

# Panel peer review of grant applications: what do we know from research in social psychology on judgment and decision-making in groups?

Meike Olbrecht and Lutz Bornmann

The allocation of resources for scientific research is determined by panel peer review. To make funding recommendations, the reviewers convene to evaluate the quality of grant applications. Many research studies in social psychology have investigated what (undesired) phenomena (such as groupthink, motivation losses, and group polarization) can occur in group judgment and decision-making. In the research on peer review, however, these phenomena have not been examined up to now. This article describes the peer review panel with the help of features used in social psychology to characterize groups (such as entitativity of groups, group task) and presents phenomena from the research in social psychology that can have an (undesired) effect on the judgment of panel groups. Measures to counteract these phenomena are discussed. The necessity of research in this area is pointed out.

**G**RANT PEER REVIEW is the principal mechanism for quality control in the funding of academic science and scientists, and its usage is increasing through the general trend towards the ‘soft money’ system (Guston, 2003). With peer review, it is customary to consult several peers of the grant applicants for evaluation of the grant application before the funding decision is made. There are two reasons to consult more than one reviewer. First, several reviewers have more resources to draw on than one reviewer. Individual reviewers are expected to have information that is unavailable to other reviewers, and in this way the funding decision can be made based on more information than would be the case with just one reviewer. This is called the *information integration function* of decision-making groups. The second reason has to do with the *consensus building function* of decision-

making groups (Nijstad, 2009). A funding decision will find greater acceptance when it is based on several reviewers’ opinions and most of the reviewers agree with the decision than when the decision is based on only one opinion.

There are generally two procedures for peer review of grant applications by several reviewers. In *individual review*, the scientists judge (in writing) the merit of a grant application *independently* of one another before the final decision-maker approves or rejects an application for funding. In *panel peer review*, a group of reviewers convene to *jointly* deliberate and judge the merit of an application before the funding decision is made. Whereas with written individual reviews the final decision-maker brings the judgments and opinions of several reviewers together to make a decision, with panel peer review the final decision-maker makes the funding decision based on the common judgment of all reviewers (in some cases the panel itself makes the funding decision directly; see Bornmann and Daniel, 2005). Thus the main difference between panel peer review and individual review is that the reviewers in panel peer review reach a common judgment *through mutual social exchange*.

Compared to individual review, panel peer review has the advantage that the reviewers are in a position

---

Meike Olbrecht (corresponding author) is at the Institute for Research Information and Quality Assurance (iFQ), Godesberger Allee 90, D-53175 Bonn, Germany; Email: [olbrecht@forschungsinfo.de](mailto:olbrecht@forschungsinfo.de); Tel: +49 228 972 73 13. Dr Lutz Bornmann is at the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany; Email: [bornmann@gv.mpg.de](mailto:bornmann@gv.mpg.de); Tel: +49 89 2108 1265.

to jointly reevaluate their arguments, to weigh up the arguments against one another, and to make distinctions between important and less important arguments. The arguments of others can stimulate reviewers to come up with arguments of their own. Individual errors and false conclusions can be discerned by the group and corrected in the discussion. With individual reviews, the decision-maker must put together all the judgments and opinions in independently prepared reviews and weigh them up against each other. The decision-maker is often a person who does not have the same expertise as the reviewers. Compared to individual review, the disadvantage of panel peer review is that it costs considerably more and takes more time. The reviewers have to work hard to evaluate the applications in preparation for the panel deliberation and then assemble at the place where the panel meetings will take place.

Peer review research in recent years (Bornmann, *in press*) has mainly examined peer review for journals (see overviews in De Vries *et al.*, 2009; Overbeke and Wager, 2003; Weller, 2001); more rarely, peer review of grant applications has been investigated (Daniel *et al.*, 2007; Lamont, 2009; Wessely, 1998). Studies on peer review of grant applications have examined mainly the fairness and predictive validity of *the judgments of individual reviewers and/or the final decision of the funding agency*. We found only five studies that dealt with panel judgment processes. However, in social psychology there has been intensive research on how groups arrive at common judgments and on what phenomena can take place that have an (undesired) effect on judgments.

The goal of this article is to present, on the basis of the findings of social psychology research on decision-making and judgment in groups, phenomena that can occur in judgment and decision-making processes in groups generally (such as groupthink or motivation losses) and that can be assumed to also play a role in panel peer review. The aim is to point out the existence of and to stimulate empirical research on these phenomena.

Before presenting the phenomena, the article summarizes the findings of the few studies available in peer review research that examined internal processes of panel peer review. The peer review panel is then characterized using features that are used in social psychology for the classification of groups. Following the description of the group phenomena, we look at measures designed to prevent the undesired effects of the phenomena and to optimize group processes.

### Empirical studies on panel peer review

The five studies that examined judgment processes in panel peer review were published within the last 10 years. All in all, the findings of the studies

---

## We found no study on panel peer review in the research in social psychology, and the findings of social psychology research have hardly found their way into studies in peer review research

---

indicate that the processes are influenced by a number of undesired group effects (such as an undue weight of reviewers who have *not* read a proposal compared to reviewers who have read it in the funding decision). However, the studies provide only isolated indications of phenomena that according to social psychology research are generally expected in group decision-making and judgment processes. We found no study on panel peer review in the research in social psychology, and the findings of social psychology research have hardly found their way into studies in peer review research. The studies summarized in the following yield information on group effects in panel peer review. In addition to the group effects, the description of the studies gives (as a side effect) an overview on different panel peer review processes used by funding agencies (which were evaluated in the studies).

In a case study Langfeldt (2001) investigated the different models of grant review at the Research Council of Norway (RCN) (Oslo, Norway). Langfeldt analyzed the process of awarding funds by the four divisions of the RCN: Medicine and Health, Culture and Society, Science and Technology, and Environment and Development. The peer review processes used by the four divisions differed in the composition and organization of panel meetings. Whereas for example the Culture and Society division had 15 discipline-based peer panels with three to five members each, the Environment and Development division had only one panel with nine members. The organization of the panel meetings was also different. For example, in the Medicine and Health division each of the panel members ranked all applications prior to the panel meeting. The chair of the panel had tables of these individual and average rankings. Panel decisions were based on discussion, average rankings, and the chairman's discretion. In contrast, in the Environment and Development division, the reviewers were assigned grant applications to evaluate prior to the panel meeting. They presented their assessments orally at the panel meeting. For each application there was in addition an advisory mail review, with that assessment being entered into the discussion. Here panel decision was based on discussion, negotiation, and/or majority rule (Langfeldt, 2001).

