (76)	2 (110)	

Numbers of applicants are in parentheses.

categories while controlling for peer review ratings. (See Table 49.) Table 49 shows that peer review ratings are far more important than citations in determining whether or not a proposal will be funded.

For example, the data on applicants in the highest quintile of citations show that the mean ratings they received on their proposals had a very strong effect on success in getting grants; with 99 percent of those receiving the highest ratings and 32 percent of those receiving the lowest ratings (a 67-point percentage difference) receiving grants. On the other hand, for applicants who received high ratings on their proposals, citations had a relatively slight influence on success in getting grants, with 99 percent of people in the highest citation category and 78 percent of people in the lowest citation category (a 21-point percentage difference) receiving grants. Further, scientists with the highest numbers of citations who receive medium ratings on their proposals are less likely to receive grants than are those with the lowest numbers of citations who receive high ratings.

The data in Table 49 are somewhat misleading, since they include data from fields employing panels. Panel judgments sometimes differ from those of mail reviewers and thus create the impression that the peer reviews were not the determining influence on decisions. Table 50 is identical to Table 49, except that it includes only those fields not employing panels. This table demonstrates even more clearly that, when ratings of proposals are controlled, citations have very little if any meaningful effect on the likelihood of receiving a grant. The only exception is in the lowest citation category for people whose proposals were highly rated. People in the very lowest citation quintile are somewhat less likely than people in the higher quintiles to receive grants even if they received high ratings. For people receiving medium

TABLE 50 Influence of Reviewers by Citations in 1974 to Work Published between 1965 and 1974 on Decision Controlling for Mean Rating (Excluding Panel Fields; All Fields Combined)

Citations to Work Published in 1965-1974	Percent Receiving Grants		
	High Mean Rating of Reviewers	Medium Mean Rating of Reviewers	Low Mean Rating of Reviewers
High	100 (90)	53 (34)	16 (19)
Medium	93 (59)	57 (56)	10 (40)
	95 (40)	57 (62)	8 (48)
Low	96 (24)	49 (51)	14 (50)
	77 (26)	43 (56)	3 (64)

Numbers of applicants are in parentheses.

or low ratings on their proposals, citations make virtually no difference in the probability of success. The data presented in this table show that citations to the work of applicants, here being used as an indicator of the extent to which their past scientific work has been highly evaluated by the scientific community, have very little if any independent effect on decisions of program directors beyond the effect, already taken into account, of reviewer ratings.

The data from the tabular analysis show that applicants whose proposals receive high ratings almost invariably receive grants and that applicants whose proposals receive low ratings almost invariably do not receive grants, regardless of other applicant characteristics we have studied. The data also indicate that applicants whose proposals receive ratings in the middle have about a 50 percent chance of receiving grants. We have been interested to know what determines grant decisions when the peer review ratings fall into this middle category. At first we thought that when peer review ratings did not provide a sufficient basis for discriminating among proposals, program directors might rely upon the characteristics of applicants and their institutions in making decisions. However, as the data in Table 50 suggest, the characteristics of the applicants, in this case citations, have no greater effect in the middle category of ratings than they do in the two extreme categories. Thus far we have not been able to find any systematic determinant of decision when the peer review scores fall into the borderline area. Apparently, these decisions depended more heavily on the judgment of the program director. Further research is necessary on how decisions are made when the peer review ratings fall into the borderline area.

TABLE 51 Influence of Citations Made in 1974 to Work Published Prior to 1965 on Decision (All Fields Combined)

Citations to Work Published Prior to 1965 ^a	Percent Receiving Grants
High	77 (86)
	69 (86)
Medium	64 (86)
Low	49 (86)
None	46 (856)

Numbers of applicants are in parentheses.

^aFor this table all applicants who had zero citations in 1974 to work published prior to 1965 were placed in the bottom category, and the remaining cases were divided into quartiles using the Z scores.

Tables 51 and 52 show the influence of citations to older work and of numbers of papers published in the last 10 years on the probability of receiving grants. Both these variables have a relatively minor effect on this probability beyond that of ratings by reviewers.

Table 53 shows the influence of rank of an applicant's current department or institution on the probability of receiving an award. Applicants from highly ranked departments are more likely to receive NSF grants than those from lowly ranked departments or from unranked or nonacademic institutions. Without any information we are able to predict 51 percent of the cases correctly. With information on rank of the applicant's current department we can predict 64 percent of the cases correctly, an increase of 13 percentage points in predictability. Here too, however, we are primarily interested in the independent influence of rank of department on decision after we have controlled for mean reviewer ratings of proposals. These data are presented in Table 54. In this table we have excluded data in fields employing

TABLE 52 Influence of Number of Papers Published between 1965 and 1974 on Decision (All Fields Combined)

Number of Papers Published between 1965 and 1974	Percent Receiving Grants
High	62 (238)
	59 (242)
Medium	53 (240)
	44 (239)
Low	41 (241)

Numbers of applicants are in parentheses.

TABLE 53 Influence of Rank of Current Department on Decision (All Fields Combined)

Rank of Current Department ^a	Percent Receiving Grants
High	74 (242)
Medium	56 (150)
Low	61 (175)
Unranked and nonacademic	43 (171)
Gamma = 0.39	38 (462)

Numbers of applicants are in parentheses.

^aWe have used nonstandardized rank of department scores.

TABLE 54 Influence of Rank of Present Department on Decision Controlling for Mean Rating of Reviewers (Excluding Panel Fields; All Fields Combined)

Rank of Present Department ^a	Percent Receiving Grants		
	High Mean Rating of Mail Reviewers	Medium Mean Rating of Mail Reviewers	Low Mean Rating of Mail Reviewers
High	94 (86)	58 (38)	6 (18)
Medium	100 (32)	62 (34)	4 (25)
Low	96 (49)	58 (60)	29 (38)
Unranked and nonacademic	94 (31)	53 (49)	7 (44)
	90 (41)	38 (78)	4 (96)

Numbers of applicants are in parentheses.

^aWe have used nonstandardized rank of department scores.

panels. The data in Table 54 indicate that, when we control for the mean rating of mail reviewers, rank of an applicant's department has virtually no influence on success in getting grants beyond that of reviewer ratings.¹⁸

The only exception to this is among applicants whose proposals receive medium ratings and who are located at unranked or nonacademic institutions. These applicants are somewhat less likely to receive grants than those receiving similar ratings on their proposals

¹⁸Note the 29 percent figure for middle department ranking in the last column, which is very much greater than the percentages for both highest and lowest departments. This shows that the assumption of linearity in regression and probit analysis probably does not hold here. However, to substantiate this, it would be necessary to have several different samples showing the same results, because these percentages are based on a small number of cases.

TABLE 55 Influence of Type of Institution on Decision (All Fields Combined)

Type of Current Institution	Percent Receiving Grants
Ph.D.-granting	55 (922)
Other	40 (278)

Numbers of applicants are in parentheses.

who are located in ranked graduate departments. However, among applicants who get either high ratings or low ratings, rank of department has no independent influence on the decisions of program directors.

Table 55 shows the influence on decisions of the institution at which an applicant is located. Scientists at Ph.D.-granting institutions are slightly more likely than those at non-Ph.D.-granting institutions to receive NSF grants.

Tables 56, 57, and 58 present the effects of rank of Ph.D. department, academic rank, and professional age on decision. Each of these variables has relatively little effect on the probability of receiving a grant.

Table 59 shows the influence of NSF funding history on decision. We have divided the applicants into those who have received NSF funds in the previous 5-year period and those who have not. Among those who have been funded by the NSF in the past 5 years, 65 percent were awarded grants on their current applications, whereas among those who had not received any NSF support in the previous 5 years, 40 percent were awarded NSF grants on their current applications. Again, a question of primary significance is the extent to which NSF granting history has an effect on the decision beyond that of the mean

TABLE 56 Influence of Rank of Ph.D. Department on Decision (All Fields Combined)^a

Rank of Ph.D. Department	Percent Receiving Grants
High	62 (374)
Medium	54 (279)
Low	49 (138)
Unranked and no information	38 (93)
	42 (316)

Numbers of applicants are in parentheses.

^aWe have used nonstandardized rank of department scores.

TABLE 57 Influence of Academic Rank on Decision (All Fields Combined)

Academic Rank	Percent Receiving Grants	
Assistant professor	45	(281)
Associate professor	49	(303)
Full professor	58	(488)

Numbers of applicants are in parentheses.

TABLE 58 Influence of Professional Age on Decision (All Fields Combined)

Professional Age	Percent Receiving Grants	
Received Ph.D. before 1970	54	(873)
Received Ph.D. after 1970	46	(285)

Numbers of applicants are in parentheses.

TABLE 59 Influence of Funding History on Decision (Excluding Panel Fields; All Fields Combined)

Funded by NSF in Last 5 Years	Percent Receiving Grants	
Yes	65	(557)
No	40	(643)

Numbers of applicants are in parentheses.

rating received by the proposal. The data are presented in Table 60, in which we have excluded data on applicants in programs employing panels. The data indicate that NSF funding history does have a slight independent effect on the decision. For example, among applicants who received high ratings on their proposals and who have been funded previously, 98 percent were funded on their current applications. This compares with 89 percent who received similarly high ratings but who had not been funded in the past 5-year period. We found similar differences among applicants whose proposals received medium ratings and low ratings. In every case, applicants who had previously been funded by NSF had a slightly greater chance of being funded on their current applications.

