CESIS
Electronic Working Paper Series

Paper No. 97

REVIEWING APPLICATIONS BY WOMEN

- critical use of additive and reasoning evaluation methods

Göran Melin

(SISTER)

September 2007

Reviewing applications by women:

Critical use of additive and reasoning evaluation methods

GÖRAN MELIN

Abstract:

Two programmes regarding research funding are investigated with respect to their evaluation processes. One programme targets individuals and the other centers of excellence. The study attempts to disclose some of the features of the selection procedures and the aim is to contribute with further understanding of the mechanisms in such selection processes, which lead to disproportionate disapproval of female applicants. The use of intuitive or reasoning evaluation methods, together with quantitatively measureable or additive methods, is found to be critical for female applicants. When funding organizations try to advance their evaluation procedures and involve more of reasoning evaluation, there is a risk that other than established main stream projects fail, including applications by women. The paper ends by proposing a hypothesis which may serve as a starting point for further empirical studies.

Swedish Institute for Studies in Education and Research
Drottning Kristinas väg 33D
SE-114 28 Stockholm
www.sister.nu

ISSN 1650-3821

SiSTER
Swedish Institute for Studies in Education and Research

Introduction

An investigation conducted by Beata Nervik (2006) about the gender differences in a handful of large Swedish scientific funding programmes included the following conclusions:

- Women apply to a lesser extent than men
- Women apply to an even lesser extent as centre-leaders
- Women are most underrepresented in Life Sciences and Medicine
- An internal selection process occurs at the universities where the number of applications are reduced
- Women are granted funding to a lesser degree than men; there is a tendency to a negative selection of women in each single step of the selection process, even if the selection process is comprehensive and transparent

A couple of the investigated programmes are individual; the research grant is applied for and given to an individual. Several other programmes are designed for the support of centers of excellence, and funding is applied for by either a research leader with one or several co-applicants, or by the hosting institution as such, normally the university where the center is to be located. Sometimes these different types of programmes can be compared, sometimes not. In the investigated Swedish programmes, a total of approximately 500 million euro (SEK 4.5 billion) has been awarded. Women have applied for 19% and received 16% of this money (Nervik, 2006). These programmes all together make up for a significant share of the available resources for carrying out research in Sweden.
There are rarely striking differences between the proportion of male/female applicants and the proportion of successful male/female applicants. When looking at one programme or one step in the selection processes at the time, there is hardly ever reason for suspicion of discrimination – the eventual differences are too small – and the few occasions when there yet seem to be a certain negative disproportion of successful female applicants, this is balanced by other occasions where there is an overrepresentation of successful women. However, as the referred investigation above indicates, in the long run, when looking more carefully into the differences in each step along the selection process, and when investigating these programmes' respective outcome all together, there is reason for suspecting a structural disfavor of female applicants in these large research programmes.
This study attempts to go deeper into two of these programmes, one individual and one regarding centers of excellence, and disclose some of the features of the selection procedures. The aim is to contribute with further understanding of the mechanisms in such selection processes, which lead to disproportionate disapproval of female applicants. The paper ends by proposing a hypothesis which may serve as a starting point for further empirical studies.

Previous results and data

The process of evaluating scientific work and applications for funding of research has indeed been investigated and debated for a long time. Doing so from a gender perspective is, however, a more recent undertaking. Following a debate in the 1970's about fairness in the peer review process, the concept of "old-boyism" was introduced, which suggests the existence of scientific networks, where those who are 'in' the networks promote each other (Mitroff & Chubin, 1979, Wessely, 1998, Woods & Wessely, 1999). The networks consist of senior scientists, who to a high degree were men in those days. Women were unseen and largely excluded.

