Peer review practices: a content analysis of external reviews in science funding

Martin Reinhart

Science Studies Program

University of Basel

martin.reinhart@unibas.ch

Acknowledgement note:

I would like to thank Sabine Maasen, Marcel Weber, Alexandra Hofmänner, Torsten Mayerhauser, Fritz Böhler, Katharina Dräger, Daniel Sirtes, Michèle Lamont and the audience at the 2007 annual meeting of the Society for Social Studies of Science in Montréal for their helpful comments and the Swiss National Science Foundation for providing support and access to their archive. Direct correspondence to Martin Reinhart, Science Studies Program, University of Basel, Missionsstrasse 21a, CH-4003 Basel. E-mail: martin.reinhart@unibas.ch

Abstract:

The primary purpose of this study is to open up the black box of peer review and to increase its transparency, understanding, and credibility. To this end, two arguments will be presented: First, epistemic and social aspects of peer review procedures are inseparable and mutually constitutive. Second, a content analysis of written reviews indicates that certain elements of peer culture from the 17th century are still active in the scientific community. These arguments are illustrated by a case study on the peer review practices of a national funding institution, the Swiss National Science Foundation. Based on the case study and the two arguments it will be concluded more generally that peer review procedures show a distinctive specificity to the reviewed objects (e.g. papers or proposals), the organisational format (e.g. panels or external reviewers), or the surrounding context (e.g. disciplinary or interdisciplinary). Scientists, administrators, and the public may conclude that appraising peer review procedures should not be done by way of general principals but should be based on concrete factual knowledge on the specific process under discussion.

Introduction

Peer review in science fulfils two main functions: quality assurance and self-regulation. In publishing, funding, and other domains of science both of these functions are fulfilled at the same time. By comparing manuscripts, project proposals, or job candidates scarce resources such as publication space, funds or employment are distributed on the basis of scientific quality assessments. Science regulates the allocation of credibility mainly through peer review.[1] There are different appraisals of how well peer review operates with regard to quality assurance and self-regulation. However, peer review fulfils both these functions independent of such an assessment.

This partly explains the ubiquity and longevity of peer review in science. For more than 300 years, peer review has played a central role in modern science and over the past 60 years it has become a self-evident part of most publishing and funding procedures. This kind of stability can only be expected when peer review is able to perform notwithstanding dissatisfaction or criticism from the part of the scientific community. However, mere operational performance is not enough for peer review to be accepted as *the* legitimate organisational form for quality assurance and self-regulation. For a long time peer culture represented the dominant frame in which the legitimacy of peer review was accepted by scientists and the public alike.

Going back to the beginnings of peer review which coincides with the beginning of modern science in the 17th century (Barnes 1934; Lock 1985; Lux 1989; Zilsel 2000), Shapin argues that "[i]t seems quite likely that small specialized communities of knowledge-makers share many of the resources for establishing and protecting truth that were current in the pre- and early modern society of gentlemen" (Shapin 1994:414). Modern science, in its beginnings, was characterized by solving epistemic and social problems at the same time. The quality of knowledge was assured ("protecting truth") by applying the social conventions of peer culture ("society of gentlemen") to science. Shapin speculates that parts of the culture of gentlemen are still relevant to current science (Shapin 1994:414), and this paper will demonstrate that epistemic and social aspects in current peer review practices are still inseparable and bear some resemblance to peer culture.

[1] In most of the conversions in Latour and Woolgar's "cycles of credit", peer review is involved. (Latour and Woolgar 1986:201)

Finding traces of peer culture in the daily practices of science may not be surprising but finding them is a precondition for understanding the overwhelming support peer review still continues to be given by scientists today (Ware 2008). Substantiating Shapin's speculations with empirical evidence can help us understand why peer review remains unchallenged in science even though peer culture has disappeared almost everywhere else in society. The disappearance of peer culture is strongly apparent e.g. in science policy where the conventional governance models have shifted towards the principles of an "audit society" (Power 1997; Lange 2009; Whitley and Gläser 2007). Along with this shift came increasing critique of peer review as non-transparent and thus inefficient. The validity of this accusation is highly contested and rendered more complicated by the possibility that transparency and efficiency may even be inversely related (Reinhart and Sirtes 2006). Despite diverging views on the measure of transparency on peer review processes should be offered to authors, reviewers, administrators or the general public, there is no disagreement on the issue of transparency for the sake of research on peer review. Accordingly, this study aims at opening up the black box of peer review and assisting in gaining a better understanding of its inner workings. Studies like these may increase the transparency, understanding and credibility of peer review for various stakeholders and will hopefully lead to discussions and decisions on peer review that are based on empirically founded insights into its strengths and weaknesses.

Epistemic and Social Aspects of Peer Review

Most research on peer review operates in the assumption that epistemic and social aspects are separable and that peer review has to be judged by its output.[2] However, this paper argues that it is precisely this separation of the epistemic and social domains that has been hampering a better understanding of peer review. Instead, it is claimed, opening the black box of peer review represents a necessary condition for peer review research and will lead to a new conception of the social and epistemic dimensions as inseparable.

Without opening the black box of peer review, one would be tempted to assume that quality assurance covers its epistemic dimension and self-regulation its social dimension respectively. Separating the epistemic (aka the rational, aka the scientific) from the social is a well known premise from "the prehistory of science and technology studies" (Sismondo 2004:1). Such a distinction offers an attractive analytical tool because it would allow studying peer review by way of its output only. Both the approach of treating science as practice that is social and

[2] Critical reviews of peer review research are Hirschauer (2004) and Wood and Wessely (1999).

epistemic at the same time, and the approach of considering peer review as more than just its output are not new. The social studies of science can look back at a large amount of work that views science as social practice that can be analysed like any other social phenomenon.[3] Furthermore, two recent works by Mallard et al. (2009) and Hirschauer (2010) have started to treat peer review as a social practice.

Mallard, Lamont, and Guetzkow (2009) interviewed and observed panelists from multidisciplinary fellowship competitions in the social sciences and the humanities to understand the ways in which epistemological differences can be negotiated in peer review. They identify different epistemological styles for arguing among panelists. Their findings suggest that peer reviewers do not agree on the criteria for assessing scientific work. However, they agree on what they define as a fair-decision: applying the most appropriate epistemological styles for the field or discipline of the proposal under review. Their findings challenge the normative literature that associates procedural fairness with the use of generalisable criteria of evaluation (Collins and Evans 2002).

Hirschauer delivers a "praxeology of 'voting' in peer review" (Hirschauer 2010:71) by analyzing written documents from and participating in the editorial decision-making process of a sociology journal. In agreement with Mallard et al., he challenges the view that peer review is based on criteria of evaluation that may be generalised. "What has become known as scientific 'criticism' is an ongoing panoptic organization of communication: in peer review, judgments themselves are judged and made public" (Hirschauer 2010:71).

In both cases, it is argued empirically that the quality criteria for assessing scientific work are inseparable from the actual social situation. However, the work of Hirschauer and Mallard et al. is open to criticism due to their choice of empirical material. For two reasons, panel peer review from the social sciences or the humanities does not provide an adequate starting point for enforcing a general argument about peer review. First, panels seem to be especially susceptible to negotiation and mediation among the panelists, therefore, an agreement on fair decision-making is a more pressing need than shared epistemic standards. It could be argued that peer review without panels is superior when it comes to enforcing epistemic standards. Second, the absence of generalisable criteria of evaluation could be expected from the

[3] Obvious starting points for reviewing the social studies of science would be introductory textbooks like Hess (1997) or Sismondo (2004) and the most recent handbook from the Society for Social Studies of Science: Hackett et al (2008).

heterogeneous disciplines of the social sciences and the humanities, compared to the natural sciences. Scepticism seems reasonable that shared epistemic standards are to be held in such heterogeneous disciplines. Furthermore, interdisciplinary panels are to be expected to show variability in epistemological styles, simply because they are interdisciplinary. In short, the results of Mallard et al. and Hirschauer may be accounted for by the specific selection of empirical material and may not be considered characteristic for peer review in general.

