



Vetenskapsrådet

KOLLEGIAL BEDÖMNING AV VETENSKAPLIG KVALITET – EN FORSKNINGSOVERSIKT



KOLLEGIAL BEDÖMNING AV VETENSKAPLIG KVALITET – EN FORSKNINGSÖVERSIKT

Rapporten är initierad av Vetenskapsrådets expertgrupp för genus.

Lena Gemzöe

**KOLLEGIAL BEDÖMNING AV VETENSKAPLIG KVALITET
– EN FORSKNINGSOVERSIKT**

Rapporten kan beställas på www.vr.se

VETENSKAPSRÅDET
Box 1035
101 38 Stockholm

© Vetenskapsrådet
ISSN 1651-7350
ISBN 978-91-7307-169-7

Grafisk Form: Erik Hagbard Couchér, Vetenskapsrådet
Produktion: Matador Kommunikation
Tryck: CM-Gruppen AB, Bromma 2010



FÖRORD

Vetenskapsrådets styrelse har ett rådgivande organ i frågor som rör genusforskning, sedan 2009 benämnd Expertgrupp för genus. Expertgruppen har till uppgift att identifiera frågeställningar av strategisk betydelse och rekommendera insatser som rör genus i forskningssammanhang. Gruppen ska också lämna förslag till beslut om finansiering inom genusområdet.

Expertgruppen har tidigare initierat uppföljningar av Vetenskapsrådets bidragsgivning till genusforskning inom olika ämnesområden. I en serie rapporter har man analyserat innehållet och inriktningen av den genusforskning som söktes pengar för hos Vetenskapsrådet åren 2004 och 2006 inom områdena humaniora samhällsvetenskap, utbildningsvetenskap och medicin, samt vilken framgång dessa ansökningar haft i jämförelse med andra.

Den föreliggande rapporten fokuserar på forskning om bedömningsprocesser som använder sig av så kallad "peer review", eller kollegial bedömning. Docent Lena Gemzöe vid Centrum för genusstudier vid Stockholm universitet har på uppdrag av Vetenskapsrådet gjort en översikt över aktuell forskning på detta tema. Rapporten tar upp forskning om kollegial bedömning i allmänhet, exempel på vad detta innebär för genusforskning och annan forskning av tvärvetenskaplig karaktär samt jämställdhetsaspekter på urvalsprocesserna.

Anna Wahl

Ordförande

Expertgrupp för genus vid Vetenskapsrådet



INNEHÅLL

SAMMANFATTNING	6
SUMMARY IN ENGLISH	9
BAKGRUND	12
GRANSKNING AV GRANSKNINGSSYSTEM SOM FORSKNINGSSOMRÅDE	15
Old boy-ism och normen om opartiskhet	16
Opartiskhet och kön	17
Är kvalitet en objektiv egenskap?	20
KVALITETSBEDÖMNINGAR I PRAKTIKEN	22
Travis och Collins om kognitiv partikularism	22
Langfeldt om beslutsprocesser i expertpaneler	26
Lamont om kvalitetsbegreppets operationalisering	29
Originalitet	31
Relevans	32
Informella kriterier	33
Pragmatisk rättvisa	36
KOLLEGIAL BEDÖMNING AV TIDSKRIFTSARTIKLAR	39
BEDÖMNING AV TVÄRVETENSKAP	43
SLUTSATSER OCH DISKUSSION	45
Kvalitet konstituerad i bedömning	45
Samband mellan kognitiv och social partiskhet	45
Diversifiering istället för standardisering	46
Sammansättning av bedömaregrupp och beslutsprocess	47
Bibliometri och kollegial bedömning	49
Vad är en god bedömning av genusforskning?	49
REFERENSER	52



SAMMANFATTNING

Föreliggande rapport är en översikt över forskning om bedömning av vetenskaplig kvalitet. Dess syfte är att sammanfatta och ringa in de viktigaste insikterna från forskning om bedömningsprocesser, främst med relevans för genus och jämställdhet. I forskningen om bedömningsprocesser ges frågan om jämställdhet – gynnande eller missgynnande av bedömda forskare på grund av kön – ett relativt stort utrymme, medan bedömning av genusforskning behandlas mer sparsamt. Dock är genusforskning inte separerad från övrig forskning och den bedöms inom samma system och med samma kriterier som övrig forskning. Översikten syftar därför till att få en överblick över forskning om bedömning av vetenskaplig kvalitet i sin helhet.

Forskningen om peer review, som jag här med en svensk term kallar kollegial bedömning, inleddes med en första fas som koncentrerade sig på att kartlägga vilka risker det finns för att bedömningssystem uppvisar olika former av partiskhet. Partiskhet förstås som att personliga, sociala eller institutionella egenskaper hos den sökande påverkat bedömningen så att den inte utförts i överensstämmelse med den gällande definitionen av vetenskaplig kvalitet. De här granskningarna av bedömningsprocesser antar i överensstämmelse med hur bedömningssystemen är uppbyggda att partiskhet kan förebyggas genom att bedömarna följer ett strikt vetenskapligt etos där man bortser från de bedömdas personliga, sociala eller institutionella egenskaper. I forskningen om eventuell partiskhet ägnas frågan om partiskhet på grund av kön ett relativt stort utrymme, bland annat som en följd av en ökad betoning på jämställdhet i forskarvärlden. Forskningen om partiskhet på grund av kön domineras av statistiska sambandsanalyser som undersöker eventuella samband mellan kön och beviljningsgrad. Sådana samband har kunnat beläggas, men forskning med denna inriktning kan inte närmare klargöra hur sådana samband uppstår.

Forskningen om kollegial bedömning har i en andra fas kommit att rikta uppmärksamheten mot begreppet vetenskaplig kvalitet och dess operationalisering i praktiken i bedömningsprocesserna. Istället för personlig, social eller institutionell partiskhet talar forskarna här om kognitiv partiskhet. I empiriska undersökningar har kvalitetsbegreppet avtäckts sina betydelser och man har påvisat hur bedömare operationaliserar kvalitetsbegreppet på olika sätt. I en studie visas hur bedömare använder sig av olika epistemologiska stilar, som hänger samman bland annat med disciplintillhörighet. Bedömare använder även informella kriterier och informella regler för att kunna fatta gemensamma beslut som uppfattas som legitima och rättvisa. Diversifieringen av kvalitetsbegreppet, liksom tendensen att definiera

kvalitet i överensstämmelse med den egna forskningen ("homofili"), har visat sig karakterisera såväl bedömning av forskningsansökningar som bedömning av manus för publicering i tidskrifter. Forskare understryker även att bedömning inte kan ses som en rent rationell aktivitet, utan att känslor är en integrerad del av bedömning. I de studier som den här rapporten behandlar fann forskarna inga skillnader mellan naturvetenskaperna och humanvetenskaperna i de här avseendena. Vidare påpekar forskarna att insikterna om hur bedömningsprocesser fungerar i praktiken (existensen av diversifierade kvalitetsbegrepp, "homofili"-tendensen och kognitiva gränsdragningar inom vetenskaperna) kan belysa de särskilda svårigheter som är förknippade med bedömning av tvärvetenskaplig forskning, ett område som ännu är relativt outforskat.

Insikterna från den andra fasen av kvalitativ forskning om bedömning av vetenskaplig kvalitet utmanar det rådande synsättet på kollegial bedömning på flera sätt. Det dominerande synsättet på bedömning i forskningspolitik och hos forskningsfinansiärer är snarare i samklang med den första fasen i forskning om kollegial bedömning. Idealet är att ett vetenskapligt etos råder som endast ser till den objektivt bestämbara egenskapen vetenskaplig kvalitet. Idealet antas vara genomförbart genom att standardisera skalor för bedömning och övervaka eventuella jävsituationer för att undvika att de sökandes personliga, sociala eller institutionella positioner påverkar bedömningen av kvalitet. Den kvalitativa forskning som diskuteras här pekar emellertid på andra kritiska punkter i kollegial bedömning och på möjliga vägar till att utveckla bedömningssystemen.

Samband mellan kognitiv och social partiskhet

Vid sidan av eventuell social partiskhet är kognitiv partiskhet en viktig faktor att ta hänsyn till vid kollegial bedömning. Analyser av sambanden mellan kognitiva gränsdragningar inom ett vetenskapligt område och den sociala sammansättningen av fältet kan fördjupa förståelsen av social partiskhet i kollegial bedömning.

Diversifiering istället för standardisering

En diversifiering i synen på vetenskaplig kvalitet bör införas i bedömningssystemen, det vill säga en explicit formulering av de diversifierade kvalitetsbegrepp som redan existerar. Ett statistiskt, standardiserat kvalitetsbegrepp kan medföra en stagnation av vetenskaplig utveckling.

Sammansättning av bedömargrupp och beslutsprocess

Sammansättningen av bedömargrupperna är den viktigaste faktorn för utgången av bedömningsprocessen. En god kännedom om fältet som ska bedömas och en större interaktion mellan bedömare och forskningsråd är önskvärd.

Bibliometri och kollegial bedömning

De kvalitativa studierna av kollegial bedömning för tidskrifter kan belysa bibliometrimåttet på kvalitet på andra sätt än den tekniskt inriktade diskussionen om bibliometri. I den mån nationella bedömningssystem som det svenska ökar betoningen på bibliometrimått och därmed beroendet av internationella, främst anglosaxiska, tidskriftsredaktioners och bedömares kvalitetsdefinitioner bör effekterna för utveckling av forskningen analyseras.

Bedömning av genusforskning

Insikterna från den kvalitativa forskningen om bedömning angående existensen av diversifierade kvalitetsbegrepp och de särskilda svårigheterna att bedöma tvärvetenskap är i hög grad relevanta för bedömning av genusforskning. Ett kännetecknande drag för genusforskningen är att den existerar både som ett teoretiskt perspektiv inom en mångfald av discipliner, och som ett eget fält. När den bedöms med inomdisciplinära kvalitetskriterier blir utfallet beroende av hur väl de inomdisciplinära kvalitetskriterierna lämpar sig för att bedöma genusforskning i området ifråga, något som kan variera avsevärt mellan olika fält. Svårigheter att bedöma genusforskning med en relevant kvalitetsdefinition kan främst finnas i områden där genusforskarna är få eller där de arbetar inom epistemologiska stilar som ligger långt från de dominerande i vetenskapsområdet ifråga. Vidare måste bedömningen av genusforskning beakta hur den inverkar på utvecklingen av genusforskningen som självständigt vetenskapsområde, snarare än som ett perspektiv inom olika discipliner. När syftet är att se till fältets utveckling i sin helhet bör kvalitets- och relevanskriterier vara definierade i genusvetenskapliga termer gemensamma för hela fältet, och förhålla sig till en internationellt definierad forskningsfront.



SUMMARY IN ENGLISH

This report presents an overview of the research on evaluating scientific quality. It aims to summarise and highlight the most important research findings on review processes, primarily in relation to gender research and gender equality. The research on review processes pays considerable attention to the question of equal opportunity for women and men – investigating gender-related favouring or disfavouring of evaluated researchers – while the evaluation of gender research is dealt with more sparsely. However, gender research is not separated from other research and is reviewed using the same system and same criteria as that used for other research. Hence, this report aims to provide an overview of the research addressing evaluation of scientific quality generally.

In an initial phase, the research on *peer review* concentrated on identifying risks for various types of bias in research evaluation.. Bias is said to exist when the applicant's personal, social, or institutional characteristics influence the review to the extent that it does not comply with the applicable definition of scientific quality. These studies of review processes assume, in agreement with how review systems are constructed, that bias can be prevented if the reviewers strictly adhere to a scientific ethos where they disregard the applicant's personal, social, or institutional characteristics. Research on potential bias gives relatively large space to gender-related bias, e.g. as a result of the increased emphasis on gender equality in the research community. Research on gender-related bias is dominated by statistical meta-analyses that examine the potential connection between gender and success rates. Although research has been able to substantiate such connections, it has been unable to clarify how they arise.

The second phase of the research on peer review directs attention to the concept of scientific quality and how evaluators operationalise it in practice. Instead of personal, social, or institutional bias, researchers here discuss cognitive bias. Empirical studies have unveiled the meaning behind the quality concept and show how reviewers operationalise the quality concept in differing ways. One study shows how reviewers use different epistemological styles, partly related to discipline affiliation. Reviewers also use informal criteria and informal rules to arrive at joint decisions perceived to be legitimate and fair. Diversification of the quality concept, as well as the tendency to define quality in a way that agrees with one's own research (homophily), has been shown to characterise both the evaluation of research grant applications and the review of manuscripts for journal publication. Researchers also

emphasise that evaluation cannot be considered a purely rational activity, but that emotions are an integral part of evaluation. In the studies addressed by this report, researchers found no differences between the natural sciences and the humanities in this respect. Furthermore, the researchers note that knowledge concerning how evaluation processes work in practice (existence of diversified quality concepts, tendency for “homophily”, and cognitive boundaries within the sciences) can illuminate the particular difficulties associated with reviewing multidisciplinary research, an area that remains relatively unresearched.

Insights from the second phase of qualitative research on reviewing scientific quality challenges the prevailing perspective on peer review in several ways. The dominant perspective towards review in research policy, and among funding institutions, corresponds more to the initial phase of research on peer review. The ideal is that a scientific ethos prevails, which looks only at the objectively determinable characteristic, scientific quality. The ideal is assumed to be achievable by standardising scales for reviewing and monitoring potential conflict-of-interest situations to avoid having the applicant’s personal, social, or institutional position affect the judgment of quality. However, the qualitative research discussed here points to other critical items in peer review and to possible ways of developing peer evaluation.

Connections between cognitive and social bias

Next to potential social bias, cognitive bias is an important factor to consider in peer review. Analyses of the connection between cognitive boundaries within a scientific field and the social composition of the field can deepen the understanding of social bias in peer review.

Diversifying instead of standardising

Diversification of the concept of scientific quality should be introduced into research evaluation, i.e. explicit formulation of the diversifying quality concepts that already exist. A static, standardised concept of quality can lead to stagnation in the advancement of science.

Evaluation panel composition and decision-making

The composition of evaluation panels is the most important factor affecting the outcome of evaluation processes. Appropriate levels of knowledge about the fields to be evaluated and greater interaction between evaluators and the research councils are desirable.

Bibliometrics and peer review

The qualitative studies of peer review for journals can illuminate bibliometric measures of quality in ways other than the technically oriented discussion of bibliometrics. Given the increasing emphasis on bibliometric measures in national evaluation systems, like the one in Sweden, – and hence the increasing dependence on quality definitions of international (primarily Anglo-Saxon) journal editors and referees – the effects of this development on the advancement of research must be analysed.

Evaluation of gender research

Findings from the qualitative research on review, as regards the existence of diversified quality concepts and the particular difficulties in reviewing multidisciplinary science, is highly relevant in evaluating gender research. A characteristic of gender research is that it exists both as a theoretical perspective in many disciplines and as its own field. When it is evaluated with intradisciplinary quality criteria, the outcome depends on how well the intradisciplinary quality criteria are suited to evaluate gender research in the area in question, something that can vary substantially between fields. The problem of evaluating gender research with a relevant quality definition is most pronounced in the fields where gender researchers are few, or where they work within epistemological styles far from the mainstream in the scientific field in question. Furthermore, the evaluation of gender research must consider how it affects the advancement of gender research as an independent scientific field, rather than as a perspective within different disciplines. When the aim is to promote advancement of the field as a whole, the quality and relevance criteria should be defined in gender-studies terms common to the entire field, and should relate to an internationally defined research front.



BAKGRUND

Syftet med föreliggande rapport, som tillkommit på initiativ av Vetenskapsrådets expertgrupp för genus, är att ge en översikt över forskning om bedömning av vetenskaplig kvalitet. Ett behov av att definiera och diskutera begreppet vetenskaplig kvalitet har påtalats av ett flertal aktörer inom forskarsamfundet, i Sverige och internationellt. Vetenskapsrådets expertgrupp för genus har under de senaste åren arbetat aktivt med frågor om bedömning. För att följa upp beredningen av ansökningar inom genusforskning har expertgruppen undersökt utfallet av beredningsprocesserna inom humaniora och samhällsvetenskap, medicin och utbildningsvetenskap med avseende på genusforskning.¹ Erfarenheterna av att främja genusforskning och hantera bedömningar av ansökningar inom området har även presenterats på seminarier med representanter för andra länders forskningsråd² och för EU:s direktorat för forskning³. De svenska erfarenheterna har väckt intresse och utgör ett viktigt bidrag till den diskussion om genus- och jämställdhetsaspekter av forskningspolitiska satsningar på excellens som förs på europeisk nivå. För att ytterligare vidga diskussionen om bedömningsprocesser har expertgruppen för genus beslutat att i ett första steg skaffa en överblick över svensk och internationell forskning inom området.

