COGNITIVE BIAS IN PEER REVIEW: A NEW APPROACH

Ulf Sandström

ulf.sandstrom@liu.se
Ulf Sandström, ISAK, Linköping University, 581 83 Linköping, Sweden

Abstract

The concept of cognitive similarity, developed by Travis and Collins (1991), is the starting point for this paper. We suggest that cognitive similarity is detectable through bibliometric analysis using bibliographic coupling (Kessler, 1963) or, as an alternative, noun phrases in title and abstract. Connected to this hypothesis is the possibility of cognitive bias in peer review. If academics tend to give higher scores to research with which the reviewer has a cognitive similarity there is a situation of cognitive bias. The design of the research project is described and the data sources available are discussed. With data on applicants and reviewers, and complemented with bibliometric identification of each individuals publications, this project will potentially give an essential contribution to our understanding of the peer review process.

Background

Essential in the classical model of research councils is that scientists, unrestricted by and independent on external factors, give advice to final decision-makers. In Sweden, active researchers constitute a majority on the council's board, i.e. in all bodies where applications are assessed and evaluated and where grants are decided upon. Committee chairmen are chosen through an electoral process involving all members of the university research community. Peer review is first and foremost used to guarantee the quality and diversity of basic research, but the method has been copied by other agencies. A specific feature of the Swedish research councils is that scientific committees are in command of the work. As a consequence, and in comparison to the situation at NSF (main U.S. financing body), program officers in Sweden have little or no influence over granting procedures, initiatives and priorities (see Langfelt et. al., 2004).

Peer review is intended to improve both the technical quality of projects in research and the credibility of the decision-making process. The independence of peers should make them more effective than internal reviewers. Nowadays it is taken for granted that peer review is fundamental to the institution of science and a symbol for the autonomy of science (Chubin & Hackett, 1990).

Inspired by the NIH (US medical research agency) Scandinavia has (to a large extent) followed the committee model for review of applications. Procedures applied at councils in Sweden are the following: Each applicant (principal investigator) submits a CV, bibliography and a research proposal. Applications are reviewed by one of several scientific committees, each covering a specified research field or discipline within the three different sub-councils (medical; natural-engineering; social and human sciences). Each application is rated by several reviewers in the committee to which the application is assigned. Reviewers grade the proposals and the "track record" with a score, and that score form the basis for the ranking later on used for funding decisions. Review committees set up a conflict-of-interest protocol stating which members of the committee that have an affiliation with applicants. Affiliation might be maternal, kinship or supervisor relations. These protocols are open for public access according to Swedish legislation.

Nepotism, or conflict of interest, is an interesting ethical problem for the research community, but that type of bias is hard to distinguish from another type of bias – *cognitive bias*. As it is hard to get access to data this aspect has seldom been the object for closer studies. Actually, there are only two studies available on that aspect of peer review and the purpose of this project is to bring in a new quantitative methodology and a large empirical material to this interesting research question.

Detecting Cognitive Bias

Bias in peer review is a crucial issue that has generated serious discussion in several scientific journals (for overview essays, see Wesseley, 1998, Bornmann & Daniel, 2005). Any type of bias would be detrimental towards strategies for scientific excellence. Although peer review functions in order to enhance the quality of research and to prevent poor research from taking place this is, of course, not always the case. Possible flaws in the peer review process have been disclosed in recent years. Fraudulent research, e.g. plagiarism, falsification or fabrication of data, is sometimes detected after the peer process. The number of known cases is quite small, but some of them reveal the possibility that there is a problem of hidden statistics.

There is surprisingly little evidence on the effectiveness of peer review from formal studies. Some studies are concerned with social biases of different types – institutional, gender, ethnical or other – in the grant peer-review procedures of research councils. Most of these come to the conclusion that the concern for social bias is overstated (Cole, 1992). Opposed to this conclusion is the standard reference for gender related studies: the Wennerås and Wold study published in Nature (Wennerås & Wold, 1997). Other studies covering several years and larger samples have not been able to show the same gender-wise discrimination (see, Sandström & Hällsten, 2008).

