Ulf Sandström & Tobias Harding:

PEER REVIEW
AS A MEASURE OF SCIENTIFIC QUALITY

Peer review is by definition the reviewing of scientific projects or project proposals by the scientist's scientific peers. According to the Royal Society this method is "to the running of the scientific community what democracy is to the running of the country." (Royal Society 1995) According to its critics however the system stigmatises the whole scientific community and conserves it from progress in terms of disciplinary composition as well as age and gender structure. The peer review system has even been accused of promoting corruption. Even a more balanced view of the system describes it as something much like Winston Churchill's view of democracy – "the worst form of government, except for all the others that have been tried." (Foltz 2000)

That peer review is the best system so far may not be a simple statement that a better system has yet to be found. Instead it may be the case that at least when it comes to research at the scientific front-line no one but the very best scientists can possibly have the necessary competence to evaluate it. In these cases the problem might very well be that evaluation is hard even for these well-established scientists, that is specialist scientists at the very front of their fields may be literally peer-less.

Peer review might be the internationally established norm but as we shall see the term in itself includes an unsuspected variation. A scientist's peer in regard to the peer review system is usually defined as an impartial expert within the relevant field (Kruytbosch 1989). This brings up questions of what is to bee defined as a scientific field and what is to be done with interdisciplinary proposals, that is proposals not classifiable in the usually disciplinarily defined scientific fields. Even more crucially: this definition brings up questions regarding the impartiality of the referees in different forms of peer review. As might be expected there is also a significant variation when it comes to the practical methods of conducting a peer review, and this variation should be expected to increase as the funding agencies are adjusting the systems to counter the

sometimes fierce critique. It is this variation as well as the variation already in existence that this overview aims to cover.

DIFFERENT FORMS OF PEER REVIEW

One of the most inclusive views of peer review is the one stipulated by Gibbons & Georgiou (1987). According to their definition any method of evaluation based on the opinion of the scientific community is a peer review. In this wide range of methods the following three major groups or categories are identified:

1. *Direct peer review* is the system most often referred to simply as "peer review". It is the method in which proposals or finished scientific work is evaluated by a pre-existent panel consisting of established experts within the discipline in question. This is in most countries the classic way of scientific evaluation.

2. *Modified peer review* is a version of the standard peer review system in which either the panel, its task definition or both has been significantly modified. A modification of the panel may consist in it being replaced with a less permanent group of referees (se below) or with a panel in which other interests than the strictly speaking scientific ones may be included. A modification of the tasks may be that the panel is not the only group involved in project decisions. The decisions might for example be made by one or several officials after hearing the opinions of referees and possibly those of other parties.

3. *Indirect peer review* is in Gibbons' and Georgiou's terminology a system in which the opinions of the scientific community guides decisions without panels of referees. The major such method that is being used nowadays is the bibliometrical analyses guiding for example the British funding agencies. Under this system funds are being awarded to those researches whose previous work has bee quoted in scientific articles to the larges extent or gets the highest scores in specific quotation indexes. The scientific community is thereby speaking indirectly through these indexes.

The indirect as well as most modified forms of peer review may bee termed impure forms of the system, as specific and known experts no longer are their sole actors. There is however also a major variation even if only the pure forms working with panels are taken into account. These differences between peer review panels can be ordered along two major scales: the scale of panel