Avgränsningarna och inriktningen i föreliggande översikt över forskning om bedömningsprocesser har gjorts utifrån expertgruppens för genus syften, men också utifrån hur forskningsområdet ser ut. Expertgruppens intresse för hur bedömningsprocesser fungerar gäller självfallet i första hand med avseende på genusforskning. Dock är inte genusforskning separerad från annan forskning och den bedöms inom samma system och med samma kriterier som övrig forskning. Den här översikten söker därför få ett grepp om forskning om peer review-bedömningar i sin helhet. Angreppssättet motsvaras av att det inte finns någon omfattande forskning som har fokus enbart på bedömningar av genusforskning. I en viss del av litteraturen nämns emellertid genusforskning som exempel och i andra fall kan forskningens resultat tolkas så att det blir relevant för frågor om genusforskning.

Forskningen om bedömningsprocesser har alltså inte särskilt beaktat genusforskning, medan däremot jämställdhet ägnats ett omfattande intresse.

¹ Vetenskapsrådet (2005, 2007a, 2007b).

² Seminarium vid Vetenskapsrådet "Att bedöma genus – en läroprocess" 3 oktober 2007.

³ Seminarium i Bryssel "Collaboration in Europe: New Arenas for Gendering Research" 9 november 2006.

Den här översiktens huvudintresse är att belysa aspekter av bedömningsprocesser med relevans för genusforskning, men eftersom de frågorna är sammanflätade med jämställdhetsfrågor på olika sätt diskuterar jag inledningsvis hur forskningsområdet förhåller sig till jämställdhet. Något bör sägas om sambanden mellan genusforskning och jämställdhet i forskarvärlden i det här sammanhanget. I Sverige idag betraktas de två områdena som separata och organisatoriskt sett är de också skilda åt: genusforskning bedrivs vid centra för genusstudier och vid enskilda institutioner medan jämställdhetsarbete i högskolan leds av jämställdhetshandläggare och grupper på samma sätt som vid andra arbetsplatser. Forskningsfinansierande organ, vars beslut får konsekvenser för jämställdheten i forskarvärlden, är underställda jämställdhetslagar i likhet med andra statliga myndigheter. Historiskt sett finns det emellertid vissa självklara samband. Genusforskningen tog form vid universiteten som en följd av att ett större antal kvinnor påbörjade högre studier och valde forskarbanan och så småningom började utveckla genuskritiska perspektiv på vetenskapen. Strävandena efter jämställdhet i forskning och vid högskolor i Sverige har också en historisk koppling till genusforskningen. En stor del av underlagen för regeringens jämställdhetspolicy för universiteten alltsedan 1975 har utformats av genusforskare.⁴ En sådan roll för genusforskningen i ett tidigt skede följde av genusforskningens karaktär av att vara det kunskapsområde som belyser och analyserar förhållanden mellan könen i samhället: i politik, kultur och arbetsliv, och därmed även inom högskolan. Genusforskningens framväxt inom forskarsamfundet har givetvis även bidragit till en ökad kvinnlig närvaro i vissa områden i högskolan.

De här övergripande sambanden mellan områdena innebär inte, som nämndes ovan, att verksamhetsområdena idag är parallella eller har en liknande utveckling. Jämställdhetsarbete i högskola och forskning kan vara önskvärt och bedrivs utan att genusforskning beaktas. Genusforskningen som fält har också, efter den första tidens närmre koppling till jämställdhet, haft ett behov av att distansera sig från jämställdhetsfrågorna. För genusforskningens vetenskapliga utveckling var det viktigt att påvisa dess betydelse just som vetenskap: det var genom den nydanande forskningen och den kritiska potentialen som genusforskningen kom att få gradvis mer fotfäste i högskolan. Att se genusforskningen endast som ett medel till jämställdhet, en chans för kvinnor att ta plats i forskarvärlden, var att begränsa dess betydelse. En ensidig koppling mellan kvinnor och genusforskning

⁴ Med UKÄ 1975:10, som utarbetades av genusforskare som Rita Liljeström, Margot Bengtsson, Christina Jonung och Karin Westman Berg inleddes ett aktivt jämställdhetsarbete inom högskolan på genusforskningens grund. Det fick sin fortsättning med SOU 1983:4, SOU 1995:110, genomförd av Ebba Witt Brattström, som bland annat ledde till inrättandet av det nationella genussekretariatet.

blundar också för det faktum att kvinnliga forskare inte nödvändigtvis sysslar med genusforskning, och att även manliga forskare deltar i genusforskningens vetenskapliga projekt.

Genusforskningens förhållande till jämställdhet i högskolan kan emellertid vara i färd med att omdefinieras, och få en liknande roll som då området formerades. Forskning om hur genus formar högskola och forskning efterfrågas i ökad grad på europeisk och internationell nivå. I exempelvis EU:s forskningsprogram ses jämställdhetsfrågor och genusforskning som nära sammankopplade. Att beakta "the gender dimension" i forskningsfinansiering betyder här ofta både att sträva efter jämställdhet och att se positivt på forskning med genusperspektiv.

Under arbetet med den här översikten har det också visat sig att ett fokus på frågor om jämställdhet och genus i en diskussion om bedömningsprocesser är fruktbart av det skälet att dessa frågor artikulerar några av de mest brännande frågorna i forskningen. Översikten kan därför vara värdefull även i sammanhang där huvudfrågan inte har att göra med beredning av genusforskning. Den kvalitativa forskning av bedömningsprocesser som vuxit fram under de senaste åren, i vilken begreppet vetenskaplig kvalitet står i centrum för analysen, ägnas ett särskilt intresse i den här rapporten.

Peer review, som jag här med en svensk term kommer att kalla kollegial bedömning, betecknar oftast ett skriftligt omdöme om kvaliteten i en enskild forskares arbete utfört av en annan forskare verksam inom fältet. Kollegial bedömning i den här formen används i bedömningar av ansökningar, sakkunnigutlåtande inför tillsättning av tjänster och bedömning av manus för publicering i tidskrifter. Kollegial bedömning på en sådan mikronivå kan skiljas från kollegial bedömning på makronivå, där hela forskningsområden, institutioner eller program utvärderas (Langfeldt 2002). Kollegial bedömning på mikronivå kan vara en del av sådana utvärderingar på makronivå. I de sammanhang där kollegial bedömning används organiseras bedömningsprocessen och beslutsfattandet på olika sätt, något som belyses i forskningen inom området. I den här översikten diskuterar jag främst forskning om bedömning av forskningsansökningar, men jag tar också upp exempel från andra områden som behandlar samma slag av problematik.

Inom Vetenskapsrådet sysslar flera personer med frågan om bedömningsprocesser. Inför arbetet med rapporten har jag sökt orientera mig om det som redan gjorts och diskuterat med berörda personer för att undvika onödiga dubbleringar. Magnus Gunnarsson har haft vänligheten att ge mig tillgång till ett större antal publikationer som han arbetat med och hans kommentarer till dessa, vilka underlättat förarbetet betydligt.



GRANSKNING AV GRANSKNINGSSYSTEM SOM FORSKNINGSOMRÅDE

Tills relativt nyligen har forskning om kollegial bedömning varit en helig ko, skriver Campanario (1998a) i en översikt över forskning om kollegial bedömning för tidskrifter. Några få heterodoxa forskare har tagit sig an uppgiften att granska granskningssystemet, men mycket av diskussionen har förts av bedömarna själva, exempelvis i vetenskapliga tidskrifters ledare och debattforum, ofta baserat på personliga erfarenheter snarare än empiriskt grundade studier. I likhet med Campanario påtalar andra forskare det faktum att det på grund av att området fortfarande är understuderat finns stora luckor i vår kunskap om hur bedömningsprocesser inom vetenskapen fungerar. De studier som utförts är dessutom ofta disciplin-specifika och utförda med olika metoder, varför resultaten är svåra att jämföra och har en låg grad av generaliserbarhet (Marsh et. al 2009).

Det är intressant att notera att forskningen om kollegial bedömning genererats av dess potentiella användbarhet – man började forska i området för att kunna försvara, kritisera och/eller förbättra bedömningssystemen, oftast för att undersöka i vilken mån systemen lever upp till kraven på objektivitet och opartiskhet (Lamont och Mallard 2005). Få studier har initierats med det rent kunskapssökande syftet att undersöka hur bedömningsprocesser bidrar till att forma vetenskapen och etablera ny kunskap som legitim. Sedan området väl öppnats för forskning framträder det emellertid tydligt, som Campanario (1998a) påpekar, att eftersom kollegial bedömning utgör en sådan central del av vetenskapen, erbjuder forskningen om bedömningsprocesser möjligheter att analysera vilka interna dynamiker som styr vetenskaplig utveckling och framväxten av ny kunskap. Sedan 1980-talet och i högre grad under senare år har forskningsområdet vuxit och frågor om sambanden mellan bedömningsprocesser och vetenskaplig utveckling har fått en mer framträdande plats, vilket jag återkommer till nedan.

Utvecklingen av vetenskapssamfundet och forskningen har också lett till att forskning om bedömningsprocesser tilldragit sig ett större intresse. Bedömning av forskningens kvalitet förekommer i ökande grad i myndigheters och finansierares utvärderingar och granskningar av forskningsprogram, institutioner och discipliner. En sådan byråkratisering och ”granskningskultur” i förhållande till vetenskaplig verksamhet gör att bedömningsprocesser blir intressanta att

studera som viktiga brytpunkter mellan vetenskap och politik (Langfeldt 2002, Strathern 2000). Den ökade användningen av kollegial bedömning och den ökade användningen av granskningar av forskning tillsammans med en relativt utbredd skepticism bland forskarna inom området om möjligheten att uppnå opartiskhet gör att behovet av att förstå bedömningsprocesserna mer på djupet har ökat (Langfeldt 2002, Lamont och Mallard 2005).

Old boy-ism och normen om opartiskhet

De studier som räknas till de första inom området genomfördes i USA på 1970-talet (e.g Merton 1973, Cole, Rubin och Cole 1978) och var inriktade på bedömningar av forskningsansökningar. Behovet av forskningen uppstod efter att politiska krav ställts på större transparens och tillförlitlighet i användningen av offentliga medel i USA vid den här tiden. I den amerikanska allmänna debatten, som var kritisk till hur offentliga medel användes inom forskningen, myntades uttrycket "old-boyism". Det syftade på att systemet för distribuering av forskningsanslag kontrollerades av en grupp äldre, etablerade manliga akademiker som var förbundna med varandra i kollegiala och/eller vänskapsnätverk. Det riktades misstankar mot kollegial bedömning som gick ut på att bedömare istället för att främja vetenskaplig kvalitet gynnade andra "old boys" eller forskare knutna till dem (Lamont och Mallard 2005).

Syftet med de tidiga större empiriska studierna var att undersöka om det fanns fog för old-boyism-kritiken som formulerats i den allmänna debatten. Studierna var därför inriktade på att undersöka i vilken grad kollegiala bedömningsprocesser levde upp till normen om opartiskhet. En sådan norm ligger till grund för antagandet att systemet gynnar endast den forskning som har den högsta kvaliteten, utan att ta hänsyn till icke-vetenskapliga faktorer som den sökandes sociala karakteristika, politiska ståndpunkter, tillhörighet till institutionella- eller vänskapsnätverk etc. Bedömnings-systemet måste stå över sådan "subjektivism" för att främja en öppen, fri vetenskaplig debatt och garantera att forskning med den högsta kvaliteten får finansiering (Merton 1973).

Frågan om i vilken mån kollegial bedömning kan sägas vara partisk har fortsatt varit en huvudfråga i forskningsområdet. En rad studier påvisar olika slag av "old-boyism" eller partiskhet i systemen (exempelvis Cole, Rubin och Cole 1978, Turney 1990, Chubin och Hackett 1990, Campanario 1998ab, Bornman 2007).

Opartiskhet och kön

Även om kön inte var en av de sociala kategorier som behandlades i de tidiga studierna av eventuell partiskhet i bedömningssystemen, så är könskategorin implicit närvarande i uttrycket "old-boyism", där det är just manliga nätverk som antas utöva makt på ett partiskt sätt. I takt med att jämställdhetssträvanden i forskarvärlden stärkts under samma period som forskningen om bedömningsprocesser vuxit fram har kön kommit att uppmärksammas som en sådan social karakteristik, som vid sidan av andra, riskerar att bidra till partiskhet och missgynna kvinnor i systemet. Man kan säga att den ökade betoningen på värdet av jämställdhet i forskarvärlden har givit bränsle åt den forskning som intresserar sig för eventuell partiskhet i bedömningssystemen.

Ett flertal studier av olika bedömningsprocesser påvisar partiskhet med avseende på kön. Ett axplock bland de oftast citerade studierna visar på komplexiteten i hur partiskhet med avseende på kön yttrar sig. Fürst (1988) genomförde en studie av sakkunniggranskningar inför professorstillsättningar i Norge och fann att samma kriterier på vetenskaplig kvalitet gavs olika innebörd vid bedömningen av de kvinnliga respektive manliga sökande till nackdel för kvinnorna. De kvinnliga sökandes produktion betecknades som antingen "smal" och "ensidig" eller "utspridd" över flera områden, medan männens verk hade "djup", "tematisk koherens" eller "bredd".

Studier av publikationer och citeringar respektive bedömningar av forskningsansökningar tycks peka på mekanismer som skapar glastak för kvinnor i delar av forskarvärlden. I en studie av antropologi visar Lutz (1990) att kön inte ger någon skillnad vad gäller antalet papers på konferenser eller att publicera artiklar i vetenskapliga tidskrifter i stort. Däremot påverkar det sannolikheten att publicera i de högst rankade tidskrifterna, vilka också är de som ger mest utslag i citeringar. På 1980-talet var 45 procent av forskarna inom området som presenterade paper på konferenser kvinnor, 30 procent av artiklarna publicerade i toptidskrifterna var skrivna av kvinnor, medan endast 20 procent av citeringarna gick till kvinnor. Lutz (1990) påvisar också att män citerade kvinnor hälften så ofta som kvinnor citerade kvinnor. Lutz påpekar att publicerings- och citeringsstatistik uppvisar ett mönster som återfinns i samhället i stort. Språkvetenskapliga studier har visat att läsares bedömning av värdet och betydelsen av en viss text (eller ett yttrande) påverkas av författarens (eller den talandes) kön.⁵

⁵ Lutz (1990) refererar till en studie (Goldberg citerad i Wikler 1975) där en artikel om forskning cirkulerats bland studenter med antingen en kvinna (Joan T. McCay) eller en man (John T. McCay) som författare. När texten ifråga hade John som författare bedömdes den betydligt mer välvilligt, bland annat för kriterier som professionell kompetens. Exempel på en liknande svensk studie återfinns i Einarsson och Hultman (1984).

I utlysningar av forskningsmedel pekar några studier på en liknande tendens, nämligen att de sista stegen i ansökningsprocesser då de slutliga besluten om tilldelning ska ske, kan vara kritiska för kvinnor. I en undersökning av bedömningar av sökande till doktorandstipendier (Graduate Fellowship) i National Science Foundation fann Brenneis (1994, 1999) att kvantitativa indikatorer spelade störst roll för att bedöma de sökande i början av beslutsprocessen. Användningen av sådana kriterier som döljer de sökandes sociala egenskaper gav bedömarna känslan av att delta i en opartisk bedömning. De sökandes sociala egenskaper, såsom genus och geografisk tillhörighet, spelade roll endast i slutet av processen då bedömarna skulle skilja sökande som låg nära varandra i vetenskaplig kvalitet från varandra.