The tabular analysis of the aggregated data on decision leads us to conclude that the primary determinant of whether an individual receives an NSF grant is peer review ratings, from both mail reviewers

TABLE 60 Influence of Number of Years Funded by NSF in Last 5 Years on Decision, Controlling for Mean Rating of Reviewers (Excluding Panel Fields; All Fields Combined)

Funded by NSF in Last 5 Years	Percent Receiving Grants		
	High Mean Rating of Reviewers	Medium Mean Rating of Reviewers	Low Mean Rating of Reviewers
Yes	98	61	16
No	89	41	5

Numbers of applicants are in parentheses.

and panels. Scientists who are highly ranked in the social stratification system of sciences are slightly more likely to receive NSF grants than those who are ranked lower; but most of the influence of eminence on decision is through its influence on the ratings rather than its influence on the program directors. In general, program directors are not substantially influenced by those characteristics of applicants that we have studied independently of the ratings given the proposals.

Post-Decision-Making Activities of the Program Director

HANDLING DECLINATIONS

The work of the program director is not finished when the funding decision has been made. He negotiates further with those who have been funded, and he must notify those who have been declined.

The Foundation has a standard form letter that goes out to declinees, but many program directors go beyond this. One program director described the situation in the following way:

On a declination, I feel that the principal investigators warrant an explanation. So, I have phoned every guy and warned him that the declination is coming and gave him the reasons why his proposal is going to be declined.

Program directors offer explanations for two reasons. First, the principal investigator wants to know what went wrong—what was substantively wrong with the proposal. Second, such an action signifies that the program director is sensitive to the principal investigator's disappointment in not being funded. Well-constructed peer reviews are helpful in explaining to applicants why they were not funded, and applicants can now request the verbatim comments of the peers.

As soon as we sent out the reject letters, the calls started coming in from applicants asking why they didn't get funded. The good reviewers really helped. They really provided constructive feedback to the applicants.

The feedback to applicants who have been turned down has significant implications for their future research activities. Program directors

Post-Decision-Making Activities of the Program Director | 151

can help turned down applicants, especially those close to the cutting point, improve their chances for future awards. One said:

A couple of times, I counseled a principal investigator who was declined on how to rewrite it, to include different kinds of data, to express it in different ways. He would need to run one more experiment in this area and then resubmit it.

Further, an expression of interest by the program director can prevent applicants from believing that there is no way for them to obtain NSF grants. Program directors can in this way influence the process of self-selection into and out of the pool of applicants.

Declinations can have different implications, since some people are declined but are given explicit or implicit information that they have a good chance to be funded in the future. Others are declined and given no relevant information, and still others are declined and given information that would lead them to believe that they would have little chance of success with a resubmission.

Program directors reported varying amounts of dissatisfaction among rejected applicants. In some cases, principal investigators felt that proposals had not been properly reviewed. One case was described in the following way:

I had one bad experience with a guy who felt he had been unfairly treated. He claimed that I picked reviewers that were not familiar with his work. I asked him who he would have picked, and he named guys I had picked. He was arguing that the NSF had something against him. I worked it out with him and he eventually said that he understood the situation.

One program director talked about how a negative decision occasionally gets validated. He said:

The only thrill I've had in this job so far is when you decline a guy and he thanks you for doing it. One guy said to me, "I just want to tell you that if I had been you, I would have done exactly the same thing. I'm going back to rethink it and I appreciate the reviewers' comments."

Many program directors point out that decreases in available funds have forced them to decline proposals that would have been approved in the past. One program director said:

I wish I could write letters outlining the reasons for declination because the implication would be that every time I decline something I have a clear reason for doing it. But that's the case in a very small proportion of proposals. Most are declined because the money wouldn't go that far.

Such cases are difficult for program directors because the proposed science would probably turn out to be pretty good work. This makes it hard for the program director to explain to a rejected applicant why he has not been funded.

An additional concern of program directors is handling declinations of eminent investigators. Eminent scientists whose proposals are declined are sometimes treated differently from less prominent scientists. First, greater attention tends to be given declinations of eminent scientists. Program directors make sure that there is a superabundance of documentation because they are more concerned about the way the decision would be reviewed by superordinates in the Foundation hierarchy. Second, careful diplomacy is used in informing the individual of the declination. In short, the process of rejecting applications is apt to be more elaborate for eminent scientists than for relatively unknown ones.

Some program directors also voiced concern about turning down eminent scientists because of possible repercussions within the scientific community. Since the scientific discipline they represent is, for many program directors, at least as important a reference group as the NSF, they feel accountable not only to their superordinates within the Foundation but also to their constituencies in the science community.

Reactions of eminent scientists to declinations vary. Some take such decisions in stride and resubmit new proposals that are approved; others attempt to put pressure on the program director. One described it this way:

There are cases where a principal investigator goes to a higher up in his university and he calls my superior in the Foundation. I don't know if they are trying to imply if I don't fund them they will take it higher.

Another person talked about a conflict that he was presently involved in over this issue. He said:

I have a case where I'm having warfare with my division director over this issue. He recommended that I fund, and I didn't want to fund it. He recommended that I fund because, "He is a marvelous guy, he is a scholar, you really can't turn him down." I gave my director the reasons and it's still unresolved.

NEGOTIATING BUDGETS

The program director also has important negotiating to do with the successful grantee. This negotiation centers on the amount of money that the principal investigator will receive from the Foundation.

Applicants are required to submit budgets with their proposals specifying how they are going to use any funds they receive. After a proposal has been accepted, the program director usually confers with the principal investigator about what the "revised" budget will be. In most cases, the revised budget will be less than the original request. There are essentially two reasons for this reduction: First, because of a general decline in resources, the program director must attempt to spread available funds around. Second, the program director must consider how much "fat" there may be in a proposed budget. One program director spoke at length about what is involved in this consideration:

You look for things like the amount of foreign travel that they might ask for (you figure that's padded), secretarial assistance, administrative assistance—people that are not crucial to the experiment—and if that's in there to a large degree, we figure that's padding, and the amount of computing time they ask for, especially if we know that there is a reasonable possibility that they can go to their dean and get free time. Ultimately, I leave it up to them. We don't tell them how to do the work, we just say, "You've asked for \$60,000 and I've got \$40,000—Do you feel you can rewrite the budget in a way where you can do the work within that \$40,000?" By and large, they will say yes.

To put this discussion of budgetary negotiation into proper perspective, let us consider how the characteristics of investigators are related to the amount and proportion of money received. The data presented in this report suggest that the characteristics of individual scientists have little effect on the ratings of proposals and on funding decisions. Critics of the peer review system have asserted that prominent scientists in lofty positions in the social system of science are more apt than their less-distinguished colleagues to receive large grants from the Foundation. We have reported above that there is no significant relationship between the amount of money scientists apply for and funding decisions. Now we want to look only at successful applicants—those who have been funded. In each of the 10 programs, these number roughly 60.

Table 61 presents the relationship between several of the characteristics of principal investigators who received funds and the amounts of money they received.¹ Consider the general pattern of findings in the

¹This analysis did not require probit estimates, since the dependent variable—amount of money received—does not have a restricted range. Therefore, we used standard Pearsonian correlation techniques. The proportion of variance explained by each variable would be the square of the correlation coefficient.

TABLE 61 Zero-Order Pearsonian Correlation Coefficients between Characteristics of Scientists on Funded NSF Research Projects and the Amount of Funds Received

Characteristics of Scientists	Program									
	Algebra	Anthropology	Biochemistry	Chemical Dynamics	Ecology	Economics	Fluid Mechanics	Geophysics	Meteorology	Solid-State Physics
Professional age	0.29	-0.28	0.12	0.14	0.28	0.25	0.24	0.11	0.06	0.12
Rank of current department	0.36	-0.24	0.24	0.25	0.33	-0.15	0.41	0.24	0.05	0.35
Citations (log of citations to work published 1965-1974)	0.23	0.26	0.51	0.23	0.08	0.14	0.15	0.07	0.24	0.33
Productivity (log of papers published 1965-1974)	0.11	0.01	0.40	0.30	0.19	0.05	0.17	-0.10	0.23	0.16
Years funded 1970-1974	0.45	-0.14	0.34	0.11	0.20	-0.05	0.31	0.01	0.25	-0.02
Mean rating	0.12	0.15	0.30	-0.03	0.00	0.11	0.08	0.24	0.32	0.35

154

10 programs. We begin with the professional age of principal investigators. Although the zero-order correlations differ somewhat in the several programs, the pattern suggests that senior scientists are, in fact, more apt to receive larger sums of money. Anthropology is the only program of the 10 that does not hold to this pattern.

The same pattern is found for the relationship between rank of current department and amounts received. Anthropology and economics are the only exceptions to the finding that scientists at the more prestigious universities receive more funds. In all 10 programs there is an association between the number of citations to a scientist's work and the amount of money he receives. Similarly, the level of scientific productivity, the granting history, and the mean rating of proposals are related to the amount of funds in such a way as to confirm the argument that, among scientists who are funded, those located in the upper strata of the stratification system generally receive larger grants.

To find out whether the more prominent scientists are favored in the amounts of funds they received independent of the sizes of their enterprises, we examined the proportions of funds requested that were actually received by principal investigators. If the scientists who are most prominent are also favored, we should find that they receive larger proportions of their requests. Table 62 presents the data relevant to this question. Again we present zero-order Pearsonian correlation coefficients between the same set of characteristics of the scientists and the proportions of funds requested that were received. Comparison of the associations in this table with those in Table 61 reveals a significant shift in the pattern of associations. There continues to be a small effect of the mean ratings of proposals on the proportion of funds received, but the correlations are somewhat less than we might expect. Among projects funded, some projects receiving relatively low peer-review ratings ask for large sums of money and receive large proportions of the requests. Correlatively, some higher-rated proposals receive lower proportions of the requested funds than do lower-rated proposals. This raises the question of whether the amounts of requested money not given to the higher-rated projects relate to fat in the budgets, or the need to spread funds around to other, less-well-regarded projects.