Langfeldt investigated how the review process affected the outcome of grant peer review. The study included grant reviews for the Research Council of Norway for 1997/98.

Data sources of Langfeldt's study were 619 applications and review documents, direct observation of panel meetings, and interviews with 25 panel members. The data material was evaluated according to the criteria and arguments that were given for evaluation and ranking of the applications. Langfeldt (2001) concluded:

rating scales and budget restrictions are more important than review guidelines for the kind of criteria applied by the reviewers. The decision-making methods applied by the review panels when ranking proposals are found to have substantial effects on the outcome. Some ranking methods tend to support uncontroversial and safe projects, whereas other methods give better chances for scholarly pluralism and controversial research. (p. 820)

Whereas ample budgets and rough rating scales raise the probability that original and controversial projects have better chances, tight budgets and fine rating scales tend to strengthen establish research.

In a further study, Langfeldt (2002, 2004) analyzed six international evaluations of Norwegian research. The panel cases encompassed different evaluation objects in different disciplines: two evaluations of research fields (natural science, humanities), two of research institutes (fields: social sciences, technology/engineering), and two research program evaluations (fields: natural science, multi-disciplinary). The panel cases differed in composition of the panel and organization of its work.

For instance, a five-member peer panel evaluated a natural science research programme; together, the panel members took a 10-day trip to the various Norwegian research units for this purpose. There was a clear division of tasks among the members of the panel. The panel members wrote separate reports on their own subject areas. Together they then composed the general conclusions section of the peer evaluation report.

In contrast, the five-member panel that evaluated the engineering research institute met three times. The first two meetings were held at the institute to be evaluated; afterwards the panel members each wrote a statement containing their impressions. Here there was no division of tasks among the panel members; they each contributed to the evaluation report where they could. Their contributions ranged in length from one to 10 pages. These drafts were discussed at the third and last meeting and edited into a final report.

Langfeldt (2002) conducted the study as a doctoral dissertation and reported the main findings in a paper (Langfeldt, 2004). The research question was:

What affects decision-making processes when research programmes, research institutions and research fields are evaluated by expert panels? (Langfeldt, 2004: 52).

The data sources of the study included the commissioning research councils' files on six panel evaluations of the RCN, semi-structured interviews with the participants in the decision-making processes, and in some cases, the participants' private notes and drafts for the evaluation reports. The evaluations were carried out in the period 1988 to 1991.

The study showed that the decision-making processes were characterized:

by clear scholarly task division, little interaction on research assessments, and mostly tacit compromises in case of disagreements among panel members. (Langfeldt, 2004: 55)

In cases where there was an overlap between the expertise areas of the panel members there was also some interaction on the research assessments. These were 'the only cases where panel members clearly disagreed on assessments of the research' (Langfeldt, 2004: 55). Langfeldt (2004) concluded that overlapping competence in panels was needed for open conflict between peers. It was also found that limited time was available for assessment of the applications, probably resulting in limited thoroughness in the assessment. Langfeldt (2004) therefore outlined the following needed basic requirements:

Two experts assessing each object under review and some time for discussing the results would be the minimum needed if expert panel evaluations are to have some function exceeding individual review reports when it comes to assessing the quality of research. (p. 60)

Obrecht *et al* (2007) examined a process for evaluating research training fellowship proposals at the Canadian Institutes of Health Research (CIHR) (Ottawa, Canada). At the time, the selection process consisted of three steps (based on the findings of this study, the process was later changed).

In a first step (pre-selection), applications for research training fellowships were sent to two panel members to be systematically rated at home (at-home review). The at-home scores were sent to the research funding agency in advance of the committee meeting, which averaged the two scores for each application. The research funding agency set minimum average rating levels. If an average at-home rating fell below this level, it was not discussed in the committee.

In the second step, the two reviewers introduced the application to the committee at the committee meeting. After discussion of the application by the committee, the two reviewers that had presented the application jointly proposed their scores

(in-committee score). After that, all members of the committee privately rated the application.

In the third step, the funding agency ranked the applications according to their assigned rating. Applications that were above a prescribed score were funded.

The researchers wanted to find out whether committee discussion of fellowship applications increased fairness, effectiveness and efficiency over and above the levels that could be attained through a structured review in which reviewers conduct assessments at home using a clearly defined protocol. They used five methods to address their research question:

1. They compared at-home and in-committee scores. The database covered 775 fellowship applications that were received in the two competitions for 157 awards offered in 2004;
2. They surveyed members or chairpersons of fellowship panels by Web-based questionnaire;
3. Six committee members participated in a focus group that discussed several topics;
4. They observed five fellowship committees in action; and
5. They 'examined administrative records to determine the amount of staff time and financial expenditures related to committee meetings' (Obrecht *et al*, 2007: 83).

The researchers concluded:

Committee discussion of fellowship applications did not appear to increase the fairness to applicants or the likelihood of achieving fellowship program objectives over and above the levels attainable with structured at-home review. (Obrecht *et al*, 2007: 87)

In the researchers' opinion, panel peer review was an inefficient use of reviewer time.

Johnson (2008) evaluated reviews ratings by approximately 14,000 reviewers for nearly 19,000 grant applications submitted to the National Institutes of Health (NIH) (Bethesda, ML, USA) in 2005. The review process at the NIH is overseen by the NIH Center for Scientific Review. Most grant applications submitted to the NIH are assigned to Integrated Review Groups (IRGs) for review, which are further subdivided into specific study sections. IRG study sections contain about 30 members and evaluate about 50 grant proposals during a meeting. From two to five reviewers ('readers') are assigned to review each proposal before the study section meets; the readers assign a score to each proposal ('pre-score'). The readers announce their pre-scores at the meeting and summarize the proposal for other members of the study section, as most of them will not have read it. After that, there is an open discussion of the proposal. At the end of the discussion:

readers state their 'post-scores' for the proposal, and all other members of the study section ... also score the proposal. (Johnson, 2008: 11076)

After these the proposal's priority score is generated by the average of the scores from all study section members.