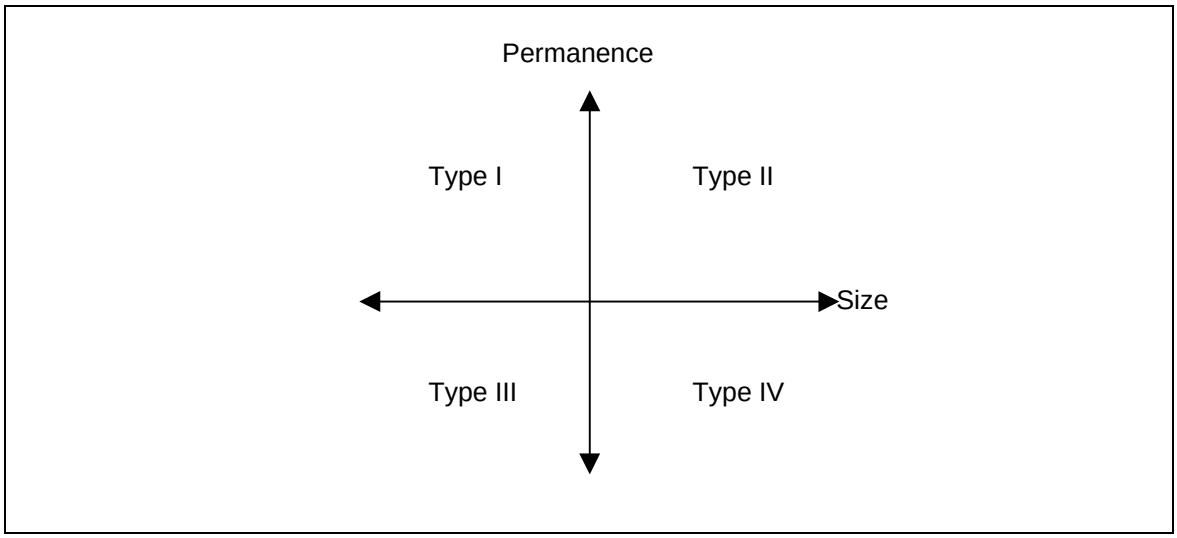

Fig. 1: Panel organization

permanence and the scale of panel size (as is illustrated by fig. 1). In many systems such as for example the Swedish ones the panels are permanent (in the Swedish Science Council they consist of highly merited peers chosen by he academic communities) and may stay in position for several years (Swedish Science Council 2002). In other systems such as the one used by the British Engineering and Physical Science Council (EPSRC) a new panel is put together for each new proposal that is to be reviewed. This is possible thanks to the large pool of available referees that has been collected by the council and that is regularly published for the benefit of the applicants (EPSRC 1997). This system makes sure that any given proposal can be reviewed by the best possible combination of reviewers wile it in a system with fixed panels might have been made by a standardized and not wholly appropriate panel (as is often the case, especially when if the proposal is an interdisciplinary one). The National Science Foundation (NSF) also uses a system of this kind. While the EPSRC gives the applicant opportunity to influence the selection of referees, however, the NSF lets programme managers make decisions after hearing peer opinions via mail. This can be held in contrast to the other major American funding agency, the National Institute of Health (NIH) in which the same decisions are made solely on the bases of classic panel review (Folz 2000).

Regardless of how permanent the peer review panels may be their reliability will always be related to the way in which new referees are appointed. Here to there is a major variation from the NSF system with its powerful deciding programme managers to the Swedish Science Council (*Veten-skapsrådet*) in which panels are appointed directly by the academic community via electoral colleges.

The size of a peer review panel matters mainly because a large panel makes it easier to represent several groups and several interests in a single panel. At the same time the importance of individual opinions are minimized, possibly leaving room for a more generalized opinion of the scientific community as such. It is important to the legitimacy of a panel that it is not preserved as a narrow vessel for self-interested elites, a risk that may be minimized – however as we shall se never terminated – by the use of larger groups. Unfortunately larger groups also have the tendency to be significantly less effective. Obviously a compromise is needed. (Sandström *et al* 1997)

The last major difference between the major forms of peer review is what is being reviewed. As has been mentioned above a peer review may be the evaluation of applications for grants *ex ante* but it may just as well be used to evaluate projects *ex post*. As implied by the term "indirect peer review" the direct method is actually often used by scientific periodicals to evaluate articles for publication (Foltz 2000) and thereby making the term "indirect peer review" apt by in reality making it a generalised indirect ex post version subject to the more direct methods.

THE CASE AGAINST PEER REVIEW

As the legitimacy of all peer review systems is firmly connected to the authority of the academic establishment most arguments against the system seems to centre on mistrust against this very establishment. The major charges against peer review are in other words charges of conservatism. There are however other charges, mainly of corruption and of simple inefficiency.

To start with the charges of conservatism there is a major suspicion against peer review that it has a strong bias against interdisciplinary work and applied sciences. This claim is based on the fact that review panels tend to consist of scientists merited within the single discipline that the panel is instituted to review. These reviewers also tend to be the ones most merited in the discipline as it has looked in times past, that is the ones who have built the existing structured and who in most cases has been working inside it, that is the ones who might be expected to be the least interested in challenging these same structures. (Royal Society 1995)

On the allegations against peer review of being biased against interdisciplinarity thorough investigations have actually been made by E.J. Rinia et al. They have studied how studies published in journals outside their home discipline (and therefore deemed to be of interdisciplinary relevance) have been treated in peer reviews. In this study no significant bias has

been detected when bibliometric success rates have been compared to success in peer review. The evidence however still remains inconclusive. (Rinia et al 2001)

On the individual level peer review has been accused of promoting men before women and of being biased against younger academics. In a 1997 (Wennerås & Wold) study of research funding in the Swedish Research Councils large discrepancies was shown in women's success rate in scientific publications and in the panels of Research Council peer review. To some degree this has been adjusted and later studies has been inconclusive as to the continuation of this bias. However the suspicion still linger that peer review in its ways of recognising excellence in reality is only reproducing the characteristics and powers of an already established scientific elite (Abdallah).