Göran Melin (2007) riktar uppmärksamheten mot de sista stegen i ansökningsprocesser för att belysa den könsmissiga snedfördelningen i tilldelning av medel till excellensmiljöer och framtida forskningsledare. Han bygger på Nerviks (2006) studie av olika forskningsfinansierings, bland annat Vetenskapsrådets, utlysningar till excellenscentra och till individuella program för forskningsledare. Nervik konstaterar att andelen kvinnor som söker sådana medel är relativt liten och dessutom sker en viss snedfördelning till kvinnornas nackdel i tilldelningen. Melin (2007) påvisar med hjälp av delstudier av två av utlysningarna att det kritiska momentet för de kvinnliga sökande sker i det sista steget i ansökningsprocessen, då vad han kallar "reasoning methods" (2007:16f) används, i dessa fall intervjuer eller hearings inför en panel. I sådana bedömningsmetoder, som ofta används i de sista stegen i bedömningsprocesser, riskerar andra faktorer än de sökandes kompetens och projektets kvalitet att påverka bedömningen, menar Melin.⁶

En av de mest citerade artiklarna som belägger partiskhet gentemot kvinnor i ansökningssystem är de svenska forskarna Wennerås och Wolds studie (1997) av ansökningar till Medicinska forskningsrådet, som beskrivits som "perhaps the most shocking and discouraging study to date on the bias against women in scientific peer review" (Marsh et. al 2009:5 citerar Sheridan 1998:4). Wennerås och Wolds studie har lett till en omfattande debatt och efterföljande studier som undersöker samband mellan beviljningsgrad och

⁶ Melin bygger sin studie på två tidigare studier, Melin och Danell (2006) och Schilling (2007) av utlysningar till strategiska forskningscentra och strategiska satsningar på yngre forskare som ska bli framtidens forskningsledare. I dessa finner han att ett slutligt moment med hearing respektive intervju där de huvudsökande ställdes mot varandra med hjälp av kriterier som "vetenskapligt ledarskap" och "strategisk relevans" (Melin 2007:12) var kritiskt för de kvinnliga sökande. I sin hypotes generaliserar han den här analysen för att göra den tillämplig på andra bedömningsprocesser (såsom de andra utlysningarna till excellensmiljöer som Nervik 2006 diskuterar) och talar då generellt om "reasoning methods". Här skapar, enligt min mening, försöket till generalisering en vaghet som minskar förklaringsvärdet. "Reasoning methods" som även inkluderar diskussion i bedömarpaneler används även i bedömningar av individuella ansökningar och det är oklart vilka, om några, bedömningsformer som är gemensamma för bedömningar av medel till excellenscentra, vilket skulle kunna förklara den könsmissiga snedfördelningen i just dessa.

kön. Några forskare framhåller att medvetenheten om att eventuell partiskhet gentemot kvinnor kan förekomma har lett till att sambandet med kön har minskat, eller med ett svenskt uttryckssätt att jämställdhetsövervakning bidrar till att partiskhet med avseende på kön minskar (e.g. Bornman 2007, se även Sandström och Hällsten 2008). För närvarande råder det oenighet bland forskarna om det slag av partiskhet som Wennerås och Wold visade är kännetecknande för bedömning av forskningsansökningar. I en färsk artikel (Marsh et al 2009) argumenterar en forskargrupp för att de med en metodologiskt nyskapande metaanalys som kan hantera heterogeniteten hos ett stort antal studier av bedömningar av ansökningar (över disciplin- och nationsgränser) kan visa att det inte finns några skillnader med avseende på kön. De skillnader som belagts i tidigare studier är visserligen statistiskt signifikanta, men så små att Marsh et al som en övergripande tes om samband mellan kön och beviljningsgrad i ansökningssystem snarare vill tala om *likhet* mellan könen än *skillnad*.⁷

Forskningsområdet om eventuell partiskhet med avseende på kön består å ena sidan av delstudier som sysslar med mindre enheter såsom ett avgränsat antal tjänstetillsättningar, en utlysning av medel eller antagning av manus till tidskrifter. Å andra sidan består forskningsområdet av större statistiska undersökningar och metaanalyser av flera sådana undersökningar. I mindre delstudier har partiskhet med avseende på kön belagts, men resultaten är svåra att göra generaliserbara. Statistiska sambandsanalyser på metanivå som strävar efter generaliserbara resultat är däremot mer sparsamma med tolkningarna av resultaten, och därmed begränsas den kunskap de ger om *hur* bedömningssystemen fungerar. Det här illustreras i Marsh et al. (2009) där en ad hoc-formulerad hypotes åberopas för att bekräfta den (enda) skillnad de fann med avseende på kön, nämligen att män gynnades i pre- och postdoktoral stipendietilldelning, men inte i forskningsansökningar. Hypotesen var att ju mer bedömarna vet om den sökandes meriter, desto mindre roll spelar externa faktorer som kön i bedömningen. Hypotesen antogs förklara skillnaden mellan kvinnor och män i de två slagen av medelstipendium. Eftersom bedömarna vet mer om de sökandes meriter och publikationer i ansökningar till forskningsprojekt, då de sökande är etablerade med formella meriter, än vad de vet om de yngre, oetablerade pre- och postdoktorala forskarna skulle kön här ge ett utslag som gynnar de unga männen.

Tillämpningen av det här resonemanget, som enligt forskarna själva är en spekulering, belyser hur den kvantitativt inriktade forskningen av partiskhet

⁷ I en tidigare meta-analytisk studie fann några av forskarna (Bornman et al 2007) en snedfördelning i förhållande till kön, men med angreppssättet i Marsh et.al (2009) uppnås en överensstämmelse med övriga forskare som försvarar tesen om likhet mellan könen.

i kollegial bedömning är i behov av en kvalitativt grundad kunskap om hur bedömningsprocesser fungerar i praktiken. I det följande ska forskning med en sådan inriktning diskuteras.

Är kvalitet en objektiv egenskap?

Langfeldt (2002) påtalar att forskningen om partiskhet i bedömnings-systemen möter en grundläggande svårighet: det finns ingen tydlig definition, varken i litteraturen eller i bedömningspraktikerna, av vad som är vetenskaplig kvalitet. Därmed finns inte heller något entydigt svar på vad som utgör en opartisk bedömning, det vill säga en bedömning som endast tar hänsyn till vetenskaplig kvalitet. På grund av vagheten i kvalitetsbegreppet är det omöjligt att slutligt "bevisa" att partiskhet förekommer. En studie som statistiskt belägger samband mellan beviljade ansökningar och universitetstillhörighet eller mellan avslag på ansökningar och kön kan anta att universitetstillhörighet respektive kön spelade roll i bedömningen på ett sätt som strider mot det vetenskapliga etos som antas styra bedömningsprocesser. Men det finns alltid en teoretisk möjlighet att kvalitet var avgörande för bedömaren. Langfeldts poäng sätter ljuset på en central punkt i den kritik som kvalitativt inriktade forskare riktat mot de tidiga studierna, en kritik som träffar även den pågående statistiskt dominerade forskningen om partiskhet. Kritiken går ut på att den här typen av forskning inte ökar förståelsen av bedömningssystemen, eftersom frågan om hur bedömnarna tillämpar kvalitetsbegreppet lämnas utforskad.

Langfeldt (2002) påvisar vidare att den här problematiken även rör den grundläggande frågan om kvalitet är en objektivt existerande egenskap i vetenskap. Hur man ser på den frågan får betydelse för förståelsen av vad som sker i bedömningssystem och även för vilka forskningsfrågor som är intressanta att ställa. Langfeldt skiljer mellan två synsätt på kvalitetsbegreppets ontologiska status. Man kan se på kvalitet som en objektiv egenskap, det vill säga en egenskap som existerar i forskningen oberoende av vad bedömaren tänker om den. En bedömning av om en viss forskning har kvalitet blir då en fråga om att bedömaren har de rätta verktygen och/eller speciella kompetens som krävs för att känna igen kvalitet. I det andra synsättet ses kvalitet som en kulturellt och socialt definierad egenskap formad av vetenskapssamfundet i bedömningsprocesser.⁸

⁸ Langfeldt kallar de två synsätten för "realism" respektive "idealism" i den klassiska filosofiska betydelsen av ontologisk status. Hon formulerar också ett tredje synsätt på kvalitet som står för ett pragmatiskt perspektiv enligt vilket begreppet "vetenskaplig kvalitet" är utan mening och alla bedömningar är lika giltiga. Normer för bedömningar enligt det pragmatiska synsättet kan vara att garantera att bedömningsprocessen är effektiv och ger otvetydiga resultat (Langfeldt 2002:46).

Det vi kan notera här är att det (lyckligtvis) inte är nödvändigt att ta ställning i den filosofiska frågan om kvalitetsbegreppets ontologiska status för att bedöma bedömningssystemen. Det räcker med att forskarsamfundet enas om vad som är en väl genomförd *bedömningsprocess*. Det är möjligt att enas om det och samtidigt ha olika uppfattningar om huruvida bedömningar av vetenskaplig kvalitet kan vara objektiva.

Langfeldts resonemang lyfter fram ett förhållande som forskningen om bedömningssystem brottas med på olika sätt. Det centrala att notera är att oavsett vilket synsätt man har på kvalitet så finns det inte något sätt att mäta vad som "verkligt" utgör kvalitet utanför bedömningssystemen. Hur ska man avgöra om en viss bedömning var opartisk och rättvis och därmed gynade den forskning som hade den högsta kvaliteten? 1970-talets forskning om bedömningsprocesser (eg. Cole et al 1978) förlitade sig på ett kvalitetsbegrepp som var definierat i bibliometriska mått. Den bästa forskningen var den som ledde till flest publikationer och flest citeringar och en bedömning av forskningsansökningar var opartisk om den lett till sådan forskning. Men som kritiker hävdade så förskjuter ett sådant synsätt frågan om opartiskhet till bedömningssystemen för artiklar till tidskrifter. Hur ska man bedöma om kollegial bedömning för tidskrifter gynnar den bästa forskningen? Även här finns inget annat sätt än att hänvisa till andra bedömningssystem. I forskning om kollegial bedömning av tidskrifter har man till exempel undersökt hur banbrytande forskares första publikationer bedömdes av tidskrifter och funnit att forskning som sedan blivit ledande inom sitt område i många fall först refuserats (här finns flera exempel på Nobelprisvinnande forskning som refuserats av ledande tidskrifter, se Campanario 1998b). Här är alltså förskjutningen i tid, och andra bedömare, den referens som möjliggör en annan bedömning. Den allmänna praxis hos forskningsfinansiärer att använda externa granskare och att ge internationella bedömare och tidskrifter företräde före nationella är ett annat sätt att komma utanför ett bedömningssystem, genom en rumslig förskjutning.

Den rådande vagheten i begreppet vetenskaplig kvalitet och insikten om att det inte finns några sätt att mäta kvalitet utanför bedömningssystemen – man kan endast röra sig inom och mellan olika bedömningssystem – tydliggör behovet att analysera de definitioner av vetenskaplig kvalitet som bedömarna använder i praktiken.



KVALITETSBEDÖMNINGAR I PRAKTIKEN

Jag har valt fyra olika undersökningar av bedömningsprocesser som jag ska återge relativt detaljerat för att ge en bild av vad forskningen i dagsläget vet om hur kvalitetsbedömningar tar form i praktiken. Det framstår som angeläget med tanke på att så gott som samtliga forskare betonar att området är understuderat, samtidigt som de statistiska undersökningarna av partiskhet behöver belysas med kvalitativt grundad forskning. De fyra undersökningarna belyser två olika aspekter av bedömningsprocesser: operationaliseringen av kvalitetsbegreppet och beslutsprocessen. I genomgången söker jag vaska fram de centrala och gemensamma dragen i bedömningsprocesser på grundval av följande studier: *Travis och Collins* (1991) som undersöker bedömning av forskningsansökningar i England, *Langfeldts* (2002) studie av utvärderingar på makronivå i Norge, *Lamont* (2009) som undersöker bedömning av forskningsansökningar i USA och slutligen *Campanario* (1998a, 1998b) som diskuterar studier av bedömning av artiklar till internationella tidskrifter.

Travis och Collins om kognitiv partikularism

De brittiska forskarna Travis och Collins (1991) utgår från Cole, Rubin och Coles studie (1978) där bedömningssystemet granskades för att bemöta anklagelserna om partiskhet.⁹ Travis och Collins påpekar att Cole, Rubin och Cole i sina försök att ringa in vad de oprecist formulerade anklagelserna om *old boyism* går ut på nämner flera olika möjliga tolkningar av partiskhet. *Old boyism* skulle kunna hänvisa till a) forskare som har en gemensam syn på sitt vetenskapsområde och uppskattar endast arbeten av dem som har en liknande syn b) sociala vänskapsnätverk som består av forskare som "vuxit upp" tillsammans, gick i samma skolor, tenderar att umgås, är av samma kön, och favoriserar varandras forskningsansökningar och c) social position, det vill säga de forskare som uppnått en hög status i systemet tenderar att favorisera ansökningar från dem som har en liknande position i systemet även om de inte har personlig kontakt med dem (Cole, Rubin och Cole 1978:33–34).

⁹ Den monografi av Cole, Rubin och Cole (1978) som Travis och Collins diskuterar presenterar resultatet av en första fas av Cole-gruppens studie. Studiens andra fas presenteras i ytterligare en monografi (Cole and Cole 1981). Forskningen är även publicerad i två artiklar (Cole, Rubin and Cole 1977 och Cole, Cole and Simon 1981). Travis och Collins (1991: 339, not 8) påpekar att det finns betydande skillnader mellan de olika publikationerna med avseende på de slutsatser som forskarna drar.

Travis och Collins menar emellertid att Cole, Rubin och Cole i sin studie inte upprätthåller distinktionen mellan en kognitiv syn på vetenskap (punkten a) och vetenskapen sedd som ett socialt system (punkterna b och c). De olika slagen av partiskhet mäts i studien genom att samband fastställs mellan rankingen av ansökningarna och rankingen av de universitet som de sökande tillhör, universitetets geografiska belägenhet, antal publikationer och citeringar, samt professionell ålder och akademisk rang. Enligt Travis och Collins kan man med dessa operationaliseringar beskriva endast partiskhet som har att göra med sociala och institutionella egenskaper hos de sökande.¹⁰

Travis och Collins kritiserar Cole-gruppens forskning för att bygga på en begränsad syn på forskning som ett system stratifierat i termer av sociala positioner. Om man istället ser vetenskapen som ett kunskapssystem format av kognitiva strukturer framträder frågan om partiskhet på ett annat sätt. Vilka effekter beslut grundade på bedömarens plats i den kognitiva strukturen får, beror på i vilken mån kognitiva gränser existerar inom och mellan vetenskapliga fält och discipliner.

I vetenskapliga områden där inga betydande kognitiva gränser existerar – där konsensus är hög – har alla forskarna en liknande uppfattning om det förhärskande paradigmet. För utvecklingen av vetenskapen spelar det ingen roll vem som löser nästa problem som paradigmet pekar på – förr eller senare kommer alla problem att bli lösta. Prestigenätverkens dynamik i bedömningssystemet kan betyda att forskaren X istället för forskaren Y får anslagen och äran för framstegen, men detta påverkar den vetenskapliga utvecklingen endast marginellt. Det slag av partiskhet som Cole-gruppen studerar kan kallas orättvis mot den enskilde forskaren, men det betyder inte nödvändigtvis att det är ett ineffektivt system för att fördela resurser sett från perspektivet att det övergripande målet är att främja den vetenskapliga utvecklingen.

I de vetenskapsområden där det finns tydliga kognitiva gränser, å andra sidan, kommer effekten på den vetenskapliga utvecklingen att bli större ju skarpare gränserna är om systemet tenderar att rekrytera bedömare från den ena sidan av en kognitiv gräns snarare än den andra. Institutionell partiskhet kommer att ha systematisk effekt på den kognitiva utvecklingen av området endast i de fall där kognitiva och institutionella gränser sammanfaller.

¹⁰ Travis och Collins hänvisar till ytterligare kritik av Cole, Rubin och Cole (1978) som påvisar att den inte besvarar frågan om kollegial bedömning missgynnar innovativ forskning (Mitroff och Chubin 1979) och att frågor om hur systemet hanterar balansen mellan olika vetenskapsområden och mellan grundläggande och tillämpad forskning inte undersöktes (Roy 1985).