Johnson (2008) proposed statistical methodology:

to account for the effect of the selection of readers on a proposal's final percentile score, quantify the uncertainty associated with the percentile scores, and demonstrate how such uncertainties can be incorporated into a decision-theoretic framework to improve the probability that the greatest proportion of top proposals are funded. (p. 11076)

Johnson (2008) concluded that the analyzed peer review system of NIH failed to account for individual bias and placed undue weight on panel members who had not read the proposals. He found that the top grants were largely unaffected by reader bias but that it did affect grants closer to the funding cut-off line.

Thorngate *et al* (2009) evaluated 306 files of grant application adjudications out of the year 2000 and 2001 of the CIHR. In addition, the researchers observed seven panel meetings to get familiar to the CIHR review process and the assessment scale. (The review process of the CIHR is described above, cf. passage of Obrecht *et al*, 2007). The central purposes of their study were to determine the value of committee discussion and the effects of external reviews in changing the ratings of reviewers (readers) who were responsible for presenting their assessments to the committee members. They coded each of the 306 files according to 74 features (e.g. use of external reviews, ratings of reviewers) to address their research purposes. In addition, they compared the ratings of the readers for one and the same application and computed various correlations of readers, private and average rates.

Thorngate *et al* (2009) found:

1. A low inter-rater reliability between the assessments of the readers. The researchers argue 'that there is at best a modest relationship between peers in how merit is judged, and that disagreement is common' (Thorngate *et al*, 2009: 117).
2. That deliberation routines occur and assessment rules change during the time of a panel meeting, which have positive consequences for some applications and negative for others.
3. When there is disagreement between readers of one and the same application, committee members are more likely to follow the assessment of the critical reader. Against the backdrop of this result, Thorngate *et al* (2009) concluded 'that committees tended toward conservatism' (p. 118).

4. That there is a high correlation between the average rating given by the readers and the average of private ratings of all committee members. These 'indicate that final committee judgments could be predicted with high accuracy from the average of assessors' judgments' (Thorngate *et al*, 2009: 119). As a consequence they ask for the additional value of committee discussion.
5. That the two readers 'were just as likely to disagree after reading four reviews as after reading three, two, one or none' (Thorngate *et al*, 2009: 119). This means that the external reviews have hardly any influence on the reader's assessment and therefore may be dispensable. (Van den Besselaar and Leydesdorff [2009] arrived in their study on the Dutch Economics and Social Research Council [Den Haag] at a similar conclusion: there is hardly any correlation between the review scores of the external reviewers and the decision of the committee.)

Altogether Thorngate *et al* (2009) came to the conclusion that the committee discussion does not lead to more valid judgments of merit than do individual judgments summed or averaged. Because of this they suggested that readers should submit their independent judgments to the funding organization before the meeting and that only proposals with great disagreement among assessments should be discussed at the panel meetings.

### Characteristics of a Panel

In this section we present features that are used in social psychology to characterize a group. A peer review panel will be characterized with the help of the features.

#### *Entitativity*

The extent to which a group will be perceived as such depends mainly on its *entitativity* (Campbell, 1958), or 'the degree to which a collection of persons are perceived as being bonded together in a coherent unit' (Lickel *et al*, 2000: 224). Entitativity is higher and lower depending on the type of group. Whereas intimacy groups (such as family members or two people in a romantic relationship) are rated as highest in entitativity, loose associations (such as the people in a cinema audience) are rated low in entitativity (Lickel *et al*, 2000). Task-oriented groups such as panels (see below) are usually perceived as high in entitativity.

#### *Group task*

Steiner (1972) proposed that there are three general dimensions on which to classify the tasks worked on by a group (in panel peer review, for example, the tasks of assessing an application and drawing up a basis for decision-making). The first dimension has

to do with potential task divisibility. *Divisible* tasks can be divided into subtasks that can be assigned to the members; *unitary* tasks are group tasks that cannot be subdivided and every member does the same task. For example, painting a house is a divisible task. Lifting a washing machine is a unitary task. Assessing a grant application is a unitary task, because the members of the panel arrive at a funding recommendation via joint discussion.

Steiner's (1972) second classification dimension is whether the goal of task fulfillment pertains to quantity (*maximizing* task) or quality (*optimizing* task). Making the best possible decision would be an optimizing task, whereas, for example, piling as many books on a stack as possible would be a maximizing task. The tasks in panel peer review are optimizing tasks, because what is decisive is the quality of the output and making the best possible judgment on an application.

Steiner's third classification dimension distinguishes between potential performance and actual performance. Potential performance is the performance of a group if the members of the group had worked on the task independently of one another. In other words, it captures the individual performance of the members of the group. Actual performance is what the group actually accomplishes. This dimension is concerned with how individual group member input is converted into potential group performance. This will differ depending on the type of unitary task. For example, panel peer review of a grant application can be a *disjunctive* unitary task, if the group assesses the complete application jointly. The group must decide on one of the several possible judgments (e.g. accept or reject for funding). With disjunctive tasks, potential performance is determined by the resources of the group's most competent member. Here the panel can only be as good as the member that proposes the 'ideal' judgment based on his/her expertise.

If the experts on the panel assess and contribute to the discussion (or the final judgment) only parts of the application, panel peer review can be an *additive* unitary task. With additive tasks, the group's performance is an equally weighted sum of the members' contributions. In addition to disjunctive and additive tasks, Steiner distinguishes two more types of unitary tasks: *conjunctive* tasks (here each member must achieve the goal, so that the potential performance is determined by the least proficient member, as in mountain climbing) and *discretionary* tasks (where the group chooses how to weight the contributions of its members to group task fulfillment, so that the potential performance results from the combination of individual performance, as in self-managed work teams like rock bands).

