CONTENT *On pages 7-19 of this issue, Lutz Bornmann and Hans-Dieter Daniel present the first comprehensive study on committee peer review for the selection of doctoral and post-doctoral fellowships. The two scientists from the Swiss Federal Institute of Technology (ETH) Zurich and the Evaluation Office of the University of Zurich, Switzerland, examined the selection process of the Boehringer Ingelheim Fonds with regard to its reliability, fairness and validity. These terms stand for the questions: (i) is B.I.F.'s selection procedure reliable; (ii) does it do justice to the applicants; and in particular (iii) does it really pick the best of the bunch? In the following article, Hermann Fröhlich the managing director of the Foundation comments on the findings.*

In the hands of social researchers

B.I.F.'s answer to Lutz Bornmann's and Hans-Dieter Daniel's evaluation

Hermann Fröhlich,
Boehringer Ingelheim Fonds, Heidesheim, Germany

• A pandemic going by the name of evaluation is making the rounds in the scientific world. B.I.F. was infected by this virus a long time ago. In 1999, to reassure ourselves that we have been doing justice to the task to which we were assigned, we asked Dr. David Evered – former secretary general of the British Medical Research Council (MRC) and adviser to numerous European organizations for research funding – to take a closer look at B.I.F. and its work.

In a »Strategic options appraisal for the B.I.F.« presented in April 2000, David Evered reached the conclusion that, during the course of the years, the Foundation has developed its own autonomous sponsoring scheme which has earned it recognition and high regard in the world of biomedical science. He used terms such as »measurable value«, »impressive impact«, and even »uniqueness«. Scientists – as he put it – would sorely miss the Foundation if it ceased to exist.

David Evered regarded B.I.F's high reputation among scientists as all the more unexpected in view of the fact that the Foundation had, and despite a considerably increased budget still has, relatively limited financial means at its disposal. He came to the conclusion that the limited resources demanded that the Foundation concentrated on a few funding programmes distinct from those of larger funding agencies. Precisely this individual profile ensured that the Foundation was »visible« and taken seriously in scientific circles. And this is still the case today.

While the Board of Trustees did not put all of David Evered's recommendations into effect, they did agree – albeit with heavy heart – to sacrifice the award for postdoctoral scientists. They were also prepared to forfeit the methods courses. Since then, B.I.F. has concentrated on providing postgraduate scholarships, travel allowances and on organizing the International Titisee Conferences. The Boehringer Ingelheim Stiftung (Boehringer Ingelheim Foundation) agreed to introduce a special scholarship programme for doctoral students of the medical faculty (cf. B.I.F. FUTURA 2002, vol.17, p. 3)

David Evered's analysis was based on discussions and meetings both with Trustees and administration, scholarship holders past and present, and representatives of the Deutsche Forschungsgemeinschaft (DFG) – in other words, with the people concerned. Spurred on by his appraisal, we decided to take things a step further and find out whether the B.I.F. selection procedure really does pick the best of the bunch! Facts and figures were needed now!

We are lucky enough to have been acquainted for many years now with the renowned research scientist Professor Dr. Hans-Dieter Daniel – who formerly worked at the Wissenschaftliche Zentrum für Berufs- und Hochschulforschung at the University of Kassel and now holds the chair for Social Psychology and Higher Education Science at the ETH Zürich. In addition, he is head of the Evaluation Office at the University of Zürich. B.I.F. fell into the hands of empirical social research at the beginning of 2001, when the Ph.D. student Lutz Bornmann, decided to base his thesis on this topic.

The results of Bornmann's three-year endeavours were published recently in Waxmann Verlag (www.waxmann.com) under the title »Stif-

tungspropheten in der Wissenschaft. Zuverlässigkeit, Fairneß und Erfolg des Peer Review« (»Foundation prophets in science. Reliability, fairness and success of peer review«). The author duly received his Ph.D. title with »summa cum laude«. Congratulations!

This work is no less than the first comprehensive study to be published on the question as to whether peer review – the assessment of professional colleagues by professional colleagues – is also a suitable method for the selection of scholarship holders. Having been extensively studied for the selection of manuscripts and grant applications, and despite occasional shortcomings, it is long since standard procedure and still generally regarded as the best of all possible worlds.

The premises offered by B.I.F. then, as now, are favourable for such a study – not least due to the fact that the Foundation maintains contact with over 90 per cent of its former scholarship holders – we know where they are and what they are doing. Bornmann's analysis investigates all of the 1,954 applications for a Ph.D. scholarship submitted between 1985 and 2000 and all 743 applications for a post-doctoral fellowship received between 1985 and 1995.

The three criteria according to which Lutz Bornmann evaluated these 2,697 applications from 15 years of scientific promotion are known in social scientific terminology as »reliability«, »fairness« and »predictive validity«. These terms stand for the questions: (i) is B.I.F.'s selection procedure reliable; (ii) does it do justice to the applicants; and in particular (iii) does it really pick the best of the bunch?