Working in the Mertonian tradition the COSUP-studies by Cole, Rubin and Cole (1978) looked upon science as a social system. They developed a design for testing the hypothesis of "old boy network" and "institutional cronyism" and other social allegations. In reaction to this study Travis and Collins wrote a pivotal paper published in 1991 (Travis & Collins, 1991, see also Mahoney, 1977), which pointed at the lack of a cognitive view in the Cole studies. They coined terms like "cognitive particularism" and "cognitive similarity" for different peer review situations. That scientists with a common view of their fields – a social network of friendship – could pose a challenge to a fair review process was apparent also in the Cole studies, but they operationalized it into measures of ranking of applicants' and reviewers' current department status and into social positions (Travis & Collins, 1991: 326).

Cognitive bias "depends on the existence of cognitive boundaries within and between scientific specialties and disciplines" (Ibid. 327; Whitley, 2000). Even in small areas of science there is a possibility of considerable cognitive variety, especially when conflicting perspectives are encountered. Accordingly, a different evaluation of an application is expected from different cognitive communities. To what extent this type of particularism would affect an ordinary evaluation is not stated, but we can expect that there would be a tendency for giving lower grades to competing areas, e.g. in an committee processing applications in economics research from two areas – the neo-classical school (A) and the innovation system school (B) – we should find a tendency, ceteris paribus, for the A-school to give lower grades to applications from the B-school and vice versa. Whether this is the case or not is an empirical question as it might very well be the other way around: you tend to be more critical towards research, and require more from research, with which you are familiar.

Methods for cognitive distance

While a lot of interest has been devoted to questions on peer review and panel decisions on grant applications, especially on features of successful applicants, less interest has been put on the cognitive dimensions (Bazeley, 1998). This is almost always explained by the lack of data or, expressed more clearly, lack of access to data. Since two Swedish researchers published the first ever analysis of peer-review scores for postdoctoral fellowship applications few studies have used the crucial variable reviewer attributes and affiliations. Bornmann and

Daniel (2007) and Niemenmaa, Hemlin & Montgomery (1995) are two exceptions. Bornmann and Daniel in their BIF-studies controlled for reviewers' attributes. These were gender, nationality and evaluation experience (number of reviewed applications over time). Apparently, none of these studies considered the cognitive dimension.

Travis and Collins used a qualitative method for their study as they had an opportunity of direct observation of ten committee meetings within SERC of the UK. We propose a more quantitative method using data from the research councils for medicine and natural sciences. The methodology should also be applied to the social sciences as they in general are said to have lower consensus (Cole, 1992). The database consists of data on each application, and data on the reviewers and their grading of the applications. This is a unique and rich material never before utilized for research purposes. In general, each committee consisting of 5-6 members process around 100 new applications each year.

Is it possible to detect cognitive bias? This question runs down into two different (in principal) researchable problems: The first being which method would be preferred for detecting cognitive similarity or dissimilarity: is bibliographic coupling a method for this purpose or would it be needed to use or complement with other methods, e.g. words from title and abstract? Both methods should be investigated and evaluated. The second problem is whether it is possible to measure cognitive distance. The map shown in Figure 1 (see below), which is a model for the design, show papers published by three different groups in question. The edges are based on bibliographic coupling normalized by Saltons' cosine and node positions are calculated using the GraphViz (www.graphviz.org) application of the Kamada-Kawai algorithm (see Kamada and Kawai, 1989). A threshold is also applied, removing edges with strength less than 0.05. With these methods a measure of cognitive distance can be established.

Additional information to the cognitive distance is the citation rate, which might further illuminate the decision taken by the committees. Citations are measured according to standard procedures for normalization. The most commonly used normalization type was developed by Schubert, Glänzel and Braun during the 1980s (1988; Glänzel, 1996). Simultaneously the Leiden group (Moed & van Raan, 1988; van Raan, 2004) developed a variant methodology with the well known "crown indicator". These normalized indicators are typically named CPP/JCS or CPP/FCS depending on whether the normalization is carried out in relation to journals or sub-fields. We apply a slightly modified version of the CPP/FCS indicator for this study. The difference is that our calculation treats all papers equal, while the Leiden version gives higher weight to papers in normalization groups with higher reference values.