By analysing more appropriate empirical material this paper sets out to test and expand Mallard et al.'s and Hirschauer's argument. A content analysis of written reviews from external reviewers in biology and medicine from the Swiss National Science Foundation (SNSF) will be presented to establish an empirical base with maximum distance to the studies referred to above. Several features of the peer review procedure at the SNSF are directly opposed to the panels studied by Hirschauer and Mallard et al. Even though the final decision rests with a panel, the reviewers themselves are not members of this panel. They write their review in complete ignorance of the panel proceedings before, during, and after the funding decision on the proposal under review. Reviewers and panel members do not interact directly, apart from standard invitations that are sent out to reviewers and their subsequent review submission to the SNSF. Administrators serve as intermediaries and are the only contact point for both parties.

This paper will put forward two arguments: First, epistemic and social aspects of peer review procedures are inseparable and mutually constitutive. This has already been shown for two case studies on panel peer review however, such a result might have been anticipated these specific cases. The validity of this argument needs to be tested for cases in which the interrelation of epistemic and social aspects is less obvious, i.e. in written, external peer reviews from the natural sciences. Second, the content analysis of written reviews will show that elements of peer culture are still active in the scientific community. Specific features of these elements will be discussed. Finally, some possibilities of extrapolating the findings of this study to more general statements on peer review will be considered.

Science Funding and the Content of Reviews

Although the importance of peer review procedures is often emphasized and even though there are a great number of studies, little is known about one of the most central elements of

peer review: the review itself.[4] It would be helpful not only to know something about the number of reviews, but also about their formal qualities, the topics mentioned and the argumentative structures deployed. The published literature addresses only some of these topics. Kretzenbacher and Thurmair (1992, 1995) report on the properties of reviews as a type of text; Guetzkow, Lamont, and Mallard (2004) investigate the usage of originality as a quality criteria; and Hartmann and Neidhardt (Neidhardt 1988; Hartmann and Neidhardt 1990; Hartmann 1990) provide a content analysis of reviews from the procedures of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) by identifying a set of eleven quality criteria. Since a search for literature on the subject yielded only these publications as relevant a short summary and discussion of each of them will be presented.

Kretzenbacher and Thurmair describe reviews by peers as a scientific text primarily from a linguistic perspective. Their text corpus consists of external reviews for the journal *Angewandte Chemie*, which they classify according to three kinds of "equivalence"[5] between the texts in the corpus: formal, functional, and pragmatic equivalence. For corpora that are assembled on the basis of pretheoretic considerations, as it is the case with peer reviews, a high equivalence of all three kinds is improbable. On the one hand, regarding formal criteria reviews differ considerably in length, (typo-)graphic design, type and sequence of parts, etc. (Kretzenbacher and Thurmair 1992:135). On the other hand, the authors find more functional equivalence regarding the content of the reviews. Most of these texts assess the content of applications by frequently using metaphors from the aesthetic realm and taking the character of a "metatext"[6] by distinguishing the textual and the representational level. In contrast, the authors find a high degree of pragmatic equivalence, as all texts are embedded in the same peer review procedure leading to unilateral anonymity and multiple addressees, which both are characteristic features of this type of text. Kretzenbacher and Thurmair state that, from a linguistic point of view, evaluative reviews are essentially situated within a communicative surrounding of scientific language in which qualitative evaluations are tabooed as they are not

[4] There are some questionnaire studies investigating quality criteria in science in general, see Hemlin (1993) and Dirk (1999).

[5] The papers by Kretzenbacher and Thurmair are in German and the term they use is "Äquivalenz". It stands for a concept from linguistics that describes the similarities and differences of texts that make up a corpus. It is a way of describing the amount of homogeneity within a text corpus.

[6] Again, metatext is a concept from linguistics being used for texts that distinguish between the textual and representational level. This is clearly the case with reviews where not only arguments about the scientific work itself are given but also about the way that work is presented in the text.

regarded as being conducive to objective representation (Kretzenbacher and Thurmair 1995:175). As a consequence reviewers lack a sophisticated and genuinely scientific lexis of evaluation outside the narrow area of single qualities that are easily objectified (Kretzenbacher and Thurmair 1995:211). They compensate for this deficit by switching between multiple assessor roles and generally using positive remarks more often than the mostly very specific, negative remarks about details.

Guetzkow, Lamont, and Mallard (2004) set their work on the level of functional equivalence by focusing on the content of reviews and, more specifically, on the meaning of *originality* as one essential quality criterion. Taking a sociological perspective they uncover multiple meanings of originality in reviews from the humanities and the social sciences. They conclude that there are significant disciplinary differences regarding the meaning of originality. On the one hand, humanists, historians, and social scientists define originality differently while also using the term for different aspects of an application. On the other hand, the authors presume a more narrow definition of the term in the (natural) sciences, but fail to support this claim by equally convincing empirical material. They also find that reviewers not only understand originality as substantive innovation, but very often regard it as a sign of the moral character of the applicant.[7] Hence reviewers tend to infer from the originality of the application to the authenticity and integrity of the applicant as a scientist.

Neidhardt and Hartmann studied the review procedures of the German Research Foundation (DFG) comprehensively and analyzed the self-regulating function of peer review along numerous dimensions and aspects from a sociological point of view. Among others, they present a content analysis of reviews from four different disciplines: psychology, electrical

[7] Guetzkow et al. define "moral" by referring to Max Weber: "In line with Weber, we treat moral qualities as those qualities that are defined as important by the ethical standards of a particular religion, institution or any other 'legitimate order'." (2004:193) They identify personal attributes of the researcher in general and moral attributes in particular as nonsubstantive factors neglected by peer review studies. Work from the sociology of knowledge supports their decision to focus on nonsubstantive factors. "For example, Shapin's (1994) study of science in seventeenth century England demonstrates that the moral virtues of scientists (defined in terms of honor, modesty, civility and courtesy) were taken as a sign that the results of their scientific experiments could be trusted." (2004:192) Nonetheless, the decision to use the term "moral" is unfortunate even though it is the only reasonable translation for Weber's German term "sittlich". It is unfortunate because the term "moral" carries connotations of "non-scientific" or "non-rational". Furthermore, Guetzkow et al. seem to associate "moral character" with what they call "nonsubstantive factors". (2004:192) When applying Weber's definition of ethical standards of a legitimate order for the case of science these connotations are to be avoided because it would be impossible for ethical standards of science to be "non-scientific". See also footnote 25.

engineering, economic theory, and political science. The reviewer's statements were coded with eleven different codes that allowed Neidhardt and Hartmann to record the frequency of usage as well as the influence on the final funding decision.[8] Regarding the frequency of usage, they found that no code is used in more than 50% of the reviews and most of the reviewer's statements are positive while negative statements are rarely articulated. They also locate considerable differences between the disciplines with respect to the relative frequency of the assigned codes. In electrical engineering, e.g., *practical relevance*[9] is a preferred criterion while in psychology questions of *method* are discussed frequently. Furthermore, there are also disciplinary differences regarding the impact of certain criteria on the recommendation of the reviewers, however, the authors consider the similarities to be more important. Neidhardt and Hartmann conclude that in none of the disciplines one criterion becomes eminent; rather, all criteria contribute in a significant way to the final recommendation and one can assume that they interact and overlap in their effects. The criterion of *reputation* receives separate attention from Neidhardt (1988:104) as a medium for regulation, because of an ongoing debate in German sociology about its role in science. He concludes that there is no evidence for Luhmann's proposition that reputation has exceptional importance in science (Luhmann 1970). However, there is also no support for Weingart's critique that the increasing focus on reputation is the crucial problem of the science system (Weingart 1970). In fact, Neidhardt finds a remarkable goodwill and civility of the reviewers towards the applicants that has to be balanced later by a much more rigorous judgment from the panel for the final funding decision.