Unfortunately these characteristics may not necessarily be the premier characteristics of good science. As we have seen the system has been accused of discriminatory habits. If this is the case this can also be assumed to have negative side effects on the scientific quality of the resulting research. This is the sum of the charge of conservatism: if these accusations are valid the choices made in a peer review system may not at all be promoting excellence but instead promote such projects that look like the elders of the academic community has expected them to look, wile more ground braking proposals and articles never get their chance. (Daniel 1993)

Finally; allegations have been made about information gained in peer review panels have been used by reviewers in their own research. These accusations range from general discussions of what is allowed and what is not to specific accusations against individuals. There have even been lawsuits made against reviewers for stealing ideas made in applications (E. Marchal in Folz 2000). It has been argued that ethical standards in the scientific community are at ebb. This may or may not be the result of shrinking public funding to the sciences leading to a larger need for academics to guard their own economy and their external incomes. (Folz 2000) It may also be that the sense of responsibility against ones academic peers are not as high as they ones were, possibly as a result of the growing science community and the possibility of a decrease in social control. Whatever the reason it is quit clear that the peer review system is based on trust and that this trust has to bee regained if the system is to survive.

ADJUSTMENTS MADE

If the claims stated above would be a full description of the reality of peer review it would be quit surprising that the system has remained strong as it has. As has already been stated the main reason may be that there simply are no realistic alternatives. One may turn to indirect measurements but there is no reason to believe that these will avoid the pitfalls of peer review as a system. On the contrary it is quit possible that it would be harder to get to the problem if the element of personal responsibility was removed from the evaluation process. There is not even any reason to believe that other methods would be immune to the equivalent problems.

This however does not necessarily imply that we have to accept the system as it is – which would be the height of conservatism – but may instead imply that the best road to a better system may lay in the improvement of existent methods rather than in extreme risk taking. Luckily several such improvements have been made worldwide and it may very well be that most of the negative side effects of the peer review system are problems that may be solved.

One important category of modified peer review includes the various forms of widened panels. A widened panel is a panel that includes not only researchers from a single discipline but may also include representatives of other sectors in society or representatives of other disciplines. In the first case the idea is to complete the strictly academic quality evaluation with more relevance oriented aspects, as a way of guaranteeing the close relationship between science and society. This aim may on the other hand just as well be met with the use of separate panels for quality and relevance evaluation. One might also consider that scientists may be just as well equipped as other citizens to evaluate the public relevance of their research, when it comes to the humanities and the social sciences probably even more so.

Adjustments have also been made in order to counteract the problems related to partiality. The impartiality of referees has been brought into focus in several countries. In smaller countries as well as generally in smaller academic fields it may be extremely difficult to actually find impartial referees, simply because the relevant scientific community is so small that all its major researcher will be more or less acquainted with each other. In Sweden's Research Council for Engineering Sciences (TFR) this problem has been counteracted with the use of international peers in its peer review system. This practice has however in turn been criticised for bringing in referees with little or no knowledge of the Swedish national research system and academic practises. (Sandström *et*

al 1997) As a result it is reasonable to believe that peer review system in smaller countries might benefit greatly from a harmonization of the European systems for peer review.

Other measures that have been taken against the partiality problems can be summed up as moving the peer review system and the academic procedures closer to the general standards of government agenises, that is, closer to standard based on the impartiality of the *Rechtstat* and of the Rule of Law. In other terms this means that academic procedures are made subject to the same kind of red tape and written rules that govern most of society but at the same time stands in sharp contrast to the informal honour society that once governed academic life (Marton 2000). According to Chubin (1994) it is vitally important that the aim, the purpose and the requirements in a peer review is totally clear to both referees and reviewed scientists. If this requirement is not met consequences might be significantly negative both to the quality of the particular evaluation and to future trust in peer review as such.

According to several organizations using peer review in research evaluation the answer to this dilemma is to formalize the previously unwritten (or in some cases non-existent) rules of peer review. As a conclusion to this discussion we will take a more thorough look at one example of this process: the rules and procedures used by the British Engineering and Physical Sciences Research Council as published in their report *EPSRC Colleges and a Guide to Peer Review 1997-99* (EPSRC 1997). In that organization peer review is to promote the qualities listed below.