Design of the study

In general, the methodological strategy (design) can be described in the following terms. At least three sets of bibliographic information are collected from each committee: First, bibliographic information for the reviewers, Second, bibliographic information for the successful applicants and third, bibliographic information for the unsuccessful applicants. They all consist of publications from each author during the last seven years. A map based on bibliographic coupling is performed per committee. This map will show the cognitive connections between all authors involved in the reviewing process (both reviewer and reviewed) either using reference information or using noun terms from title and abstract. A measure of cognitive distance is the result of the mapping procedure. This measure will be used in order to detect tendencies of cognitive bias.

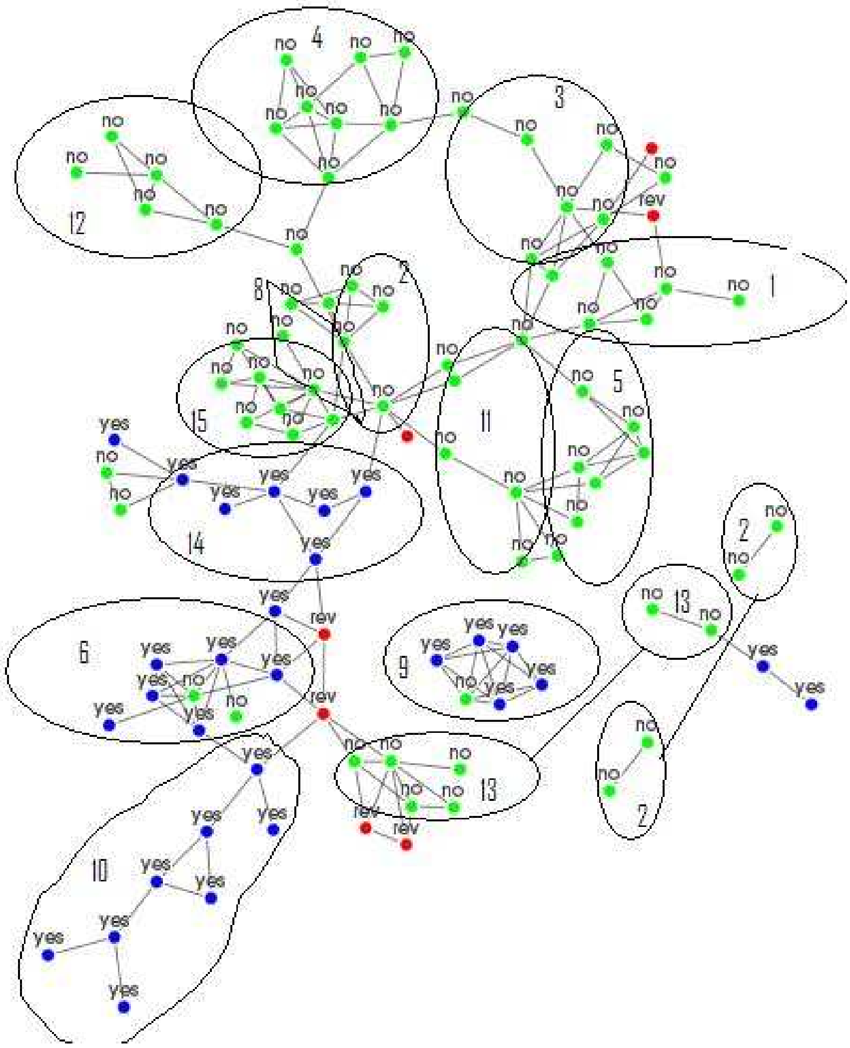

Figure 1: Cognitive distance between reviewers ("rev") and applicants (approved nodes – "yes"; rejected – "no"). Lines are bibliographic connections between nodes.

Preliminary results

The interpretation of science maps is never clear-cut and easy. Figure 1 is based on a round of fourteen groups reviewed by seven "international" peers in the area of natural science. (Therefore, this test case is slightly dissimilar to the normal situation of Swedish peer review.) We find that all reviewers have a more or less close connection to "their" cognitive similar groups. Obviously, there is a sharp distinction between the northern and southern part of the map and the approved groups are at the southern end. Ranking of groups (1–14) indicates the bibliometric value per group, i.e. group 1 have a CPP/FCS of 2.00 and group 14 have 0.89. Reviewers are at 1.40. Overall, we find the best performing groups in the northern part but the most successful groups in the south.