The extent of agreement among experts is the gauge as to how reliable the process of assessment actually is. Experience from scientific funding and other areas shows that the more complicated the issue and the more aspects which are to be considered, the less the experts will agree in their recommendations. What could, after all, be more complex than the draft of a scientific project that ventures – or should venture – onto new territory?

At B.I.F., three channels are involved in the selection of the scholarship holders. Provided the effort is justifiable in terms of time and money, a staff member interviews the applicant and writes a report paying particular attention to his/her curriculum vitae and personality. The Foundation requests an external reviewer, chosen on the basis of his scientific expertise, to scrutinize and submit a report on the proposed project and on the laboratory in which the applicant plans to work. The Trustees of the Foundation, seven scientists of international repute, make the final decision.

As reliable as DFG and NSF

• The extent of agreement between the Foundation's staff, reviewers and the Board of Trustees cannot be expressed as a single figure. However, Lutz Bornmann is of the opinion that it is astonishingly high – and a glance at table 6 on page 57 of his book confirms this. This phenomenon can be explained only partly by the fact that if the reviewer's evaluation is already accessible, it may influence the member of staff in his assessment, not of the applicant personally, but possibly of his application as a whole. The Board of Trustees, on the other hand, has access to both the assessment of the external reviewer and the staff member's report. In spite of these limiting observations, the outcome is extraordinary.

This conclusion is all the more convincing, given that the administration and reviewers have a free hand in making their recommendations, while members of the Board of Trustees must make theirs according to the financial framework available. At every Board meeting, the Trustees have to reject scholarship applications that they would have preferred to approve. While the external reviewers recommend 62 per cent of the applications for approval, the Foundation's staff endorses 43 per cent, 10 per cent of which they strongly recommend. The Board of Trustees has to reduce the number of suggestions to 25 per cent.

Despite considerable pressure to reach a decision, the extent of agreement among the Trustees is significant. Decisions are made in two or three rounds. In the first round, 10 per cent of all the applications are classified as clearly eligible for promotion, while 66 per cent are rated as definitely rejected. The agreement among the Trustees is thus 76 per cent. Social scientists classify a value of between 71 per cent and 80 per cent as a moderate agreement.

The DFG, to draw a comparison, has an agreement rate of 82 per cent for grant applications, the Heart and Stroke Foundation in Canada 73 per cent, while the National Science Foundation (NSF) in USA has an agreement rate of 68 per cent. So, when it comes to the »reliability« of our selection of scholarship holders we are in the very best company!

Nationality plays no role, but gender and discipline do

• What about the »fairness« of our selection procedure? How does it stand up to the test? In other words, do certain characteristics, which cannot be ascribed to science, influence the selection procedure? To find out whether this is the case, Lutz Bornmann hurled himself at the Ph.D. students, pooling their qualifications and attributes to mould a »typical« applicant. The common-or-garden B.I.F. scholarship applicant is male, German and a biologist. He proposes to carry out his project at a German university, and attained his degree at the age of 26 with a final grade of 1,4 (best possible mark: 1,0). He attended more than one university during the course of his studies. He submitted two letters of recommendation with his application and the latter was endorsed by both external reviewer and Foundation's staff. The above criteria having been met, the probability that he will be allocated a scholarship is 50:50.

In a second step, Bornmann altered one of the attributes of his phantom applicant, while the other attributes remained unchanged. First the good news: if the applicant is not German,

his prospects of success are just as high. Now the somewhat more unsettling results: if the applicant is female, her prospects fall by 17 per cent. Should the applicant be a chemist, they drop by 25 per cent. If, however, he intends to work at a Max-Planck-Institute, his chances rise by 17 per cent.

In an alternative calculation of the variables – which is less easy to explain and therefore not presented in detail here – the deviations are less dramatic. However, these results cannot be made seem better than they are either! For years now, we at B.I.F. have observed that our female applicants and chemists generally fare less well, while Ph.D. students in Max-Planck-Institutes had rather better cards. Now we have facts and figures!

So do certain criteria, which have nothing to do with science at all, influence our decisions? To try to answer this rather important question, Lutz Bornmann transposed his statistics onto the group of post-doctoral applicants. And lo and behold, everything was hunky-dory! None of the variables nationality, sex, discipline or proposed research institute showed any statistically significant deviation from the norm. This observation partly restores our somewhat shattered self-confidence. Having said which, it still doesn't divulge the causes!

The question of »women in B.I.F.« and their putative disadvantage has been a matter of concern to us at B.I.F. for some years now, as can be read in B.I.F. FUTURA 1998, vol. 13, p. 159 and p. 237. Here it is proposed that in science, as indeed in all other walks of life and despite all efforts to the contrary, the sex-specific differences prevailing in the professional and private world are deeply rooted in social expectations and subsequent education. In other words, personal reasons have a greater say in a female applicant's choice of promotion theme and desired place of work than for her male counterpart. The difference between men and women disappears only at the next step, the post-doctoral phase, when the decision to pursue a career in academic research has definitely been made.

