Peer Review System: How to Hand Out Money Fairly

"People told me the system was all political and very unfair," says a young biologist of her first attempt to seek a research grant from the National Institutes of Health (NIH). "Women told me it was biased against women. Also I had had a fight with my thesis adviser, who wrote a letter to the NIH trying to prejudice them against me. He was quite a famous man, but they paid no attention to him. I didn't have any trouble getting a grant. I was overwhelmed by the amount of time they spent talking with me about my proposal and the amount of energy they put into considering it."

The description, admittedly from a satisfied customer, exhibits two not uncommon attitudes toward the peer review system: the expectation, based on hearsay, that getting a grant may not depend on scientific merit alone, and the experience that the review procedure appears to be thorough and equitable. The peer review system is not the only route by which the federal government dispenses monies for science, but for basic research, particularly of the kind supported by the NIH and the National Science Foundation (NSF), it is the dominant mechanism. A third major patron of biological research, the Department of Agriculture, does not rely heavily on the peer review system, but may do so in the future: a committee of the National Academy of Sciences which was highly critical of agricultural research recently recommended peer review as the principal remedy (see *Science,* 5 January). The following sketch of how the system operates is based largely, although not exclusively, on discussions with members of the molecular biology study section, one of about 50 peer review committees operated by the NIH's Division of Research Grants. (The molecular biology section probably does not differ greatly from other study sections, but the peer review system is a complex process and the description below rests on a narrow data base.)

Peer review is a system whereby nongovernment scientists advise the government how to give away its money to their fellows, the rationale being that, as a jury of peers, they are the best qualified people to decide whose research is worth supporting and whose is not. Those who serve on study sections spend between 1 and 3 weeks in reading grant applications before each of the three sessions a year. Except for a per diem during the actual meeting, they receive no financial reimbursement, although there are compensations of a different sort; membership in a study section carries a certain prestige, and the time spent reviewing applications is a thorough, if arduous, way of keeping abreast with one's field. Those asked to serve on study sections must work at a nonprofit institution, be still active in research, and be prominent in their field.

According to NIH administrators, each of the NIH's study sections contains 10 to 15 members, most of whom serve for a 4-year period, thereby assuring a reasonably steady turnover. Each section is staffed by a full-time executive secretary who is an NIH scientist knowledgeable in the field covered by the section. The executive secretaries select the new members of their study sections and, in addition to individual qualities, try to attain a disciplinary as well as geographical balance. There are various constraints to exclude the more obvious sources of bias. Sections may have only one member from a particular institution (but campuses of the University of California, say, count as single institutions). Grant applications of members are channeled to different study sections (or, if no suitable section exists, to a special committee). Members customarily absent themselves when applications from colleagues at the same institution come up for discussion, or if for any other reason they feel uncertain of giving an unprejudiced opinion.

A distinctive feature of the NIH system, compared with that of the NSF, is that the study sections are not concerned with the funding of an application and thus, in theory, can concentrate exclusively on scientific merit. When a grant application arrives at the NIH, it is scanned by a referral officer in the Division of Research Grants, who allocates it simultaneously both to the most relevant study section and to the appropriate institute of the NIH.

The study section decides on the scientific merit of the application, without regard to its practical relevance, and assigns a priority score to the applications that are approved. Each institute ranks the applications assigned to it in order of the priority scores given by the study sections and generally funds them from the top downward until the money runs out. Each institute decides at what level to draw its payline. In general, between 50 and 60 percent of grant applications submitted to the NIH are approved, and in fiscal year 1972 roughly 60 percent of those approved were funded. (The funding ratio has ranged from 75 percent in the early 1960's to less than 50 percent in recent years.)

The funding decision is made by the institute's advisory council, which consists of both scientists and public figures. The advisory council may not fund any application that a study section has disapproved, but for reasons of "high program relevance" the council may lift an application with a low priority score above the payline. Councils may also strike out low-relevance grants above the payline, but are said to do so only rarely.