It is mainly the peer review process that has been in focus. The fundamental question is whether the peer review process is 'fair'. When problematizing and elaborating on this, various angels can be taken. The gender perspective is one, and in this respect it has been shown that female applicants systematically have been disfavored in peer review processes for research funding (Wennerås & Wold, 1997). It has also been shown that the kind of rating scales used matter for the outcome, as do evaluation of applications in panels or individually without panel meetings (Langfeldt, 2001). Fine rating scales, in contrast to rough rating scales as well as majority decisions in the panels, lead to less heterogeneity among the approved applications, according to Langfeldt (2001). Well established research tend to get funded while innovative and risky projects tend not to. Langfeldt furthermore makes another observation: members of review panels tend not to care too much about the guidelines that they are supposed to follow in their evaluations, while reviewers who are mailed the applications and mail back their evaluations, pay much more attention to the guidelines. She concludes somewhat bluntly:

"The guidelines given to the panels had little effect on the criteria they emphasized, whereas mail reviewers were more consciously attempting to write reviews in accordance with the guidelines. Put more clearly, it seems that the panels do as they like, whereas mail reviewers do as they are told – or, at least, mail reviewers phrase their reviews more in accordance with the guidelines, to make sure they have influence on the ranking of proposals." (Langfeldt, 2001:835) She also notes that reviews made by administrative staff are less likely to promote unconventional and perhaps risky projects, while individual expert scientists may be more prone to do so now and then.

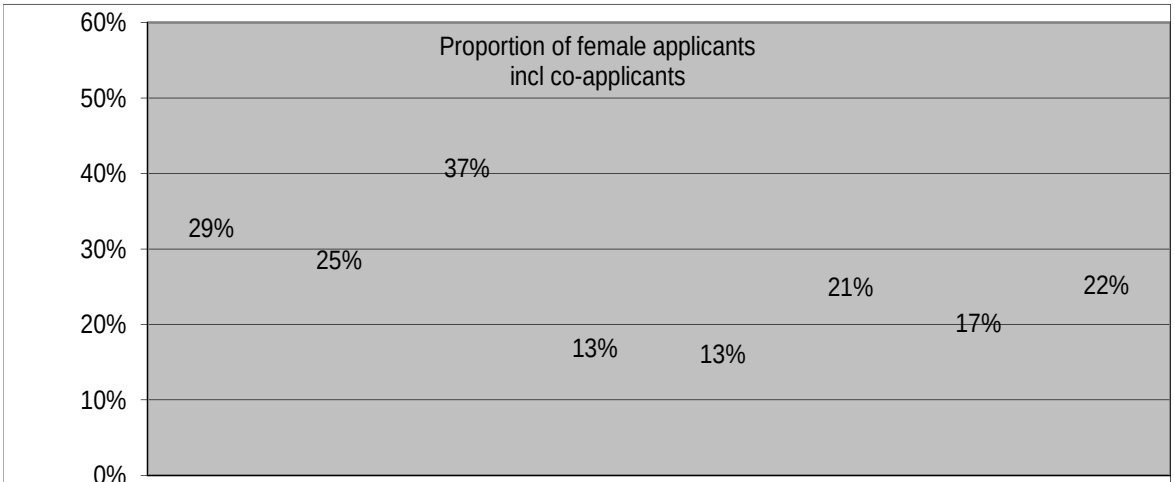
With particular respect to peer review for research grants, Langfeldt (2006) again observes how reviewers act contrary to their task in certain situations. Other criteria than pure scientific quality (whatever that is defined to be) are often neglected, resulting in a generally bad outcome for applications with another focus or orientation than the established main stream, like for instance interdisciplinary projects. She calls for a better understanding of how to select reviewers and setup peer review panels.

Langfeldt's findings have implications on the study at hand here, and as we will see, on how we can understand and interpret the results. We will come back to this after a presentation of the included data and an analysis of them.

Figure 1 presents the programmes that Nervik (2006) investigated. The first programme, Swedish Research Council – ind. projects, refers to the regular individual research grants that this main national funding body provide, in all areas of science and humanities.

Ingvar I and II (Individual Grant for the Advancement of Research Leaders) is a high profile individual grant programme set up by The Swedish Foundation for Strategic Research (SSF), with the purpose to select the future top layer of research leaders and provide them with a major grant, unlimited in its usage, relatively early in their careers. Ingvar I consisted of a SEK 10 million grant (app. €1.1 million) during six years, given to 20 successful candidates, while Ingvar II consisted of a SEK 6 million (app. €650.000) grant during four years, given to 18 successful candidates.