In 8 of the 10 programs, young principal investigators receive a higher proportion of requested funds than do their older colleagues. In most programs, rank of current department has either a negative correlation or a very small positive correlation with proportion of funds received. Biochemistry is the only program in which highly cited principal investigators were significantly more likely to receive a high

TABLE 62 Zero-Order Pearsonian Correlation Coefficients between Characteristics of Scientists on Funded NSF Research Projects and the Proportion of Requested Budget Actually Received

Characteristics of Scientists	Program									
	Algebra	Anthropology	Biochemistry	Chemical Dynamics	Ecology	Economics	Fluid Mechanics	Geophysics	Meteorology	Solid-State Physics
Professional age	0.24	-0.17	-0.19	-0.32	-0.12	-0.28	-0.16	-0.11	-0.02	0.13
Rank of current department	-0.03	-0.01	0.08	-0.25	0.02	-0.04	-0.24	-0.12	0.02	-0.08
Citations (log of citations to work published 1965-1974)	0.01	-0.14	0.43	-0.03	0.00	-0.25	-0.26	-0.04	0.12	0.01
Productivity (log of papers published 1965-1974)	0.09	0.03	0.36	0.01	-0.02	-0.11	-0.27	-0.08	0.17	0.10
Years funded 1970-1974	0.13	-0.01	0.28	0.21	0.17	-0.21	-0.11	0.05	0.08	0.18
Mean rating	0.07	0.06	0.29	0.08	-0.21	0.29	0.07	0.33	0.38	0.07

156

proportion of requested funds. In general, the correlations suggest that, at most, only a small amount of the variance in the proportions of requested funds that are received can be explained by any one of these characteristics of principal investigators.

These tables represent first attempts to understand the influences on the amounts of money and the proportions of requested funds received among funded projects. But many questions remain unexamined here. For instance, we do not have data on the types of grants being funded—the extent to which some projects are labor-intensive and others are more dependent on equipment. Further, we do not know what types of cuts are made in budgets, what kinds of justification are made for the cuts, and the consequences for scientists of budgetary decisions. All these questions go far beyond these data, but all require further study. For the moment we can only tentatively conclude that there is no strong evidence that scientists of high rank receive strikingly disproportionate shares of NSF funds.

Another question about the distribution of NSF funds is whether some large and prestigious universities receive “disproportionate” amounts of money. Unfortunately, data we have collected cannot be used to answer this question. The unit of analysis for our research has been the investigator applicant, not the institution. Therefore, the relatively low correlations between rank of current department and amount of money received reported in Table 61 do not tell us anything about the amount of money going to particular institutions. The data on the correlation between rank of current department and proportion of funds received reported in Table 61 suggest that, if indeed some institutions are receiving a “disproportionate” amount of money, it is probably because they have more scientists applying and these scientists apply for larger grants.

MONITORING RESEARCH PERFORMANCE

Aside from the work directly involved with the decision-making process, the program directors have one other important task. They must determine how NSF resources given to investigators are being used. Thus, they must monitor the work of principal investigators. This is extremely important not only because the Foundation demands that funds be used for the purposes for which they were requested and given, but also because many principal investigators seek renewals of their grants.

The program directors monitor grant activity in a number of different

ways. The most obvious strategy is evaluating the publications that come out of the funded work. As one program director put it:

On renewals, people will go through and review what they [applicants] have done on previous proposals. They will list their publications, and usually I am familiar enough with the areas to know if it is really good or significant, or if they're just publishing for the sake of it.

Although publications do provide some indication of progress, some program directors are reluctant to deny requests for renewals on the basis of low publication rates. As one said:

The only question I have is, "Is he going to stop working?" As long as there is evidence that he is working, that's O.K. Most people feel they would rather not publish, than publish crap. Most of the people to whom we give continuing grants we feel are like that.

Evaluation of work is also done by visiting research settings. Such site visits provide information about both the principal investigators and the work. This was described in the following way:

In a site visit, we make sure that the principal investigator is not sitting in a room talking to us, but that the rest of the staff is there in the room and they all have a chance to say something, and the grad students are there too. We don't want a snow job. We want to get the complete picture of what is going on.

Our intensive interviews with program directors, panel members, and other members of the NSF hierarchy suggested that there is no systematic attempt to monitor the outcomes of research grants. The scientific community is heavily relied upon to give feedback to the Foundation in the form of evaluations on subsequent grant applications. Almost nothing is done with the informal records of output that accumulate in the files of program directors. Recently, the NSF has taken steps to systematize the collection of information about research publications based on NSF-sponsored research. At the termination of grants, principal investigators must fill out forms listing all research publications produced by funded activity. To date, this information is not available to program directors for their use in evaluating the results of their prior funding decisions. Moreover, publications resulting from grants frequently come out after formal termination of the grants.

APPENDIX

A

Criteria for Selection of Grants

NOTE: The content of the following pages (through page 171) consists of facsimiles of materials used by the National Science Foundation in collecting information used as part of the data base of this report.

We have referred to the criteria that reviewers are asked to use in evaluating grant proposals. In a publication entitled, "Criteria for the Selection of Research Projects by the National Science Foundation," the NSF (1975b) lists them. They are grouped in four categories:

Category A

Criteria relating to competent performance of research—the technical adequacy of the performer and of his institutional base:

1. The scientist's training, past performance record, and estimated potential for future accomplishment.
2. The scientist's demonstrated awareness of previous and alternative approaches to his problem.
3. Probable adequacy of available or obtainable instrumentation and technical support.

Category B

Criteria relating to the internal structure of science itself:

4. Probability that the research will lead to important discoveries or valid, significant conceptual generalizations within its field of science or (in the most favorable cases) extending to other fields as well.

5. Probability that the research will lead to significant improvements or innovations of investigative method—again with possible extension to other fields of science.

Category C

Criteria relating to utility or relevance:

6. Probability that the research can serve as the basis for new invention or improved technology.
7. Probable contribution of the research to technology assessment—that is, to estimating and predicting the direct and indirect, intended and unintended effects of existing or proposed technologies.
8. Identification of an immediate programmatic context and uses of the anticipated research results.

Category D

Criteria relating to future and long-term scientific potential of the United States:

9. Probable influence of the research upon the capabilities, interests, and careers of participating graduate students, postdoctoral associates, and other junior researchers.
10. Probability that the research will lead to radiation and diffusion, not only of technical results, but also of standards of workmanship and a tradition of excellence in the field.
11. Anticipated effect upon the institutional structure of U.S. science.

According to this document, the first three criteria are “‘vigorously applied” in all programs and the last three are always “‘kept in view.”” The document further discusses the special emphases that characterize particular programs. In the sphere of basic research, the emphasis is “‘overwhelmingly” on the criteria of scientific merit (4 and 5). Consideration is also given to the utility criteria (6 and 7) in a more general way and “‘considerable direct weight” is given to criteria 9-11. In addition, the following are said to be encouraged by the Foundation:

1. Participation in research by graduate students.
2. Open publication of research results in the standard literature.
3. Widest possible access to unique facilities for interested and competent scientists.
4. Emphasis upon originality, elegance, and economy of method in university research.

NATIONAL SCIENCE FOUNDATION
PROPOSAL RATING SHEET

Reviewer: _____ Proposal No.: _____

Investigator: _____

Institution: _____

Please return to: _____

If possible by: _____

Comments (Continue on additional sheet if necessary)

OVERALL RATING Signature of reviewer: _____

 EXCELLENT Other suggested reviewers (optional): _____ VERY GOOD GOOD FAIR POOR

Verbatim but anonymous copies of reviews will be sent only to the principal investigator/project director on request. Subject to this NSF policy and applicable laws, including the Freedom of Information Act, 5 USC 552, reviewers' comments will be given maximum protection from disclosure.

NSF Form 173, Jan. 1976

The staff of the National Science Foundation frequently needs to call on scientists and other specialists outside the Foundation for advice and guidance in appraising the many requests it receives. For this reason I am asking you to assist us in evaluating the scientific merit of the project described in the enclosed proposal.

The criteria for evaluation and the details of current policies and procedures are described in the enclosed instructions for reviewers. Some of the policies and procedures have changed as of January 1, 1976, so please read the instructions carefully even if you have reviewed proposals for us in the past. Please make special note of the new policy on providing verbatim copies of reviews. We would like to add to our lists of highly qualified reviewers. Your suggestions, especially of young, women, and minority scientists would be appreciated. Thank you for participating in the review processes. Your assistance is vital in assuring that each proposal is evaluated fairly, critically, and promptly.

Sincerely yours,

J. CHRISTOPHER CORDARO
Program Director for Genetic Biology

Enclosures

INSTRUCTIONS FOR REVIEWERS OF GENETIC BIOLOGY PROPOSALS

Your critical review of the merits of the project will be of great value. Aspects that may enter into such an evaluation include the relative importance, feasibility, and conceptual framework of the project. You may also wish to consider the capacity of the investigators to conduct successfully the contemplated project, as evidenced by their description of developments in the field; their presentation of the proposed project; and their training and research contributions. If the facilities described seem inadequate, this should be noted.

Please enter your overall rating of the scientific merit of the proposal at the bottom of the rating sheet, using the following guidelines:

EXCELLENT: Highly meritorious and deserving of top priority for funding. This rating should be reserved for truly excellent proposals but should be used without hesitation when warranted.

VERY GOOD: Proposals considered superior, both for the intrinsic merit of the project and the ability or potential of the investigator, but having secondary priority.

GOOD: Quality sufficiently high to warrant consideration for support but definitely with tertiary priority.

FAIR: Unsupportable in present form; might merit consideration for support if resubmitted with major changes.

POOR: Unsupportable.

Check two adjacent boxes if your rating falls between two categories. Budgetary aspects should not enter into the rating, but comments on the budget are helpful.

Duplicate rating sheets are enclosed, together with a preaddressed envelope for the return of one of these sheets to us by *the date stamped* in the upper right-hand corner. You may keep the other sheet if you wish.