Priority Scores

When an application reaches a study section, it is assigned by the executive secretary to two or three section members for primary, in-depth review. Each member has about 20 applications per session for which he is primary reviewer, but he is expected to read all the other applications as well. About 100 applications are discussed during the 3-day meeting period. As each application comes up, the primary reviewers give their opinions first, and it is then put up for discussion. If the two primary reviewers agree, their decision usually settles the matter quite quickly. If they disagree, it is up to each to convince the other members. After the debate, which may take between 15 minutes and 3 hours, a vote is taken on whether to approve, disapprove, or defer to the next decision. (Many applications on which the section may want more information are deferred in order to allow a site visit to the applicant.) The priority scores assigned by the section members are tallied by the executive secretary (who does not vote) and are later normalized to a specified mean and standard devi-

ation that is common to all study sections. (This means that each institute is dealing with a more comparable set of priority scores.) Priority scores range from 1.0 (most meritorious) to 4.5 (least meritorious).

Priority scores are recorded on a summary form, or "pink sheet," on which the executive secretary also writes detailed notes of the primary reviewers' comments and the points made by other members in discussion. Although the pink sheet is kept confidential, an applicant who asks to see the verdict on his application will receive an extract, or paraphrase, of the pink sheet from the executive secretary. Even the extracts can run to five pages of single-spaced typing and contain a fine level of detail: for example, "You seem to have overlooked the claim of references X and Y that the two sodium fluxes were balanced. Apparently you have not studied these papers adequately."

The pink sheet also contains provision for the "executive secretary's note," a mechanism whereby, if he considers that one member has for some particular reason given a low-priority score, the secretary can mention the reason and recalculate the average without the score he differs with. The funding institute can choose whichever of the two scores it prefers.

Whether for better or worse, the peer review system operated by the NSF is considerably less formalized than that of the NIH. The NSF does not have to worry about the practical relevance of the research it supports, and its peer review committees in practice make what is virtually a funding decision at the same time they determine scientific merit. There are no priority scores. The NSF panels vote to approve or decline a grant (the ratio between the two is about half-and-half). The approved grants are then assigned to a high or a low category (currently 90 percent of the NSF molecular biology panel's applications are assigned to the high category—in general, the proportion is set so as roughly to correspond with the funds available). The NSF tries to support all of the applications in the high category and will fund some in the low category if any money remains.

Counterpart peer review committees of the NIH and NSF generally pass similar verdicts on applications submitted to both agencies. Administrators say this is evidence of the fairness of the system. There is, however, one case on record of a scientist who submitted the same pair of grants to the two agencies: the application that the NSF funded the NIH rejected, and vice versa.

The flavor of what goes on in a study section is hard to describe, not least because the NIH refused this reporter's request to attend one. (On the grounds of protecting the ideas set forth in an application, study section meetings are exempted from the recent presidential order opening all government advisory committees to the public.) Rumors of cronyism are not uncommon, but scientists who serve on the study sections, including some who were inclined to believe such rumors before becoming members, say the system just does not work this way. The practice of having the two primary reviewers of an application defend their opinion against the other members encourages a real debate, as well as careful homework on the part of the reviewers. "Before I became a member, I often thought that, if there was someone on the study sections who was prejudiced against you or your line of work, he could do a lot of damage," says Gordon H. Sato, professor of biology at the University of California, San Diego. "I think now this is not likely to happen. When a member gives a hard sell in one direction or another, I would find myself making an adjustment, and then I would be surprised to find the rest of the study section had made the same adjustment."

According to George N. Eaves, executive secretary of the NIH molecular biology study section, the members are never impersonal about judging an applicant's work and keep in mind how their decisions may affect a man's career. The 4 to 5 hours a primary reviewer may spend studying an application, says Walter Eckhart of the Salk Institute, La Jolla, "is done not so much because of a sense of responsibility or what the other members may think of your presentation, but because one knows that for the applicant it's a matter of life or death." The system is open to the discovery of redeeming features. One poorly written application was given a high-priority score because a panel member happened to have read in manuscript an outstanding paper the applicant had written on another subject; the member persuaded the others to give the applicant the benefit of the doubt. Another member recalls how he switched the majority to support of an application that all agreed was poor science—the applicant claimed to have cured a disease by injection of a substance she refused to identify—but that he felt should be given a chance because the animals seemed in fact to have been cured.