Figure 1. Proportion of female applicants including co-applicants (where applicable) to a sample of high profile funding programmes in Sweden.



	Swedish Research Council - ind.projects	Ingvar I	Ingvar II	SRC I	SRC II	Linné programme	Berzelii Centers	VINN Excellence center 2005
% Women	29%	25%	37%	13%	13%	21%	17%	22%
Women	1206	126	149	15	29	201	37	200
Tot	4115	504	401	112	230	940	222	923

Source: Nervik (2006)

The Swedish Foundation for Strategic Research is also the originator of the two following programmes in Figure 1: SRC I and II (Strategic Research Centers). This is a similarly high profile programme targeting the establishment and support of internationally leading centers of excellence in Sweden. The programmes have a mix of initiatives built into them, but in brief, the first round took place in 2002 and awarded SEK 400 million (app. €45 million) to six centers. An additional amount of SEK 500 million (app. €55 million) was awarded to other projects within the frame of the programme. The second round took place in 2005 and awarded SEK 800 million (app. €90 million) to 18 centers during a five year period.

The Linné programme is a programme directed towards the establishment and support of research groups or smaller centers of excellence, organized by The Swedish Research Council. The grants amount to approximately €1 million per year during up to a ten year period.

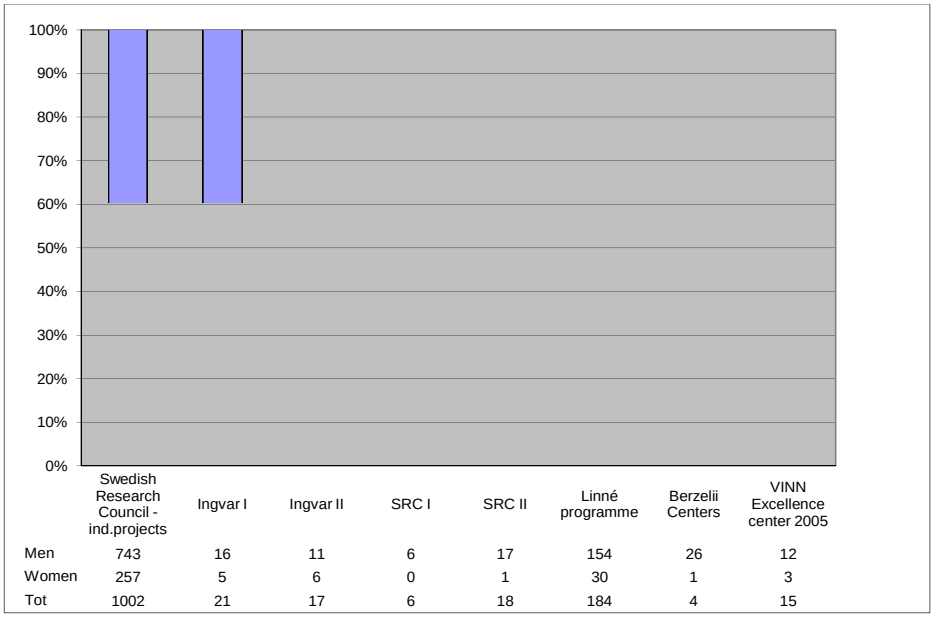
The Berzelii Centers are much larger, and the economic support is accordingly larger. Four centers have received some SEK 170 million per center (app. €20 million) over a ten year period. The Swedish Research Council co-finances the Berzelii Centers with the Swedish Governmental Agency for Innovation Systems (Vinnova).

The VINN Excellence Centers is a funding programme solely within the regime of Vinnova, and has had two rounds, one in 2003, giving support to four centers, and one in 2005, giving support to fifteen centers. Each center gets SEK 70 million (app. €8 million) during ten years. Substantial additional funding by the host institution is requested. Only the 2005 round is included in Figure 1.

Figure 1 shows the proportion of female applicants to these abovementioned programmes. The proportion fluctuates in the range of about one sixth to one third of the applicants. This proportion is fairly in accordance with the number of women in the system, on various professorship positions. Individual programmes have the highest share of female applicants, and applications regarding large centers of excellence have the lowest share. This pattern gives support to some of Nervik's (2006) conclusions, as presented in the introduction of this paper.