En central fråga blir då hur ofta det sker, det vill säga hur ofta en viss skolbildning har sin tillhörighet på en viss institution eller ett universitet. Travis och Collins hävdar med stöd i vetenskapsociologi att så ofta är fallet – vetenskapliga synsätt formas i social interaktion i vetenskapliga miljöer och det är därför att förvänta att starka sociala nätverk i en vetenskaplig miljö leder till stora kognitiva likheter bland forskarna. Det skulle betyda att skydd mot institutionell partiskhet i bedömningssystemen också skulle skydda mot kognitiv partiskhet. Men, fortsätter de, även om det finns starka skäl att på en makronivå förvänta sig överensstämmelser mellan vetenskaplig miljö och kognitivt innehåll, så gäller inte alltid detsamma på mikronivå. Det kan finnas stora skillnader mellan enskilda forskares vetenskapliga utgångspunkter trots att de tillhör samma vetenskapsområde, skillnader som inte nödvändigtvis sammanfaller med universitet eller institutioner. Travis och Collins drar slutsatsen att skydd mot institutionell partiskhet endast kan ge ett begränsat skydd mot kognitiv partiskhet.

Travis och Collins genomförde själva en studie av tio kommittémöten för bedömning av naturvetenskaplig grundforskning inom SERC, den största forskningsfinansieraren i Storbritannien. Kommittéerna behandlade forskningsansökningar från enskilda forskare eller forskargrupper, vanligen för två- eller treårsperioder. De flesta ansökningarna gällde grundforskning, medan vissa innehöll ett element av tillämpad forskning, då de föll inom forskningsområden som bedömdes ha långsiktiga strategiska eller ekonomiska implikationer. I bedömningsprocessen graderades först ansökningarna av bedömarna i skriftliga utlåtanden, vilka kommittéerna sedan tog ställning till. I sin studie använde sig Travis och Collins av kvalitativa metoder i form av ett fältarbete med deltagande i kommittémötena som också bandades.

Travis och Collins beskriver i artikeln några fall av vad de kallar kognitiv partiskhet som de fann i sin studie. Där illustreras hur kommittéledamöterna handskas med fenomenet på olika sätt.

I ett fall var samtliga sakkunnigutlåtanden positiva och resulterade i den högsta eller nästhögsta poängen på ansökan. I diskussionen i kommittén kom emellertid en diskussion upp där det framkom att endast en liten del av forskarna inom området ifråga (10%) ansåg att man borde fortsätta att forska inom det här området. En kommittéledamot kommenterade att man lätt skulle kunna få området att självdö (genom att inte ge anslag):

”Interesting! You could actually kill the subject quite easily...”. (Travis och Collins 1991:332).

I den fortsatta diskussionen konstaterades det att alla de sakkunniga tillhörde skolan ifråga och på frågan varför man inte skickat ansökan till en bedömare som tillhörde den majoritet inom området som ansåg att detta inte

var värt att forska om blev svaret att man hade försökt men att forskarna med den åsikten avböjt. Trots dessa insikter beslöt kommittén att behålla de sakkunnigas gradering av projektet, eftersom de inte ansåg det vara möjligt att påbörja vad som i praktiken skulle varit att rangordna olika ämnen eller vetenskapsområden i förhållande till varandra. En sådan diskussion bordlades till ett senare möte.

Travis och Collins konkluderar att kommittéledamöterna i det här fallet var bekymrade över att en kognitiv partikularism existerade hos åtminstone tre av de fyra sakkunniga. Dessa identifierades som en grupp forskare som hade ett särskilt intresse av att det här projektet genomfördes, i en sådan grad att det kunde ses som partiskhet. Kommittén var ovillig att ta sig an problemet trots att den kognitiva partiskheten hos de sakkunniga och kommitténs beslut möjligen ledde till vad som av andra bedömare kunde ses som att hålla liv i ett döende fält.

I ett annat exempel (från en annan kommitté) identifierade kommittén på samma sätt en hög kognitiv likhet mellan de sakkunniga som samtliga rankat ansökan ifråga högt. Här beslöt dock kommittén att gå emot de sakkunniga och på grundval av sin egen expertis inte ge projektet samma höga värde. Kommittén bedömde här således att graden av kognitiv likhet mellan de sakkunniga grundad på deras tillhörighet till samma vetenskapsområde som den sökande diskvalificerade deras bedömning av projektets vetenskapliga kvalitet. I diskussionen enades man om att här fanns en "...strong community[...] supporting its own..." (Travis och Collins 1991:334).

I ytterligare ett fall hade ett projekt som först rankats högt skickats ut till två andra sakkunniga som rankat projektet lägre. Här rörde sig diskussionen kring i hur hög grad man skulle låta de senare sakkunnigutlåtandena "sänka" projektet. En kommittéledamot påpekade att man gav anslag till andra projekt som rankats högt i en första omgång och som inte hade utsatts för ytterligare en bedömning. I det här fallet påtalades inte öppet i diskussionen att de olika sakkunniga tillhörde olika skolor. Den informationen fick emellertid Travis och Collins utanför mötet vid lunchen. En av forskarna sa att han bett kommittésekreterariatet att skicka ut den här ansökan till ytterligare sakkunniga därför att de ursprungliga sakkunniga alla tillhörde en "klubb för inbördes beundran". Med den här informationen kunde Travis och Collins kartlägga fältet och komma fram till att ansökan var relevant för två separata kognitiva gemenskaper. Ansökan hade skickats endast till bedömare tillhöriga den ena av de två miljöerna, medan den senare förfrågan om utlåtande hade skickats till bedömare i den andra miljön vilka rankat ansökan lägre.

Det ursprungliga valet av sakkunniga var inte medvetet partiskt. Forskningsfinansiärens sekretariat var helt enkelt omedvetet om vetenskapsområdets kognitiva karta. Travis och Collins framhåller att den här typen av

partiskhet inte är möjlig att spåra med hjälp av statistisk sambandsanalys eller att upptäcka för den utomstående observatören. De poängterar dock att inga slutsatser kan dras från deras fältarbete om hur utbredd det de kallar kognitiv partikularism är, eller i vilken grad effekterna är skadliga för vetenskaplig utveckling. Exempelen är avsedda att illustrera hur fenomenet är verksamt i modern vetenskap och hur intrikata besluten om vad som skall räknas som "fair play" i områden där det finns kognitiv mångfald är.

Som Langfeldt (2002) påpekar så råder det delade meningar om ifall det Travis och Collins kallar kognitiv partiskhet eller kognitiv partikularism ska räknas som partiskhet. Det är just bedömarens expertis och specialisering i ett område som är grunden för hennes syn på vetenskaplig kvalitet. Granskaroenighet i bedömningsprocessen kan ses som ett tecken på att gruppen av bedömare har en bred sammansättning och att processen fungerar opartiskt, det vill säga utan tysta överenskommelser om hur ansökningarna ska rankas. Oavsett om vi ser kognitiv partikularism som partiskhet eller inte så påvisar Travis och Collins studie, menar jag, ett intressant förhållande, nämligen att bedömarna inte hade en överenskommen metod för att hantera granskaroenighet och skolbildning – de agerade på olika sätt i de tre exemplen. Den operationalisering av vetenskaplig kvalitet som avgjorde utfallet formulerades inte explicit – bedömargruppen bestämde sig på grundval av den egna synen på kvalitet för att acceptera de andra sakkunnigas bedömning i det första fallet, för att ge en lägre ranking i det andra fallet och i det tredje för att kalla in andra sakkunniga.

Langfeldt om beslutsprocesser i expertpaneler

I den norska forskaren Liv Langfeldts studie (2001, 2002, 2004) står själva beslutsprocessen i expertpanelers bedömning av kvalitet i fokus. I sin studie undersöker hon beslutsprocesser i sex expertpaneler tillsatta för att utvärdera norsk forskning. Panelerna representerade ett brett urval av vetenskapsområden, organisationsformer och syfte för utvärderingen för att möjliggöra generella slutsatser om beslutsprocessers centrala egenskaper. De valda fallstudierna innefattade utvärderingar av två forskningsområden, två forskningsinstitut och två forskningsprogram. Alla vetenskapliga områden var företrädna (dvs medicin, naturvetenskap, ingenjörskonst, samhällsvetenskap och humaniora). Tre paneler bestod främst av utomnordiska experter medan de övriga tre var nordiska.

Till skillnad från Travis och Collins deltog Langfeldt inte på expertpanelernas möten, utan hennes material bestod av forskningsrådets skriftliga material om utvärderingarna, djupintervjuer med samtliga experter och, i vissa

fall, deras privata anteckningar och utkast. Ytterligare en skillnad mellan Langfeldts och Travis och Collins studie är typen av kvalitetsbedömning. Travis och Collins undersökte kollegial bedömning på mikronivå, bedömningar av enskilda ansökningar om forskningsanslag. Langfeldts studie gäller bedömningar på makronivå av forskningsfält, institut och program. I expertpanelernas uppdrag ingick bland annat att bedöma den forskning som producerats inom respektive fält, institut eller program. Langfeldts fokus i studien är dessa kvalitetsbedömningar.

Samtliga panelers arbete karakteriserades av en arbetsdelning så att en expert bedömde den del av fältet/institutets/programmets produktion som låg närmast hans eller hennes forskning. Vidare karakteriserades panelernas arbete av liten interaktion och diskussion så att i praktiken kom den expert som genomfört en viss del av utvärderingen att ensam stå för innehållet. Langfeldt undersöker olika möjliga gruppeffekter på beslut, vilket hon definierat som en del av sin undersöknings syfte motiverat av tanken att en expertpanel som sammanträder ett visst antal gånger ska ge något mer än en enskild forskares bedömning. På grund av arbetsdelningen uppträdde mycket få gruppeffekter. Istället karakteriserades beslutsprocesserna av tysta överenskommelser. Såväl grunden för beslutsfattandet som kriterierna för bedömning av forskningens kvalitet var uttalade. Langfeldt karakteriserar de tysta beslutsprocesserna hon fann i utvärderingarna som "sounding" (jfr Olsen 1972), det vill säga processer med en stark tonvikt på att nå konsensus, och en systematisk användning av vaghet och ett undvikande av att definiera klara alternativ som ett medel till det. En ledamot karakteriserar beslutsfattandet såhär:

In practice there is not very much confrontation in these [kinds of] meetings. If there had been some confrontation then the whole committee would have been [unable] to work. So we had to trust one another's judgments. And when we disagreed with one another we disagreed quietly and mildly, and whoever was in the minority would give way, I think. (Langfeldt 2004:56).

I de fall där oenighet, som skulle kunna bero på kognitiv partikularism, uppstått gjordes tysta kompromisser i betydelsen att de skilda ståndpunkter som uttryckts inte inkluderades i slutrapporten. Paneldeltagarnas uppfattning om vilka oenigheter som fanns och hur man uppnått kompromisser skilde sig åt och alla var inte medvetna om att det fanns olika ståndpunkter. I endast ett av fallen var alla deltagarna medvetna om att det funnits en öppen kontrovers och vad den handlade om, eftersom man i det fallet öppet hade förhandlat fram en lösning (en förhandling som inkluderade anklagelser om personlig partiskhet).

Vidare fann Langfeldt att regler och kriterier för kvalitetsbedömningarna av forskningen var uttalade. De diskuterades inte i panelerna och beskrevs

inte i slutrapporterna, informanterna var inte heller förmögna att beskriva dem fullödigt när de tillfrågades. Paneldeltagarna formulerade kriterier för att bedöma vetenskaplig kvalitet utan problem på förfrågan i intervjuerna, men de kunde inte förklara hur de använde kriterierna när de utvärderar forskning. Operationaliseringen av kriterierna framstod därför som helt grundad på tyst kunskap.

När paneldeltagarna tillfrågades om ifall utvärderingarna var opartiska framkom mönstret att flera paneldeltagare menade att slutsatserna i rapporterna var objektiva, att de själva alltid uttalade sin uppriktiga mening, att de inte hade några band till någon bland de bedömda enheterna, och var fria att ha vilken ståndpunkt som helst, eller att de var förmögna att inta en självständig hållning. Å andra sidan, menade några paneldeltagare att andra paneldeltagare hade personlig kännedom om några av de utvärderade och därför var mindre kritiska i sina bedömningar än resten av panelen (i ett fall motsatsen: att försöka straffa fiender), att andra paneldeltagare hade vissa förutfattade meningar som de försökte bevisa under bedömningen eller främjade sina särintressen i vissa fält. Bedömarna nämnde även allmänna problem gällande objektivitet i kollegial bedömning – att sådana bedömningar är ett slags konststart som med nödvändighet är subjektiv, grundad på personliga intryck, eller renommé. Vidare påtalade paneldeltagarna tidsbristen som en faktor som gjorde det svårt att gå på djupet och utföra en grundlig bedömning.

Langfeldt påtalar att den oklarhet angående vad som utgör goda eller acceptabla bedömningspraktiker som hon fann hos sina informanter återfinns i forskningen om kollegial bedömning. Hon pekar här på att de olika slag av partiskhet som diskuterats har olika karaktär och – i linje med exempelvis Travis och Collins – vill hon skilja på partiskhet som har att göra med forskarens personliga och/ eller sociala egenskaper å ena sidan och å andra sidan kognitiv partiskhet.

De kontroverser som nämndes i hennes studie handlade om olika forskningsområdets värde samt om tillämpad forskning kontra grundforskning. Vetenskaplig oenighet och "partiskhet" var centrala för att förstå utgången av beslutsprocesserna och Langfeldts slutsatser kan sägas vara i linje med Travis och Collins betoning av kognitiv oenighet.

Langfeldt understryker att förekomsten av anklagelser om partiskhet av olika slag i en bedömningsprocess inte är detsamma som faktisk partiskhet, men hon konkluderar ändå att de sex expertpaneler hon undersökte inte levde upp till krav på opartiskhet. Hon påpekar att det inte fanns någon mekanism i processen som kunde skydda mot eller förebygga möjlig partiskhet hos enskilda paneldeltagare. Den strikta arbetsdelningen och avsaknaden av interaktion mellan deltagarna minskade möjligheterna för en sådan kon-

troll av partiskhet, kognitiv eller personlig. Hon menar att bristerna inte kan ses som specifika för just dessa sex, utan härrör ur den internationellt förekommande modell för expertpanelbedömning som följdes och de outtalade reglerna för sakkunnigas pragmatiska inställning till uppdraget.

I Langfeldts studie av beslutsprocesser i expertpaneler är kvalitetsbegreppets vaghet och de tysta överenskommelserna bestämmande för processen – kvalitetsbegreppet är fortsatt höljt i dunkel. I den studie jag ska diskutera härnäst är forskarens ambition att öppna bedömningsprocessens ”svarta låda” och skärskåda kvalitetsbegreppet.

Lamont om kvalitetsbegreppets operationalisering

Michèle Lamont har, delvis tillsammans med medforskarna Joshua Guetzkow och Grégoire Mallard, genomfört kvalitativa studier (intervjuer och observation) av bedömningsprocesser i tjugo paneler för samhällsvetenskaplig och humanistisk forskning knutna till fem olika forskningsfinansierare i USA under 2000-talet (Lamont 2009). Panelerna som Lamont studerat är mångvetenskapligt sammansatta och bedömer ansökningar från hela fältet av humanistiska och samhällsvetenskapliga studier.

Liksom Travis och Collins uppmärksammar och analyserar Lamont de kognitiva elementen i bedömningsprocessen. Hennes angreppssätt fördjupar analysen av hur kognitiva skillnader i bedömningar ska förstås. Studien är av särskilt intresse för en förståelse av bedömning av genusforskning eftersom den utförts i mångvetenskapligt sammansatta paneler för humaniora och samhällsvetenskap, där merparten av genusforskning bedöms (även i Sverige).