The main points of the above mentioned studies on peer review for this paper's subsequent argument, are as follows: Taking the absence of a scientific lexis of evaluation serves as a starting point to argue that this lack is compensated by a certain kind of usage of quality criteria in peer reviews. Quality is either assigned to the project itself or to the applicant, following the insight that reviewers tend to interpret originality as a sign of the (moral) character of the applicant. Regarding methods, the eleven codes from the content analysis will serve as a source for an adjusted and enlarged coding scheme.

[8] Neidhardt and Hartmann based their codes on theoretical considerations but also on interviews with reviewers and preliminary tests. Their eleven codes are: *qualification/reputation*, *previous work*, *scientific relevance*, *practical relevance*, *theory*, *method*, *feasibility in general*, *research plan*, *costs*, *unspecific statements on relevance/theory/method*, and *other aspects*. (Hartmann 1990:102)

[9] To avoid ambiguities the names of codes used in content analysis are in italic.

The Swiss National Science Foundation

The data presented and analyzed here originate from the archives of the Swiss National Science Foundation (SNSF). They include external reviews of project applications in the investigator-driven research program in the disciplines biology and medicine for the year 1998. This year was chosen because it would allow analysing the predictive validity of the peer review procedure (Reinhart 2009). The SNSF is the largest public funding organization in Switzerland and states its main purpose in the advancement of basic research in all disciplines. In 2006, approximately 80% of the CHF 500 million budget was used to support basic research with the largest part being spent on investigator-driven projects.[10] Due to the small size of Switzerland, the SNSF is a small funding organization. For example, in comparison to the U.S. National Science Foundation it has only a tenth of the financial means, but the approval rate of 67% is considerably higher than the NSF's one of 23%.[11] Despite the small relative size, the SNSF is a typical national science funding body comparable to national funding organizations in other countries. However, Switzerland, and thus also the SNSF, is exceptional in that it steadily takes the top spot in international bibliometric rankings with the highest number of citations per researcher (Prathap 2010).

In 1998, the SNSF received 635 applications for funding in biology and medicine. The typical decision-making process can be reconstructed from archival documents and statements from employees. First, grant proposals can be submitted by meeting two deadlines each year (in spring and in autumn) and will be assigned by the administrative office to one of the members of the National Research Council. This person is the designated expert consultant for the proposal. The members of the council are currently active researchers who are appointed for four-year terms.[12] As expert consultants they suggest suitable external reviewers to the administrative office, who will then solicit reviews. On average, a consultant receives three reviews and summarises them in a recommendation for the council. This recommendation also contains the consultant's opinion and a funding priority score. The recommendation is

[10] The SNSF is also supporting what is called "targeted research". These programmes are either politically initiated and contribute to solutions to problems of national significance (National Research Programmes) or support infrastructure vital for the development of science in Switzerland (National Centers of Competence in Research).

[11] These statistics and more can be found in: (Schweizerischer Nationalfonds 2007; National Science Foundation 2006). For a general discussion of approval rates in science funding see Reinhart (2009).

[12] The constitution of the SNSF states that members of the council are internationally recognized researchers with an intimate knowledge of the scientific landscape in Switzerland. (Schweizerischer Nationalfonds 2002)

sent to the 24 council members who will then decide collectively on funding at their next meeting.[13]

Division of Labour and Structural Ignorance

Explicating the decision-making process more accurately and indicating ways the epistemic and the social are intertwined requires highlighting at this point two characteristics of the decision-making process. Analysis of the organizational structure of the SNSF reveals a form of division of labour, and this division of labour is accompanied by a structural and institutionalized form of ignorance. It is obvious from the short description of the decision-making process given above that there is a considerable amount of division of labour.[14] Its structure can be diagnosed as distributing and organizing a sizable number of actors according to temporal, spatial and epistemic dimensions. The temporal and spatial organisation of the actors will be called division of labour while the epistemic aspects are constituted by the notion of structural ignorance.[15] Structural ignorance refers to the fact that every role in the decision-making process (reviewer, expert consultant, research council) is equipped with the competence to evaluate some part of the content of the grant proposal, while at the same time ignoring other aspects. Reviewers, expert consultants, and research council can claim expertise and authority for certain aspects of a proposal and this claim is respected by other actors. Observation reveals, for example, that reviewers focus on questions of originality, consultants on questions of the career of the applicant, and the council on questions of cost. This demonstrates, not surprisingly, that epistemic aspects interrelate with the temporal and spatial organization. Structural ignorance then refers to those characteristics of the decision-making process that trim the task for individual actors to a tractable size and complexity.

[13] The information in this paragraph is extracted from the archival data and the annual report 1998 (Schweizerischer Nationalfonds 1999).

[14] Since all the participants are capable of overviewing more-or-less the whole process it would be an exaggeration to call this *high-grade* division of labor. "More-or-less" refers to the situation that all of the participants know the order and the content of the steps in the decision-making process without necessarily having access to and insight into the procedures.

[15] In order to distinguish these two notions one can also refer to the history of sociology, e.g., to classical ideas from Mannheim or Merton. According to their sociologies of knowledge science generates knowledge within a socio-cultural context that remains in the background without significantly influencing the process of finding knowledge. This is not to say that the context cannot determine, for example, the topics science deals with, which is what Merton (1938) argues in his thesis. The focus in this paper about peer review is less on context and more on organization and process. This is especially true for the concept of structural ignorance.

While segmenting the decision problem into manageable tasks, structural ignorance also bears the risk of dropping relevant aspects between and outside of tasks. This risk is somewhat aggravated by the fact that at the end of the decision-making process there is no instance to check the whole process for completeness or consistency. The only actor to be in such a position is the expert advisor, who would be overstrained if he or she had to evaluate the whole process while at the same time being a central part of it.

A telling example for this downside of structural ignorance is the procedure taken in cases where the research council is not ready to decide on an application but wants the applicant to answer further questions first. The final decision is subsequently based on these answers without checking them for consistency with the application or the reviews that can now be two to six months old. On the upside, structural ignorance allows a decision on funding in a reliable and step-by-step process. The decision-making process reduces complexity to such a degree that the council can make a final decision collectively in an average time of seven minutes per application. Considering the complexity of scientific projects and the evaluation process incorporating professional, scientific, and budgetary concerns of numerous actors, this seems to be an extraordinarily difficult task to be completed in only seven minutes. This accomplishment is achieved by the specific integration of social and epistemic aspects through division of labour and structural ignorance.

Content Analysis of Reviews – What Do Reviewers Talk About?

What are the characteristics of the external reviews as written documents in a decision-making process in which they arguably represent the most important input? In order to answer this question, a content analysis was performed based on a corpus of 212 reviews out of 68 randomly selected applications from the year 1998 in biology and medicine.[16] On average, there are 3.1 reviews per application in that sample. The length of the texts ranges from three short sentences to multiple typed pages and is on average 500 words long, which amounts to approximately one page of text.

The content analysis was performed by two independent coders with the software package Atlas.ti (Lisch and Kriz 1978; Kelle, Prein, and Bird 1995; Merten 1995; Kuckartz 2007). Several pre-tests were conducted in order to improve the intercoder reliability by developing and standardising suitable codes. The codes are based on Hartmann und Neidhardt's (1990)

[16] Reinhart (2009) analyses reliability, fairness, and validity of the SNSFs peer review process based on the same sample.

coding scheme, as mentioned above. However, during the course of the pre-tests they were modified, adapted, and expanded to the corpus.[17] Our coding scheme contains 22 different categories (see appendix for details) that were applied to all 212 reviews, resulting in 3109 coded segments of text and thus on average 15 coded text segments per review (absolute frequency). Since every single code can appear multiple times in one document, every review contains on average ten *different* codes (frequency per review). Table 1 lists the 22 codes and their frequencies accordingly. The columns "Valency" and "Valency in %" will be discussed in the next section.