While the approved groups are coherent several of the rejected groups are more dispersed. This indicates that they consist of people from several research teams and were put together in order to attract funding.

Group 13 falsifies the hypothesis of a cognitive bias as they are surrounded by reviewers. But we might save the supposition due to the low performance (1.13) of the group. A test using the distance measures (data not shown) indicates that the cognitive bias hypothesis should be dropped. Probably there are a number of differences due to topical areas, funding history,

application design etc. that might better explain the differences between groups. Also, the strength between reviewers when they form a group has to be considered.
Next step would be to develop the method for different interpretation cases and to apply it, if possible, to a larger group of committees using a design that tries to eliminate the structural effects.

References

Bazely P (1998). Peer review and panel decisions in the assessment of Australian Research Council project grant applicants. Higher Education 35: 435–452.

Bornmann L & Daniel HD (2005). Selection of research fellowship recipients by committee peer review. Scientometrics, 63(2): 297–320.

Bornmann L & Daniel HD (2007). Gatekeepers of science—Effects of external reviewers' attributes on the assessments of fellowship applications. Journal of Informetrics 1(1): 83–91.

Boyack KW & Klavans R (2006). Identifying a better measure of relatedness for mapping science. Journal of the American Society for Information Science and Technology 57 (2): 251–263.

Boyack KW, Klavans R, Börner K (2005). Mapping the backbone of science. Scientometrics, 64: 351–374.

Chubin DE & Hackett, EJ (1990). Peerless science: peer review and U.S. science policy. SUNY Press.

Cole S (1992). Making science: between nature and society. Cambridge, Mass.: Harvard Univ Pr.

Cole S, Rubin L & Cole JR (1977). Peer review and the support for science. Scientific American 237: 34–41.

Glänzel, W. (1996). The need for standards in bibliometric research and technology. Scientometrics 35:167 – 176.

Kamada, T & Kawai, S (1989). An algorithm for drawing general undirected graphs. Information Processing Letters, 31(1):7–15.

Kessler, MM (1963). Bibliographic coupling between scientific papers. American Documentation, 14, 10- 25.

Klavans R & Boyack KW (2005). Mapping world-wide science at the paper level. Proceedings of ISSI 2005, pp. 426–436.

Langfelt L (2006). The policy challenges of peer review: managing bias, conflict of interest and interdisciplinary assessments. Research Evaluation 15(1):31–41.

Langfelt L, Berglund F & Maasen P (2004). Søknadsbehandling og habilitetsregler i andre lands forskningsråd. Rapport från NIFU till "Smithutvalget". Utgitt av Norges forskningsråd. Oslo 2004.

Mahoney MJ (1977). Publication prejudices: an experimental study of confirmatory bias in the peer review system. Cognitive Therapy and Research, 1(2): 161–175.

Moed H. & van Raan A.F.J. (1988). Indicators of research performance: applications in university research policy. In: van Raan, (Ed.) Handbook of Quantitative Studies of Science and Technology. Amsterdam: North-Holland, pp. 177–206.

Niemenmaa P, Hemlin S & Montgomery H (1995). Peer review protocols of research applications in psychology II. Report from the Dept of Psychology, Stockholm University, No 794, March.

van Raan AFJ (2004). Measuring Science: Capita Selecta of Current Main Issues. (Eds.) Moed et. al. Handbook of Quantitative Science and Technology Research: The use of publication and patent statistics in studies

Sandström U & Hällsten M (2008). Persistent nepotism in peer review. Scientometrics 74(2): 175–189.

Schubert A. Glänzel W. & Braun T (1988). Against absolute methods: relative scientometric indicators and relational charts as evaluation tools. In: van Raan, (Ed.) Handbook of Quantitative Studies of Science and Technology. Amsterdam: North-Holland, pp. 137–176.

Travis GDL & Collins HM (1991). New light on old boys: cognitive and institutional particularism in the peer review system. Science, technology & Human Values 16 (3): 322–341.

Wennerås, C. & Wold, A. (1997). Nepotism and Sexism in Peer Review. Nature 387, 341-343.

Wessely S (1998). Peer review of grant applications: what do we know? Lancet 352: 301–305.

Whitley, R. (2000). The intellectual and social organization of the sciences. Second Edition (First ed. 1984). Oxford: Oxford University Press.