Lamont och hennes medforskare fann att bedömarna använde sig av en eller flera av fyra olika epistemologiska ”stilar” i sin argumentation för eller mot forskningsansökningar. Bedömarna kunde använda sig av flera stilar i sin argumentation, men använde ofta en stil mer frekvent än andra. Forskarna fann tydliga skillnader mellan humanister, samhällsvetare och historiker (de sistnämnda kategoriserades separat på grund av disciplinens hybridstatus mellan humaniora och samhällsvetenskap) med avseende på hur ofta de olika epistemologiska stilarna användes. De fyra stilarna som forskarna fann var förståelseorienterad (comprehensive), konstruktivistisk, positivistisk och utilitaristisk (Mallard, Lamont och Guetzkow 2009, Lamont 2009). Den *förståelseorienterade* stilen tillämpar ett tolkande eller teoretiskt perspektiv för att förstå sociala fenomen. Argument i linje med denna stil stödjer humanistisk och samhällsvetenskaplig forskning som lägger vikt vid historisk och kulturell kontext. I ansökningar värderas *verstehen*, sinne för detaljer

och specifik kontext. Stilen var den mest frekventa bland alla bedömare. Argument i linje med denna stil åberopades av 86 procent av de humanister som intervjuades, 71 procent av samhällsvetarna och 78 procent av historikerna.¹¹ Den *konstruktivistiska* stilen användes av bedömare som önskar ”ge röst” åt minoriteter eller andra underprivilegierade grupper. Stilen värderar reflexivitet, det vill säga att uppmärksamhet ges åt forskarens identitet och position och hur dessa inverkar på den teoretiska analysen. Stilen används med förkärlek av bedömare som bedriver socialt eller politiskt engagerad forskning. Konstruktivistiska argument förekom mest frekvent bland humanister (28 procent) och historiker (29 procent), medan en mindre andel (14 procent) av samhällsvetarna använde sig av denna epistemologiska stil. Den *positivistiska* stilen värderar generaliserbarhet och hypotesprövning. Den används oftast av samhällsvetare (57 procent) och i mindre grad av historiker (23 procent), medan ingen av humanisterna använder en sådan stil i sina bedömningar. Den *utilitaristiska* stilen slutligen liknar den positivistiska i sin värdering av hypotetisk-deduktiv metod, men den värderar endast tillämpningsbar kunskap. Den användes av endast 4 procent av historikerna och ingen av humanisterna medan 19 procent av samhällsvetarna använde utilitaristiska argument i sina bedömningar.

Konstruktivistisk och förståelseorienterad epistemologisk stil var således mest förekommande i ansökningsomgångar inom historia och humanistiska ämnen, medan förståelseorienterad och positivistisk stil dominerade i samhällsvetenskaperna.

Lamont (2009) diskuterar även skillnader i vad hon kallar bedömningskultur mellan de enskilda discipliner som ingick i panelerna hon studerade. Hon presenterar sex av de elva discipliner som var representerade bland hennes respondenter: filosofi, *English* (litteraturvetenskap), historia, antropologi, statsvetenskap och ekonomi. De discipliner inom vilka bedöarna var mest eniga om definitionen av vetenskaplig excellens var historia och ekonomi. I historia vilade den rådande samsynen kring kvalitet på en uppfattning om vad ett gott hantverk i utförandet av empirisk forskning är. I ekonomi härrörde konsensus från en kognitiv konsolidering kring matematiska verktyg. De skilde sig däremot åt i inställningen till om vetenskaplig kvalitet är en objektivt existerande egenskap i ansökningar/forskning. Medan ekonomerna beskrevs, av andra och sig själva, som kapabla att tydligt rangordna projekt med hög kvalitet eftersom kvalitet ses som en objektivt existerande egenskap i forskning, så tenderade historiker att erkänna nödvändigheten av förhandling och diskussion för att avgöra vad

¹¹ Procenttalen anger andelen intervjuade bedömare som använde sig av den epistemologiska stilen ifråga. Varje bedömare kunde använda sig av mer än en epistemologisk stil. Se tabell 5.6 i Lamont 2009: 177.

som är kvalitet. Historiker, liksom litteraturvetare, var mer benägna än ekonomer att fråga efter ”vems bedömningskriterier som blir universaliserade till att vara disciplinära kriterier”.

Litteraturvetenskapen var den disciplin av de sex där konsensus kring kvalitetsdefinitioner var lägst. Flera bedömare ifrågasatte själva begreppet excellens och noterade förekomsten av skilda uppfattningar om hur kvalitet i ämnet ska bedömas. Likaså i antropologi fanns en liknande betoning på kvalitet som kulturellt konstruerad snarare än objektivt existerande.

Statsvetenskap aspirerade på en konsensus liknande den i ekonomi, men präglades av konflikter kring den högre värderingen av kvantitativ forskning i jämförelse med kvalitativ. Filosofi slutligen betecknades både av andra bedömare och av forskningssekreterare som en problemdisciplin, då filosoferna ansåg att endast de själva är kapabla att bedöma kvalitet i ansökningar för filosofisk forskning, delvis därför att relevans och originalitet måste mätas med disciplinspecifika mått.

Med Lamonts studie har vi kommit något närmare ett exempel på hur kvalitetsbegrepp tar form i praktiken. De skilda epistemologiska stilarna och bedömningskulturerna ger upphov till kognitiva skillnader i bedömningar. Bedömarna i panelerna använde sig inte av ett enhetligt kvalitetsbegrepp utan kvalitet framstod som i hög grad diversifierad. Det skedde vidare genom bedömarnas tolkningar av de olika kvalitetskriterier som tillämpas i ansökningsförfarandet. Det fanns en bred samstämmighet i vilken vikt bedömarna lade vid de mest använda kvalitetskriterierna i bedömningsprocesser: originalitet, relevans, metod och genomförbarhet. De kriterier som nämndes oftast av bedömarna, och som fanns med i finansierarnas formella kriterier var originalitet och relevans, följt på ett visst avstånd av metod och genomförbarhet. I den vikt som lades vid kriterierna förekom inga tydliga skillnader när bedömarna placerades in i de tre grupperna, humanister, historiker och samhällsvetare. Däremot varierade betydelsen av kriterierna originalitet och relevans på följande sätt.

Originalitet

Lamont påpekar att originalitet i litteraturen om kollegial bedömning definieras enligt en naturvetenskaplig syn på originalitet, formulerad i Mertons (1973) tidiga studie, som upptäcker som leder till ny vetenskaplig kunskap. Huruvida en sådan definition av originalitet karakteriserar förståelsen av begreppet även i humaniora och samhällsvetenskap är utforskat. Det Lamont och hennes kollegor fann pekar på varierade betydelser av originalitet (Guetskow, Lamont och Mallard 2004, Lamont 2009). De identifierade följande definitioner av begreppet: användning av en ny infallsvinkel eller

angreppssätt (approach), en ny teori, en ny metod, nya data, ett nytt ämne, forskning i ett understuderat område och forskning som producerar nya resultat. Hur begreppet användes varierade mellan discipliner på så sätt att humanister och historiker gav företräde åt originalitet i betydelsen ny ”approach” medan humanister också betonade originalitet i data, t ex användandet av nya manuskript, brev eller författare. Samhällsvetare värderade originalitet i metod högst, men de värderade också originalitet i mer varierande betydelser, såsom originalitet i approach, teori eller ämnesval.

Relevans

Fyra av de fem forskningsfinansiärerna i studien inkluderade relevans (”significance”) som ett bedömningskriterium. Bedömnarna använde relevans i olika betydelse, med en skiljelinje mellan scholarly (intellektuell eller teoretisk) relevans å ena sidan och å den andra social eller politisk relevans, enligt Lamont (2009). För enkelhetens skull benämner jag på svenska dessa två slag av relevans för vetenskaplig relevans respektive samhällsrelevans.

Hur bedömnarna avgjorde en ansökans vetenskapliga relevans varierade. Den kunde bedömas i förhållande till hur välutforskat ett forskningsområde var. En annan betydelse var i vilken grad forskningen ifråga kunde leda till generaliserbara slutsatser eller alternativt en bedömning av de teoretiska implikationerna av en specifik studie för mer övergripande frågor. En för ”smal” ansökan kunde exempelvis avvisas för att den inte bedömdes ha implikationer för annan forskning inom närliggande områden. I dessa omdömen menar Lamont att bedömnarna uttrycker smakpreferenser som är grundade i deras specifika expertis, smakpreferenser som varierar över fältet och ger olika tolkningar av vetenskaplig relevans.

Samhällsrelevans tolkades på främst två skilda sätt. Samhällsrelevant forskning kunde vara forskning som ger instrumentell kunskap eller forskning som syftar till att ge röst åt underprivilegierade grupper. De två tolkningarna av samhällsrelevans var associerade med den utilitaristiska respektive konstruktivistiska epistemologiska stilen. Fördelningen av de olika epistemologiska stilarna över det humanistiska och samhällsvetenskapliga fältet påvisade en intressant skillnad i hur forskningens samhällsrelevans tolkades av olika forskare. Humanister tillämpade den konstruktivistiska stilen mer än samhällsvetare, det vill säga för humanister var forskning som synliggjorde underprivilegierade eller stigmatiserade grupper ett legitimt bidrag till vetenskapen i högre grad än för samhällsvetare. Samtidigt fanns en tendens hos humanister att se ner på forskning som strävade efter att vara tydligt tillämpningsbar. Samhällsvetarna å andra sidan hade en starkare

tendens att använda den utilitaristiska stilen och var mer villiga att stödja projekt som var direkt relevanta för policy-skapande.

För en del bedömare var tillämpningen av kriteriet samhällsrelevans problematisk då det ansågs implicera en instrumentell syn på vetenskapligt kunskapsökande. De såg som sin uppgift att försvara vetenskapens autonomi i förhållande till politiska och ekonomiska krafter och avvisade kriteriet samhällsrelevans överhuvudtaget (Lamont 2009).

Ett flertal bedömare använde även kriteriet klarhet eller tydlighet (*clarity*) som var inkluderat i några, men inte alla, finansiärers formella bedömningskriterier. En klar skrivstil och förmågan att framställa sin ansökan med pregnans och tydlighet ansågs av ett flertal bedömare vittna om förmåga till vetenskaplig analys och en nödvändig förutsättning för att kunna komma ifråga för finansiering.

Informella kriterier

Kunskap om operationalisering av kvalitetsbegreppet av det slag som Lamonts studie ger belyser ytterligare vad tidigare forskning kallat "kognitiv partiskhet", men Lamonts studie komplicerar bilden. I de bedömningspaneler hon studerat arbetar forskarna med diversifierade kvalitetsbegrepp och en stor variation i tolkningarna av kriterierna på kvalitet, och istället för tydliga kognitiva gränser mellan discipliner visar Lamont på ett komplicerat samspel i forskningsfältet mellan olika epistemologiska stilar och tillämpning av en mångfald av kriterier. Bilden av vad "kvalitet" och "partiskhet" är komplicerats än mer av användningen av informella kriterier.

Utöver de formella kriterier som finansiärerna anger och bedömarna hänvisar till tillämpade bedömarna i Lamonts studie en rad informella kriterier, dock inte alltid på ett medvetet sätt. De informella kriterierna var ofta knutna till den förmåga som en bedömare kallade intuition eller "att ha öga för det". En ordförande i en panel talade om de informella kriterierna, inte som idiosynkratiska eller nyckfulla och därför oförenliga med en rättvis process, utan som en viktig del av uppgiften att identifiera sann excellens:

"There are these intuitive aspects to it, and at the same time, I believe we're doing social science. These aspects have aesthetic sides to them, like "This is a very elegant and striking proposal". This is an important part of the individual and group processes. That doesn't mean that... we don't have standards. I think that those two things can go together quite well, especially if you're looking for things that are just a little bit outside the norm. You're actually looking for some of those qualities in the spark, the godlike qualities in the proposal." (2009:188).

De "gudalika" kvaliteterna i ansökningar, tecknen på excellens, är alltså knutna till informella kriterier, vilka tillämpades i lika hög grad av bedömarna i de tre områdena humaniora, samhällsvetenskap och historia. Bedömarna från discipliner som ser subjektivitet som en korrumpierande kraft som ger upphov till bias (exempelvis ekonomer) visade inte mindre benägenhet att bedöma vilka sökande som har "det". De informella kriterierna användes inte endast vid tillfällena då de formella kriterierna inte var tillräckliga för att ge utslag, utan var en integrerad del av hela bedömningsprocessen. Lamont identifierar följande informella kriterier på vetenskaplig excellens:

Tecken på intelligens. Ansökan läses så att den anses säga något om den sökandes intelligens, den sökande beskrivs med ord som artikulerad, kompetent, intelligent eller begåvad. Två tredjedelar av bedömarna använde sig av "tecken på intelligens" i bedömningarna av ansökningar. En bedömaren menade att en sökandes intelligens kan avläsas i "the subtlety and complexity with which the project is framed" (2009: 189). Hur bedömarna utläste tecken på intelligens kunde också tolkas med ledning av hur de beskrev sina egna intellektuella hjältar, ofta forskare med förmåga att beskriva och tolka komplexa fenomen.

Elegans och kulturellt kapital. Elegans, förstått som ett slags perfektionism i ansökans form, en förmåga att formulera en så välskriven ansökan att den kan sägas ha litterära kvaliteter värderades av humanister och historiker, medan samhällsvetare betonade det mindre. Med avseende på kulturellt kapital lade bedömarna vikt vid kulturell bredd och en förtrogenhet med "kultur", att med lätt hand uppvisa en kulturell bildning. Lamont anger att det informella kriteriet elegans ofta är svårt att skilja från uppvisandet av ett kulturellt kapital som är ojämnt distribuerat över disciplingränser, universitet och mellan sociala klasser. Hon hänvisar här till Bourdieus *Homo Academicus* där han hävdar att kulturellt kapital som innehas av en övre medelklass-elit präglar akademien. Att värdera elegans och uppvisning av kulturellt kapital kan därför innebära att likställa vetenskaplig excellens med en tillhörighet i övre medelklass. Enligt en bourdieusk logik skulle därmed de sökande som konstrueras som "briljanta" vara de som känner sig klassmässigt mest hemma i den akademiska världen, exempelvis barn till akademiker och forskare, och därför kan uttrycka sig "elegant".

"Spännande" och "intressant" forskning. Omdömen om vad som är "spännande" och "intressant" forskning och dess motsats, vad som är "tråkigt", användes i bedömningsprocesserna. Omdömena förekom oftare bland humanister än bland historiker och samhällsvetare. Bedömarna från discipliner där tolkning spelar stor roll, humanister, antropologer och kulturhistoriker är troligen mer bekväma med att låta en sådan ståndpunkt spela roll i bedömningen, då förmågan att tolka är deras främsta analytiska verktyg, menar

Lamont. Bedömare från icke-tolkande discipliner å andra sidan var ibland mycket kritiska till att använda vad som är "intressant" eller "cool" som ett bedömningskriterium.

Moraliska egenskaper hos den sökande. Ungefär 40% av bedömarna hänvisade till moraliska egenskaper hos de sökande när de bedömde ansökningar, vilket enligt Lamont är tillräckligt många för att se dessa omdömen som en normal aspekt av bedömningsprocesser. Hängivenhet, hårt arbete, ödmjukhet, autenticitet och mod var de egenskaper som värderades. De sökandes självpresentation, förmågan att visa upp ett akademiskt och moraliskt jag, var således en viktig del i definitionen av excellens över disciplingränserna. En bedömare svarar på Lamonts fråga: "To what extent do you feel that you are rewarding a moral self or a different intellectual identity?" att: "We're unquestionably doing that. Unquestionably. A substantial part of the way the committee works is by its response to the person, and what you very nicely call the 'moral self'. That is, these are not quantifiable reactions. These are not reactions that you could necessarily document. But there's no question". (Lamont 2009: 198).

Att karaktärsbedömningar spelar roll i vetenskaplig bedömning är i samklang med en annan studie av Lamont, Kaufman och Moody (2000). I studien undersöktes vilka kriterier på vetenskaplig excellens som användes av de sökande till ett prestigefyllt stipendium. Här spelade självpresentation och uppvisande av värderade moraliska egenskaper roll på ett sätt som liknade bedömarnas resonemang om moraliskt goda egenskaper hos de sökande – ett slags deltagande i samma moraliska diskurs kan skönjas. Bedömningar av de sökandes moraliska egenskaper tycks spegla bedömarnas idealiserade bild av vad som ger det akademiska livet dess moraliska värde, menar Lamont: hängivenhet, ödmjukhet, autenticitet och en syn på den akademiska banan som ett djupt känt kall.