How have these women succeeded with their applications to these programmes? Figure 2 presents the outcome of the application rounds. While the success of women in the individual funding programmes corresponds well with the proportion of female applicants, this is not the case when we look at support to centers of excellence. Women are repeatedly underrepresented as receivers of the funding, compared to their proportion as applicants. The SRC I and II, Berzelii and VINN Excellence Centers taken together, support 47 centers where 7 women stand as main applicants. In the first round of the Linné Centers, 20 centers have been granted. The rector of the host university has to stand as main applicant, why it is interesting to look at the co-applicants instead in this programme. 16% of the co-applicants were women.

Figure 2. Comparison between granted applications by men and women.



	Swedish Research Council - ind.projects	Ingvar I	Ingvar II	SRC I	SRC II	Linné programme	Berzelii Centers	VINN Excellence center 2005
Men	743	16	11	6	17	154	26	12
Women	257	5	6	0	1	30	1	3
Tot	1002	21	17	6	18	184	4	15

Source: Nervik (2006)

Two of the funding programmes that Nervik (2006) has studied, have been further investigated: the Ingvar I programme and the SRC II programme. Both are programmes within the realm of the Swedish Foundation for Strategic Research, SSF, and aim at supporting the highest layer of people/centers within natural science, medicine and technology. Again, Ingvar I is an individual programme while SRC II is a programme for the support of centers of excellence. The respective selection processes have been carefully scrutinized and compared, with particular reference to the proportion of female applicants. Both programmes have been investigated separately and the results from those investigations have been published elsewhere (Melin & Danell, 2006; Schilling, 2007). Melin & Danell (2006) as well as Schilling (2007) contain information on gender issues, although this has not been in focus in any of those studies. This paper combines some of the findings from the two studies and uses their already available results and data in order to highlight gender related circumstances. Thus, this paper synthesizes from the two previous studies.

Findings

In both programmes, the Swedish Foundation for Strategic Research (SSF) has made great efforts to let the applications undergo a careful evaluation process. Review of the applications has been done in different ways and in several steps. Regarding Ingvar I, 504 individual short applications came in, so called pre-applications. These were reviewed in "strategy groups" with competence in selected main areas of science. The reviews made by

the strategy groups were then considered and judged upon by an evaluation committee at SSF. 115 applicants were selected for further evaluation by international field experts – a typical peer review procedure. Each application was evaluated by at least three experts, and each expert got at least four applications. Panels at SSF furthermore evaluated the strategic potential of the projects.

The expert reviews were again considered and judged upon by the evaluation committee at SSF. As a result, 43 applicants were invited to submit full applications. 40 applicants choose to do so. The same international experts were again used to evaluate the full applications. One or two acted as main evaluators, and another two-four experts graded the applications due to several criteria. The grading and evaluations were weighted together by the evaluation committee at SSF. All 40 remaining applicants were also called to a panel interview. The interview started by a presentation of the planned research by the applicant. After considering the evaluations and grading of the full applications, as well as the performance at the interview, 20 successful applicants were selected for the prestigious research grant.

The share of women along this process is presented in Table 1. From the original share to the first selection step there is no difference at all. The share decreases slightly in the next step, and again slightly in the final step.