1. Scientific and technological quality
    a. The importance of the area of research in which the project is being undertaken.
    b. Awareness of the current state of the art and related work in progress.
    c. Novelty/originality of the work proposed and its timelines.
    d. How competitive is the project in the UK and internationally? Does the project offer the possibility for a significant breakthrough?
2. Ability to undertake research.
    a. Track record/standing/previous work in the field of the principal investigator.
    b. Track record/standing/previous work in the field of the research team.
    c. Does the team have the skills required to undertake the project?
    d. Other factors e.g. if a new lecturer, you may wish to consider the lecturer's potential to undertake the project.
3. Viability and planning
    a. Are objectives, milestones and deliverables clearly defined and realistic?

b. Are the proposed methodologies sound and the techniques suggested appropriate?
c. Where appropriate, is an adequate project management structure in place?
d. Is the necessary research infrastructure to support the project in place?
e. Is the timescale proposed realistic?
f. Is the project achievable?

4. Relevance to beneficiaries
    a. Likely impact/importance of research to industry, commerce, government, academia etc.
    b. What contribution, if any, will the project make to UK competitiveness or quality of life?
    c. Does the research have any relevance to Foresight recommendations?
    d. For collaborative projects, is there an appropriate range/mix of collaborators? Is the input, financial and otherwise, from the collaborators?
    e. Have mechanisms to ensure potential exploitation been identified?
    f. Are the dissemination paths adequate?
5. Resources and cost effectiveness
    a. Is the level and choice of recourses reasonable for the project proposed? Have these been adequately justified?
    b. Does the project offer a good value for money?

Note that the EPSRC is a council for applied sciences and that the requirement of practical relevance as well as organizational and economical aspects under this model is included in the evaluation made by peer review. It may be argued that scientific peers themselves are experienced in the management of projects and therefore competent to evaluate these aspects. It might also be argued that they are better acquired that non-scientists to judge possible outcome of the project and therefore able to evaluate its possible societal relevance. In many versions of the peer review system (such as the one until recently used in the now closed Swedish *Forskningsrådsnämnden*) questions of relevance are however not considered to be best judged by scientific peers but instead better left to bureaucrats, politicians or representatives of the relevant sectors of public life.

CONCLUSION

It is clear that no perfect list or evaluative manual can cover all academic fields and at the same time be specific enough to guide referees in their practical work. That is however not the point. The idea is instead to make the rules of any specific field clear to all involved parties and to strengthen the legitimacy of the process by making shore that everyone follows these rules impartially.

This might also be the best final remark in this short overview: peer review may very well be the best possible system for scientific evaluation of quality but if so, this is not to say that we have found one universal solution to all problems but instead as a direct result of the peer review systems ability to variation and adaptability. In the end it is perhaps simply an acceptance of the fact that maybe scientists are themselves the people best accepted on to recognize got science, while on the same time the modifications made in resent times represents an acceptance of the impression that scientists are mere citizens and needs to be guided just as much as everyone else.

REFERENCES

Abdallah, Laila 2002: "Kvinnor, forskning och karriärhinder" in Sandström, Ulf (red.) *Det nya forskningslandskapet*, SISTER skrifter 5, Nya Doxa, Nora.

Daniel, H.-D. 1993: *Guardians of Science, Fairness and reliability of peer review,* VCH, Weinheim.

Engineering and Physical Sciences Research Council (EPSRC) 1997: *EPSRC Colleges and a Guide to Peer Review 1997-99,* Polaris House, North Star Avenue, Swindon.

Chubin, 1994: "Grants Peer Review in Theory and Practice", *Evaluation Review,* Vol 18 No 1, Feb. 1994, Sage Publication Inc.

Foltz, Franz A. 2000: *The Ups and Downs of Peer Review: Making Funding Choices for Science.*

Gibbons, Michael & Luke Georghiou1987: *Evaluation of Research, A Selection of Current Practices,* OECD, Paris.

Kruytbosch, Carlos E. 1989: *The role and effectiveness of peer review,* Science Indicators Group, National Science Foundation, Washington D.C.

Marton, Susan Gerard 2000: *The Mind of the State, The Politics of University Autonomy in Sweden 1968-1998,* Department of Political Science, University of Gothenburg.

Royal Society 1995: Peer Review - An assessment of recent developments, ”www.royalsoc.ac.uk/policy/index.html” 2002-03-11.

Rinia, E.J., Th.N. van Leeuwen, H.G. van Vuren, A.F.J. van Raan: Influence of interdisciplinarities on peer-review and bibliometric evaluation in physics research in *Research Policy* 30 (2001).

Swedish Science Council (Vetenskapsrådet) 2002: “www.vr.se” 2002-07-31.

Sandström, Ulf et al 1997: “Does Peer Review Matter?” in *Peers on Peers, Allocation Policy and Review Procedures at the Swedish Research Council for Engineering Sciences*, TFR.

Wennerås, C. & Wold, A. 1997: ”Nepotism and Sexism in Peer-Review”, *Nature*, vol. 387, pp. 341–343.