Table 1 Frequency and valency of the assignment of codes to text segments in reviews. Absolute valency counts as – (negative), 0 (neutral), and + (positive) and relative valency counts in percent. (Valency total does not add up to the total under absolute frequency, 3109, because for 27 coded text segments no valency could be inferred.)

Code	Absolute Frequency	Frequency per Review	Valency			Valency in %		
			-	0	+	-	0	+
Priority	203	80 %	23	25	128	11	26	63
Summary	248	75 %	0	248	0	0	100	0
Methods	244	66 %	86	34	124	35	14	51
Originality	216	66 %	59	12	145	27	6	67
General Project	197	60 %	50	44	103	25	22	52
Research Plan	205	57 %	96	18	91	47	9	44
Feasibility	185	55 %	43	27	115	23	15	62
Theoretical Relevance	189	53 %	33	10	146	17	5	77

[17] The coding scheme of Hartmann and Neidhardt was somewhat limited by only using codes for aspects that were deemed scientific or rational in traditional philosophy and sociology of science. See e.g. (Lakatos 1970; Merton 1973). This theoretically grounded premise has been called into question by the last 30 years of science studies and cannot be taken for granted anymore. For example, the dichotomy of rational vs. social aspects is often problematized and sometimes completely rejected, see e.g. (Shapin 1994; Longino 2002). As a consequence, we tried to allow for aspects that traditionally are not understood as rational but still appear often in the corpus by inductively expanding the coding scheme. The content analysis thus also covers aesthetic (*presentation*), reflexive (*reviewer*), and procedural (*priority*) aspects with corresponding codes. See Langfeldt (2001:823) for a similar approach.

Qualification	150	52 %	15	14	121	10	9	81
State-of-Research	188	50 %	0	188	0	0	100	0
Previous Work	156	50 %	11	29	116	7	19	74
Topicality	128	45 %	2	5	121	2	4	95
Reputation	118	42 %	6	11	101	5	9	86
Practical Relevance	123	41 %	13	7	103	11	6	84
Presentation	168	38 %	123	4	41	73	2	24
Co-Applicant	98	32 %	7	12	79	7	12	80
Rest	87	30 %	16	45	26	18	52	30
General Impression	58	24 %	12	2	44	21	3	76
Environment	52	21 %	5	5	42	10	10	81
Costs	53	16 %	18	10	25	33	19	47
Reviewer	26	11 %	21	3	2	81	12	8
Previous Environment	17	8 %	0	2	15	0	12	88

Two preliminary remarks have to be made about these frequencies regarding the pragmatic embeddedness of these texts and possible influences on the usage of certain quality criteria. First, the frequency of quality criteria and thus the frequency of corresponding codes depend, among other things, on how the reviewers interpret their role. This interpretation, in turn, depends on explicit requests by the funding organization attached to the invitation to review. In the case of the SNSF the reviewers receive a standardized covering letter which, according to the archival documents, comes in two varieties. The short version is a simple letter inviting the recipient to kindly review the attached application. The long version, in addition, explains the purpose of the SNSF and its reviewing procedures. More importantly, the long version also asks the reviewer to specifically pay attention to the following aspects of the application: "originality and topical interest of the work", "suitability of the methods", "past performance of the applicant", and a "priority of funding" on a scale of high/medium/low". Since both versions of the covering letter were rarely archived, because they were either regarded as obvious or unimportant, a statement about their relative frequency is difficult to make. A

reasonable hypothesis seems to be that the long version of the covering letter was only sent to first time or international reviewers, since they would be less familiar with the SNSF and its procedures. It is therefore impossible to exclude a possible influence that the long version of the covering letter might have had on the usage of quality criteria in the reviewer's arguments even though the explicitly mentioned criteria are common and probably self-evident in peer review procedures of this sort. Contrary to this possible distortion, it must be emphasized that reviewers are free of any constraints regarding the composition of their reviews and the documents show that they are in fact making good use of this freedom. As explained above, Kretzenbacher and Thurmair made the same observation and termed it low formal and medium functional equivalence. Langfeldt similarly concludes: "directions to reviewers are of limited importance" (Langfeldt 2001:826).

The second remark concerns the fact that numerous codes are assigned to more than half of all the reviews while in Hartmann and Neidhardt's study this was not the case for any code (Hartmann and Neidhardt 1990:422). External reviewers for the SNSF use a wider range of quality criteria when reviewing compared to the reviewers of the German DFG. Without any further data for comparison, we have to assume that this difference originates in differential organisational structures of the DFG in comparison to the SNSF. The DFG had permanent appointed reviewers (Fachgutachter) and external reviews were only requested if required (Sondergutachten). It would be obvious to assume that these permanent internal reviewers needed to employ less quality criteria explicitly in order to justify their evaluation.

After these preliminary remarks, the content of table 1 needs to be explained and discussed. To gain a first impression, the extreme cases on the frequency list will be considered, the codes that were assigned very often and the ones that were assigned very rarely. On top of the list in Table 1 we find *priority*, coding text passages that recommend the application as a whole for funding or for rejection. As mentioned above, we cannot estimate how often the reviewers are explicitly requested to assign a funding priority. Nonetheless, we can assume that reviewers are aware that a funding recommendation is expected of them. Given this, it seems remarkable that 20% of reviewers give no overall recommendation for funding even though this would be the most obvious function of a review. Since ignorance seems an unlikely explanation for this fact, two other possibilities have to be considered: These 20% of reviewers are either deliberately choosing not to assess the priority or they find themselves unable to issue such a recommendation. Either way the multilevel organisation of the decision-making process absorbs these cases with the consultant or the council interpreting

the review in order to derive a recommendation. In turn, some reviewers might even be aware of this absorbing quality of the process and thus feel less obliged to recommend a priority explicitly.

Summary can be found in second place and codes for passages that give an account of the content of the application without evaluating it. The high frequency of *summary* is inconsistent with the remark by Kretzenbacher and Thurmair that "pure information or communication as a non-evaluative speech act almost never occurs in peer reviews" (1995:188, translated from German by the author).[18] They interpret this as one of the main differences between the two text types, peer review and review in general as e.g. book review or published review paper. Based on the available data, it seems impossible to attribute this discrepancy either to disciplinary differences (they are analysing data from a chemistry journal), or to differences between editorial and funding peer review. Regarding the explanation by disciplinary differences, there is some evidence from Hartmann's study (1990:112) that scientific fields differ in their weighting of quality criteria, but her results are confined to evaluative statements. The same can be said for the results of Guetzkow et al. (2004), which show disciplinary differences in the usage of the quality criterion *originality*. These results should warn us from abolishing the disciplinary explanation prematurely. Nonetheless, the explanation by differences in the procedures of editorial and funding peer review that affect the text types respectively appears more proximate. If the text types review in general and peer review for editorial purposes differ in their usage of non-evaluative statements, then peer reviews for funding purposes are to be regarded as different from these two types of text. It is different from the review in general by using evaluative statements less often and by stemming from a pragmatic context that results in one-sided anonymity and multiple addressees. In contrast to peer reviews for editorial purposes, it uses non-evaluative statements frequently and its topic is not scientific results but future scientific work.

Analyzing the content of the text passages coded as *summary* allows us to draw some conclusions on the non-evaluative statements and their function in text strategy. On the one hand, summaries of the content of an application can mostly be found at the beginning of reviews, emphasizing the main points of the subsequent review. The reviewer addresses the consultant directly and contextualizes the evaluating parts of the text, thus allowing him or her to qualify and place the review within other possibly rivalling reviews. By this the reviewer is

[18] The original passage reads: "Reines Informieren oder Mitteilen als nicht bewertende Sprechhandlung kommen in Peer Reviews fast nie vor…".

taking the consultant seriously as an expert colleague and rejects the role of an unquestionable authority in the decision-making process. Consultant and council are thereby put in a position to decide beyond a purely arithmetic majority of the reviewers' recommendations. In fact, they can balance different opinions against each other and add new criteria to the deliberation, which is, according to the documents, what they are actually doing.[19] On the other hand, introductory or informing parts of the text are strategically used to build up authority. Reviewers apply a summary to demonstrate their ability to read, understand, and put the application in a disciplinary context for which they can be considered experts. This strengthens the observation that reviewers do not see themselves as experts with unquestionable authority but rather as colleagues (peers) to the members of the council who need to maintain their authority.