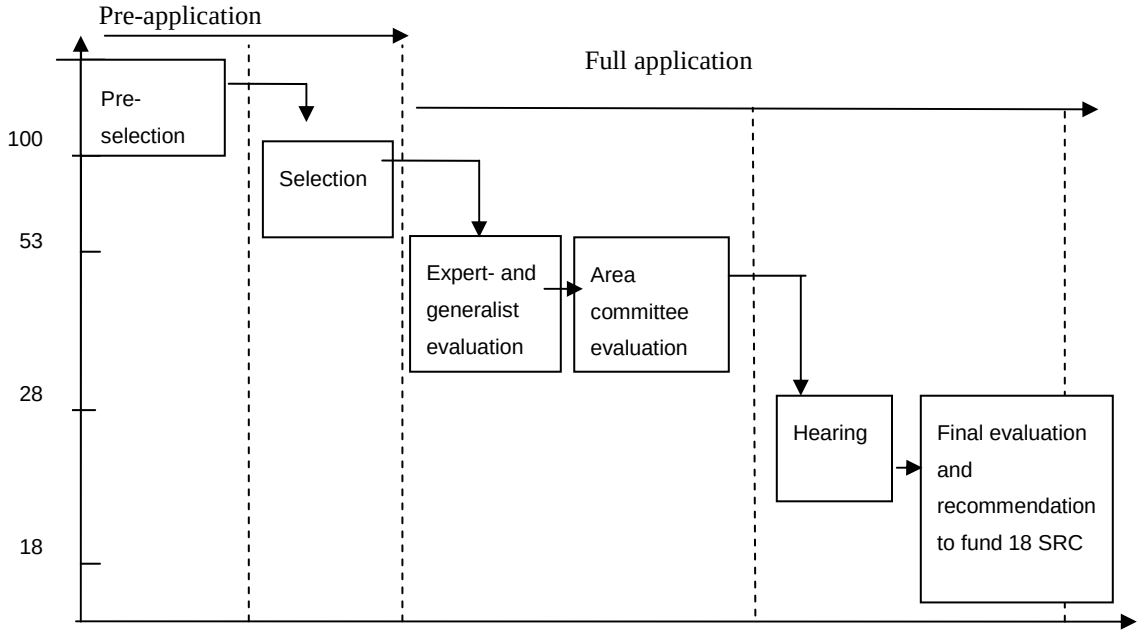
Table 1. Ingvar I: Share of men and women in the selection steps.

No. applications	**504**	**115**	**43**	**20**
Men	74 %	74 %	77 %	80 %
Women	26 %	26 %	23 %	20 %
Total	*100 %*	*100 %*	*100 %*	*100 %*

Source: SSF archives

The evaluation procedure regarding the Strategic Research Centers is similar but yet even more complex. Figure 3 summarizes the different steps.

Figure 3: Summary of the evaluation process regarding SRC II. X-axis: Evaluation process (time); Y-axis: No. of applications.



Source: Schilling (2007)

All applications were evaluated by the members of the evaluation committee at SSF as a first step. A rating scale was used, and in addition a "security grading" was made. This was a grading of how certain the evaluator felt that he or she had made a correct judgment. After this, the rating and grading were weighted together by the committee. 80 applications were selected, and in dialogue with the host universities, another 20 applications were selected as well, summing up to 100. These applications were evaluated by three panels, looking at

educational and competence aspects, research aspects and industrial- and societal aspects. The panels used numerical rating scales. Again, a security grading was made as well. 53 applications were selected from this step in the process as qualified for submitting full applications. Again the successful applications were selected in dialogue with the universities.

The submitted 53 full applications were discussed by the evaluations panels, from different perspectives. One panel used foreign experts, another made site visits to each university (15 universities were involved), and yet another used national experts and evaluated the strategic relevance in the applications. The panels were furthermore reorganized into field area or competence panels, with particular knowledge in different scientific areas. From this new perspective additional evaluations were made. The panels produced written reports and added together their numerical ratings. Many of the evaluation criteria were quantitative and easily comparable. Others, however, were much more uncertain and perhaps untraditional. They were of a qualitative character and thus had to be interpreted. This rather complex step in the evaluation process contains both the weighting together of measurable evaluations as well as judgments of various impressions and interpretations. Furthermore, this was done in some dialogue and negotiation with the host universities, which naturally had their own priorities and agendas.

28 applications were ranked by the evaluation committee and selected for the final step, a hearing. However, seven of these were considered absolutely certain for funding and needed not go through the hearing step, why 21 hearings were held. During the hearing, the planned center was presented to the panel and committee members. Scientific leadership and strategic relevance was included as important components. The hearing was graded numerically by the panel and committee members. Finally, the evaluation committee met once again in order to judge on the evaluations from the hearings. Careful discussions were held. In the end, 18 centers were decided to be given funding.