The Foundation receives proposals in confidence and is responsible for protecting the confidentiality of their contents. For this reason, we ask that you refrain from copying, quoting, or otherwise using material from this proposal. If you believe that a colleague can make a substantive contribution to the review, please consult me before disclosing either the contents of the proposal or the applicant's name. When you have completed your review or if, for some reason, you find yourself unable to respond to this request, please destroy the proposal.

Verbatim copies of reviews, ratings, and associated correspon-

dence will be sent to the principal investigator/project director on request. The copies will not contain your name, the name of your institution, or names that might constitute an invasion of the privacy of others. Subject to this Foundation policy and applicable laws, including the Freedom of Information Act, 5 USC 552, your participation as a reviewer and the content of your review will be given the maximum protection from disclosure.

The Foundation will publish annually a list of the names and addresses of persons who have reviewed proposals. Individuals will not, however, be identified with specific proposals. In this way the Foundation can publicly acknowledge your service as a reviewer and at the same time protect the confidentiality of your comments.

I am enclosing a proposal for your review.

The success of the Foundation's effort to strengthen scientific development through the wise use of public funds rests, fundamentally, upon the interest of individual scientists in sharing with us the responsibilities for insuring the scientific integrity of our programs. The careful and prompt evaluation of grant proposals is a vital part of this effort. Your assistance is, therefore, greatly appreciated.

Please comment as freely and with as much detail as possible. Your comparison of the merits of the enclosed proposal with others you may have reviewed in the past, and a comparison of the merits of the individual investigators with others in the field, would be of considerable value. I would also appreciate your suggestions of other possible reviewers.

The Foundation receives proposals in confidence and is responsible for protecting the confidentiality of their contents. For this reason, we ask that you refrain from copying, quoting, or otherwise using material from this proposal. If you believe that a colleague can make a substantive contribution to the review, please consult me before disclosing either the contents of the proposal or the applicant's name. When you have completed your review or if, for some reason, you find yourself unable to respond to this request, please destroy the proposal.

Verbatim copies of reviews, ratings, and associated correspondence will be sent to the principal investigator on request. The copies will not contain your name, the name of your institution, or material that might constitute an invasion of the privacy of others. Subject to this Foundation policy and applicable laws, including the Freedom of Information Act, 5 USC 552, your participation as a reviewer and the content of your review will be given the maximum protection from disclosure possible.

The Foundation will publish annually a list of the names and addresses of persons who have reviewed proposals. Individuals will not, however, be identified with specific proposals. In this way the Foundation can publicly acknowledge your service as a reviewer and at the same time protect the confidentiality of your participation.

Sincerely yours,

B. R. AGINS
Program Director for Applied Mathematics and Statistics

TO REVIEWERS OF MATHEMATICAL SCIENCES PROPOSALS:

Experience has shown that a somewhat discursive review, which describes the various pros and cons that have gone into making up the final judgment, is the most helpful for comparing the relative merits of proposals.

It is important that ratings shall reflect your opinion of the technical merits of the proposal and the qualifications of the investigators but shall be independent of the amount requested. However, separate recommendations concerning the proposed budget and general magnitude of the research program will be appreciated, since budgets of all proposals are adjusted before a grant is awarded.

To aid in the uniform rating of proposals, the following definitions of the rating terms are supplied.

EXCELLENT: Important research problems being undertaken by qualified mathematicians who can be expected to make substantial progress; presents an opportunity for a major contribution to basic knowledge; should be supported regardless of budgetary limitations.

VERY GOOD: Does not quite measure up to the previous category insofar as the research problem may be of somewhat lesser significance, or there may be some question concerning the ability of the investigator to carry through the complete program; should be supported.

GOOD: Worthwhile research being undertaken by competent mathematicians, but more routine in nature; may be supported if funds are available.

FAIR: Research proposal has serious deficiency that decreases probability of successful completion; not deserving of support in present form.

POOR: Inappropriate, technically unsatisfactory, or of a purely routine nature; not deserving of support.

Unless the reviewer specifically requests otherwise, certain review comments may be anonymously supplied to the proposer for his assistance and guidance.

The time and effort of reviewers is a vital part of the contribution to mathematical research made by the National Science Foundation. We thank you very much for your assistance.

Proposals submitted to the Engineering Division of the National Science Foundation are evaluated with the assistance of outside reviewers who are competent in the appropriate technical areas.

Primary emphasis is placed on the technical merit of the proposal and the probable contribution of the work to the understanding of fundamental engineering principles. We are generally not concerned with specific applications or with answers to particular problems.

We would like to request your evaluation of the enclosed proposal. The proposal should be treated as a privileged document. The Foundation receives proposals in confidence and is responsible for protecting the confidentiality of their contents. For this reason, we ask that you refrain from copying, quoting, or otherwise using material from this proposal. If you believe that a colleague can make a substantive contribution to the review, please consult me before disclosing either the contents of the proposal or the applicant's name. When you have completed your review or if, for some reason, you find yourself unable to respond to this request, please return the proposal in the enclosed franked envelope.

Verbatim copies of reviews, ratings, and associated correspondence will be sent to the principal investigator/project director on request. The copies will not contain your name, the name of your institution, or names that might constitute an invasion of the privacy of others. Subject to this Foundation policy and applicable laws, including the Freedom of Information Act, 5 USC 552, your participation as a reviewer and the content of your review will be given the maximum protection from disclosure.

The Foundation will publish annually a list of the names and addresses of persons who have reviewed proposals. Individuals will not, however, be identified with specific proposals. In this way the Foundation can publicly acknowledge your service as a reviewer and at the same time protect the confidentiality of your comments.

A rating sheet, our definition of the rating terms, and a franked envelope for return of the proposal and your evaluation are enclosed. We hope that you will be able to undertake this task for us. If you are unable to complete the review or feel it is not close enough to your technical area, perhaps you can suggest names of other reviewers.

Any assistance you may be able to provide in making the evaluation will be appreciated.

Sincerely yours,

CHARLES A. BABENDREIER
Program Director, Structural Materials and Geotechnical Engineering
Engineering Division

TO REVIEWERS OF NSF ENGINEERING DIVISION PROPOSALS:

The Engineering Division supports meritorious research that promises to advance significantly the basic engineering capabilities of the country. Such research in engineering includes the investigation of principles and techniques involved in the design of engineering systems and the analysis and synthesis of processes and systems that contribute to the mastery of the environment.

The following definitions of our standard rating terms are supplied for your guidance in evaluating the attached proposal. Reviewers are asked to interpret the definitions broadly rather than in a restricted way.

EXCELLENT: The problem is very important and well defined in the proposal. The investigators are highly competent and fully capable of doing the job. Strongly deserves support.

VERY GOOD: The problem is important and adequately defined in the proposal. The investigators are competent, and the research will contribute to their growth. Deserves support.

GOOD: The problem may be important, but the research is not well defined. The approach is routine but might contribute to graduate education or developing the potential of the principal investigator.

The proposal is marginal in its present form.

FAIR: The problem is probably unimportant or not well enough defined in the proposal to allow evaluation. The approach is questionable.

The proposal is not deserving of support in its present form.

POOR: The problem is unimportant, subprofessional, or has been solved by others.

Note: The Foundation will be pleased to receive comments regarding the proposed budget, but this factor should not influence your overall merit rating, since the budgets of all proposals are adjusted prior to awarding a grant.

TO: REVIEWERS OF RESEARCH PROPOSALS FOR THE
NATIONAL SCIENCE FOUNDATION, DIVISION OF SOCIAL
SCIENCES

CRITERIA OF JUDGMENT: Primary emphasis should be placed on the scientific merit of the project. The importance of the scientific problems addressed in the proposal, the conceptual framework of the project, the feasibility of the research strategy, and the competence of the investigator are all appropriate criteria for judgment.

BUDGET: Comments on the appropriateness of the budget for carrying out the work proposed are solicited, but budget aspects should not influence your overall merit rating. Budgets can be adjusted by the NSF staff if an award is to be made.

CONFIDENTIALITY OF PROPOSALS: Please treat the proposal as a privileged document; no material from it should be copied, quoted, or otherwise used without permission of the author, nor should the author's name be revealed in connection with this proposal. If you believe that a colleague can make a substantive contribution to the review, please consult before disclosing either the contents of the proposal or the applicant's name. When you have completed the review, or if you are unable to do it, please destroy the proposal.

CONFIDENTIALITY OF REVIEWS: To acknowledge our debt to reviewers and to inform the public, the Foundation will publish annually a list of the names and addresses of all persons who have assisted us by providing reviews. There will be no identification with specific proposals.

Verbatim copies of reviews, ratings and associated correspondence will be sent, on request, to the principal investigator/project director. These copies will be edited to remove your name and institutional identification and to avoid compromising the privacy of third parties whom you might mention but may not be retyped or otherwise changed to remove other information which might reveal your identity. Your comments and identity will, of course, be available to Foundation staff processing the proposal and to its Advisory Panels. Except for these Foundation policies, your participation as a reviewer and the content of your review will be given maximum protection from disclosure under applicable laws, including the Freedom of Information Act, 5 USC 552.

Any letter accompanying your review will be considered part of it and will be available to the investigator also, subject to the limitations mentioned above.

Criteria for Selection of Grants | 171

ENCLOSURES: Along with one copy of each proposal are three rating sheets—two copies to be returned to us in the return postage-paid envelope provided and one copy for your files. The review would be most helpful to us if the rating sheet could be returned by the date indicated. **PLEASE SIGN ONLY THE ORIGINAL. Please return original and one copy.**

APPENDIX
B | Sample, Data,
 and Methods

SAMPLE

In selecting the 10 programs for study we considered the form of peer review system used. We included some programs that used only *ad hoc* mail reviews and some that used a combination of *ad hoc* mail and panel reviews. We also considered the disciplinary context of the program. We wanted to sample programs in the physical, biological, and social sciences and in engineering. The other criterion was the size of the program. We decided to include the larger rather than the smaller programs because they would include a more significant portion of all NSF activity. The 10 programs selected for analysis were: algebra, anthropology, biochemistry, chemical dynamics, ecology, economics, fluid mechanics, geophysics, meteorology, and solid-state physics. Of these, algebra, chemical dynamics, fluid mechanics, meteorology, and solid-state physics use only *ad hoc* mail review; the others use panel and mail reviews. In the geophysics section no data are kept on the ratings by panel members.