Charges of Cronyism

The peer review system as practiced by the NIH is not without its critics, chief among whom is probably Julia T. Apter, a qualified physicist and physician at St. Luke's Medical Center, Chicago. Apter became interested in the system's treatment of women and found out, she says, that "the mechanisms responsible for excluding women were also excluding the men—because of this cronyism." Evidence of cronyism, as Apter sees it, is her estimation that many members of NIH advisory committees, study sections included, are serving their second, third, or fourth 4-year terms. This is denied by the associate director of the Division of Research Grants, S. Stephen Schiaffino, who says that for study sections, at least, NIH policy is to make no reappointments, either to the same or to a different study section, if a suitable replacement can be found.

Another of Apter's criticisms is that many members of study sections are chairmen of departments whose administrative duties must leave them little time for research—hence they cannot be considered the "peers" of the scientists whose applications they are judging. Schiaffino, however, says that virtually all members of study sections, chairmen included, are actively engaged in research (the exception is where an administrator is specifically required to review grants for big centers).

Apter is concerned in particular about the low representation of women on study sections. To which Schiaffino replies that the number of women members has doubled during the last 18 months (it is now 10 percent).

The allegation of cronyism—that the peer review system consists of a group of committees whose members hand out grants to each other and to their friends—is one that is almost impossible to substantiate or refute. Those who allege cronyism cite the fact that certain large institutions, such as the Harvard–Massachusetts Institute of Technology complex, are well represented on the study sections and on the list of successful applicants. NIH and NSF staff respond, in effect, that they have to go to good places to get good people. Since each study section covers

a small area of biology or medicine, it is not surprising if members know one another by other means. From the outside, it is impossible to say if ties of acquaintanceship ever influence the award of a grant; but if the peer review system works as NIH staff and study section members say it does, then such influence seems unlikely to be common. For one thing, it would be difficult to affect significantly the votes of a 15-member committee on a special-interest issue without arousing antagonism. For another, it is hard to see that members would spend up to 9 weeks a year of their own time studying applications in preparation for a bout of horse trading. Study sections are probably open to certain nonscientific considerations—such as giving a young applicant an extra chance, say, and maybe old ones too—but the "tradition" of the sections, as members describe it, is averse to making special pleas for one's friends.

Another criticism of the system is the suggestion that an applicant be able to monitor the review process, either by attending the meeting when his application is being discussed or by having the opportunity to rebut criticisms on the pink sheet before his application comes before an advisory council. NIH and NSF staff say such procedures would inhibit free discussion, turn the review into an adversary proceeding, and increase its administrative complexity.

Grantsmanship

"No amount of methodological sophistication or grantsmanship can bring to life a sterile thought," warns an NIH brochure designed for the edification of applicants. The only kinds of grantsmanship that NIH officials concede are effective are the literary virtues of clarity and succinctness. The art is more usually understood to mean dressing up an idea so as to increase its fundability. Take the case of the botanist said to have won a grant from his local American Cancer Society to study the induction of flowering; he avoided all mention of flowering in his application, describing the project solely in terms of the manipulation of nucleic acids. Other examples of grantsmanship, some successful, some not, include the following reported instances.

► An application to study the biochemical turnover of collagen in the uterus was turned down by a study section but came back the next session with a revised rationale: to study the effect of air pollution on the turnover of collagen in the uterus.

► A project to measure the *p*H of mitochondria was submitted in five separate versions, one using heart cells, one using cancer cells, and so forth, the applicant's intention being to target each version to a different institute. (What he had failed to allow for was that all five versions landed in the same study section.)

► Applicants studying a basic cellular process will propose to do so in cancer cells rather than normal cells, even when normal cells would be better scientifically.

► A scientist interested in the natural pigmentation of cells will write up his application so as to stress those aspects to melanoma cells.