Table 2 presents the share of women along the steps in this complex evaluation process. The share of female main applicants falls from 12% to 6% during the process. Indeed, the number of female main applicants is low in absolute numbers, why it may be relevant to also include co-applicants. When doing so, the share of women remains steady through the process.

Table 2. SRC II: Share of men and women in the selection steps.

No. applications	**230**	**100**	**53**	**28**	**18**
Men	88 %	87 %	91 %	93 %	94 %
Women	12 %	13 %	9 %	7 %	6 %
Total	*100 %*	*100 %*	*100 %*	*100 %*	*100 %*
No. co-applicants					
Men	84 %	84 %	85 %	-*	85 %
Women	16 %	16 %	15 %	-	15 %
Total	*100 %*	*100 %*	*100 %*	-	*100 %*
All					
Men	85 %	85 %	86 %	-	86 %
Women	15 %	15 %	14 %	-	14 %
Total	*100 %*	*100 %*	*100 %*		*100 %*

Source: SSF archives; Schilling (2007). *Data is unavailable for this step.

Discussion

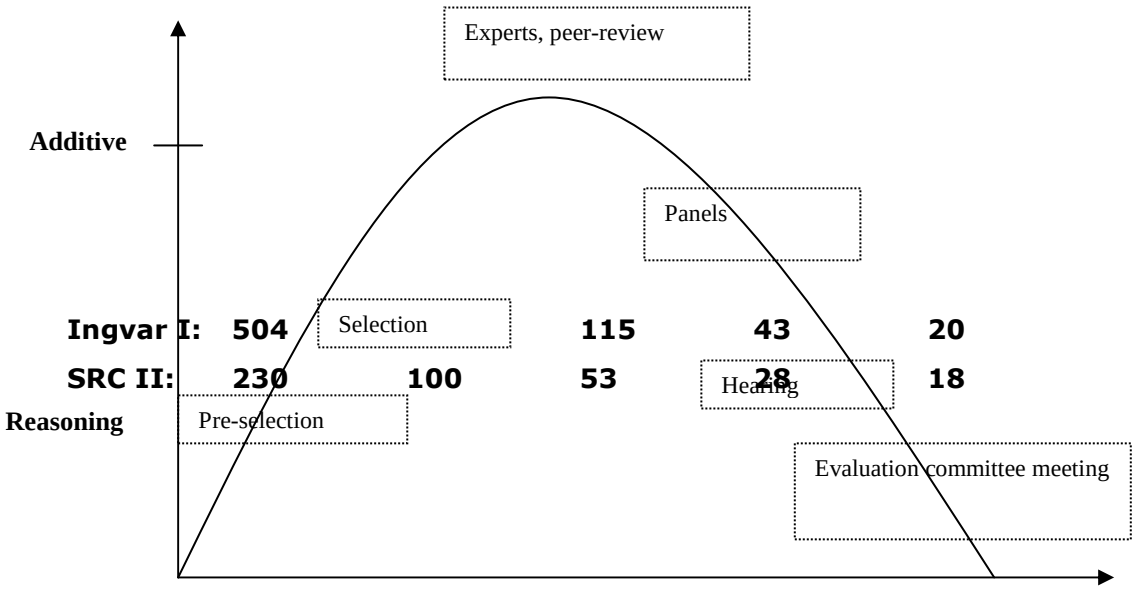
There are two separate studies which we build on in this paper. One concerns an individual funding programme (Melin & Danell, 2006), one concerns institutional support to centers of excellence, through large block grants (Schilling, 2007). Both programmes contain careful and advanced evaluation processes where the funding organization, The Swedish Foundation for Strategic Research, has tried to combine traditional peer review with other complimentary evaluation methods in order to select the best individuals/centers that meet the requested criteria of outstanding scientific performance and strategic potential.

Both evaluation processes have started with a pre-selection of the applications that have been submitted to the funding organization. This pre-selection has been made by one or several evaluation committees in-house. Regarding the programme for centers of excellence, this was followed by another selection, where experts in three different panels were used for a more careful review of the remaining applications. Site-visits and negotiations with the host universities were part of this work. Regarding the individual programme, traditional peer review by international experts was conducted after the pre-selection of submitted applications. An evaluation of the strategic potential was also made in this step.