#### *Norms, roles, and shared cognitions*

In groups there are norms that the group members share and that prescribe how members must behave.

**In groups there are norms that the group members share and that prescribe how members must behave. The norms serve as guidelines for attitudes and behavior and thus perform an important regulating function**

The norms serve as guidelines for attitudes and behavior and thus perform an important regulating function. These norms can develop specific to the group, or they can have a general character. Merton (1973) described the 'ethos of science' as the norms held to be binding on scientists, the norms that prescribe what is appropriate and correct. Applied to peer review, Merton's norms of disinterestedness (the scientist forgoes all forms of personal gain), organized skepticism (knowledge must always be scrutinized), and universalism (knowledge claims should be judged impersonally, independently of their source) are mainly decisive (Bornmann, 2008). Whereas norms prescribe how each individual group member should behave, roles serve to specify how persons in a specific position should behave. For example, in their role as reviewers, scientists' task is to assess the quality of grant applications. In this role scientists do not have to perform administrative tasks such as organizing the review process.

The term shared cognitions refers to a common understanding among group members regarding various aspects of the group, such as regarding the specific role and the skills and abilities of every member. If the group has *transactive memory* (see here Liang *et al.*, 1995), there is shared knowledge of which group member has what knowledge. Through this, the individual members of the group in principle have access to knowledge that they themselves do not have but other members have. For example, if a certain subject-matter expertise is required for assessing a part of a grant application, the peer review group can access the expert in a targeted manner. In this way the group members can optimally exploit expertise available within the group for their collective work.

**Phenomena that can occur in group judgment and decision-making**

Social psychology has examined in-depth how judgments and decisions are made in task groups and what phenomena (such as groupthink) can occur. In the following a number of these phenomena are presented that can be assumed to have an (undesired) effect on judgment processes also in panel peer review.

To explain a process by which a group may make faulty decisions, Janis (1972, 1982) coined the term 'groupthink'. Janis studied foreign policy decisions made by the US Government that later proved to be wrong (such as the Bay of Pigs invasion) and looked at what the decision processes in these cases had in common. Based on analysis of documents such as the minutes of the policy-makers' meetings, Janis (1972, 1982) traced the faulty decision-making to dysfunctional thinking and interaction patterns, or groupthink, in the policy-making groups. Groupthink is:

a mode of thinking that people engage in when they are deeply involved in a cohesive in-group, when the members' strivings for unanimity override their motivation to realistically appraise alternative courses of action. (Janis, 1982: 9)

Janis (1972, 1982) developed a model that:

aimed to explain why a group of competent people [such as high-level government policy makers] sometimes make such disastrous decisions. (Nijstad, 2009: 140)

Numerous studies have been conducted on the phenomena of groupthink. Park's (1990) review of the empirical research on groupthink revealed that most of the studies only partly confirmed the theory. In a more recent review of groupthink research, Esser (1998) concluded that further empirical studies were needed. Because of the lack of unequivocal support in previous research, Nijstad (2009) recommended viewing the groupthink theory 'as a heuristic model rather than a validated theory' (p. 144).

According to Janis (1982), the danger that the phenomenon of groupthink will happen in a decision-making group is especially great, if:

1. There is high group cohesion — that is, a 'high degree to which the members value their membership in the group and want to continue to be affiliated' (p. 245);
2. The group has structural faults, including insulation of the group from outsiders, a lack of leader impartiality, a lack of adequate decision procedures, and high group member homogeneity; and/or
3. The group is in a stress situation.

The groupthink phenomenon could occur in panel peer review, if, for example:

1. A group of scientists work together jointly assessing applications on one and the same panel for a long time;
2. There are no formal guidelines for the judgment process;

3. The panel is very homogeneous, made up of scientists all belonging to the same 'school'; and
4. The panel is under pressure to discuss and judge many grant applications within a short period of time.

### Motivation losses

Motivation losses refer to the reduction group members' motivation to contribute towards task fulfillment. The research on group performance has identified mainly three types of motivation losses, of which *free riding* and the *sucker effect* are especially relevant for panel peer review (*social loafing* is less relevant here, for it occurs mainly where individual contributions are *not* identifiable).

*Free riding* (Kerr and Bruun, 1983) describes a reduction in effort by group members who see their individual efforts as having little influence on group performance. There is a danger that free riding will occur especially with disjunctive tasks, such as assessing a grant application. This holds mainly for large groups, for the larger a group is, the larger the probability is that another member will be better at working on a disjunctive task. This phenomenon could happen in panel peer review, if, for instance, on a panel with a number of older and more experienced scientists there are also a few young and inexperienced scientists who believe that their contributions to assessment of the application are not needed.

The *sucker effect* (Kerr, 1983) refers to a loss of motivation in group members if they perceive or expect that other members are capable of contributing to the group but will not. To prevent being taken advantage of by the others in this way, they then reduce their own efforts. For example, if members of a panel saw that the weaker performance of other members was not dependent on a lack of ability but was instead due to a lack of effort, they themselves could begin to exert less effort. There is then a danger that these members will be willing to accept poorer group performance so as to prevent others from free riding on their efforts and thus putting them in the sucker role.

### Group polarization

Group polarization is the tendency to take a more extreme position after participating in a group discussion than before the discussion, and the position will be in the direction of the average tendency of the individual positions before the discussion (Isenberg, 1986; Moscovici and Zavalloni, 1969). Two important approaches explain group polarization: *persuasive argumentation* (Burnstein and Vinokur, 1977), and *social comparison processes* (Sanders and Baron, 1977).

- *Persuasive argumentation* starts out from the assumption that an individual position results from

the number, direction and persuasiveness of the arguments that a person has. In a discussion the different group members exchange arguments. If the members present new and credible arguments that support the position that is already the strongest in a person, this can make the person willing — due to the increasing number of supporting arguments — to take a more extreme position than prior to the group discussion (Burnstein and Vinokur, 1977).