Within each program, in selecting cases for quantitative analysis we included only applications that required a new set of reviews in fiscal year 1975. We included only those applications requesting funds for new research projects. Using the basic research grant requiring a set of reviews as the unit of analysis, we chose from each program a systematic random sample of 50 grant applications that were successful and 50 that were unsuccessful. We excluded all committed renewals because the projects had been funded for more than 1 year, if funds were available. In these cases, new reviews are not obtained. We also

excluded dissertation grants, travel grants, and conference grants because the criteria used in evaluating applications for these are apt to differ significantly from those used in evaluating applications for basic research.

We stratified the sample of applicants on the dependent variable, decision (i.e., 50 percent successful and 50 percent unsuccessful). We did so for two reasons. First, we wanted to make sure that the amount of variance in the dependent variable was not significantly greater in one field than in the others, since differing amounts of variance on the dependent variable could affect the size of the correlation coefficients. Second, we wanted enough acceptances and rejections to ensure statistically accurate pictures of both. Since we stratified on the dependent variable, we cannot generalize our results to all the applications in the 10 programs without using a weighting procedure because in some of the programs, our stratified random sample over-represents or under-represents successful applications. Presumably, successful and unsuccessful applications may differ in some significant ways.

Since some of the program samples over-represented or under-represented successful applicants, we wanted to make sure that this did not affect the substantive conclusions reached. In order to do this, we examined the algebra program, in which the proportion of acceptances in the sample differed by the widest margin from that in the population, repeating the entire analysis using a weighted sample that corresponded to the population. The results of this analysis were virtually identical with those reported for the algebra program throughout the report. As an example we present below the results of 20 regression equations that compare the estimated R^2 from point probit analysis for both the unweighted and weighted samples:

TABLE B-1 Proportion of Variance Explained on Ratings by Nine Independent Variables: Algebra Program

Independent Variable	Unweighted Sample	Weighted Sample
Citations to recent work	0.06	0.07
Citations to older work	0.03	0.01
Number of papers published	0.00	0.01
Rank of present department	0.07	0.05
Rank of Ph.D. department	0.04	0.03
Type of current institution (Ph.D. or not)	0.01	0.00
Professional age	0.00	0.00
Academic rank	0.03	0.03
Years funded 1970-1974	0.05	0.05
All nine variables	0.17	0.16

The probit estimates for the other tables did not differ by more than a point in the comparison of the weighted and unweighted samples. We may conclude that stratification on the dependent variable probably had no substantive influence on the probit findings reported in this report, and examination of the tabular data and ordinary regression analysis yielded the same conclusion.

The final sample in each program includes slightly more than 50 grants in each category. When we initially selected the sample and requested the jackets, some could not be located or were not available. Substitutions were made. Later, however, we obtained the missing jackets and added these to the files. A comparison of the missing jackets with the ones given to us on a first request revealed no systematic differences between the two. Thus, the final sample in each field generally contains somewhat more than 100 applications. Many of the proposals involve more than one principal investigator. For these cases we collected data on all the principal investigators.

The principal investigator, rather than the grant, is used as the unit of analysis in the quantitative analysis. Thus, if a particular grant had more than one principal investigator, each was treated as a separate unit. The analysis, then, focuses on the probability of different types of scientists receiving NSF grants. We thought that perhaps the probability of a principal investigator getting a grant would be affected by the status of his coprincipal investigator. We used two statistical comparisons to see whether our conclusions were affected by the way in which coprincipal investigators were treated. First, we compared two separate correlation matrices for the variables in the data set—one including grants with more than one principal investigator, the other omitting those grants. The correlations differed by no more than a few points. In the second comparison we identified each grant with the most eminent principal investigator, using citations to past work as the indicator of eminence. Then, with the grant as the unit of analysis, we repeated our correlation matrix and, once again, found that the correlations with the dependent variable, decision, differed by no more than a few points. Since 19 percent of the grants had more than one principal investigator, this is an expected conclusion. Using the principal investigator as the unit of analysis did not distort the results in any substantive way.

We encountered one other problem in selecting the sample. Some of the principal investigators who appeared on our list of declined applications also appeared on our list of granted applications or were current recipients on committed renewals. Inclusion of these investigators could distort the results, since the criteria in awarding a scientist a concurrent NSF grant might differ from those used in deciding whether

or not he or she should be funded at all (NSF Peer Review Study, 1975a, p. 8). Therefore, we excluded from the sample all principal investigators on declined proposals who were concurrently being funded on other NSF grants in the program. When such cases turned up in our sample, we substituted the next grant on the list of rejected applications.¹

The two tables below provide information about the level of activity in the programs we studied.

Table B-2 presents data on the actions taken by each of the 10 programs in fiscal year 1975. The first column shows the total number of actions taken by the program. This varies from a high of 419 for anthropology to a low of 140 for fluid mechanics. The second column shows the number of new research grants applied for and the last column shows the proportion of these applications that were funded. This varies from a low of 30 percent for algebra to a high of 63 percent for geophysics. This success rate may be an indicator of the amount of "proposal pressure" on each program. The extent to which such pressure influences the peer review system is not analyzed in this report. It is, however, a significant question and requires further research.

Table B-3 presents the financial statistics for each of the 10 programs for fiscal year 1975. Some of these statistics are based on estimates from our sample because the NSF did not have the data available. The first column shows the total amount of money spent by each program. The second column shows the amount of money applied for in new grant applications. The third column gives the proportion of new grant money applied for that was awarded. This proportion is a measure of the dollar pressure on the program. This varies from a low of 23 percent for chemical dynamics to a high of 56 percent for meteorology. The

¹It turned out that we were not completely successful in eliminating from the list of declined applicants all scientists currently receiving NSF support. We were able to check only whether these scientists were currently receiving NSF support in the particular program we were studying. Thus, for example, we could only tell whether a scientist turned down in chemical dynamics was currently receiving money from the chemical dynamics program. We could not tell whether that scientist was receiving money from some other program. We later found that several of our declined scientists were, indeed, receiving money from some other NSF programs. This applies to only a small number of the 1,200 scientists in the sample. We also were not able to determine whether principal investigators on declined projects were concurrently receiving funds from other government or nongovernment agencies. In some fields, such as biochemistry, there is a stronger possibility of this than in others, such as mathematics. Further, we do not know how often projects are declined explicitly because of alternative sources of support. Review of the proposal jackets suggests that this number is probably quite small.

TABLE B-2 Proposal Activity for 10 Programs, FY 1975

Program	(1) Total	(2)	(3)	(4)	(5)		(6)
		New Grants Awarded	Continuing Grants	Dissertation/Travel/Conference Grants	Declinations		Proportion of New Grants Funded (Excluding Dissertation Travel/Conference Grants), %
					Grants	Dissertation/Conference	
Algebra	308	61	104	0	143	0	30
Anthropology	419	77	44	88	111	99	41
Biochemistry	378	122	120	10	125	1	49
Chemical Dynamics	236	51	92	1	91	1	36
Ecology	285	82	42	22	106	33	44
Economics	269	75	55	10	121	8	38
Fluid Mechanics	140	59	22	4	54	1	52
Geophysics	185	91	40	0	54	0	63
Meteorology	186	81	43	8	54	0	60
Solid-State Physics	293	71	114	6	101	1	41

176

TABLE B-3 Financial Statistics for 10 Programs, FY 1975

Program	(1)	(2)	(3)	(4)	(5)	(6)
	Total Funds Awarded ^a (New and Continuing)	Total New Grant Funds Applied for ^b	Proportion of New Funds Granted Out of Total Funds Applied for	Proportion of Total Funds Committed	Mean Value of New Award ^c	Proportion of New Funds Granted to Awardees
Algebra	1,963,005	2,547,653	0.28	0.63	11,897	0.73
Anthropology ^d	3,041,698	6,267,613	0.31	0.36	25,138	0.81
Biochemistry	6,865,056	10,812,705	0.32	0.50	28,368	0.59
Chemical Dynamics	4,590,300	7,060,231	0.23	0.64	32,100	0.74
Ecology	3,143,400	5,912,246	0.35	0.34	25,350	0.76
Economics	4,389,970	8,831,890	0.29	0.42	33,769	0.75
Fluid Mechanics	2,099,439	3,363,205	0.45	0.27	25,919	0.85
Geophysics	3,584,946	5,001,117	0.50	0.31	27,366	0.74
Meteorology	4,766,436	5,553,036	0.56	0.35	38,439	0.82
Solid-State Physics	5,752,945	6,048,056	0.37	0.61	31,097	0.79

177

^aThe data on proposal activity are drawn from "Program Officers Workload Status Report Fiscal Year 1975 Proposals and Actions" (an NSF document). Although this report contains funding information, certain limitations prohibit us from using this information. First, it only provides the amount received by grant recipients and the amount applied for by declined applicants, and second, there is no specification of the time period for successful grants, unsuccessful requests, and continuing awards, making it impossible to determine how funds are distributed on a yearly basis.

Because of these limitations, all figures presented in this table are estimates based on our sample. Total new awards was computed by multiplying the mean award for the sample (for each program) by the total number of new awards given by the program. Total continuing awards was computed by multiplying the mean award for the sample by the total number of continuing awards. (We used the new-award mean, since we cannot determine the value of continuing awards from NSF data.)