Only after this was the remaining applicants invited to submit full applications. These were reviewed and evaluated more thoroughly, again in panels regarding the programme for centers of excellence, followed by a hearing for a selected share of those remaining applicants. Regarding the individual programme, foreign experts evaluated the full applications, and all applicants were called to a panel interview.

In his study, Shilling (2007) elaborates on the idea of established versus non-established evaluation criteria, and combines this with so called additive and reasoning evaluation methods. When applying his model to yet another programme than the one that it was originally developed for, the model has to be slightly re-made and simplified. As a consequence, it may be more general, thus showing patterns and features that may be of some relevance and value to other funding programmes and evaluation processes than the two ones that are presented in this study. The re-made model is presented in Figure 4.

Figure 4. Ingvar I and SRC II: Additive and reasoning evaluation criteria through the evaluation process. X-axis: Evaluation process (time) with figures of the number of applications in Ingvar I and SRC II. Re-make of a model proposed by Schilling (2007:40).



Through both evaluation processes, there has been a usage of both numerical ratings of various kinds, as well as other "intuitive" or "qualitative" evaluation methods. Schilling calls the first kind "additive", and the latter kind "reasoning" (Schilling, 2007:38). There is no clear boundary between them; it is a sliding scale where one may dominate over the other in any given step along the evaluation process. Most often, both are present to some extent in each step. The inverted u-shaped curve in Figure 4 is stressed in order to be more typical – it should perhaps not be quite this steep. There are differences in the process between the two investigated programmes and what apply well to one may not apply perfectly to the other. Below, a general interpretation of the evaluation process is described with reference to the proposed model in Figure 4.

The first step in the process, with pre-selection, is characterized by a high degree of reasoning evaluation methods, as a large number of pre-applications containing limited

amount of information are to be evaluated. Whether the evaluator got a "good sense" for the project is likely to be crucial for a positive outcome of the application. Rating scales were used, but it is reasonable to assume that sense and intuition played an important role in this step. Committee meetings were also held were the applications and the evaluations were discussed.

Later in the process, the applications were exposed to more careful expert evaluation. This type of evaluation is characterized by a rather high degree of additive methods, with strict rating scales and rather transparent evaluation criteria. This step is comparable to peer review evaluation.

While many funding programmes simply judge on the outcome of expert evaluations, yet other steps were introduced in these two programmes. Such additional steps may become increasingly frequent in funding programmes internationally as well. Through a hearing or an interview, the funding organization wanted to meet the applicants and listen to a presentation. This part of the evaluation process is clearly characterized by a high degree of reasoning methods. And at the end of the process, committee meetings were deciding on the final outcome.

The combination of additive and reasoning evaluation methods is assumed to be used in many evaluation processes. The different weight that additive or reasoning methods are given in different steps along the process, as suggested in Figure 4, is also of plausible correspondence to other similar evaluation processes. Given that the reader agrees with this, at least to some extent, a hypothesis can be suggested regarding women's expected success in such evaluation processes.

We have seen from the findings of Nervik (2006) that women are less successful than men when applying for support from individual funding programmes, and they are even less successful when applying for support from institutional programmes supporting centers of excellence. However, when scrutinizing the individual programme at hand here, we have seen that women perform on par with men in the expert evaluations but are less successful in later stages of the evaluation, where reasoning evaluation methods dominate. The key mechanism may be the usage of additive/reasoning evaluation methods, where it is the degree of reasoning methods that are critical for the female applicants. Such methods were much more at play in the evaluation process regarding centers of excellence than in the programme for individual grants. However, they were dominating at the end of the individual programme, and at this step the proportion of women fell.
A hypothesis can be formulated as follows:

> The higher the degree of reasoning evaluation methods in an evaluation process, the higher the risk for women to face exclusion which is not related to their competence or the quality of the proposed project. The higher the degree of additive evaluation methods, the lower the risk for women to face exclusion which is not related to their competence or the quality of the proposed project.