- *The social comparison approach* starts out from the assumption that group members tend to compare themselves to other members. They have the need to see themselves in a positive light and to obtain agreement and approval from others (Goethals and Zanna, 1979). At the same time, they desire to differ from others. If members of the group ascertain that the others have about the same opinion, they are prepared to stand out from the others by taking a more extreme position than the majority of the group — in the direction in which the majority of the group members are tending anyhow (Myers, 1978). In this way they can expect to be judged positively by the group despite their different opinion.

According to Nijstad (2009) '*group polarization* has since been shown in a variety of contexts, including ... jury decisions' (p. 37). We therefore assume that it occurs also in panel peer review. Here, the mean of the individual opinions after the group decision would be more positive or more negative than the mean of the members' individual opinion before the panel discussion — and they would shift to the direction that the panel members favor initially.

### Conformity pressure

When consensus in a group:

comes under the dominance of conformity, the social process is polluted and the individual at the same time surrenders the powers on which his functioning as a feeling and thinking being depends. (Asch, 1955: 34)

A smaller subgroup conforms to a larger subgroup, because it feels subjected to real or supposed pressure to conform. Just how strong conformity pressure can be was demonstrated by the famous experiments by Asch (1955, 1987) in the 1950s. Asch showed that group members having equal hierarchical standing frequently changed their opinions to conform to the majority opinion when it was obviously false. This conformity of group members seems to occur mainly with difficult tasks. Baron *et al* (1996) conducted an experiment in which the participants had to complete tasks of differing levels of difficulty. Baron *et al* (1996) found that people tended to conform more to the opinion of others when task difficulty was high and concluded that people



---

**Asch showed that group members having equal hierarchical standing frequently changed their opinions to conform to the majority opinion when it was obviously false. This conformity of group members seems to occur mainly with difficult tasks**

---

tend to rely more on others' judgments when it is difficult for them to check their own judgment against reality. As assessing a research grant application is a comparatively complex task and reviewers frequently form different judgments on one and the same grant application (Daniel *et al*, 2007), there could be a tendency among reviewers (who actually hold a minority opinion in the panel or who cannot easily assess an application due to its subject matter) to conform to the majority opinion in the panel.

Looking at the reasons for the occurrence of conformity, Deutsch and Gerard (1955) distinguished between informational and normative social influence. *Informational social influence* occurs in a group when individuals agree with the group opinion because the arguments presented have convinced them (Eagly and Chaiken, 1993; Kaplan and Miller, 1987). In panel peer review, then, when a reviewer changes his or her private opinion because of the (convincing) arguments of the other reviewers on the panel, this is informational social influence according to social psychology. In *normative social influence* individuals go along with the judgment of the majority, because they want to be accepted by the group. This conformity does not necessarily mean that they also change their private beliefs (public compliance). In panel peer review, reviewers could go along with the majority opinion because they fear that their own judgment on a grant application could make them look bad in front of the other members. The study by Obrecht *et al* (2007) found such a tendency:

Many committee members do not like to contradict the accepted wisdom of their committee even though they may not agree with it. (p. 90)

*Information sampling*

A group has at its disposal both unshared information (information that is known to only a few persons or only one person in the group) and shared information (information available to many persons in the group). Experiments in social psychology found that groups that must make a task-related decision predominately exchange shared information and hardly exchange unshared information (see

Shaw and Penrod, 1962; Stasser and Stewart, 1992; Stasser *et al*, 1995, 1989; Stasser and Titus, 1985). This is called *sampling bias* in favor of shared information (Nijstad, 2009), which increases with an increasing amount of decision-relevant information.

Further, unshared information that is mentioned in the discussion is picked up and repeated less often by the group than shared information. Although the advantage of a group is supposed to be that judgments and decisions can be made based on a better information basis due to the exchange of unshared information, unshared information is exchanged insufficiently.

The influence of shared and unshared information on the decision-making process in a group has been investigated mainly using an experimental condition called *hidden profiles* (Stasser and Titus, 2003). In this constructed decision-making condition, task-relevant information is distributed among the members of the group such that no group member can make the correct decision based on his/her information alone. The optimal task solution can be achieved only through pooling the unshared information. A number of studies showed that most groups fail at solving hidden profiles (Stasser and Birchmeier, 2003). This is mainly because shared, or common, information has a bigger impact on group choice than unshared information (Gigone and Hastie, 1993; 1997).

Nijstad (2009) gives two reasons for the occurrence of this *common-knowledge effect*. First, members of the group generally see shared information as more credible and more relevant than unshared information (Chernyshenko *et al*, 2003). This is because, among other things, the information is found credible by several persons, which for the individuals increases the reliability of the information. The risk of the information being false is seen as lower than if the information is available to a single person only. Second, people are reluctant to change an initial preference, and they defend it (advocacy) — even if information is brought into the discussion that suggests another alternative (e.g. Postmes *et al*, 2001). The reason is that people evaluate positive information concerning a preferred alternative as more positive and credible than positive information concerning a non-preferred alternative (the opposite is true for negative information) (Greitemeyer and Schulz-Hardt, 2003).

There is, of course, no situation comparable to a constructed hidden profile in the panel peer review process. Nonetheless, both shared and unshared information is available to the individual reviewers. If information sampling took place in a panel, the panel members would discuss mainly things that everyone already knows. Information known to individuals only would be mentioned not at all or would not be taken into consideration when arriving at a judgment. The positive expectation associated with a panel discussion would not be fulfilled — namely, that the group performs better than any



member could perform alone, because the members share their different information and arguments. Reviewers are chosen for the panel because they are meant to bring mainly *unshared* information into the discussion of research grant applications.

### Measure for optimizing group processes

As shown above, phenomena occur in task groups that can have a negative influence on the quality of judgments and decisions. Enhancement of the effectiveness of research panels could be accomplished by greater attention to these phenomena, given the time pressures under which panels ordinarily work. Certain measures have been proposed to counteract the phenomena. Three measures have been investigated in numerous studies and found to be effective. They relate to: the organization of a group meeting; moderation of group discussion; and the composition of the group.