^bThis figure is the sum of the mean value of requests by successful and unsuccessful applicants in the sample multiplied by the actual number of applicants in each category.

^cThis figure is a result of dividing the estimated total amount of new funds granted by the number of recipients of new grants.

^dDissertation grants and requests are not included in these figures. In this field, \$431,800 was distributed for dissertation support. This represents 12 percent of the program's total budget. Forty-four percent of requests for the support were granted.

Two other programs, ecology and economics, also provide dissertation support, but the amounts involved are quite small (less than 2 percent of total budget). Conference and travel funds are also granted by other programs but this activity is quite minimal.

influence of such dollar pressure on the operation of peer review requires further research.

The fourth column of Table B-3 shows the proportion of program budget spent on committed renewals. This tells us the extent to which the various programs commit money to relatively long-term projects. The reasons for such decisions and the consequences should be further analyzed.

The fifth column shows the mean dollar amount for new awards, and the last column shows for grant recipients the amount of money awarded divided by the amount applied for. This varies from a low of 59 percent to a high of 82 percent. Most of these percentages are around 75 percent, indicating that the average grant recipient gets about three-fourths of what he applies for. The low figure for biochemistry deserves further research. It may be that in this program an effort is being made to "spread the money around." The reasons for this decision and the consequences should be analyzed.

DATA

After selecting the sample we collected data on the principal investigators. The data sources included: the grant jacket, which contained the peer review material, a curriculum vitae of the principal investigator, and a list of his publications; *American Men and Women of Science* (AMWS); the rankings of graduate departments published by the American Council on Education (ACE); and the *Science Citation Index* (SCI). The data contained information on the status of the principal investigators in the sciences; indicators of their scientific track record; peer review ratings; and decisions. The variables constructed from these data are:

- Chronological age of principal investigator.
- Professional age of principal investigator (number of years since acquiring Ph.D.). In this report we have divided the applicants into those who obtained their Ph.D. prior to 1970 and those who obtained it in 1970 or after.
- Prestige ranking of Ph.D. department (obtained from the 1964 ACE rankings). Depending upon how many departments were included in the ACE survey, we divided them into four or five groups. In those parts of the analysis in which we could not use pair deletion, we assigned a low score to all cases for which the applicant's department was not

included in the ACE survey or for which the applicant worked in a nonacademic setting. We did this because, *on the average*, unranked departments and nonacademic institutions have lower prestige than ranked departments. For those parts of the analysis in which we were able to delete cases with missing data, we compared the results using deletion with those using a low score for missing data. In all cases the results were substantively similar.

Measuring ranks of academic departments was relatively difficult in geophysics and meteorology. The subject closest to geophysics included in the ACE survey was geology, and meteorology was not included at all. For those two fields we sent questionnaires to members of the National Academy of Sciences and asked them to rate all the departments for which we needed information. We received usable responses from 15 geophysicists and 9 meteorologists. For these two programs, rank of Ph.D. department was obtained from scores computed from these questionnaires.

- Type of present institution (this was coded dichotomously as a Ph.D.-granting institution or all other institutions).²
- Rank of current academic department (for those teaching in Ph.D.-granting institutions we used the 1969 ACE rankings). This variable was coded the same as rank of Ph.D. department. For those parts of the analysis in which we could not use pair deletion, we assigned all missing cases, scientists in unranked institutions, and scientists in nonacademic institutions the lowest possible score. This is based upon the assumption that most of these scientists were employed at institutions with lower prestige than ACE-ranked graduate departments. Scores of geophysicists and meteorologists were determined from the questionnaires returned by NAS members.
- Academic rank (coded as follows: 1 = researcher; 2 = postdoctoral; 3 = instructor, lecturer; 4 = assistant professor; 5 = associate professor; 6 = full professor).
- Amount of money applied for in first year of grant.
- Number of single-authored papers published between 1965 and

²At first we coded the principal investigators as coming from Ph.D.-granting institutions, other academic institutions, government agencies, private industry, foundation or other private research institutes, and other institutions. It turned out that there were so few scientists in each of the categories other than Ph.D.-granting institutions that the basic distinction was simply whether or not the scientist worked at a Ph.D.-granting or some other institution. Therefore, we decided to dichotomize this variable.

1974 (obtained from the source index of the *Science Citation Index*, 1965-1974).³

- Number of first-authored papers published between 1965 and 1974 (obtained from the source index, sci).
- Number of second-authored papers published between 1965 and 1974 (obtained from the source index, sci).
- Total number of papers published between 1965 and 1974 (the sum of variables 9-11). For the analysis we used a log transformation to the base 10. For anthropology and economics we used a log transformation of the productivity index.
- Number of citations in 1974 to work published between 1970 and 1974 (obtained from 1974 sci).⁴
- Citations in 1974 to work published between 1965 and 1969 (obtained from 1974 sci).
- Number of citations in 1974 to papers on which the author was not first author published between 1965 and 1974 (obtained from 1974 sci).
- Total number of citations to work published between 1965 and 1974 (the sum of variables 13-15). We used a log transformation to the base 10. Citation data for anthropology and economics do not include citations to papers on which the author was not first author.
- Citations in 1974 to work published prior to 1965 (obtained from 1974 sci). We used a log transformation to the base 10.
- Number of years between 1970 and 1974 in which the principal investigator has received funds from the NSF.
- Rating given the application by the program director.
- Ratings given the application by each mail reviewer and by each panel member, where such data were recorded.
- Type of institution at which each reviewer was located.
- Prestige of the reviewer's department for those reviewers at Ph.D.-granting institutions.
- Geographical locations of reviewer and applicant.
- Decision: accept = 2, decline = 1.

Data were set up on two different files for each program. On one file the applicant was the unit of analysis; on the other the unit of analysis

³For anthropology and economics the publication data were obtained not from the source index of the *Social Science Citation Index*, since this goes back only to 1972, but from curricula vitae contained in the grant jacket. For these two fields we used a productivity index. We classified publications into one of the following weighted categories: books weighted 10; edited books weighted 4; professional articles weighted 2; chapters in books weighted 2; applied articles and reports weighted 1.

⁴Citations for anthropologists and economists were obtained from the 1974 volume of the *Social Science Citation Index*.

was the pair of applicant and mail reviewer. The latter file was used in sections 2 and 3 and the former in section 5.

METHODS⁵

We have used probit regression analysis to deal with the problem raised by using the dichotomous variable, decision, as our dependent variable.

Theoretically, and in an ideal world, all proposals could be ordered unambiguously from best to worst, and that ordering is what we would like to be able to predict. But in the real world, even if that ordering were possible, it would be in some sense irrelevant, for at some point on the continuum, a break must be made between proposals to be funded and those to be refused. Indeed, it is this funded-refused dichotomy (decision) that we have frequently used as our dependent variable.

Clearly, however, we must conceive of the funded proposals as ranging from excellent to barely fundable and the refused proposals as ranging down from practically fundable to clearly rejectable. This distinction between an unobserved continuum of possible ratings and the observed dichotomous variable (decision) was formally incorporated in the probit regression models that were fit to the data in this study, as discussed below.

In order to specify the probit model, first introduce Y_1, Y_2, \dots, Y_n as independent random variables that can assume arbitrary real values. The random variable Y_j is an unobservable rating for the purpose of the program director's decision making. It is not the same as the reviewers' ratings but is the variable in the program director's mind, summarizing the information he has about j in terms of a numerical index that determines whether the grant is above or below the critical cutoff value that is called μ_j below. Next let Z_1, Z_2, \dots, Z_n be the observed decisions defined by the rule

$$Z_j = \begin{cases} 0 & \text{if } j\text{th proposal is refused} \\ 1 & \text{if } j\text{th proposal is funded} \end{cases}$$

Z_1, Z_2, \dots, Z_n are defined in terms of Y_1, Y_2, \dots, Y_n according to

⁵Professor Burton Singer, Department of Mathematical Statistics, Columbia University, and Professor Judith Tanur, Department of Sociology, State University of New York, Stony Brook, worked on this description of probit analysis. Professor Singer produced the mathematical description of the probit statistics used in this report.

$$Z_j = \begin{cases} 0 & \text{if } Y_j < \mu_1 \\ 1 & \text{if } Y_j \geq \mu_1 \end{cases}$$

where μ_1 is an unobserved cutoff level at which a proposal is either funded or refused. (We can set $\mu_1 = 0$ without loss of generality in the present study.)

Now introduce the linear regression model

$$Y_j = \sum_{i=0}^k \beta_i X_{ij} + u_j$$

where $X_{0j} \equiv 1$; X_{ij} = value of i th independent variable for j th proposal (principal investigator). For the j th individual, the X_{ij} represent the values of different variables such as reviewer rating, citation index, rank of university. The u_1, u_2, \dots, u_n are independent, normally distributed random variables with $E(u_j) = 0$, $\text{Var}(u_j) = 1$, $1 \leq j \leq n$; and $\beta_0, \beta_1, \dots, \beta_k$ are coefficients to be estimated. In terms of this specification we can write the probability of being refused or funded as

$$\text{Prob}(j\text{th proposal is refused}) = \text{Prob}(Z_j = 0) = \Phi\left(-\sum_{i=0}^k \beta_i X_{ij}\right) \quad (1)$$

where

$$\Phi(x) = \int_{-\infty}^x \frac{e^{-w^2/2}}{\sqrt{2\pi}} dw$$

and

$$\text{Prob}(j\text{th proposal is funded}) = \text{Prob}(Z_j = 1) = 1 - \Phi\left(-\sum_{i=0}^k \beta_i X_{ij}\right)$$

The nature of the linearity assumption (Eq. (1)) and the normality assumption for the error terms u_j , as well as the question of whether it holds strictly (sometimes, even approximately), and our reasons for using the probit model, are discussed in section 5, page 110.