Homofili. Utöver epistemologiska stilar och informella kriterier, eller kanske snarare som en röd tråd genom hela bedömningspraktiken, identifierar Lamont en mekanism i bedömningsprocessen som hon kallar för "homofili", vilken kan fungera som en mer eller mindre medveten bedömningsgrund.¹² Det är tendensen att se excellens i den forskning (och den forskare i linje med ovanstående) som mest liknar den forskning man själv ägnar sig åt eller skulle vilja åstadkomma. En bedömare är medveten om tendensen: "I see scholarly excellence and excitement in this one project on food, possibly because I see resonance with my own life, my own interests, who I am, and other people clearly don't. And that's always a bit of a problem, that excellence

¹² "Homophily" definieras av Lamont (2009: 8) som "an appreciation for work that most resembles one's own".

is in some ways...what looks most like you.” (2009: 130). I intervjuerna berättar bedömare upprepade gånger hur deras preferens för ett visst projekt har att göra med deras personliga smak och erfarenheter. Att de inte fann någon anledning att dölja detta tolkar Lamont som att bedömningen av vad som är ”intressant” forskning (ett av de informella kriterierna) för dessa bedömare inte kan göras på ett abstrakt sätt skilt från vad som fångar deras intresse på ett personligt plan. Vissa bedömare var medvetna om att det som fångar ens intresse inte nödvändigtvis är detsamma som excellens, och försökte använda andra kriterier för de projekt de uppfattade som ”tråkiga” för att ge dem en chans. ”Homofilin” kom emellertid oftast till uttryck genom att bedömare klädde sina preferenser för vissa projekt i universalistiska termer, exempelvis genom att tolka kriteriet ”originalitet” i överensstämmelse med den egna forskningens anspråk på originalitet.

Pragmatisk rättvisa

Hur kan bedömarna med sina skilda bedömningskulturer och diversifierade kvalitetsbegrepp komma fram till gemensamma beslut som uppfattas som legitima och rättvisa? Med den frågan riktar Lamont uppmärksamheten på själva beslutsprocesserna.

Den viktigaste regeln för att få beslutsprocesserna att åstadkomma vad som uppfattas som pragmatisk rättvisa är det Lamont kallar kognitiv kontextualisering. Det innebär att bedömarna låter den kvalitetsdefinition som värderas i den sökandes disciplin avgöra utfallet snarare än kvalitetsdefinitionen i bedömarens disciplin. Det betyder att bedömarna i paneler tillämpar disciplinrelevanta kriterier och respekt för kollegernas inomdisciplinära expertis. Regeln utgör en motvikt mot bedömarens egna uppfattningar om vad som är intressant forskning (utanför den egna disciplinen) och innebär att man erkänner existensen av diversifierade kvalitetsbegrepp. På samma sätt är brott mot regeln om kognitiv kontextualisering den vanligaste orsaken till de få förekommande fallen av verkliga konflikter i bedömningsprocesserna. En bedömare som inte visar respekt för en kollegas expertis inom dennes eget område eller försöker få igenom sin bedömning på områden där hon/han inte har expertis anses bryta mot en viktig informell regel. Lamonts beskrivning har en parallell i Travis och Collins iakttagelse av bedömarens hantering av kognitiv partikularism som en tvekan att gå in och upphäva kollegors bedömning och Langfeldts beskrivning av tyst konsensus i expertpaneler.

Vid sidan av kognitiv kontextualisering tillämpas ytterligare informella regler i beslutsprocessen. *Formering av allianser* sker när bedömare med liknande preferenser och definitioner av excellens gör gemensam sak i argumentationen för ansökningar, vilket bedömare medgav att de själva ägnade

sig åt och att de såg sina kollegor göra detsamma. De flesta bedömare uppgav i intervjuerna tydliga personliga, intellektuella eller teoretiska beröringspunkter med vissa bedömare och inte med andra, men såg inte allianser som korrupta eller illegitima. *Kohandel* innebär att underlätta för andra bedömare att nå sina mål med förhoppningen att de ska återgälda stödet. Kohandel uppfattas av en del bedömare som ett undergrävande av meritokratin eftersom de kor man handlar med inte nödvändigtvis har samma värde, och den som vinner kanske gör det ”på grund av politik” och inte värde (kvalitet). *Strategisk röstning* uppfattas av de flesta bedömare som en normal del av processen. Den sker genom att man rankar ansökningar inte enbart utifrån en kvalitetsbedömning, utan med tanke på hur rankingen påverkar andra ansökningars chanser. Den kan ske antingen genom att man rankar ett mediokert eller kontroversiellt projekt högre än vad som är motiverat för att öka dess chanser eller motsatsen, att man rankar ansökningar lägre än de förtjänar utifrån en kvalitetsbedömning för att öka chanserna för andra ansökningar att gå igenom. Den första typen av strategisk röstning ansågs accepterbar, medan den andra inte sågs som legitim av de flesta bedömare. Strategisk röstning och kohandel sker vanligtvis i slutet av processen då bedömare ska välja mellan ansökningar som alla kan ha olika brister, men brister av olika slag som kan vara svåra att bedöma i förhållande till varandra. I det här momentet spelar även tidsfaktorn in och bedömarens taktiska skicklighet och panelens maktdynamik får större betydelse än tidigare i processen.

Slutligen fann Lamont ytterligare faktorer som inte var direkt relaterade till innehållet i ansökningarna som påverkade beslutsprocessen. För att lyckas med en argumentation för en ansökan måste bedömare etablera en auktoritet i förhållande till sina medbedömare och hantera maktdynamiken i panelen. Auktoritet är ojämnt fördelat i bedömningspaneler och påverkas, inte enbart av vetenskaplig expertis, utan av etnicitet, genus och senioritet, aspekter som i sin tur samspelar med personlighet, redan existerande nätverk, institutionell tillhörighet och anseende. Lamont påtar att personlighet och självpresentation kan spela stor roll i beslutsprocessen. I ett fall nämnde bedömare en kollega med en stark personlighet som utövade ett stort inflytande på utgången av beslutet i en viss panel. Om denne bedömare bestämde att han inte gillade en ansökan i hans eget område hade den inga chanser att gå igenom. Ett antal bedömare nämnde genus i kombination med personlighet som en faktor som påverkade maktdynamiken i paneler: några kvinnliga bedömare påtalade att en mer krigisk argumentation eller bombastisk självpresentation än de själva kände sig bekväma med ibland verkade öka chanserna för framgång. Trots att de flesta bedömares syn på vad som är en god bedömare och kollega inte inrymmer sådana karaktärsdrag kan en aggressiv, envis och målmedveten personlighet ge utdelning, konkluderar Lamont.

Den viktigaste slutsatsen att dra av det Lamont visar oss i den kollegiala bedömningens svarta låda är att excellens eller vetenskaplig kvalitet inte är begrepp med konstant, enhetligt innehåll. Kvalitetsbegreppet är diversifierat och informella kriterier spelar roll för identifiering av "sann excellens". Lamonts analys visar att bedömningsprocesser är interaktiva processer där känslor, identitet och självrepresentation är integrerade delar av bedömningen. Kollegial bedömning kan därför inte beskrivas som en rent kognitiv process, som potentiellt kan korrumpas av extra-kognitiva faktorer.

Den mekanism som Lamont kallar "homofili", tendensen att definiera excellens som det som mest liknar den egna forskningen, illustrerar väl hur kognition, känslor och forskarens identitetsuppfattning är sammanvävda i bedömningsarbetet. Som vi ska se i nästa studie är begreppet tillämpningsbart på andra slag av kollegial bedömning.



KOLLEGIAL BEDÖMNING AV TIDSKRIFTSARTIKLAR

De studier jag diskuterat hittills har som empiriskt fokus bedömningsprocesser som de tar form i mångvetenskapligt sammansatta paneler, antingen för att bedöma forskningsansökningar (Travis och Collins 1991, Lamont 2009) eller retrospektiva utvärderingar av hela forskningsmiljöer (Langfeldt 2001). För att förstå hur fält av vetenskaplig kunskap formeras är emellertid ett annat sammanhang minst lika viktigt, nämligen bedömning av artiklar inför publikation i vetenskapliga tidskrifter. Med den ökade betoningen på bibliometriska mått på vetenskaplig excellens och som grund för medeltilldelning förskjuts frågan om hur vetenskaplig excellens främjas till frågan om hur kollegial bedömning i tidskrifter fungerar. Man bör alltså ställa samma frågor angående kollegial bedömning av tidskriftsartiklar som de man ställer om bedömning av ansökningar om medel och tillsättande av tjänster. Fungerar kollegial bedömning av tidskrifter som den mekanism som garanterar att den bästa forskningen gynnas? Är systemet rättvist?

Till skillnad från kollektiva bedömningsprocesser organiserade av forskningsfinansiärer eller av politiska myndigheter, vilka har en viss grad av offentlighet, så är bedömning av tidskriftsartiklar i hög grad undandraget offentlighet. Bedömningarna sker i enskildhet i bedömarens (*referee*) arbetsrum eller hem, och i tidskrifternas slutna redaktionsrum. Det innebär att studiet av bedömningar för tidskrifter är behäftade med andra slag av metodproblem än forskning om relativt sett mer öppna beslutsprocesser. De studier som existerar på området koncentrerar sig på olika aspekter av kollegial bedömning av tidskrifter i olika vetenskapsområden och ger inte några tydliga svar på hur processerna fungerar i sin helhet. I en forskningsöversikt undersöker Campanario (1998a, 1998b) om några generella slutsatser kan dras utifrån existerande delstudier av kollegial bedömning för tidskrifter.

De flesta tidskrifter använder sig av två eller flera bedömare, men ett system med en ensam första bedömare existerar också och används främst i naturvetenskap. I systemet med en granskare innebär ett positivt omdöme från denne oftast att redaktörerna säger ja till publicering. Om omdömet är negativt bör redaktörerna, enligt gällande professionella kod, skicka artikeln till ytterligare en bedömare. Forskare inom området har bland annat intresserat sig för bedömningssystemets tillförlitlighet (*reliability*) enligt den klassiska definitionen av vad som gör en metod tillförlitlig, nämligen att den ger samma resultat när olika personer använder den och/eller när samma person använder den vid olika

tidpunkter. Systemets låga grad av tillförlitlighet är väl dokumenterad. Cicchetti (Campanario 1998a; Cicchetti 1991) utförde en grundlig studie av systemet och fann en mycket låg grad av tillförlitlighet. I ett flertal av de studier som han använde sig av var överensstämmelsen i utfall mellan olika bedömare mindre än vad slumpen skulle åstadkomma. Liknande resultat finns i andra studier: även om bedömarna är ense om ifall en artikel ska avslås eller publiceras är de det på olika, ibland motstridiga grunder. Ofta består granskaroenigheten emellertid i att olika granskare bedömer och fokuserar på olika aspekter av artikeln, snarare än att vara ense på särskilda punkter. Precis som i fallen av kognitiv oenighet i bedömningar av ansökningar kan den ringa överensstämmelsen mellan olika bedömare av tidskriftsmanus tolkas på olika sätt. Mångfalden av synpunkter kan ses som ett tecken på kvalitet i processen, redaktörer kan vilja ha olika experters synpunkter på olika aspekter av manus, skriver Campanario. Om man jämför med expertpanelers sammanjämkningar av bedömningar som är ett resultat av diversifierade kvalitetsbegrepp, så är det i manusbedömningen redaktörerna som tar emot de i hög grad diversifierade bedömningarna utförda av granskare i enskildhet och som sedan utför arbetet att sammanjämka dem till ett beslut. I den här processen sker den avgörande operationaliseringen av vad som är vetenskaplig excellens.

Ytterligare en faktor som bidrar till låg tillförlitlighet i den kollegiala bedömningen är avsaknaden av riktlinjer och normer för bedömning av manus. Manusbedömare förses inte med en uppsättning objektiva kriterier som de uppmanas använda sig av, vilket gör att en typ av brist i ett manus kan ses som oacceptabel för en bedömare, men kan passera i en annan bedömares granskning, konstaterar Campanario. Man kan anta att diversifieringen av kvalitetsbegreppet kan bli större än i paneler, eftersom bedömarna inte konfronteras med varandras kvalitetsdefinitioner eller med riktlinjer. En studie av 429 bedömare för ledande samhällsvetenskapliga tidskrifter uppvisade avsevärda skillnader mellan granskare för samma tidskrift, varför forskarna drog slutsatsen att tidskriftsnormer och riktlinjer inte är lika avgörande för publicering som faktorer knutna till bedömarens ifråga (Campanario 1998a; Kerr, Tolliver och Petree 1977). I en annan studie var syftet att belysa granskaroenigheten djupare. Siegelman (Campanario 1998a; Siegelman 1991) fann att de stora skilljelinjerna mellan bedömarna gällde deras uppfattning av som utgör kvalitet i en vetenskaplig artikel.

En annan parallell till Lamonts studie är tendensen bland manusbedömare att gradera de manus högre som överensstämmer med deras egna vetenskapliga trossatser och teoretiska perspektiv, det vill säga i enlighet med den mekanism som Lamont benämner "homofili". I en studie där bedömarna värderade en samhällsvetenskaplig undersökning så att den bekräftade de egna ideologiska ståndpunkterna kunde man se det Lamont också belagt, strategin att klä sitt

positiva gillande av viss forskning i universalistiska termer, exempelvis som styrka i metod eller liknande (Campanario 1998b:282). Campanario understryker att sådana resultat inte är så överraskande eftersom forskare involverar sina jag i de teorier de arbetar med och ibland är villiga att försvara dem in absurdum, trots evidens som vederlägger teorierna ifråga. Här hänvisar Campanario till en belysande insikt av Boring (1964). Denne påpekar att en teori som byggt upp en forskares självbild har blivit en del av denne och att överge den är lika med självmord eller åtminstone en akt av självstypning. Ytterligare en studie påvisar att bedömare har en tendens att gilla manus som överensstämmer med de teoretiska paradigmen och metoder som är i svang vid de universitet där de själva är verksamma eller doktorerade. Tendensen till "homofili" i sitt negativa uttryck, det vill säga att bedömare helt avvisar manus som inte stämmer med deras egen forskning kan få än allvarligare konsekvenser. I en starkt specialiserad vetenskap är det troligt att bedömare då och då är konkurrenter till manusförfattarna och därför kan ha intresse av att refusera manus. Ett talande exempel fanns i en studie där en bedömare yttrat att "if this were true I would have already described it" (Campanario 1998b: 285; Riggs 1995: 225).

Campanario påpekar vidare att till skillnad från bedömningar av ansökningar så ska den bedömda personens tidigare meriter inte spela roll för bedömningen av manus – texten bör bedömas isolerad från författaren. Studier visar emellertid att i de sakkunnigas bedömningar förekom rikligt med emotionellt färgade argument och bekräftar alltså Lamonts argumentation för att bedömning är en emotionell process. En del redaktörer berättar att de ibland måste tillrättavisa sakkunniga på grund av deras personangrepp i sina bedömningar. Campanario hänvisar vidare till studier som påvisar samband mellan positiv manusbedömning och tillhörighet till samma universitet som bedömaren. I sin genomgång blandar Campanario studier som påvisar kognitiva beröringspunkter mellan bedömare och manusförfattare med studier som påvisar institutionella band dem emellan, vilka som litteraturen klargjort utgör olika slag av partiskhet, även om de kan vara svåra att skilja åt (jfr diskussionen av Travis och Collins ovan).

Campanarios genomgång är utförd elva år tidigare än publiceringen av Lamonts studie och refererar studier utförda under de föregående decennierna av ett annat slag av kollegial bedömning än den Lamonts diskuterar. Parallellerna till Lamonts studie – förekomsten av diversifierade kvalitetsbegrepp, avsaknad av riktlinjer, hög granskaroenighet, "homofili" och bedömarnas känslor och subjektivitet som en integrerad del av bedömningen – utgör goda skäl för att anta att dessa mekanismer är generellt förekommande i kollegial bedömning. Finns några liknande paralleller mellan kollegial bedömning i paneler och på tidskriftsredaktioner?