1. A group meeting should be organized such that mainly unshared information is brought up during discussion (Larson *et al.*, 1994; Winquist and Larson, 1998). A proven way to encourage the exchange of unshared information is to give the members sufficient time for discussion. Beyond that, members of the group should be assigned responsibility for different task areas (in panel peer review, reviewers could be assigned responsibility for different parts of an application, for instance). Because unshared information brought into the discussion by an expert has a greater probability of being accepted than information that is offered by a person not having the same expertise, transactive memory should be established (Stewart and Stasser, 1995; Wegner, 1987). This could be done in panel peer review, for instance, by introducing the specific expertise of each group member at the start of the panel meeting.
2. The group's work on a task should be moderated (see here Langfeldt *et al.*, 2010) (panel peer review of grant applications could be moderated by a member of the funding agency, for example). The moderator should divide the discussion and judgment process of the group into two components: information search first, followed by integration and decision. This helps to insure that all relevant information is mentioned and taken into consideration in the discussion (Kerr and Tindale, 2004). The moderator should also see that all members of the group are included in working on the task and have sufficient say (e.g. to counteract motivation losses and to encourage the pooling of unshared information). If members are aware that their input is indispensable, the tendency towards free riding will be counteracted. It has also been found to be favorable to have the moderator, as a formalized procedure, to ask members to play the role of devil's advocate for a certain length of

time (see Herbert and Estes, 1977). The member playing the role of devil's advocate is instructed to present counter arguments to every position and to offer constructive criticism. This creates an atmosphere in which others are also willing to put forward their objections. Use of the devil's advocacy technique is advisable mainly when it becomes apparent that a group is reaching an over-hasty consensus.

3. As to group composition, group members should be as diverse as possible. Group diversity leads to different points of view and perspectives (key word: unshared information) in the discussion (see e.g. Brodbeck *et al.*, 2002; Frey *et al.*, 1996). Diversity stimulates *divergent thinking* and prevents (over-hasty) *convergent thinking* (Nemeth, 1986). For panel peer review, *informational diversity* (unique specialist knowledge, insights, and abilities as well as work experience) is more important than *social category diversity* (differences in demographic characteristics). Although selecting reviewers according to the principle of complementarity (for example, choosing a generalist and a specialist) will lower inter-reviewer agreement, validity can be increased considerably, as the decisions about a proposal will be based on much broader information. The disadvantage of diverse groups is that members can 'trust each other less, are less willing to cooperate closely, have more conflicts' (Nijstad, 2009: 169).

### Discussion

If the panel peer review process is designed such that all of the reviewers undertake to judge a complete grant application (reviewers are not assigned parts of the application to judge), the potential performance of the panel will be determined by the reviewer having the greatest competency. In the best case, a member of the panel proposes the 'ideal' judgment on a grant application. However, Nijstaad (2009) makes a plea for expanding the concept of potential performance to include the possibility that a group succeeds in achieving better performance than the best member and defines potential performance as:

---

**If the panel peer review process is designed such that all of the reviewers undertake to judge a complete grant application, the potential performance of the panel will be determined by the reviewer having the greatest competency**

---

a group's potential given individual member resources and the demands of the group task. Potential performance in that case is what individuals could possibly achieve (given their individual resources and task demands), and how these possible individual achievements relate to group performance. (Nijstad, 2009: 53)

This means that if a panel as a whole has more resources than the best individual and if group composition makes for a very effective combination of individual resources, then the panel's performance can be higher than its most competent member (coordination gains). In this case, the quality of the group composition determines the quality of the judgment. However, even if a panel of reviewers has an optimal composition, there is still a danger that conformity pressure will lead to a suboptimal judgment. If the reviewers have different opinions, pressure for a consensus could result in acceptance of the majority position without adequate consideration of deviating opinions. If there are additional unfavorable conditions in a peer review process, such as inadequate decision procedures and insufficient time for discussion, there is also a risk of groupthink.

Whether (and to what extent) the phenomena discussed in this article actually occur in panel peer review is an open question. There is a great probability that they do. Only with empirical studies we will arrive at valid conclusions. Some examples for interesting research questions to test in future studies are:

1. Does the size of the panel group increase the number of reviewers who do not take part in the discussion and decision-making process?
2. Does the same increase appear when raising the number of reviewers with different academic status?
3. Do reviewers conform to the majority opinion in the panel even they don't agree with it?
4. Is there a tendency in the panel to take a more extreme position after participating in the panel group discussion than before?
5. To what extent are shared and unshared information discussed in a panel meeting?
6. Is there a tendency in long-standing panel groups to groupthink?

In social psychology there are mainly three ways of investigating these (and other) research questions (Nijstad, 2009).

*Qualitative research* can consist in *case studies*, such as interviews with open-ended questions and participative observation. In these studies, one or a few entities, such as panels, are studied extensively. The advantage of these studies is that they can produce a detailed picture of a few review processes. The disadvantage is the small number of cases, which makes the generalizability of the findings unclear. A problem found mainly with participative observation is the problem of access. For one, peer

review is a confidential process, where discussion contents are not meant to become public. For another, it is to be feared that participation of observers disturbs the judging. Besides case studies, another way to conduct qualitative research on panel peer review is *content analysis*, which Janis (1972) used in a study on groupthink, for instance. Using content analysis (see here Stemler, 2001), the review process can be investigated based on meeting minutes, reviews, notes, draft of a final decision, and so on. However, use of content analysis requires that the review process is well documented in writing.

*Surveys* are a second possible method for investigating group processes. For instance, reviewers that have carried out reviews for a funding agency for a certain period of time can be surveyed on their experiences with panel peer review by means of a structured questionnaire. The advantage of surveys is that a large number of reviewers can be included (thus making generalizable statements possible) and that surveys can be used to obtain quantitative relations between variables. The disadvantage of the survey method is that the data are subjective. Reviewers are asked to report on their own attitudes, feelings, behaviors and performance. In addition, it is not possible to establish causality with survey studies. Correlations do not allow the drawing of causal conclusions.