However, to empirically substantiate this hypothesis, Lutz Bornmann would have had to climb into the ring with the Trustees to assess the candidates' projects and laboratories, which would be a fruitless endeavour. And yet precisely these are the decisive criteria when panning for the top candidates out of the excellent set of applications already recommended by reviewers and administration.

The deviations between the disciplines are considerably easier to account for than the »women« question. The Boehringer Ingelheim Fonds is a Foundation dedicated to the promotion of basic research in biomedicine, helping to discover the very essence of human life. It is clear that considerably more biologists than chemists are dedicated to this aim. Chemists are therefore obviously more liable to seek and find sponsoring by the Fonds der Chemischen Industrie (Fonds of the Chemical Industry).

In the same way, we have no problem explaining why Max-Planck-Institutes have the edge over universities. The working facilities and financial conditions generally (!) available at Max-Planck-Institutes are something of which most (!) scientists in universities can only dream! So it is hardly surprising that they attract those up-and-coming young academics that have set their sights high. And they are precisely the students that B.I.F. too has cast its eye upon. For an analysis of this, see »Immer dieselben?«(always the same names?) B.I.F. FUTURA 1991, vol. 6, p. 5.

PNAS, EMBO Journal, Nature, Cell, Science...

• There is no doubt that the questions of reliability and fairness must also be taken seriously. However, things don't get really exciting until it gets down to examining the validity of our selection. Do we really sort the chaff from the wheat, albeit at the expense of the occasional good grain? A first indication that this is the case could be gleaned from the fact that only two per cent of our scholarship holders failed to complete their Ph.D. thesis. At the British Wellcome Trust, eight per cent of the scholarship holders do not see their Ph.D. project through. However, these figures alone were hardly going to be enough to make a case. So to conclusively determine the success of our selection procedure, Lutz Bornmann turned his attention to those scholarship holders who have remained in academia. As is customary in such disciplines, he consulted the list of publications. All in all, Bornmann analysed 2,039 scientific articles from 120 former scholarship holders. These appeared in 508 journals, whereby 36 journals contained 10 or more entries. The greatest number of articles was published in the Proceedings of the National Academy of Sciences (USA), followed by the Journal of Biological Chemistry, the EMBO journal, Nature, Development, Cell and Science. A return beyond our wildest expectations! Things become even clearer when he calculates how often the articles by scholarship holders were cited in their disciplines. Here Lutz Bornmann based his calculations on the 1,949 papers published between 1991 and 2000 in journals classified by the Institute for Scientific Information (ISI, Philadelphia, USA) into three groups: »multidisciplinary«, »molecular biology and genetics« and »biology and biochemistry«. He recorded how often the article had been cited by the end of 2001 and compared these figures with the average frequency with which an article is usually cited in these journals. For each of the ten years in question, he determined three, i.e. a total of 30, so-called»crown indicators«.

Passed with distinction – no reason to rest on our laurels!

• Surely anyone still questioning the efficacy of our selection procedure must now finally admit defeat. 21 of the 30 crown indicators established by Lutz Bornmann for B.I.F. turned out to be between very good and excellent, 7 rated higher than average, while only two were of average standard. The question which drove us into David Evered's arms in 1999 and which has not given us peace of mind

since, has, finally, and according to the rules of empirical social research, been decided in our favour. Yes! We do justice to our task and achieve our objective. We provide financial assistance, in Germany and beyond, for the best up-and-coming young scientists in biomedicine – admittedly not for all, but for as many as we possibly can. »We«, that is to say, first and foremost the Trustees, whose standards with respect to the quality of applicant, proposed project and work group have stood the test of time. May we, once again, thank them for their selfless commitment!

In February 2004, the secretariat of the Foundation presented each of the Trustees with a copy of Lutz Bornmann´s findings with the dedication: »Nothing is so good that it can't be made even better«. In fact, Bornmann himself made a point of ensuring that success does not go to our heads by recommending »Measures for optimizing committee peer review«.

Some of Bornmann's suggestions have already been implemented. These include improved electronic processing of applications and supervision of the selection process. We have also introduced a pre-selection system, in which the Trustees have the last word. In putting certain other suggestions into effect, progress has actually beaten us. In 2003, for instance, 60 per cent of the scholarships were awarded to women. Only time will tell if this marks the beginning of a new trend. Further suggestions are still being pondered in our hearts. Whether or not they can be brought into effect will depend, among other things, on whether the amount of effort involved can be justified. Promotion of research versus administration costs! His advice is certainly constructive and worth considering, placing us as it does in the debt of Lutz Bornmann, Hans-Dieter Daniel and empirical social research. To repay this debt will be our duty and our pleasure!