The coefficients β_0, \dots, β_k were estimated by maximum likelihood—maximizing the likelihood function

$$L(Z_1, \dots, Z_n | \beta_0, \dots, \beta_k) \\ = \prod_{j=1}^n \left[\Phi\left(-\sum_{i=0}^k \beta_i X_{ij}\right) \right]^{1-Z_j} \left[1 - \Phi\left(-\sum_{i=0}^k \beta_i X_{ij}\right) \right]^{Z_j}$$

using the computer program NPROBIT of McKelvey and Zavoina (1975).

The fit of the model (Eq. (1)) to the data was assessed by calculating the statistic R^2 defined below. This measure of goodness of fit was motivated by the following considerations (see McKelvey and Zavoina, 1975, for further details):

Given the estimated coefficients $\hat{\beta}_0, \dots, \hat{\beta}_k$, define the fitted (i.e., estimated value of) unobserved values \hat{Y}_j according to

$$\hat{Y}_j = \sum_{i=0}^k \hat{\beta}_i X_{ij}$$

and introduce the residuals

$$e_j = Y_j - \hat{Y}_j$$

By analogy with ordinary linear regression analysis, we define the total sum of squares to be

$$S_{\text{total}}^2 = \sum_{j=1}^n (Y_j - \bar{Y})^2, \text{ where } \bar{Y} = n^{-1} \sum_{j=1}^n Y_j$$

Then writing

$$(Y_j - \bar{Y})^2 = [(Y_j - \hat{Y}_j) + (\hat{Y}_j - \bar{Y})]^2$$

and replacing \bar{Y} by its approximation,

$$\hat{Y} = \frac{1}{n} \sum_{j=1}^n \hat{Y}_j,$$

we introduce the plausible (for large sample size, n) approximation to S_{total}^2 ,

$$\hat{S}_{\text{total}}^2 = \hat{S}_{\text{explained}}^2 + \hat{S}_{\text{residual}}^2 \quad (2)$$

where

$$\hat{S}_{\text{explained}}^2 = \sum_{j=1}^n (\hat{Y}_j - \hat{Y})^2$$

and

$$\hat{S}_{\text{residual}}^2 = \sum_{j=1}^n (Y_j - \hat{Y}_j)^2$$

(The formula corresponding to Eq. (2) in ordinary linear regression is exactly satisfied without the carets, and it suggests the present approximation. In the probit model, Eq. (2) without carets is *not* satisfied exactly.) Finally, we define

$$\hat{R}^2 = \frac{\hat{S}_{\text{explained}}^2}{\hat{S}_{\text{total}}^2}$$

Since

$$\frac{1}{n} \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 \xrightarrow{n \rightarrow \infty} 1 \text{ in probability}$$

under the model (1), for large n we may approximate \hat{R}^2 as

$$\hat{R}^2 \approx \frac{\hat{S}_{\text{explained}}^2}{\hat{S}_{\text{explained}}^2 + n}$$

This \hat{R}^2 was utilized in the calculations of the present study, and for simplicity is referred to in section 5 simply as the probit R^2 . It gives an estimate of its population counterpart, a measure of the proportion of the variance of the Y_j about \bar{Y} that is explained by the X_{ij} in terms of the probit model.

As McKelvey and Zavoina (1975, p. 112) point out, the partitioning of the sum of squares (Eq. (2)) is not really legitimate for finite n . However, it is a plausible conjecture that

$$S_{\text{total}}^2 \sim \hat{S}_{\text{total}}^2, n \rightarrow \infty$$

although a rigorous proof is not currently in hand. McKelvey and Zavoina also mention a number of possible limitations of this approach including the likelihood of overestimation by \hat{R}^2 of its population counterpart. Nevertheless \hat{R}^2 , as defined above, has proved a plausible and useful quantity in other social science applications and yields conclusions in the present study that are compatible with the tabular analyses.

The probit regression program we used did not allow for paired deletion of cases with missing data. Therefore, on all variables except rank of doctoral department and rank of current department, we used the variable mean for cases with missing data. For the two rank-of-

department variables we assigned all cases with "missing" data a low score.

Decision is, of course, a dichotomous dependent variable. And it was specifically for use with this dependent variable that probit was chosen. Nonetheless, it is possible to compare what the results would have been had we used regular regression analysis with those obtained from the probit analysis. A comparison is presented in Table B-4. The first two columns show the proportion of variance in decision explained by the nine characteristics of the applicants in each of the 10 programs, using both probit techniques and ordinary linear regression techniques.

As can be clearly seen from these statistics, the probit estimates are consistently higher than are the regression estimates. McKelvey and Zavoina have pointed out that the probit estimates may be too large. However, we would also expect the probit R^2 to be larger than the R^2 for regression analysis with a dichotomous dependent variable (some value less than 1), provided the normality assumption is better than the assumption that Φ can be replaced by a linear function over the domain of values of

$$\sum_i \beta_i X_{ij}$$

Although the size of the R^2 estimated by probit procedures and the proportion of variance explained by regression procedures differ, the relative magnitude of the results in the 10 different programs is approximately the same. For example, when probit is used, we find the

TABLE B-4 Variance on Funding Decision Explained by Characteristics of Principal Investigators and Ratings of Reviewers—Comparison of Probit Analysis and Regression Analysis

Program	R^2		Mean Rating of Reviewers (and Panels Where They End)		R^2	
	Probit	Regression	Probit	Regression	Probit	Regression
Algebra	0.34	0.24	0.76	0.48	0.84	0.53
Anthropology	0.17	0.11	0.83	0.55	0.86	0.59
Biochemistry	0.51	0.33	0.86	0.57	0.86	0.62
Chemical Dynamics	0.39	0.24	0.92	0.54	0.96	0.60
Ecology	0.40	0.28	0.77	0.40	0.86	0.48
Economics	0.39	0.27	0.78	0.60	0.83	0.60
Fluid Mechanics	0.37	0.20	0.58	0.38	0.71	0.45
Geophysics	0.36	0.23	0.49	0.30	0.70	0.44
Metereology	0.24	0.17	0.92	0.55	0.94	0.57
Solid-State Physics	0.70	0.48	0.70	0.47	0.91	0.61

smallest proportion of variance explained by the nine individual variables in the program of anthropology and the largest proportion explained in solid-state physics. Exactly the same results are obtained when we use regular regression analysis. In the second column of Table B-4 we compare the estimated proportion of variance explained by the ratings of reviewers (and panel members where panels were used), using probit techniques and using regression techniques. Once again we find that the probit estimates are considerably higher than the regression estimates. In fact, the higher the correlation is in regression analysis, the greater the difference seems to be between the results from the regression analysis and those from the probit analysis. Substantively, the important part of Table B-4 is the last column, which shows how much additional variance is explained by the nine characteristics of the applicants when we add them to an equation containing the mean rating of the proposal. In algebra, for example, all nine variables and the mean rating explain 84 percent of the variance, using probit, where the mean rating alone explains 76 percent, using regression. Thus, the nine applicant characteristics increase the amount of variance explained by 8 points. Using regular regression techniques, we find that the increase in the amount of variance explained when the nine individual variables are added into the regression equation is 5 points. In general, whether we use the probit analysis or the regression analysis, the nine applicant characteristics add relatively little variance to the explained variance after mean ratings are entered into the equation.

Thus, we may conclude that both analytic techniques would lead to a similar conclusion, that is, that under the linearity assumption (Eq. (1)), mean reviewer rating of all the variables for which measurements are available is the most significant single determinant of whether or not a grant is made and that among the variables for which we have measurements, the nine characteristics of applicants do not add substantially to the predictability of whether or not a grant will be made.

The probit procedure we used is also useful for analyzing variables with a restricted range, such as reviewer ratings. (It was unnecessary to perform a separate Tobit analysis on the data.) Although conceptually the proposals could be ranked from "most meritorious" to "least meritorious," in fact the great majority of funded proposals are rated as "excellent" (1), "very good" (2), or "good" (3). This means that there is a problem of making the dependent variable discrete.

Our probit treatment of ratings of proposals was analogous to that for decisions. In particular we assume a continuum of unobserved possible

ratings Y and cutoff points $\mu_1 < \mu_2 < \mu_3 < \mu_4$ such that a proposal is rated in the following response categories⁶:

- 1 = excellent if $Y \leq \mu_1$
- 2 = very good if $\mu_1 < Y \leq \mu_2$
- 3 = good if $\mu_2 < Y \leq \mu_3$
- 4 = fair if $\mu_3 < Y \leq \mu_4$
- 5 = poor if $Y > \mu_4$

For the discussion that follows define $\mu_0 = -\infty$, $\mu_1 = 0$, $\mu_5 = +\infty$, and then let Y_1, Y_2, \dots, Y_n be independent ratings represented by the linear regression model

$$Y_j = \sum_{i=0}^k \beta_i X_{ij} + u_j \quad 1 \leq j \leq n \quad (3)$$

where u_1, \dots, u_n satisfy the same hypotheses as in the model of equation (1). Then represent the observed ratings by the dummy variables.

$$Z_{jk} = \begin{cases} 1 & \text{if } j\text{th rating is in the } k\text{th response category} \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} 1 & \text{if } \mu_{k-1} < Y_j \leq \mu_k \\ 0 & \text{otherwise} \end{cases}$$

Then the probability of the j th rating being in the k th response category is given by

$$\text{Prob}(Z_{jk} = 1) = \Phi\left(\mu_k - \sum_{i=0}^k \beta_i X_{ij}\right) - \Phi\left(\mu_{k-1} - \sum_{i=0}^k \beta_i X_{ij}\right)$$

where, again,

$$\Phi(x) = \int_{-\infty}^x \frac{e^{-w^2/2}}{\sqrt{2\pi}} dw$$

⁶The above numbering of response categories is for indexing purposes only; it is not to be viewed as an assignment of numerical scores.