Our findings go hand in hand with results presented by Langfeldt (2001). She observed a less strict review behavior in panels compared to "mail reviews", i.e. reviews by experts who were not present in panels but made their evaluations where ever they had their office, and mailed them back to the funding organization. Several other observations of hers indicate that in review situations where any other than quantitative rating scales and transparent decision procedures were used, a certain bias or disobedience versus the evaluation guidelines was easily occurring. This may be interpreted along the terms suggested by our results: that it is the usage of reasoning evaluation methods that is problematic, with

reference to female applicants, but also with reference to other sides of applications and application rounds, like for instance interdisciplinarity, or geographical or linguistic fairness. The evaluation processes investigated here include a high degree of reasoning evaluation methods with significant opportunities for intuitive interpretations and negotiable outcome.
Would the hypothesis above be proven correct, some policy implications can be foreseen. Although the combination of additive and reasoning evaluation methods is valuable and promising, and perhaps necessary in general, it is important to be aware of the downsides of a high degree of reasoning methods. It is commonly assumed that network structures, differences in language between men and women, differences in the choice of research topic, among others, are components which disfavor women when applying for research funding. When, for instance, hearings and interviews are introduced in the evaluation process, such differences between men and women can be even more emphasized and thus be of increased disfavor to women. Both Shilling (2007) and Melin & Danell (2006) present findings where the very hearing or interview situation is commented upon by people who were present. The way a confident and convincing performance "on stage" mattered was questioned by some and it is possible that a project idea can seem better than it actually is due to a good appearance before the audience. There may indeed be situations when such presentation skills are of high value, but the meaning of the evaluation methods should be problematized and discussed, as well as the overall outcome of the increased usage of reasoning methods. The downsides associated with the concept of "old-boyism" has to do with hidden communication in closed networks, values and decision making structures, where other criteria than the openly viewed ones are at play. The same downsides are given more free space to influence the evaluation of applications when reasoning evaluation methods increase compared to additive evaluation methods.

It is not the reasoning evaluation methods as such that are improper, and they should not be abandoned. In fact a combination of reasoning and additive evaluation methods is healthy and shows acceptance to the fact that evaluation processes always include a reasoning component. However, reasoning methods ought to be used with care and with an awareness of the downsides that so easily arise, with potential negative results for other than established main stream projects, including applications by women.

Acknowledgement: The design of this study has been developed together with Dr. Peter Schilling. Valuable comments and advice on the manuscript draft have kindly been provided by Dr. Lars Geschwind.

REFERENCES

Langfeldt, L (2001): "The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome", *Social Studies of Science*, Vol 31, No. 6, pages 820-841.

Langfeldt, L (2006): "The policy challenges of peer review: managing bias, conflict of interest and interdisciplinary assessments", *Research Evaluation*, Vol 15, No. 1, pages 31-41.

Melin, G, Danell, R (2006): "The top eight percent: development of approved and rejected applicants for a prestigious grant in Sweden", *Science and Public Policy*, Vol 33, No. 10, pages 702-712.

Mitroff, I, Chubin, D (1979): "Peer Review at the NSF: A Dialectical Policy Analysis", *Social Studies of Science*, Vol 9, No. 2.

Nervik, B (2006): "Ansökningar till strategiska forskningscentra. En jämförande undersökning av könsfördelningen i ansökningarna till strategiska centra respektive traditionella projektbidrag", Power-Point presentation, [Online] Available at: http://www.idas.nu/filer/Strategiska_centra_okt-06.ppt, *IDAS, www.idas.nu*, July 2007.

Schilling, P (2007): "Att välja ut strategiska forskningscentra. En analys av Stiftelsen för strategisk forsknings bedömningsprocess", *SISTER Reports*, 2007:66.

Wennerås, C, Wold, A (1997): "Nepotism and sexism in peer-review", *Nature*, Vol 387, pages 341-343.

Wessely, S (1998): "Peer review of grant applications: what do we know?", *Lancet*, Vol 352 pages 301-305.

Wessely, S, Wood, F (1999): "Peer review of grant applications: a systematic review" in Godlee and Jefferson (eds.) (1999): *Peer Review in Health Sciences*, London: BMJ Publishing Group.