The third way to investigate group processes is the *experiment*, where a situation is created that is under the control of the researcher. Participants in an experiment are *randomly* assigned to different conditions. With experiments, the effects of manipulations by the researcher can be studied. The advantage of experiments is that researchers can test causal relations, for differences in the participants' behavior can be attributed to the manipulation in the experiment. The disadvantage of the experiment is that in many situations it is not possible to assign people randomly to certain conditions. For example, funding agencies might not allow experiments to be conducted in which randomly chosen grant applications are reviewed under different conditions or grants are allocated at random.

Because each of these three methodological approaches has its advantages and disadvantages, it makes sense to investigate panel peer review using a combination of several approaches. If several case studies on different procedures are available, the results of the studies can be compared and discussed together in one study, as Bornmann *et al* (2010) have done for the peer review process at the European Molecular Biology Organization (Heidelberg, Germany) and the Dutch Economics and Social Research Council.

## References

- Asch, Solomon E 1955. Opinions and social pressure. *Scientific American*, **193**(5), 31–35.

- Asch, Solomon E 1987. *Social Psychology*. Oxford, UK: Oxford University Press.
- Baron, Robert S, Joseph A Vandello and Bethany Brundsmann 1996. The forgotten variable in conformity research: impact of task importance on social influence. *Journal of Personality and Social Psychology*, **71**, 915–927.
- Bornmann, Lutz 2008. Scientific peer review: an analysis of the peer review process from the perspective of sociology of science theories. *Human Architecture: Journal of the Sociology of Self-Knowledge*, **6**(2), 23–38.
- Bornmann, Lutz in press. Scientific peer review. *Annual Review of Information Science and Technology*.
- Bornmann, Lutz and Hans-Dieter Daniel 2005. Selection of research fellowship recipients by committee peer review: analysis of reliability, fairness and predictive validity of Board of Trustees' decisions. *Scientometrics*, **63**(2), 297–320.
- Bornmann, Lutz, Loet Leydesdorff and Peter van den Besselaar 2010. A meta-evaluation of scientific research proposals: different ways of comparing rejected to awarded applications. *Journal of Informetrics*, **4**(3), 211–220.
- Brodbeck, Felix C, Rudolf Kerschreiter, Andreas Mojzisch, Dieter Frey and Stefan Schulz-Hardt 2002. The dissemination of critical, unshared information in decision-making groups: the effects of pre-discussion dissent. *European Journal of Social Psychology*, **32**, 35–56.
- Burnstein, Eugene and Amir Vinokur 1977. Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology*, **13**, 315–332.
- Campbell, Donald T 1958. Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, **3**, 14–25.
- Chernyshenko, Oleksandr S, Andrew G Miner, Michael R Baumann and Janet A Sniezek 2003. The impact of information distribution, ownership; and discussion on group member judgment: the differential cue weighting model. *Organizational Behavior and Human Decision Processes*, **91**, 12–25.
- Daniel, Hans-Dieter, Sandra Mittag and Lutz Bornmann 2007. The potential and problems of peer evaluation in higher education and research. In *Quality Assessment for Higher Education in Europe*, ed. A Cavalli, pp. 71–82. London: UK Portland Press.
- De Vries, Dennis R, Elizabeth A Marschall and Roy A Stein 2009. Exploring the peer review process: what is it, does it work, and can it be improved? *Fisheries*, **34**(6), 270–279.
- Deutsch, Morton and Harold B Gerard 1955. A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, **51**, 629–638.
- Eagly, Alice H and Shelly Chaiken 1993. *The Psychology of Attitudes*. New York, NY, USA: Harcourt Brace Jovanovich.
- Esser, James K 1998. Alive and well after 25 years: a review of groupthink research. *Organizational Behavior and Human Decision Processes*, **73**(2/3), 116–141.
- Frey, Dieter, S Schulz-Hardt and D Stahlberg 1996. Information seeking among individuals and groups and possible consequences for decision-ranking in business and politics. In *Understanding Group Behaviour*, eds E Witte and J Davis, pp. 211–225. Hillsdale, NJ, USA: Erlbaum.
- Gigone, Daniel and Reid Hastie 1993. The common knowledge effect: information sharing and group judgment. *Journal of Personality and Social Psychology*, **65**, 959–974.
- Gigone, Daniel and Reid Hastie 1997. The impact of information in small group choice. *Journal of Personality and Social Psychology*, **72**, 959–974.
- Goethals, George R and Mark P Zanna 1979. The role of social comparison in choice shifts. *Journal of Personality and Social Psychology*, **37**, 1469.
- Greitemeyer, Tobias and Stefan Schulz-Hardt 2003. Preference-consistent evaluation of information in the hidden profile paradigm: beyond group-level explanations for the dominance of shared information in group decisions. *Journal of Personality and Social Psychology*, **84**, 322–339.
- Guston, David H 2003. The expanding role of peer review processes in the United States. In *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, eds P Shapira and S Kuhlmann, pp. 81–97. Cheltenham, UK: Edward Elgar.
- Herbert, Theodore T and Ralph W Estes 1977. Improving executive decisions by formalizing dissent: the corporate devil's advocate. *Academy of Management Review*, **2**, 662–666.
- Isenberg, Daniel J 1986. Group polarization: a critical review and meta-analysis. *Journal of Personality and Social Psychology*, **50**(6), 1141–1151.
- Janis, Irving L 1972. *Victims of Groupthink: a Psychological Study of Foreign-Policy Decisions and Fiascos*. Boston, MA, USA: Houghton Mifflin.
- Janis, Irving L 1982. *Groupthink*. Boston, MA, USA: Houghton Mifflin.
- Johnson, Valen E 2008. Statistical analysis of the National Institutes of Health peer review system. *PNAS – Proceedings of the National Academy of Sciences of the United States of America*, **105**(32), 11076–11080.
- Kaplan, Martin F and Charles E Miller 1987. Group decision making and normative versus information influence: effects of types of issue and assigned decision rule. *Journal of Personality and Social Psychology*, **53**, 306–313.
- Kerr, Norbert L 1983. Motivation losses in small groups: a social dilemma analysis. *Journal of Personality and Social Psychology*, **45**, 819–828.
- Kerr, Norbert L and Steven E Bruun 1983. The dispensability of member effort and group motivation losses: free rider effects. *Journal of Personality and Social Psychology*, **44**, 78–94.
- Kerr, Norbert L and R Scott Tindale 2004. Group performance and decision making. *Annual Review of Psychology*, **55**, 623–655.
- Lamont, Michèle 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA, USA: Harvard University Press.
- Langfeldt, Liv 2001. The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science*, **31**(6), 820–841.
- Langfeldt, Liv 2002. *Decision-making in expert panels evaluating research: constraints, processes and bias*. Doctoral dissertation, University of Oslo, Norway.
- Langfeldt, Liv 2004. Expert panels evaluating research: decision-making and sources of bias. *Research Evaluation*, **13**(1), April, 51–62.
- Langfeldt, L, B Stensaker, L Harvey, J Huisman and D F Westerheijden 2010. The role of peer review in Norwegian quality assurance: potential consequences for excellence and diversity. *Higher Education*, **59**(4), 391–405.
- Larson, James R, Penny G Foster-Fishman and Christopher B Keys 1994. Discussion of shared and unshared information in decision-making groups. *Journal of Personality and Social Psychology*, **67**, 446–461.
- Liang, Diane W, Richard Moreland and Linda Argote 1995. Group versus individual training and group performance: the mediating role of transactive memory. *Personality and Social Psychology Review Bulletin*, **21**(4), 384–393.
- Lickel, Brian, David L Hamilton, Grazyna Wierzchowska, Ami Lewis, Steven J Sherman and A Neville Uhles 2000. Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, **78**, 223–246.
- Merton, Robert K 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, USA: University of Chicago Press.
- Moscovici, Serge and Marisa Zavalloni 1969. The group as a polarizer of attitudes. *Journal of Experimental Social Psychology*, **12**, 125–135.
- Myers, David G 1978. Polarization effects of social comparison. *Journal of Experimental Social Psychology*, **14**, 554–563.
- Nemeth, Charlan J 1986. Differential contributions of majority and minority influence. *Psychological Review*, **93**(1), 23–32.
- Nijstad, Bernard A 2009. *Group Performance*. East Sussex, NY, USA: Psychology Press.
- Obrecht, Michael, Karl Tiberius and Guy D'Aloisio 2007. Examining the value added by committee discussion in the review of applications for research awards. *Research Evaluation*, **16**(2), June, 79–91.
- Overbeke, John and Elizabeth Wager 2003. The state of the evidence: what we know and what we don't know about journal peer review. In *Peer Review in Health Sciences*, eds F Godlee and T Jefferson, pp. 45–61. London, UK: BMJ Books.
- Park, Won Woo 1990. A review of research on groupthink. *Journal of Behavioral Decision Making*, **3**, 229–245.
- Postmes, Tom, Russell Spears, Khaled Sakhel and Daphne de Groot 2001. Social influence in computer-mediated communication: the effect of anonymity on group behaviour. *Personality and Social Psychology Bulletin*, **27**, 1243–1254.
- Sanders, Glenn S and Robert S Baron 1977. Is social comparison irrelevant for producing choice shifts? *Journal of Experimental*