Next on the list are *methods* and *originality*. Both are distinguished as being relevant quality criteria in the everyday language of science. They are also among the most often mentioned quality criteria in Hartmann's study, where originality is termed a "quality criterion immanent to science" (1990:102, translated from German by the author) and coded among others as *scientific relevance*.[20] Therefore, it is no surprise that these two criteria occur very frequently but for the following argument it is important to keep in mind that the denotation of the two terms is oppositional. *Methods* relates to technical capabilities of the applicant that are laboriously learnt as well as correctly and precisely reproduced and applied. Furthermore, reviewers discuss methods in context of the proposed work where they are seen as means to an end. For example, they will suggest that a certain method is not instrumental for achieving the promised results or that the applicant is not competent enough to achieve the results via

[19] It is remarkable in this context that consultant and council heavily favor one of these two options. In the course of the deliberation, they position themselves as experts for aspects of the application that were neglected by the reviewers. The authority of the reviewer remains untouched by not qualifying or interpreting his or her statements, which in turn relieves the council of the burden to argue against the reviews. This seems in line with the observation that consultant and council put the same weight on all reviews independent of length, level of detail, or elaborateness of argument. They regard the reviews not as texts that need to be interpreted besides their obvious face-value, which contradicts the often articulated assumption that the reviews have to be deciphered for hidden messages like letters of recommendation in personnel offices.

[20] This applies to three of the four disciplines studied by Hartmann. In psychology, economic theory, and political science, both criteria are always among the top four or top five. Electrical engineering is an exception where *methods* and *scientific relevance* rank very low, because of a strong focus on *practical relevance*. A further elaboration of this exception seems futile, since electrical engineering is in many ways a special case in Hartmann and Neidhardt's study (Neidhardt 1988:21).

this method. By contrast, statements about originality mostly refer to properties of the project that are described and understood as creative or elusive. Frequently recurring terms in this context are: original, new, fascinating, unprecedented, interesting, singular, imaginative, path-breaking, or exciting. Thus the semantic fields of the terms originality and methods are quite distinct and their elements could be described as creative or elusive in the case of the former and rigid or instrumental in the case of the latter. This polarity opens up a space for statements that allow parts of the application to be assigned qualities from a broad spectrum ranging from creative, elusive to rigid, instrumental. Following up on the remark by Kretzenbacher and Thurmair that science lacks a vocabulary of evaluation, this gives a first hint as to how this deficit is compensated.

On the other end of the list, the least assigned codes are *environment*, *previous environment*, *costs*, and *reviewer*. The two codes *environment* and *previous environment* refer to statements about present or past institutional affiliations, be they the university, the department, the institute, the research team, or the laboratory. Even combining the two codes would still leave them very low on the frequency list. It has to be assumed from this that reviewers place little importance on local surroundings in their assessment of an application. This can be interpreted in two ways: They might consider the council to be more competent in this matter or they regard the environment to be a criterion of low importance compared to other criteria.

Equally rare are statements about *costs* of a proposed project, which, prima facie, is astonishing since all the participants in the decision-making process are well aware that costs must play some role in the allocation of funds. This is even more astonishing considering the fact that Hartmann and Neidhardt find costs as the quality criterion mentioned third most. In light of the organization of the DFG decision-making process, which is based on permanent reviewers and thus exhibiting less division of labour, the difference in the SNSF seems to be that assessment of the costs is not expected from the reviewers but from the consultant and the council. While the quantitative content analysis does not confirm this hypothesis, there are at least singular cases supporting this kind of explanation. If reviewers explicitly mention costs in their review, it is not unusual that consultant and council ignore these statements or at least assign them less weight than statements about other quality criteria.

As mentioned above, reviewers often employ a text strategy to establish or secure a position of professional authority. This strategy as a reflexive discussion of the reviewer's role remains an implicit one. Furthermore, it most likely has to stay implicit to achieve its purpose. Since the code *reviewer* is assigned only rarely, there seems to be little explicit self-reflection in the

reviews about the role of the reviewer. In the cases where the reviewers address their role as reviewers it is in problematic form, admitting either incompetence or conflict of interest regarding the topic of the application. In return we find almost no statements of reviewers about why they consider themselves to be suitable experts. From this, the following mutual attribution of competence and expertise between the SNSF and the reviewers can be inferred: The invitation to review assigns a considerable amount of authority to the reviewer. This attribution is considerable because the reviewers sometimes answer by declining the invitation in admitting incompetence or conflict of interest. The advance of authority is also large enough to prevent reviewers from feeling the need to emphasize their competence and expertise. They also dispense with such a self-proclamation because they see themselves as professional colleagues to the members of the research council so that this kind of action is deemed inappropriate.[21] But in return the attribution of authority is also not excessive as reviewers still consider it necessary to employ text strategies to secure their authority even though these text strategies remain implicit. This interplay of mutual expectations performs not only the function of assigning a role to the reviewer but also regulating and controlling it.[22] As an explication of a control mechanism this only concerns the role of the reviewer. As far as the mutual relation of consultant and council is concerned more data and more work is needed. We can infer from the organization of the decision-making process that the consultant is probably less exposed to similar mechanisms of control. Nevertheless, it would be premature to suspect that consultants are not subjected to any control mechanisms and thus can act as they please, since a quantitative analysis identifies the external reviews as the most dominant predictor for the funding decision. (Reinhart 2009)

Valency Analysis of Reviews – How Do Reviewers Rate Applications?

[21] This statement may seem inconsistent with earlier remarks but in fact it is not. Reviews are addressed to multiple recipients thus displaying the anticipation of multiple expectations which in turn can appear as inconsistency. A full account following Kretzenbacher und Thurmair's multiple assessor roles and multiple addressees (1995:211) would have to trace all these interactions one by one to disentangle all these statements that appear as contradictions. The main point here was to emphasize the interplay of mutual expectations and for this purpose such an extensive discussion seemed unnecessary and so only the interaction between reviewer and research council is discussed.

[22] Reviewer anonymity and intransparency of the process lead to periodical discussions in the literature on peer review about the possibilities for more control and transparency, see (Reinhart and Sirtes 2006).

Since peer reviews are types of text that contain explicit positive or negative assessments, it is possible to perform a content analysis of valency in addition to frequency[23] (Schnell, Hill, and Esser 1995:411). For this purpose, we added a mark to every coded text passage reflecting the nature of the assessment as positive, negative, or neutral. Table 1 lists the codes according to their valencies.

Two codes were excluded from being marked with valency indicators: *summary* and *state-of-research*. Both were deliberately designed as codes to be used when a summarising statement or an account of the state-of-research was present which was of purely informative nature without giving an assessment of any part of the application. For all other coded statements, we can see from table 1 that they are mostly positive (63%), rarely negative (24%), and even more rarely neutral (13%).[24] Hartmann (1990:123) discovered a similarly uneven distribution in favour of positive assessments.

What are the extreme cases in table 1 that appear as negative or positive in an above average number of cases? Among the most often negatively assessed codes are *presentation*, *research plan*, *methods*, and *costs*.[25] *Presentation* is the most noticeable of them by appearing in very few reviews (38%) while at the same time being used as negative more frequently than the others (73%). Comments about formal aspects of an application like writing style, clarity, or level of detail are thus specifically employed to break the basically positive tone of the reviews. The other three mostly negative aspects – *research plan*, *methods*, and *costs* – are illustrated in the appendix. They share the feature that they could be assigned to the type of quality criteria mentioned above as rigid and instrumental.