Som nämndes ovan så är bedömning av tidskriftsartiklar undandraget offentlighet och utförs av enskilda bedömare och tidskrifternas redaktörer, varför beslutsprocesserna ser annorlunda ut. För att belysa huruvida bedömare och redaktörer är de mest kvalificerade att bedöma manus inom sitt område, det vill säga huruvida de är experter inom det vetenskapliga område som tidskriften representerar, har forskarna gått tillväga på olika sätt. Ett sätt är att undersöka redaktörernas egen produktionsnivå och förekomst i citeringsindex. Studierna ger inte entydiga svar: en del tidskriftsredaktörer har även själva hög publikationsnivå och citeringstal, medan de i andra fall citeras mer sällan än de tidskrifter de är redaktörer för. Frågan är om bibliometrimåttet mäter deras kapacitet att bedöma manus; i en studie drar forskarna slutsatsen att det krävs andra kvaliteter för att fungera som tidskriftredaktör än för att vara en forskare med hög produktionsnivå som blir flitigt citerad. Lindsey (Campanario 1998b; Lindsey 1977) fann två typer av redaktörer: en hårt arbetande som bedömer ett stort antal manus och har avgörande inflytande på besluten om publicering och en redaktör eller redaktionsmedlem som i kraft av att vara en erkänd, framstående forskare som ofta citeras tjänar som tidskriftens frontfigur, men deltar mycket lite i det konkreta redaktionsarbetet. Det betyder att tidskrifternas roll som grindvakter för sitt vetenskapsområde legitimeras av en erkänt excellent forskare men själva grindvaktsarbetet utförs ofta av någon annan. Vilka är då de verkliga grindvakterna?

Hur redaktörer och granskare utses är föga utforskat, enligt Campanario, som menar att mystik omger denna process. Ett antal studier har funnit att partikularistiska kriterier, såsom tillhörighet till institution eller till ett visst universitet spelar roll när redaktörer ska utses, snarare än enbart grad av expertis. En tydlig tendens, liknande den ovan beskrivna med avseende på förhållandet frontfigur/redaktör, är att de mest framstående forskarna inom sitt område är för upptagna för att tjänstgöra som bedömare och det är därför inte givet att de mest framstående individerna i ett visst område fungerar som grindvakter. Redaktionerna måste söka efter mindre etablerade, yngre bedömare och några redaktörer gör detta med avsikt. En sådan praktik är emellertid inte nödvändigtvis skadlig för kvaliteten på bedömningarna; enligt studier är sambandet mellan kvaliteten på bedömningarna och akademisk status hos bedömarna inte det förväntade, de bästa bedömningarna görs ofta av doktorander. Det förekommer också att seniora forskare med tung arbetsbörda lägger ut de bedömningsuppdrag de åtagit sig på yngre kollegor och doktorander. Om detta är allmänt förekommande bland tidskrifternas bedömare blir en något förvånande slutsats av Campanarios genomgång att i den mån de akademiska finansieringssystemen förlitar sig på bibliometriska kvalitetsmått tycks de därmed ha lagt en hel del av ansvaret för vetenskapernas utveckling i händerna på doktorander.



BEDÖMNING AV TVÄRVETENSKAP

Slutligen ska ytterligare en aspekt av bedömningsprocesser, nämligen bedömning av tvärvetenskap, diskuteras. Eftersom genusforskning ofta har en tvärvetenskaplig karaktär är diskussionen särskilt relevant för den här rapportens syften. Vilka speciella krav ställs på bedömningar av tvärvetenskap? Finns några särskilda former av partiskhet som bedömningsprocesserna riskerar att ge upphov till vid bedömning av tvärvetenskap?

Grit Laudel och Gloria Origi (2006), redaktörer för ett temanummer av tidskriften *Research Evaluation* som ägnas bedömning av tvärvetenskaplig forskning, påpekar att tvärvetenskaplig forskning allt oftare är föremål för bedömning och utvärdering, men ännu finns få empiriska studier som undersöker just dessa bedömningsprocesser och ingen samlad syn på vilka kriterier som bör användas i bedömning.

Begreppet vetenskaplig kvalitet och hur det tillämpas, både som explicit formulerat kriterium och som ett intuitivt mått, utgör en central fråga i forskningen om bedömning av tvärvetenskap. Irwin Feller (2006) och Liv Langfeldt (2006) beskriver i sina respektive studier hur kvalitetskriteriet kan förvandlas till ett instrument som hämmar tvärvetenskap genom att de etablerade disciplinernas kvalitetsdefinitioner består. Med Travis och Collins analysätt riskerar tvärvetenskap att utsättas för kognitiv partiskhet i högre grad än en-disciplinär forskning. En sådan användning av kvalitetsargumentet ser Feller i funktion i tre undersystem av det amerikanska nationella finansieringssystemet, vilket i hans tolkning lett till att tvärvetenskaplig forskning fått nedskurna finansieringsmöjligheter. Langfeldt påpekar att bedömning av tvärvetenskaplighet lägger till ytterligare en dimension till frågan om hur risk för institutionell, personlig och kognitiv partiskhet ska förebyggas, eftersom förhållandet mellan de disciplinära kvalitetsdefinitionerna måste hanteras. Hon ställer frågan i vilken grad och hur kollegial bedömning kan främja eller hämma tvärvetenskaplig och annan icke-konventionell forskning. Langfeldt resonerar kring ett större antal studier, dels egna studier där vissa data om bedömning av tvärvetenskap framkommit, dels andras studier. Sammantaget innefattar materialet studier av bedömningsprocesser med stor bredd, både på makro- och mikronivå. Hon drar bland annat slutsatsen att kollegial bedömning av forskningsansökningar är det kritiska momentet för tvärvetenskaplighet, eftersom det är en konkurrensinriktad kontext där tvärvetenskapliga ansökningar konkurrerar med disciplinbaserade ansökningar. Om kvalitetskriterierna är disciplinbaserade riskerar det att missgynna tvärvetenskapliga ansökningar. Resultaten

av olika undersökningar divergerar, i några undersökningar fann man att tvärvetenskapliga ansökningar missgynnades, i andra inte. Här tycks finnas en skillnad mellan en mer vanligt förekommande tvärvetenskaplighet mellan discipliner där det finns en kontaktyta och det Langfeldt kallar radikal tvärvetenskaplighet. Radikal tvärvetenskaplighet är när perspektiv från minst två discipliner som inte normalt interagerar med varandra används i forskning, och det således inte finns en separat grund för bedömning. Mer ordinär tvärvetenskaplighet missgynnas inte lika lätt i kollegial bedömning med disciplinbaserade kvalitetskriterier som radikal tvärvetenskaplighet.

Medan tvärvetenskapliga ansökningar kan möta svårigheter beroende på hur starka grindvaktartendenserna i området ifråga är kan utfallet bli ett annat i retrospektiva utvärderingar på meso- eller makronivå. I dessa är tvärvetenskaplig forskning oftare en separat kategori och konkurrerar inte med disciplinär forskning. Dessa utvärderingar syftar inte heller till att fördela resurser och det är vanligt att man rekryterar bedömare som har en positiv inställning till hela området och är respekterade av de som utvärderas – om ett tvärvetenskapligt område, institution eller program ska utvärderas, söker man efter bedömare som uppskattar och har kompetens i tvärvetenskaplig forskning och inte efter bedömare som har en intolerant inställning till tvärvetenskap.

För de tvärvetenskapliga ansökningarnas möjlighet att beviljas är allokering av resurser och valet av bedömare viktigare än själva bedömningsprocessen, menar Langfeldt. Som policyinstrument för att främja radikal tvärvetenskaplig forskning är därför en särskild budget för tvärvetenskaplighet och en separat sammansatt bedömning mer effektiv än att förlita sig på den konventionellt sammansatta bedömningsprocessen medan en mer ordinär tvärvetenskaplighet kan bedömas utan kognitiv partiskhet i många fall beroende på bland annat val av bedömare.



SLUTSATSER OCH DISKUSSION

Den här översikten har syftat till att sammanfatta och ringa in de viktigaste insikterna från forskning om bedömningsprocesser. Den kvalitativa forskning som diskuteras här pekar på ett antal kritiska punkter som är gemensamma för kollegial bedömning såsom den fungerar idag som en internationellt förekommande modell för bedömning. I det följande ska jag lyfta fram några av dessa kritiska punkter och de förändrade synsätt på kollegial bedömning som forskningen i området uppmanar till. Avslutningsvis berör jag kort hur några av de generella slutsatserna kan tillämpas på bedömning av genusforskning.

Kvalitet konstituerad i bedömning

En grundläggande fråga som forskningen om kollegial bedömning aktualiserar är den om kvalitetsbegreppets ontologiska status. Inledningsvis i den här översikten påpekade jag att man inom forskarsamfundet inte måste avgöra ifall kvalitet i forskning är en objektivt existerande egenskap för att kunna utforma välfungerande bedömningsystemen. Däremot är det viktigt att understryka att det inte finns några andra sätt att bestämma kvalitet än *inom* bedömningsystem. Ett sådant synsätt främjar en uppmärksamhet på själva bedömningsprocessen, där forskarsamfundet måste enas om vilka definitioner och operationaliseringar av vetenskaplig kvalitet som ska gälla och hur en rättvis bedömningsprocess ska gå till.

Samband mellan kognitiv och social partiskhet

Den kvalitativa forskningen om kollegial bedömning öppnar en intressant möjlighet att analysera sambanden mellan kognitiv gränsdragningar inom ett vetenskapligt område och den sociala sammansättningen av fältet. Den kan därmed fördjupa förståelsen av partiskhet i kollegial bedömning, en förståelse som kan användas för att tolka den kvantitativa forskningens resultat och förfina de hypoteser den arbetar med. Travis och Collins påvisande av hur institutionell partiskhet i vissa fall kan tolkas som kognitiv partiskhet är ett exempel, men andra aspekter av ett områdes sociala sammansättning, exempelvis geografisk eller köns- och åldersmässig, kan också ha samband med kognitiva gränsdragningar inom fältet ifråga.

Frågan om vilka kognitiva gränsdragningar, om några, som sammanfaller med en viss könsmässig sammansättning av vetenskapsområden är till stora delar utforskad. Den omfattande jämställdhetsstatistiken i forskarvärlden visar entydigt att det finns stora variationer i fördelningen av kvinnor och män inom olika vetenskapsområden, där den stora skiljelinjen går mellan de, i stora drag, manligt dominerade naturvetenskaperna och kvinnodominansen inom, delar av, humanvetenskaperna. På en sådan makronivå sammanfaller därför troligtvis gränsdragningar mellan olika definitioner av vetenskaplig kvalitet med en gränsdragning mellan vetenskapsområden som också har en könsbaserad aspekt. Lamont (2009) ger ett exempel från humanvetenskaperna och den fördelning av epistemologiska stilar hon fann där. Studien visade att den konstruktivistiska epistemologiska stilen användes främst av kvinnliga bedömare.

En fråga att undersöka vidare är om det finns fler sådana överensstämmelser mellan fältens könsmässiga sammansättning och dominerande epistemologisk stil, exempelvis inom naturvetenskaperna där kvinnor och män fördelar sig på olika sätt inom olika områden. En sådan typ av samband mellan sociala och kognitiva kategorier skulle kunna belysa de mönster som jämställdhetsstatistiken uppvisar. Man kan anta att forskningsprojekt (genusforskning eller andra) formulerade i en konstruktivistisk epistemologisk stil har svårare att placera sig i topp i bedömningspaneler och forskningsområden där andra stilar, exempelvis positivistiska är dominerande. Detsamma gäller då för andra epistemologiska stilar som representerar en minoritet inom ett område, eller för projekt som är radikalt tvärvetenskapliga enligt Langfeldts definition.

Lamont et. al (2009) efterlyser mer forskning om hur genus och etnicitet påverkar användningen av olika epistemologiska stilar och uppfattningar om rättvisa bedömningsprocesser, något som är viktigt även för internationella jämförelser av kollegial bedömning.¹³

Diversifiering istället för standardisering

Den tredje kritiska punkten där forskningen föreslår ett förändrat synsätt är ifråga om synen på definitionen av vetenskaplig kvalitet. De forskare vars studier jag diskuterat föreslår samtliga förändringar i motsatt riktning av

¹³ Lamont med medforskare använder dock inte termen etnicitet utan ras (race) och syftar då på afroamerikanska forskare och bedömare. Jag väljer här termen etnicitet, då den föredras i svensk vetenskaplig text. Vidare kan noteras att Lamont (2009) även diskuterar bedömarnas medvetenhet om klassaspekter av bedömning och analyserar, i Bourdieus efterföljd, ett av de informella kriterier som bedömare i hennes studie använde (kriteriet elegans och kulturellt kapital) som klassrelaterat.

den rådande trenden i forskningspolitik. Istället för standardisering av skolor och tillit till *ett* enhetligt bedömningssystem – eller till och med centralisering av bedömning till en enda myndighet – föreslår forskarna diversifiering. Forskarna föreslår alltså en explicit formulering av de diversifierade kvalitetsbegrepp som redan existerar. Travis och Collins pekar ut som den största faran med ett statistiskt, standardiserat kvalitetsbegrepp att forskarna för att öka sina chanser att få pengar gör sina ansökningar mer konservativa och konventionella än de skulle vilja, vilket i förlängningen medför en stagnation av vetenskaplig utveckling. De menar att tvärvetenskaplig forskning, frontlinjeforskning, kontroversiella områden och riskfyllda nya forskningsuppgifter löper större risk att drabbas av kognitiv partiskhet än mainstream-forskning. Därför pläderar de för en mångfald av forskningsfinansiärer och bedömningsstilar. För stor koncentration av finansiering är farlig, hävdar de. Lamont (2005:22) argumenterar på samma sätt för det hon kallar "context-sensitive peer review", dvs kvalitativ, finiskalig bedömning med kvalitetsdefinition anpassad till området ifråga. I linje med samma tänkesätt menar Langfeldt (2006) att stringens och grundlighet bör anpassas till syftet. En för stringent och grundlig bedömning enligt konventionella mått kan exempelvis radera ut innovativ, radikalt tvärvetenskaplig forskning.

Vid sidan av en analys av kvalitetsbegreppet framstår användningen av kriteriet relevans som viktigt att kartlägga. Hur används kriterierna vetenskaplig respektive samhällrelig relevans inom olika vetenskapsområden? I vilken mån använder bedömarna relevansbegreppet för att uttrycka en preferens för vilka områden som bör beforskas? I vilken fas av beslutsprocesserna är relevanskriteriet viktigt, i själva kvalitetsbedömningen eller i den beslutsfattande fasen? Langfeldt (2006) påpekar att bedömare i sin värdering av vad som är relevant att forska om även hänvisar till generella icke-akademiska värdeomdömen, men hur det påverkar utgången av bedömningsprocesser är svårt att avgöra. En bättre kännedom om hur bedömare använder relevanskriteriet kan öppna för en dialog med forskningspolitikens och samhällets önskemål om vad forskarna bör forska om.

Sammansättning av bedömargrupp och beslutsprocess

Den fjärde kritiska punkten gäller synen på bedömargrupperna och beslutsprocesser. Istället för att följa konventionella, inte genomtänkta eller godtyckliga sätt att komponera bedömargrupper och därefter sätta tilliten till att standardiserade kriterier ska främja kvalitet, menar forskarna att sam-

mansättningen av bedömargrupperna är den viktigaste faktorn för utgången av bedömningsprocessen. Travis och Collins menar att då bedömare inte kan lösas från sina specifika intressen och kognitiva preferenser när de tjänar som bedömare – vilket vore den bästa lösningen – är den näst bästa lösningen en större interaktion mellan bedömare och forskningsråden som sätter ihop bedömargrupperna. En god kännedom om fältet måste vara vägledande för sammansättningen. Forskarna föreslår även återkoppling från de bedömda som ett led i bedömningsprocessen. Här finns exempel på att när så skett har alla upplevt processen som mer rättvis. Det kan tilläggas att återkoppling från de bedömda också kan ses som ett led i att upprätthålla den kännedom om fältet som krävs för att kunna forma fungerande bedömargrupper.

Langfeldt (2006) framhåller att en bättre förståelse av effekterna av olika organisationssätt av bedömningsprocesser möjliggör medvetna val om organisationssätt. Olika policymål – som att stödja innovativ frontforskning, stödja uppbyggandet av tvärvetenskapliga fält, stöd till att få forskning till en internationellt definierad forskningsfront, stöd till samhällsrelevant forskning, eller främjandet av konkurrens mellan de sökande – nås bäst genom olika bedömningsprocedurer. Diversifiering innebär då att man utser olika slag av bedömare för olika syften. Ifråga om tvärvetenskap menar Langfeldt (2006) att det i vissa fall kan finnas bättre alternativ än kollegial bedömning eller att kompletterande metoder kan användas. Om det finns en önskan att främja tvärvetenskaplig, innovativ forskning och man vill undvika den traditionella bedömningsprocessens risker med godtycklig utgång på grund av skarpa kognitiva gränser kan ett alternativ vara att sätta ihop en kommission som representerar de olika inriktningarna/berörda forskarna och höra deras åsikter (exempelvis om deras forskning redan är kollegialt bedömd i andra sammanhang).