TABLE B-5 Comparison of Results Using Probit and Regression Analysis: Dependent Variable = Rating^a

Program	Citations to Recent Work		Rank of Current Department		NSF Funding History		All Nine Independent Variables	
	Probit R^2	Regression R^2	Probit R^2	Regression R^2	Probit R^2	Regression R^2	Probit R^2	Regression R^2
Algebra	0.09	0.06	0.07	0.07	0.06	0.05	0.19	0.17
Anthropology	0.01	0.00	0.00	-	0.00	-	0.05	0.04
Biochemistry	0.19	0.16	0.09	0.07	0.07	0.06	0.25	0.20
Chemical Dynamics	0.15	0.14	0.03	0.02	0.06	0.05	0.18	0.16
Ecology	0.01	0.01	0.02	0.02	0.02	0.02	0.05	0.06
Economics	0.10	0.08	0.14	0.13	0.09	0.08	0.23	0.21
Fluid Mechanics	0.03	0.03	0.12	0.10	0.01	0.01	0.20	0.17
Geophysics	0.08	0.07	0.03	0.03	0.01	0.01	0.10	0.09
Meteorology	0.09	0.08	0.06	0.05	0.02	0.02	0.15	0.14
Solid-State Physics	0.08	0.08	0.09	0.08	0.07	0.06	0.19	0.17

^aRelationship is negative.

The cutoff points μ_2, μ_3, μ_4 , and the regression coefficients $\beta_0, \beta_1, \dots, \beta_k$ were estimated using NPROBIT (see McKelvey and Zavoina, 1975) to maximize the likelihood function

$$L(\{Z_{jk}\} | \beta_0, \dots, \beta_k, \mu_2, \mu_3, \mu_4) = \prod_{j=1}^n \prod_{k=1}^5 \left[\Phi\left(\mu_k - \sum_{i=0}^k \beta_i X_{ij}\right) - \Phi\left(\mu_{k-1} - \sum_{i=0}^k \beta_i X_{ij}\right) \right]^{Z_{jk}}$$

With the estimates $\beta_0, \beta_1, \dots, \beta_k$ at hand, the goodness of fit of the regression model is assessed by the same procedure as described for the case when decision was the dependent variable. The increased complexity of this case, compared with the earlier one involving only one response, means that the limitations of the \hat{R}^2 estimates are probably now greater.

In Table B-5 we compare the results of a probit analysis with those of a regression analysis when the dependent variable was rating. Table B-5 shows the proportion of variance explained by citations to recent work, by rank of current department, by the NSF funding history of the applicant, and by the nine variables we have used to characterize applicants. The table presents the results of a total of 40 ordinary least-squares regressions and 40 probit regression equations. As can be seen by this comparison, there is no great systematic difference between the results of the two procedures. In all cases the probit results are either exactly the same as the regression results or slightly higher. (See earlier comments about this.) Comparison of the amount of variance explained by the sum of the nine variables produces remarkably similar results. For example, in algebra the probit analysis explained 19 percent of the variance on ratings using the nine variables characterizing applicants, and the regression analysis explained 17 percent of the variance on ratings. Because the two methods yielded essentially identical results and the same substantive conclusions, we have used the regression technique, which is simpler and more widely understood. (For discussion of probit analysis and its uses, see page 110, section 5.)

Bibliography and References

- Allison, P. D., and J. A. Stewart. 1974. "Productivity Differences Among Scientists: Evidence for Accumulative Advantage." *American Sociological Review* 39(4):596-607.
- American Men and Women of Science (12th ed.). 1973. New York: Jacques Cattell Press.
- Biomedical Science and Its Administration*. 1965. A Study of The National Institutes of Health. Supt. of Documents, U.S. Government Printing Office, Washington, D.C.
- Cain, G. G., and H. W. Watts. 1970. "Problems of Making Policy Inferences from the Coleman Report." *American Sociological Review* 35:222-42.
- Carter, G. 1974. *Peer Review, Citations and Biomedical Research Policy: NIH Grants to Medical School Faculty*. Santa Monica, Calif.: Rand Corporation Report 1583-HEW.
- Carter, A. M. 1966. *An Assessment of Quality of Graduate Education*. Washington, D.C.: American Council on Education.
- Cole, J., and S. Cole. 1972. "The Ortega Hypothesis." *Science* 178(4059):368-74.
- Cole, J., and S. Cole. 1973. *Social Stratification in Science*. Chicago: University of Chicago Press.
- Cole, S. 1977. "Scientific Reward Systems: A Comparative Analysis." Forthcoming in Robert Alun Jones, ed., *Research in Sociology of Knowledge, Sciences, and Art*. Vol. I. Greenwich, Conn.: Jai Press.
- Cole, S., and H. A. Zuckerman. 1976. "The Use of ACE Ratings in Research on Science and Higher Education." Prepared for planning conference on assessment of the quality of graduate education programs in the United States, September 27-29, Woods Hole, Massachusetts.
- Cole, S., J. Cole, and L. Dietrich. 1976. "Measuring Consensus in Scientific Research Areas." In Yehuda Elkana *et al.*, eds., *Science Indicators*. New York: Wiley Inter-science.
- Committee on Government Research of the House of Representatives. 1964. *Report*. Washington, D.C.: U.S. Government Printing Office.

- Committee on Science and Technology of the U.S. House of Representatives. 1976. *National Science Foundation Peer Review*. Washington, D.C.: U.S. Government Printing Office.
- Douglas, C. D., and J. C. James. 1973. "Support of New Principal Investigators by NIH: 1966 to 1972." *Science* 181:241-44.
- Finney, D. J. 1952. *Probit Analysis* (2nd ed.). Cambridge, England: Cambridge University Press.
- Goode, W. J. 1967. "The Protection of the Inept." *American Sociological Review* 32:5-19.
- Greenberg, D. 1967. *The Politics of Pure Science*. New York: New American Library.
- Greenveld, L., N. Koller, and N. C. Mullins. 1975. "The Advisers of the United States National Science Foundation." *Social Studies of Science* 5:343-54.
- Gustafson, T. 1975. "The Controversy over Peer Review." *Science* 190(4219):1060-66. Institute for Scientific Information. 1974. *Science Citation Index*. Philadelphia.
- Kuhn, T. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Laveck, C. C., et al. 1974. "Recipients of Research Grants for NICHD: Do Age, Sex, Type of Degree Affect Funding Chances?" *Pediatrics* 53(5):706-11.
- Leege, D. C. 1975. "Is Political Science Alive and Well and Living at NSF: Reflections of a Program Director at Midstream." Paper delivered at annual meeting of American Political Science Association, San Francisco, California.
- Liebert, R. J. 1976. "Productivity, Favor, and Grants among Scholars." *American Journal of Sociology* 82(3):664-73.
- McKelvey, R., and W. Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology*, 4:103-20.
- Merton, R. K. 1942. "Science and Technology in a Democratic Order." *Journal of Legal and Political Science* 1:115-26.
- Merton, R. K. 1957. *Social Theory and Social Structure*. New York: The Free Press.
- Merton, R. K. 1968. "The Matthew Effect in Science: The Reward and Communication Systems of Science." *Science* 199:55-63.
- Morrison, D. E., and R. E. Henkel, eds. 1970. *The Significance Test Controversy*. Chicago: Aldine.
- National Science Foundation. 1975a. "Peer Review and Proposal Evaluation, Staff Study." *NSF Peer Review Study*. Washington, D.C.: NSF Administration Directorate.
- National Science Foundation. 1975b. *Criteria for the Selection of Research Projects by the National Science Foundation* (NSB-74-300, February 1975).
- National Science Foundation. 1975c. *An Analysis of the Geographical Distribution of NSF Awards as Compared with other Selected Indicators*. Washington, D.C.: NSF Administration Directorate.
- National Science Foundation. 1975d. *NSF Management Statistics*. Washington, D.C.: NSF Administration Directorate.
- Office of Management and Budget. 1973. "The National Institutes of Health and National Institute of Mental Health Peer Review System." OMB Issue Paper. Unpublished.
- Pfeffer, J., G. R. Salancsek, and H. Leblebici. 1976. "The Effect of Uncertainty on the Use of Social Influence in Organization Decision Making." *Administrative Science Quarterly* 21:227-45.
- Price, D. 1963. *Little Science, Big Science*. New York: Columbia University Press.
- Roose, K. D., and C. J. Andersen. 1971. *A Rating of Graduate Programs*. Washington, D.C.: American Council on Education.
- Small, H. G. 1974a. *Characteristics of Frequently Cited Papers in Chemistry*. Philadelphia: Institute for Scientific Information.

- Small, H. G. 1974b. *Report on Citation Counts for National Science Foundation Grant Recipients and Non-Recipients*. Philadelphia: Institute for Scientific Information.
- Useem, M. 1976. "State Production of Social Knowledge: Patterns in Government Financing of Academic Social Research." *American Sociological Review* 41(4):613-29.
- U.S. Office of Science and Technology. 1965. *Biomedical Science and its Administration: A Study of the National Institutes of Health*. Washington, D.C.: U.S. Government Printing Office.
- Vivona, D., and D. DoVan Quay. 1973. "Comparability of the American Cancer Society and the National Institutes of Health Peer Review System." *Cancer Research* 33(4):919.
- Wilson, M. K. 1975. "The Top Twenty and the Rest: Big Chemistry and Little Funding." *Annual Review of Physical Chemistry* 26:1-16.
- Wirt, J. G., A. J. Lieberman, and R. E. Levien. 1974. *R & D Management: Methods Used by Federal Agencies*. Santa Monica, Calif.: Rand Corporation Report R-1156-HEW.
- Zuckerman, H. 1977. *The Scientific Elite*. New York: Free Press.
- Zuckerman, H., and J. Cole. 1975. "Women in American Science." *Minerva* XIII(1):82-102.