Table 1 shows *topicality*, *environment*, *reputation*, *practical relevance*, *qualification*, and *co-applicant* as the codes most often associated with positive statements. This prima facie heterogeneous group can be structured along the dimensions mentioned above. The dimension between the poles rigid, instrumental and creative, elusive has already been explicated. An additional, second dimension is based on the work of Guetzkow et al. and their interpretation that reviewers "value original work because they regard it as a sign of the moral character of

[23] Valency analysis is sometimes also called "evaluative assertion analysis" (Osgood 1959).

[24] To compute these numbers the codes *summary* and *state-of-research* were excluded because they were defined to have no valency.

[25] In fact *reviewer* would be the most often negatively assessed code (81%) but will be omitted from the following discussion because it codes not for statements about the project or the applicant but for those concerning the reviewer itself. In the same way as the already omitted codes *summary* and *state-of-research*, they are only indirectly instrumental to an understanding of the assessment by the reviewers.

the researcher" (Guetzkow et al. 2004:192). There is an underlying distinction separating attributions by the reviewers either to the researcher or the project.[26] Thus, the second dimension would be situated between the poles of the researcher and the project. The two dimensions together cover both the semantic fields as well as the carriers of these attributions of properties and allow statements by the reviewers to be situated within a two-dimensional space. Applying this to the six mainly positive codes shows them either to be attributions to the researcher (*reputation*, *qualification*, *co-applicant*, *environment*) or to be semantically close to the pole creative, elusive (*topicality* and to a lesser degree *practical relevance*). In contrast, the mainly negative codes are situated in the opposite corner by being attributions to the project as well as semantically close to the pole rigid, instrumental (*research plan*, *methods* and *costs*).

Results

Based on the preceding arguments the findings from the content analysis may be summarized as follows: The review process of applications for scientific projects is challenged by the problem that scientific language lacks a sophisticated and genuine lexis of evaluation. The empirical analysis presented here suggests that this deficit is compensated by reviewer statements that draw on a two-dimensional field ranging from rigid, instrumental to creative, elusive in the semantic dimension, and from the researcher to the project in the attributive dimension. Furthermore, there is an uneven distribution within this two-dimensional field in which positive evaluations are mainly phrased as creative, elusive properties of the researcher and negative ones as rigid, instrumental properties of the project.

Several reasons can be offered to explain why further empirical data is expected to support these results. As mentioned above, existing empirical studies are in part consistent with the presented argument presented here. Guetzkow et al. find a remarkable reinterpretation by reviewers of the kinds of criteria that are usually considered to be scientific or rational as

[26] Guetzkow et al. describe this attribution as reviewers regarding a quality criteria as a sign of the moral character of the researcher. There is ample material in our data from the SNSF that could tellingly be interpreted in the same manner. But since Guetzkow et al. are analyzing data from fellowship programs exclusively it seems obvious to expect quality criteria to be attributed to researchers. For the case of project funding the same attribution seems far less obvious. Therefore I would suggest first and foremost not to ask to what part of the researcher quality is attributed but more fundamentally what are possible entities to which quality can be attributed. Why do reviewers in generally tend to view the quality of a project as an indicator for the quality of the researcher and not just, plain and simple, as the quality of the project?

personal (and moral) properties of the applicant. This corresponds to the observation from funding peer review for projects where quality criteria are personalized in an analogous manner. In addition, we also find an interference of these attributions with the valency of the associated assessment. This, in turn, resonates with Neidhardt's conclusion that reviewers display goodwill and civility towards the applicant, an attitude that he sees rooted in the norm not to speak negatively about colleagues (Neidhardt 1988:119). In simpler terms, framing positive evaluations as *genius of the author* and negative evaluations as *technical shortcomings of an application* might be a general feature of peer review.

Conclusions

Two conclusions can be drawn from the analysis: First, the results support the hypothesis that epistemic and social aspects of peer review practices are inseparable. A pattern of evaluative statements may be discerned, characterised by positive remarks mainly phrased as creative, elusive properties of the applicant and negative remarks phrased as rigid, instrumental properties of the project. This pattern demonstrates that communicative practice in peer review always combines the epistemic with the social, by assessing scientific work according to quality criteria and by employing these criteria in a way adequate to the social situation. Otherwise, we would be left to assume that projects are rejected because they are bad projects or else are funded because they were submitted by good researchers. This is highly implausible and also not what is suggested by the empirical material. Rather, this pattern in evaluative peer review statements considers two norms of scientific culture at the same time: colleagues are not to be spoken negatively about and assessment must be held according to disciplinary standards of quality.

Second, studies that have treated peer review as an inherently social and epistemic practice have only been able to offer a few general features. Apart from the pattern in evaluative statements discussed above, Mallard et al. (2009) offer the *negotiation of fairness criteria* and Hirschauer (2010) the *panoptic organisation of communication*. Both of these features are not present in the material analysed for this study. Negotiation of fairness criteria is absent because external written reviews are shaped in a social context that differs from orally presented panel evaluations. External experts lack face-to-face contact and have limited options for repeated interactions so that negotiation of any kind is highly unlikely. However, negotiation of fairness criteria in the form discussed by Mallard et al. may very well take place in the second phase of the SNSF's decision-making process, the council members' panel meeting.

A similar observation can be made for the panoptic organisation of communication even though Hirschauer employs this term in a very general manner: "Peer review can be understood as the whole process that reaches from author's initial writing and scrupulous self-editing, through colleagues' informal and formal comments up to the journal readers' curious, polemical and jealous readings. What has become known as scientific 'criticism' is an ongoing *panoptic* organization and surveillance of communication" (Hirschauer 2010:96). Such a definition of peer review excludes, among others, evaluation procedures for research funding like the one discussed in the present study. In most instances, funding peer review is characterised by carefully managed visibility conditions. The question of who gets to see, read, and criticise a proposal from the writing stages right up to the funding decision is regulated by different forms of non-transparency and confidentiality. Most importantly, proposals for research funding are never published and never judged by a public audience and therefore the surveillance of communication cannot be termed panoptic. At least in the general sense, in which Hirschauer equates scientific criticism with a panoptic organization of communication, funding peer review cannot be termed panoptic.

However, in a limited sense and closer to Foucault's (1977) example of Bentham's prison, funding peer review is indeed panoptic. Just like Bentham's prison reserved a privileged observation spot for the prison guards, funding peer review reserves a privileged observation spot for the gatekeepers of science. Gatekeepers like the SNSF's council members have complete transparency of the decision-making process and can observe and judge the work of the applicant, the external reviewers, and the administration while the observed actors are prevented from communicating with each other. There is a power differential associated with this social arrangement but at least two mechanisms prevent the gatekeepers form abusing their position in an arbitrary manner. First, gatekeepers working as a panel are accountable to each other and will, for this reason, negotiate fairness criteria (Mallard et al. 2009) that limits arbitrary decision making. Second, gatekeepers are embedded in an organisational context that subdivides the decision making process into several steps. Such forms of organisation restrict the range of action for all the actors involved even those in alleged positions of privilege (Reinhart 2009). In this limited sense peer review can be termed a panoptic organisation of communication without having to narrow the definition of peer review to editorial peer review.

The primary objective of this study was to open up the black box of peer review and to increase its transparency, understanding and credibility. It is not without irony that we end up

granting a prison-like panoptic social arrangement where we were hoping to find scientific criticism. However, the lesson here is not that peer review works like a prison and hope for rational and fair decisions is unrealistic. Rather, the decisions from peer review are accepted as more or less rational and fair by the community because peer review structures visibility, communication and criticism in an appropriate organisational form. Accepting that epistemic and social aspects of peer review are mutually constitutive means always relating the rationality of peer review decisions to the social arrangement from which they emerge. Organisational aspects are central to this relationship and the present study aimed to clarify one such linkage. The value of such studies for science policy and the regulation of peer review procedures is that they oppose approaches that treat all peer review procedures alike. Peer review comes in a many organisational forms, embedded in specific contexts and delivering results accordingly. More research is needed to shed light on the many forms of peer review and to provide a sound foundation for reaching more general conclusions.