Vidare kan man konkludera att beslutsprocessens inverkan på utgången måste ges uppmärksamhet. Samtliga studier påvisar att steget från kvalitetsbedömning till beslut om tilldelning av medel utgör en kritisk punkt. I den mån man anser att strategisk röstning och kohandel inte är förenligt med ett främjande av kvalitet kan ett första steg vara att tydligare skilja på kvalitetsbedömning och beslut i processen. Det kan ske genom att andra personer än bedömarna fattar beslut på grundval av tydliga principer för beslutsfattandet. Ett intressant fynd från Campanarios studie är att enligt tidskriftsredaktörer utför unga forskare, doktorander eller post-doktorer ofta bättre bedömningar än etablerade, äldre forskare. Etablerade forskare som representerar ett helt forskningsområde som de kanske personligen varit med om att bygga upp har också större skäl att blanda samman en kognitiv bedömning med en önskan att främja den egna forskningen, understött av "homofili"-tendensen. En möjlighet vore att i större utsträckning använda

unga forskare som bedömare och grunda besluten om fördelning av medel på medvetet formulerade principer om en rättvis fördelning som representerar för olika inriktningar i fältet kommit överens om, oberoende av kvalitetsbedömningen.

Bibliometri och kollegial bedömning

En femte kritisk punkt gäller bedömningssystemens användning av bibliometri. Diskussionen om bibliometri domineras av en diskussion av dess teknikaliteter, medan den kvalitativa forskningen kan belysa bibliometrimåttet på andra sätt. Lamont menar att bibliometriska mått hotar att skapa "lock-in"-effekter, det vill säga man permanentar det fenomen som man mäter, t ex genom att mäta excellens förhindras framväxten av nya poler av excellens. Kvantitativa, standardiserade mätmetoder verkar lättare skapa lock-in effekter än kontext-sensitiv, finskalig kvalitetsbedömning, menar hon.

En viktig kritisk punkt för vidare diskussion är överensstämmelserna mellan studierna av kollegial bedömning för tidskrifter och studierna av bedömning av ansökningar, exempelvis vad gäller den långtgående diversifieringen av kvalitetsbegreppet, förekomsten av informella kriterier och bedömarens vetenskapliga smakpreferenser. Campanario påtalar i sin studie av bedömning för tidskrifter risken för nationell partiskhet. I den mån nationella bedömningssystem som det svenska ökar betoningen på bibliometrimått och därmed beroendet av internationella, främst anglosaxiska tidskriftsredaktioners och bedömares kvalitetsdefinitioner bör effekterna för utveckling av forskningen analyseras. En möjlighet är att diversifiering som strategi för att främja vetenskaplig kvalitet i viss mån även borde vara en strategi för nationell diversifiering (i ett internationellt perspektiv). Ett intressant exempel är det franska akademiska systemet som inte anpassar sig till det amerikanskt dominerade systemet för kollegial bedömning och därmed inte heller till de dominerande kvalitetsdefinitionerna inom det systemet (se Lamont et al 2005).

Vad är en god bedömning av genusforskning?

Historiskt sett har genusforskningens kognitiva karta sammanfallit med en förändring i den sociala sammansättningen av forskarkåren. Ökningen av antalet kvinnliga forskare var en förutsättning för att genusforskningens kritiska perspektiv skulle ta form. Genusforskningens framväxt är också ett exempel på hur en ökad pluralism i forskarkåren, i det här fallet en könsmissig sådan, kan leda till förnyelse och utveckling av nya teoretiska per-

spektiv. Genusforskningens framväxt har också kastat nytt ljus över hur kön som social kategori inverkar på utvecklingen av vetenskaperna. Att genusforskningen historiskt sett burits fram av kvinnliga forskare betyder att den, åtminstone till en början, löpte risk att utsättas för partiskhet grundad på kön. Med etableringen av genusforskningen som vetenskapligt fält, med tillkomsten av manliga genusforskare och med en ökad medvetenhet om vikten av jämställdhet i forskning är förutsättningarna för en bedömning av genusforskning utan social eller kognitiv partiskhet idag goda. Svårigheter att bedöma genusforskning med en kontext-sensitiv kvalitetsdefinition kan främst finnas i områden där genusforskarna är få eller där de arbetar inom epistemologiska stilar som ligger långt från de dominerande i vetenskapsområdet. Samtidigt varierar naturligtvis förutsättningarna för en god bedömning av genusforskning mellan olika nationella kontexter, i förhållande till graden av vetenskaplig etablering och även inställning till jämställdhet.

Utgångspunkten för den här översikten är att en god bedömningsprocess också kan bedöma genusforskning på ett bra sätt. Av de insikter som den kvalitativa forskningen om bedömning ger är de som gäller bedömning av tvärvetenskap särskilt relevanta för genusforskning och borde kunna tillämpas för att förfina kollegiala bedömningar av densamma.

Ett kännetecknande drag för genusforskningen är att den existerar som ett teoretiskt perspektiv inom en mångfald av discipliner, inom såväl humanvetenskaperna som delar av naturvetenskaperna. Med en inomdisciplinär bedömning av genusforskning bedöms den i förhållande till rådande kvalitetsdefinitioner och definitioner av vetenskaplig relevans. Här liknar den annan tvärvetenskaplig eller mångvetenskaplig forskning, men äger möjligen en mer långtgående tvärvetenskaplighet än andra fält då den existerar inom såväl human- som naturvetenskaperna. I vissa områden utgör genusforskning vad Langfeldt (2006) kallar ordinär tvärvetenskaplighet, medan den i andra är ett exempel på radikal tvärvetenskaplighet. Tillämpningen av inomdisciplinära kvalitetskriterier får troligtvis olika effekter beroende på slaget av tvärvetenskaplighet.

Genusforskningens utformning som fält utgör ytterligare en utmaning för kollegial bedömning. Genusforskning existerar både inom andra discipliner och som ett eget fält och bör därför också bedömas i förhållande till utvecklingen av hela fältet. Den forskning som är excellent med en inomdisciplinär kvalitetsdefinition kan mycket väl också vara excellent när den bedöms i förhållande till genusforskning som självständigt vetenskapsområde. Men det är troligt att det uppstår skillnader ifråga om vilken forskning som har högst vetenskaplig relevans, det vill säga som ligger i forskningsfronten och har högst vetenskaplig prioritet för att utveckla området som helhet. I takt med att tjänster och akademiska titlar införts i ämnet genusvetenskap

utförs redan i omfattande grad kollegiala bedömningar av genusforskning, inte som ett perspektiv inom en disciplin, utan som ett självständigt vetenskapsområde. Om bedömningsprocessen ska anpassas till policy-målet, som Langfeldt (2006) föreslår, innebär det att sammansättning av bedömargrupp och utformning av kvalitetskriterier troligtvis bör skilja sig något åt i förhållande till om målet är att främja inomdisciplinär genusforskning eller att utveckla genusforskningen som självständigt vetenskapsområde. För det senare policy-målet bör kvalitets- och relevanskriterier vara definierade i genusvetenskapliga termer gemensamma för hela fältet, och förhålla sig till en internationellt definierad forskningsfront.

REFERENSER

- Boring, E G (1964): "Cognitive dissonance. Its use in science". *Science* 185:680–85.
- Bornman, Lutz (2007): "Bias cut: Women, it seems often get a raw deal in science – So how can discrimination be tackled". *Nature*, Vol 445 (7127), page 566.
- Bornman, Lutz, Mutz, Rüdiger and Daniel, Hans-Dieter (2007): "Gender differences in grant peer review: A meta-analysis". *Journal of Informetrics*, Vol 1 (3) pages 236–238.
- Brenneis, Donald (1994): "Discourse and Discipline at the National Research Council: A Bureaucratic Bildungsroman". *Cultural Anthropology* 9: 23–36
- Brenneis, Donal (1999): "New Lexicon, Old Language: Negotiating the 'Global' at the National Science Foundation". I Marcus, George (red): *Critical Anthropology Now*, Santa Fe N M: School of American Research Press., pp. 123-146.
- Campanario, Juan Miguel (1998a): 'Peer review for journals as it stands today – Part 1'. *Science Communication*, 19(3), pages 181–211.
- Campanario, Juan Miguel (1998b): 'Peer review for journals as it stands today – Part 2'. *Science Communication*, 19(4), pages 277–306.
- Chubin, D E and E J Hackett (1990): *Peerless Science*. State University of New York Press: New York.
- Cicchetti, D. (1991): "The reliability of peer-review for manuscript and grant submissions: A cross-disciplinary investigation". *Behavioural and Brain Sciences* Vol 14, pages 119–135.
- Cole, Jonathan R and Stephen Cole (1981): *Peer Review in the National Science Foundation: Phase two of a study*. Washington DC: National Academy Press.
- Cole, Stephen, Rubin, L and Jonathan R Cole (1977): "Peer review and the support of science". *Scientific American* 237:34–44.
- Cole, Stephen, Rubin, L and Jonathan R Cole (1978): *Peer Review in the National Science Foundation: Phase one of a study*. Washington DC: National Academy Press.
- Cole, Stephen, Jonathan R Cole and G.A. Simon (1981): "Chance and consensus in peer review". *Science* 214: 881–86.
- Einarsson, Jan och Tor G Hultman (1984): *God morgon pojkar och flickor. Om språk och kön i skolan*. Stockholm: Liber.
- Fürst, Elisabeth (1988): *Kvinner i Akademia – inntrengere i en mannskultur?* Oslo: NAVFs sekretariat for kvinneforskning.
- Guetzkow, Joshua, Lamont, Michèle och Grégoire Mallard (2004): What is Originality in the Humanities and the Social Sciences? *American Sociological Review* 69: 190–212.
- Kerr, S, Tolliver J och D. Petree (1977): "Manuscript characteristics which influence acceptance for management and social science journals". *Academy of Management Journal* 20:132–41.
- Lamont, Michèle, Jason Kaufman and Michael Moody (2000). The Best of the Brightest: Definitions of the Ideal Self among Prize-Winning Students. *Sociological Forum* 15:187–224.

- Lamont, Michèle and Gregoire Mallard (2005): *Peer Evaluation in the Social Sciences and the Humanities Compared: The United States, the United Kingdom, and France*. Report prepared for the Social Sciences and Humanities Research Council of Canada.
- Lamont, Michèle (2009): *How Professors Think. Inside the curious world of Academic Judgment*. Cambridge, London: Harvard University Press.
- Langfeldt (2001): "The decision-making constraints and processes of grant peer review, and their effects on the review outcome". *Social Studies of Science*, 31(6), pages 820–841.
- Langfeldt, Liv (2002): *Decision-making in Expert Panels Evaluating Research: Constraints, Processes and Bias*. Oslo. NIFU och Unipub.
- Langfeldt, Liv (2004): "Judging quality. Expert panels evaluating research: decision-making and sources of bias". *Research Evaluation*, vol 13, no 1, April, pages 51–62.
- Langfeldt, Liv (2006): "Risk avoidance. The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments". *Research Evaluation*, vol 15, no 1, April
- Laudel, Grit and Gloria Origg (2006): "Introduction to a special issue on the assessment of interdisciplinary research". *Research Evaluation*, vol 15, no 1, April.
- Lindsey, D. (1977): "Participation and influence in publication review proceedings: A reply". *American Psychologist* 31:799–804.
- Lutz, Catherine (1990): "The Erasure of Women's Writing in Socio-cultural Anthropology". *American Ethnologist* 17:611–627.
- Mallard, Grégoire Lamont, Michèle och Joshua Guetzkow (2009): "Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review". *Science, Technology and Human Values*.
- Marsh, H.W, L. Bornmann, L. and R. Mutz, H-D Daniel och A. O'Mara (2009): "Gender effects in the Peer Reviews of Grant Proposals: A Comprehensive Meta-Analysis Comparing Traditional and Multilevel Approaches". *Review of Educational Research*, Online July 6.
- Melin, Göran (2007): "Reviewing applications by women: Critical use of additive and reasoning evaluation methods". *SISTER Working paper 2007:68*
- Melin, Göran och Richard Danell (2006): "The top eight percent: development of approved and rejected applicants for a prestigious grant in Sweden". *Science and Public Policy*, 33 (10): 702–712.
- Merton, Robert (1973/1957): "Priorities in Scientific Discovery: A Chapter in the Sociology of Science" I Storer, Norman (red): *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: Chicago University Press.
- Mitroff, I.I och D. E. Chubin (1979): "Peer Review at NSF: A dialectical policy analysis". *Social Studies of Science* 19:199–232.
- Nervik, Beata (2006): "Ansökningar till strategiska forskningscentra. En jämförande undersökning av könsfördelningen i ansökningarna till strategiska centra respektive traditionella projektbidrag", Power-Point presentation, [Online]. Tillgänglig på [http://www.idas.nu/filer/Strategiska centra okt-06.,IDAS](http://www.idas.nu/filer/Strategiska_centra_okt-06.,IDAS), www.idas.nu, JUki 2007.
- Riggs, J E (1995): "Priority, rivalry and peer review". *Journal of Child Neurology* 10:255–56.

- Roy, Rustum: (1985): "Funding science: The real defects of peer review and an alternative to it". *Science, Technology and Human Values*, 10- 73–81.
- Travis, G D L and H M Collins (1991), "New light on old boys: cognitive and institutional particularism in the peer review system". *Science, Technology and Human Values*, 16(3): 322–341.
- Sandström, Ulf och Martin Hällsten (2008): "Persistent nepotism in peer-review". *Scientometrics*, 74: 2.
- Schilling, Peter (2007): "Att välja ut strategiska forskningscentra. En analys av Stiftelsen för strategisk forsknings bedömningsprocess". *SISTER Reports* 2007:66.
- Siegelman S S (1991): "Assassins and zealots. Variations in peer review". *Radiology* 178: 637:42.
- Strathern, Marilyn (red.) (2000): *Audit Cultures. Anthropological studies in accountability, ethics and the academy*. New York: Routledge.
- Turney, J (1990): "End of the peer show?" *New Scientist*, sid. 38–42.
- Vetenskapsrådet (2005): Hillevi Ganetz, Genusvetenskapliga projektansökningar inom humaniora och samhällsvetenskap – en uppföljning av Vetenskapsrådets beredning och utfall år 2004. *Vetenskapsrådets rapportserie* 15:2005.
- Vetenskapsrådet (2007a): Gunilla Carlstedt, Genus i projektansökningar inom medicin – en uppföljning av Vetenskapsrådets beredning och utfall år 2004. *Vetenskapsrådets rapportserie* 1:2007.
- Vetenskapsrådet (2007b): Susanne Andersson och Lars Jalmert, Genus i forskningsansökningar inom utbildningsvetenskap – en uppföljning av Vetenskapsrådets beredning och utfall år 2004. *Vetenskapsrådets rapportserie* 8:2007.
- Wennerås , C. and Agnes Wold (1997): 'Nepotism and sexism in peer review'. *Nature*, Vol 387, sid. 341–343.

Rapporten "Kollegial bedömning av vetenskaplig kvalitet – en forskningsöversikt" är en översikt över forskning om kollegial bedömning av vetenskaplig kvalitet (så kallad peer review). Syftet är att sammanfatta och ringa in de viktigaste insikterna från forskning om bedömningsprocesser, särskilt med relevans för genusforskning och jämställdhet.

Rapporten är initierad av Vetenskapsrådets expertgrupp för genus.



Klarabergsviadukten 82 | Box 1035 | 103 78 Stockholm | Tel 08-546 44 000 | vetenskapsradet@vr.se | www.vr.se

Vetenskapsrådet är en statlig myndighet som ger stöd till grundläggande forskning av högsta vetenskapliga kvalitet inom alla vetenskapsområden. Utöver forskningsfinansiering arbetar myndigheten med strategi och analys samt forskningskommunikation. Målet är att Sverige ska vara en ledande forskningsnation.

ISSN 1651-7350
ISBN 978-91-7307-169-7