- Social Psychology*, **13**, 303–314.
- Shaw, Marvin E and William Penrod 1962. Does more information available to a group always improve group performance? *Sociometry*, **25**, 377–390.
- Stasser, Garold and Zachary Birchmeier 2003. Group creativity and collective choice. In *Group Creativity: Innovation through Collaboration*, eds P B Paulus and A Nijstad, pp. 85–109. New York, NY, USA: Oxford University Press.
- Stasser, Garold and Dennis D Stewart 1992. Discovery of hidden profiles by decision-making groups: solving a problem versus making a judgment. *Journal of Personality and Social Psychology*, **63**, 426–434.
- Stasser, Garold and William Titus 1985. Pooling of unshared information in group decision making: biased information sampling during discussion. *Journal of Personality and Social Psychology*, **48**, 1467–1478.
- Stasser, Garold and William Titus 2003. Hidden profiles: a brief history. *Psychological Inquiry*, **14**(3&4), 304–313.
- Stasser, Garold, Laurie A Taylor and Coleen Hanna 1989. Information sampling in structured and unstructured discussion of three- and six-person groups. *Journal of Personality and Social Psychology*, **57**, 67–78.
- Stasser, Garold, Dennis D Stewart and Gwen M Wittenbaum 1995. Expert roles and information exchange during discussion: the importance of knowing who knows what. *Journal of Experimental Social Psychology*, **31**, 244–265.
- Steiner, Ivan D 1972. *Group Process and Productivity*. New York, NY, USA: Academic Press.
- Stemler, Steve 2001. An overview of content analysis. *Practical Assessment, Research and Evaluation*, **7**(17).
- Stewart, Dennis D and Garold Stasser 1995. Expert role assignment and information sampling during collective recall and decision making. *Journal of Personality and Social Psychology*, **69**, 619–628.
- Thorngate, Warren, Robyn M Dawes and Margret Foody 2009. *Judging Merit*. New York, NY, USA: Psychology Press.
- Van den Besselaar, Peter and Loet Leydesdorff 2009. Past performance, peer review and project selection: a case study in the social and behavioral sciences. *Research Evaluation*, **18**(4), October, 273–288.
- Wegner, Daniel M 1987. Transactive memory: a contemporary analysis of the group mind. In *Theories of Group Behavior*, ed. B Mullen, pp. 185–205. New York, NY, USA: Springer.
- Weller, Ann C 2001. *Editorial Peer Review: Its Strengths and Weaknesses*. Medford, NJ, USA: Information Today.
- Wessely, Simon 1998. Peer review of grant applications: what do we know? *Lancet*, **352**(9124), 301–305.
- Winquist, Jennifer R and James R Larson 1998. Information pooling: when it impacts group decision making. *Journal of Personality and Social Psychology*, **74**, 371–377.