Bibliography

Barnes, Sherman B. 1934. “The Scientific Journal, 1665-1730.” *The Scientific Monthly* 38:257-260.

Collins, Harry M., and Robert Evans. 2002. “The Third Wave of Science Studies: Studies of Expertise and Experience.” *Social Studies of Science* 32:235.

Dirk, Lynn. 1999. “A Measure of Originality: The Elements of Science.” *Social Studies of Science* 29:765-776.

Foucault, Michel. 1977. *Discipline and Punish: The Birth of the Prison*. New York: Vintage.

Guetzkow, Joshua, Michèle Lamont, and Grégoire Mallard. 2004. “What Is Originality in the Humanities and the Social Sciences?.” *American Sociological Review* 69:190-212.

Hackett, Edward J., Olga Amsterdamska, Michael Lynch, and Judy Wajcman, eds. 2008. *The Handbook of Science and Technology Studies*. Third Edition. Cambridge, Mass: The MIT Press.

Hartmann, Ilse. 1990. *Begutachtung in der Forschungsförderung - Die Argumente der Gutachter in der deutschen Forschungsgemeinschaft*. Frankfurt am Main: Rita G. Fischer.

Hartmann, Ilse, and Friedhelm Neidhardt. 1990. “Peer Review at the Deutsche Forschungsgemeinschaft.” *Scientometrics* 19:419-425.

Hemlin, Sven. 1993. “Scientific Quality in the Eyes of the Scientist. A Questionnaire Study.” *Scientometrics* 27:3-18.

Hess, David J. 1997. *Science Studies: An Advanced Introduction*. London / New York: New York University Press.

Hirschauer, Stefan. 2010. “Editorial Judgments: A Praxeology of 'Voting' in Peer Review.” *Social Studies of Science* 40:71-103.

Hirschauer, Stefan. 2004. “Peer Review Verfahren auf dem Prüfstand: Zum Soziologiedefizit der Wissenschaftsevaluation.” *Zeitschrift für Soziologie* 33:62–83.

Kelle, Udo, Gerald Prein, and Katherine Bird. 1995. *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice*. London: Sage Publications.

Kretzenbacher, Heinz L., and Maria Thurmair. 1995. “"... sicherlich von Interesse, wenngleich...". Das Peer Review als bewertende Textsorte der Wissenschaftssprache.” Pp. 175-216 in *Linguistik der Wissenschaftssprache*, edited by Heinz L. Kretzenbacher and Harald Weinrich. Berlin: de Gruyter.

Kretzenbacher, Heinz L., and Maria Thurmair. 1992. “Textvergleich als Grundlage zur Beschreibung einer wissenschaftlichen Textsorte: Das Peer Review.” Pp. 135-146 in *Kontrastive Fachsprachenforschung*, edited by Klaus D. Baumann and Hartwig Kalverkämper. Forum für Fachsprachenforschung.

Kuckartz, Udo. 2007. *Einführung in die computergestützte Analyse qualitativer Daten*. 2nd ed. Wiesbaden: VS Verlag für Sozialwissenschaften.

Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." *Criticism and the Growth of Knowledge* 4:91–195.

Lange, Stefan. 2009. "Neuer gesellschaftlicher Legitimierungsbedarf der Wissenschaft in Deutschland – Veränderungen in der Wissenschafts-Governance am Beispiel der Universitäten." in *Exzellenz durch Steuerung? Neue Herausforderungen für das deutsche und das französische Wissenschaftssystem, Discussion Paper*, edited by Effi Böhlke, Pascale Laborier, Andreas Knie, and Dagmar Simon. Berlin: Wissenschaftszentrum für Sozialforschung.

Langfeldt, Liv. 2001. "The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome." *Social Studies of Science* 31:820-841.

Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. 2nd ed. London: Sage Publications.

Lisch, Ralf, and Jürgen Kriz. 1978. *Grundlagen und Modelle der Inhaltsanalyse: Bestandsaufnahme und Kritik*. Reinbek bei Hamburg: Rowohlt.

Lock, Stephen. 1985. *A Difficult Balance: Editorial Peer Review in Medicine*. Philadelphia: iSi Press.

Longino, Helen E. 2002. *The Fate of Knowledge*. Princeton University Press.

Luhmann, Niklas. 1970. "Selbststeuerung der Wissenschaft." Pp. 232-252 in *Soziologische Aufklärung 1: Aufsätze zur Theorie sozialer Systeme*. Köln, Opladen.

Lux, David S. 1989. *Patronage and Royal Science in Seventeenth-Century France: The Académie dePhysique in Caen*. Ithaca: Cornell University Press.

Mallard, Grégoire, Michèle Lamont, and Joshua Guetzkow. 2009. "Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review." *Science Technology Human Values* 34:573-606.

Merten, Klaus. 1995. *Inhaltsanalyse: Einführung in Theorie, Methode und Praxis*. 2nd ed. Opladen: Westdeutscher Verlag.

Merton, Robert K. 1938. "Science, Technology, and Society in Seventeenth Century England." *Osiris* 4:360-632.

Merton, Robert K. 1973. "The Normative Structure of Science." Pp. 267-279 in *The Sociology of Science*.

National Science Foundation. 2006. *OLPA Congressional Affairs/Congressional Update - January 11, 2006*. http://www.nsf.gov/about/congress/109/highlights/cu06_0112.jsp (Accessed March 27, 2008).

Neidhardt, Friedhelm. 1988. *Selbststeuerung in der Forschungsförderung: Das Gutachterwesen der DFG*. Opladen: Westdeutscher Verlag.

Osgood, Charles E. 1959. "The Representational Model and Relevant Research Methods." Pp. 33–88 in *Trends in Content Analysis*, edited by Ithiel de Solla Pool. Urbana: University of Illinois Press.

Power, Michael. 1997. *The Audit Society: Rituals of Verification*. Oxford University Press.

Prathap, Gangan. 2010. "An iCE Map Approach to Evaluate Performance and Efficiency of Scientific Production of Countries." *Scientometrics*. http://dx.doi.org/10.1007/s11192-010-0192-x (Accessed June 2, 2010).

Reinhart, Martin. 2009. "Peer Review of Grant Applications in Biology and Medicine. Reliablity, Fairness, and Validity." *Scientometrics* 81:789-809.

Reinhart, Martin, and Daniel Sirtes. 2006. "Wieviel Intransparenz ist für Entscheidungen über exzellente Wissenschaft notwendig?." *IfQ Working Paper* 27-36.

Schnell, Rainer, Paul B. Hill, and Elke Esser. 1995. *Methoden der empirischen Sozialforschung*. 3rd ed. München: R. Oldenbourg Verlag.

Schweizerischer Nationalfonds. 1999. *Jahresbericht 1998: Bewilligte Beiträge*. Bern: Schweizerischer Nationalfonds.

Schweizerischer Nationalfonds. 2007. *Jahresbericht 2006*. Bern: Schweizerischer Nationalfonds.

Schweizerischer Nationalfonds. 2002. "Stiftungsurkunde / Statuten." http://www.snf.ch/SiteCollectionDocuments/por_org_statuten_d.pdf (Accessed August 14, 2007).

Shapin, Steven. 1994. *A Social History of Truth: Civility and Science in Seventeenth-Century England*. University Of Chicago Press.

Sismondo, Sergio. 2004. *An Introduction to Science and Technology Studies*. Oxford: Blackwell Publishing.

Ware, Mark. 2008. *Peer Review: Benefits, Perceptions and Alternatives*. London: Publishing Research Consortium.

Weingart, Peter. 1970. "Selbststeuerung der Wissenschaft und staatliche Wissenschaftspolitik." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 22:567-592.

Whitley, Richard, and Jochen Gläser, eds. 2007. *The Changing Governance of the Sciences*. Dordrecht: Springer.

Wood, Fiona, and Simon Wessely. 1999. "Peer Review of Grant Applications: A Systematic Review." Pp. 14-31 in *Peer Review in Health Sciences*, edited by Fiona Godlee and

Tom Jefferson. London: BMJ Books.