"Since the initial review of a research grant application is for scientific merit only, an applicant can gain nothing by distorting his actual intentions in anticipation of the program interests of the institutes," says Eaves.* As an executive secretary, he is in a position to know, and certainly the cruder forms of grantsmanship, and those enlisted to

* G. N. Eaves, "Who reads your project-grant application to the NIH?," *Fed. Proc.* **31**, No. 1 (1972).

David, PSAC Exit Predicted

Drastic changes in the White House science advisory system, rumored for months, seemed to be imminent as this issue went to press on 2 January. Knowledgeable sources said that Presidential Science Adviser Edward E. David, Jr., would resign within days, and that the resignations of the President's Science Advisory Committee (PSAC) would also be accepted soon.

The President was said to have had a favorable opinion of David, who was allegedly offered a lower ranking role in energy affairs. However, he will take an executive post with industry instead. It is not known whether a successor will be appointed.

The departure of the well-regarded David and the disappearance of PSAC, which under the Johnson and Kennedy administrations represented the views of the highest echelons of the scientific community, cast a shadow over the Office of Science and Technology (OST). One possibility is that a reduced OST might be eventually absorbed into the much larger Office of Management and Budget.

The departure of PSAC, the preeminent science council, and of David, the most highly placed science appointee, signals, at the least, a wish by the Administration to decentralize science in government. It also means that there will no longer be a special niche for scientists in the White House.

When PSAC last met on 18 and 19 December, the members were asked to submit their resignations, apparently as a pro forma move, just as some 2000 high-ranking government officials (including David) had in November. Knowledgeable sources, however, said that the PSAC resignations will in fact be accepted. In PSAC's present form, the chances of survival seem slim indeed.

PSAC's relatively diminished role in recent months may have been due in part to a shadow science cabinet of Republican scientists who have made regular but unofficial inputs to key Presidential aides ever since Nixon was first elected in 1968. The group continues to be active according to sources close to it. However, its exact membership is not known, although it is said that these trusted advisers are among the members of the Science and Engineering Council to Support the President which surfaced just before last November's elections (*Science* 27 October).* Evidently, as the official White House science apparatus decentralizes—or diminishes—the unofficial advisers could find themselves playing a larger role.—D.S.

* Members of the Science and Engineering Council in Support of the President were: William O. Baker, Z. Dave Bonner, Robert Charpie, Clyde Cowan, Henry Eyring, Kurt Glaser, Richard Godwin, Martin Goland, Lawrence A. Goldmuntz, Patrick E. Haggerty, H. Richard Johnson, Willard F. Libby, Gordon J. F. MacDonald, William G. McMillan, Richard Morse, George Mueller, Howard K. Nason, William Nierenberg, Bernard M. Oliver, Thomas Pownall, Simon Ramo, Warren Ruderman, S. Fred Singer, Athelstan Spilhaus, Edward Teller, Howard Turner, O. G. Villard, Jr., Dean A. Watkins, Eugene Wigner.

aid a scientifically poor proposal, are likely to be self-defeating. "If an individual tries to relate his project to cancer it must be assigned to the NCI [National Cancer Institute]," Eaves says, "but if it is too distorted the members of the study section may not understand it and it may not be funded."

But there are cases when angling a grant can help. The National Institute of General Medical Sciences (NIGMS), which is a major supporter of basic research, was recently funding only the top 15 percent of grants assigned to it (priority scores of about 1.5 or less). An applicant who could angle a basic project toward a richer institute, such as the NCI, clearly had a better chance of being funded. Disparity between the paylines of different institutes reflects social and political decisions to support research on one kind of disease rather than another. The mismatch between these decisions and the scientific merit of applications in various areas is, from the scientists' point of view, the major injustice, such as it is, in the peer review process. Study sections are not unaware of this glitch in the system. According to one member, sections are occasionally shocked to find that an application with a priority score of 1.6, say, has not been funded by the NIGMS. The study section cannot reassign the application but may give it a higher priority score the next time around. (This is not without cost: a batch of high scores from a study section will lead to a general lowering of all its scores in the normalization process.) Study sections are not supposed to know the paylines being worked to by the institutes, but in practice this information seems to be available.

Peer Review System under Attack

Some study section members admit that "accidents" occur—usually when an outstanding application is assigned to a penurious institute—but quickly add that the peer review system is the best there is. Members perceive the system to be under attack from enemies both without and within. "University administrators, especially from scientifically less active institutions, feel they ought to get more money than they do, and politicians want the grants to be distributed on a pork barrel basis. The system has to be defended as strongly as possible from such pressures," says Leonard S. Lerman of Vanderbilt University. The enemy within is the policy, partly imposed on the NIH, of routing an increasing volume of funds through the contract mechanism and big center grants. Contracts differ from grants chiefly in that they are awarded for targeted research, the goal of which is specified by the NIH. Contracts are awarded by a variety of review processes, the commonest being a mixed committee of NIH administrators and outside scientists. (This is a peer review process insofar as the administrators are considered the peers of the applicants.) Center grants are lump sums awarded to centers specializing in a specific disease; about half go through the NIH's study section system, and half, at the wish of the institutions concerned, are reviewed by local committees.

Devotees of the peer review system contend that the growing reliance on alternative mechanisms of support weakens or even subverts the peer review system, as well as giving the public less value for money. A common objection to the contract mechanism is that, for various reasons, the scrutiny of applications is less rigorous than in the peer review process. Sato, who serves on both a study section and an NCI contract committee, says, "I have the feeling that the outside scientists have a lesser input on contract decisions than they do with grants. This is because the applications for contracts are without doubt inferior to applications for grants, yet a higher percentage can be funded. So there is not an awful lot of soul-searching, as there is with grants." According to P. Roy Vagelos, chairman of the department of biological chemistry at the Washington University School of Medicine, St. Louis, "The NCI's enormous contract programs are in some instances just bad. A man whose ideas would be worth a modest grant from a study section often receives inflated support through the contract mechanism." (According to James A. Peters, NCI deputy scientific director for etiology, the NCI's contracts receive a fully adequate peer review from outside scientists, who are no mere window dressing. On the basis of cases where the same project has been studied on a grant and on contract, he believes that contracts give as good value for money as do grants.)

Study section members also compare the big center grants unfavorably with the peer review system, chiefly because the distribution of research grants at the local level allows the entry of the very politics and prejudices that the peer review system is designed to exclude. "I and my colleagues feel very strongly that we would prefer to have our grant applications in the hands of our peers at a distance rather than our peers at home," says Charles G. Cochrane of the Scripps Clinic and Research Foundation, La Jolla. Big center grants also produce an umbrella effect: "It's much easier to bring in poor quality grants because other parts of the package are good," says Jerome Gross of the Massachusetts General Hospital. And Vagelos feels that the peer review system remains by far the best protection for the taxpayer's dollar: "The NIH's push towards big centers is going in the wrong direction because it takes the primary review process out of the hands of scientists. The peer review system, by encouraging the independence of young scientists, has built up the type of science which is really great in this country. Once you have local administrators giving out the money, scientific judgment plays a very much smaller part. If the review of big center grants is to be done in a manner similar to the review of contracts, then the quality control guaranteed by the peer review system will be lost."

Is the peer review system really under attack? "One might get that impression because of the tremendous growth in the contracts program in the last few years," says Schiaffino. But grants reviewed by the study section system still account for roughly the same amount of funds ($461 million in 1968; $437 million in 1970). Outgoing NIH director Robert Q. Marston, who is said to be concerned about the image of the contracts program, has recently appointed a committee to see if contracts should not be reviewed by a method similar to that of the study sections. Whatever the committee's recommendations, its subject of inquiry touches on a wider issue, the proper balance between free and targeted research, which will be an important decision point for Marston's successor.

Although other mechanisms of support may be more appropriate in certain kinds of targeted research, none is obviously better than the peer review system when it comes to the assessment of pure scientific merit. The peer review system may not in practice always attain ideality, but the intense loyalty it commands from those who have participated in its operation is probably a not insignificant guarantee of the system's overall fairness.

—NICHOLAS WADE