Zilsel, Edgar. 2000. “The Sociological Roots of Science.” *Social Studies of Science* 30:935.

Appendix

The content analysis was performed with a coding scheme consisting of 22 codes listed below. For every code a short description of what it codes for is presented, followed by one or more typical text segments (in English and German) and where necessary a description of how borderline cases were handled.

Topicality

codes for explicit statements about the topicality or the timeliness of the project. Implicit remarks on topicality can be descriptions of a growing and increasingly noticed line of research or a central but unsolved question or problem the project is approaching.

"aktuell; of interest; topical; hot; timely; cutting edge; modern; up-to-date"

Statements about topicality often appear alongside statements about originality and the terms "cutting edge" and "of interest, interesting" can be borderline cases. The codes are assigned according to the examples given.

Originality

codes for statements about originality that are in most cases explicitly identified by the reviewer or else belong to the semantic field described with the given examples.

"originell, neu, faszinierend, unprecedented, unique, imaginative, Pionierarbeit, exciting, interesting"

Theoretical Relevance

codes for statements about the relevance of the proposed project for theoretical knowledge or for the development of the scientific field.

"I believe that this work will open up new and important ground in our understanding of hostparasite interactions; Es beinhaltet ein wissenschaftlich höchst interessantes Problem der Grundlagenforschung; I see this work as somewhat basic rather than merely applied, leading to more knowledge about control in anesthesia"

Practical Relevance

codes for statements about the practical use of the project or possible application areas outside of science. In medicine these are obviously very often remarks about possible cures for diseases.

"of great clinical relevance for the practice of surgery and anesthesia; will have a major impact on apple breeding and cultivation; have little significance for an average anaesthesiologist"

Statements about applications within science e.g. tools for further research could also be coded as *theoretical relevance* and have to be handled on a case-by-case basis.

Methods

codes for statements about methods that are sometimes explicitly addressed with the term "method" but more often are discussing methodological aspects in technical terms belonging to the semantic field of "method".

"Die im Gesuch angesprochenen Diagnoseverfahren werden nicht näher erläutert; Auf der anderen Seite bestehen Zweifel, ob die Methodik genügend ausgefeilt ist, um brauchbare Antworten auf die gestellten Fragen zu finden; selbst wenn – wie in dieser Studie vorgeschlagen HRV – Bestimmungen mit Troponin T kombiniert werden sollen – wird man zeigen müssen, dass diese Kriterien mehr bringen als simple klinische Scores"

Feasibility

codes for statements about the feasibility or viability of the project. These statements are very often close to statements coded as *costs*, *qualification*, *reputation*, *environment*, *methods*, or *research plan* and are only coded as *feasibility* if they explicitly mention feasibility or if the statement is too general to be coded with one of the other mentioned codes. Statements about feasibility are often presented as reports of the success or failure of similar earlier projects.

"answers can be obtained; they have the personnel resources to carry out the project; this work is in continuity with the previous work; since the specific techniques to be used are not described in detail, it isn't possible to judge what is the likelihood of success; the present proposal would be feasible and could be completed within the time frame proposed within the application"

Research Plan

codes for specific statements about the research plan that focus on the temporal sequence and compilation of the proposed works. Not qualified are statements of general nature and/or statements about specific methods. In these cases using more than one code was avoided.

"Without careful attention to the research design, however, this study could/will produce a biased result; the size of this undertaking, as well as the variation of methods applied, is unprecedented; for this 3-year project a reduction in both the number and texting of methods needs to be considered"

Presentation

codes for statements about formal aspects of the application such as: writing style, presentation, comprehensibility,… Statements about formal aspects are sometimes hard to distinguish from statements about content. In these cases statements may be coded with more than one code.

"the most confusing grant application I have ever seen; well presented; this is an extremely well written grant application; the specific techniques to be used are not described in detail, it isn't possible to judge what is the likelihood of success"

Since unclear presentation often leads to doubts about the feasibility of the project, both codes *feasibility* and *presentation* are used in such cases.

General Impression

codes for general statements about the project that cannot be assigned to any other code and/or express a subjective impression of the reviewer.

"to my delight, the proposal was even better than I'd expected; there are certain weaknesses; I have read the grant application with great interest and enthusiasm"

Statements on the borderline to *presentation* will be coded as *general impression* if they focus on the reading experience of the reviewer ("es war eine Freude zu lesen"). If they focus on the writing- and presentation-style in the application then they are coded as *presentation* ("it is well written").

General Project

codes for general statements about the project that cannot be assigned to other codes and/or relate to the basic research idea or direction of research.

"The proposed project is sound; this is a generally very interesting project; although this proposal is strong and worthy of funding, there are certain weaknesses that may be more serious than the author has implied; I think this is the direction where the research should go"

For statements describing the reviewer's reading experience *general impression* is used.

Previous Work

codes for statements about the applicant's previous work but not for statements expressing the reviewer's estimation of the applicant's reputation or standing. Statements about previous work are very often based on personal acquaintance or familiarity with the work of the applicant.

"Ich habe ihn als fleissigen und seriösen Arzt und Forscher kennengelernt; past performance of the applicant seems excellent; his group has produced a number widely cited and influential publications; in recent years he has published an enormous number of important publications on the depth of anaesthesia and gas monitoring"

Statements about *previous work* are often also statements about *qualification* or *feasibility* and vice versa. These are mostly hard to separate and are coded with multiple codes accordingly.

Qualification

codes for statements about the professional qualification of the applicant in relation to the proposed project. Assessements based on personal acquaintance or familiarity with the work of the applicant are frequent (see *previous work*)

"Er ist ohne Einschränkung zur Leitung und Durchführung des beschriebenen Forschungsprojektes qualifiziert; zum beantragten Thema erfolgte allerdings noch keine Publikation"

Reputation

codes for statements about the applicant that are of a general kind (in contrast to *qualification*) and/or describe the applicant's reputation in the scientific community. Also qualified are general statements about publication success, as e.g. "er publiziert nur in den besten Zeitschriften".

"Dr. [x] gilt in der 'Kinderanästhesieszene' als initiativer und kluger Kollege; his publications routinely appear in the best journals"

Statements are coded as *previous work* if they are based on personal acquaintance or familiarity with the work of the applicant.

Co-Applicant

codes for statements about a possible co-applicant or other researchers listed as cooperative partners for the project.

"Beim Zweitgesuchsteller habe ich Fragezeichen; The CV of Dr [x] (whom I do not know!) indicates clearly a general interest in the field, and read like a business advert; I see [x]'s collaboration as the key to making this project cutting edge science"

Environment

codes for statements about groups, laboratories, institutes, departments, or universities where the project will be performed.

"The project might benefit from the clinical environment in Bern; auch hier liegen von der Gruppe schon veröffentlichte Ergebnisse vor"

Previous Environment

codes for statements about groups, laboratories, institutes, departments, or universities where the applicant was active.

"Die Klinik in London, in der er sich aufhielt, hat tatsächlich auf diesem Gebiet international einen guten Ruf."

If the statement focuses on the applicant or the proposed project it is coded as *qualification* or *environment* respectively.

Costs

codes for statements about financial and material input.

Reviewer

codes for statements about the reviewer himself or herself. Very often these are statements about their qualification or conflict of interest in relation to the application.

"since I am not a cell biologist I may not be the best person to judge the work; Ich habe aber selbst wenig Erfahrung auf dem Gebiet"

Summary

codes for statements summarizing the content of the application without giving an assessment.

State-of-Research

codes for statements describing the state of knowledge or the state of research in a given field.

Priority

codes for statements about funding priority. Most of these are answers to the explicit request from the SNSF for a funding priority on the scale low/medium/high. But other statements beyond this scale qualify also.

"I would give to the funding of the project a high priority; verdient die volle Unterstützung"

Rest

codes for statements that seem relevant but cannot be assigned to any of the other codes.