

Scores rapportserie 2020:1

# Tvetydigheter, oenighet och slumpens diskreta inflytande i peer review processen

Lambros Roumbanis

# SCORE

STOCKHOLM CENTRE FOR  
ORGANIZATIONAL RESEARCH

# Tvetydigheter, oenighet och slumpens diskreta inflytande i peer review processen

Lambros Roumbanis  
lambros.roumbanis@score.su.se

Scores rapportserier 2020:1  
ISBN 978-91-88833-06-8

Stockholms centrum för forskning om offentlig sektor  
106 91 Stockholm



2020-02-03

# Tvetydigheter, oenighet och slumpens diskreta inflytande i peer review processen

Lambros Roumbanis  
SCORE, Stockholms universitet

## Sammanfattning

Den moderna vetenskapen genomsyras av kollegial bedömning – peer review – som på olika sätt fungerar som en intern granskningsprocedur. När anslag ska fördelas till nya forskningsprojekt spelar denna sakkunnighetsbaserade instans en central roll i att generera kollektiverade beslut. Bedömningen av forskningsansökningars kvalitet har dock visat sig innehålla ett ganska stort mått av tvetydigheter och slump som på ett ofrånkomligt sätt inverkar på de beslutsunderlag som skapas. Trots att peer review kretsar kring att de sakkunniga tillsammans eftersträvar konsensus, präglas denna konsensus likväl av stora variationer och oenighet om hur vissa ansökningar ska värderas. Genom att närmare studera hur oenighet kommer till uttryck och hanteras inom olika beredningsgrupper är det emellertid möjligt att få en fördjupad förståelse för den subtila relationen mellan slump och konsensus i peer review processen.

Följande studie baseras på observationer gjorda på tio av Vetenskapsrådets beredningsgrupper inom natur- och teknikvetenskaper. Den metodologiska utgångspunkten som valts består i att illustrera ett antal fall av oenighet, för att därigenom belysa skillnaderna i hur man kom fram till en praktisk lösning. En problematik som särskilt kommer beröras har att göra med det faktum att oenighet påfallande ofta leder till att bedömarna slår ut varandras favoritansökningar. Detta tar i sin tur på den mer övergripande frågan om hur den enskildes expertkänsla i vissa situationer ställs i direkt motsättning till gruppens samlade expertis, något som kanske inte alltid medför de mest tillfredställande besluten i termer av innovation och risk.

## Förord

Denna rapport utgör det sista bidraget i mitt projekt om peer review, akademiskt omdöme och forskningsfinansiering. Projektet har anknytning till det på Score sedan flera år tillbaka etablerade forskningstemat, "Kunskapens organisering." Min grundambition har från allra första början varit att försöka sprida lite nytt ljus över en i vår tid väletablerad företeelse inom den akademiska världen – den kollegiala bedömningen av forskningsansökningar.

Värderingen av nya projektidéer både möjliggör och begränsar vilka vägar forskningen tar in i framtiden. Hur denna typ av bedömningsprocess går till i praktiken och hur resurser slutligen fördelas utgör därför en ytterst viktig policyfråga. Över hela världen satsas årligen gigantiska summor pengar på forskning och innovation. Likväl finns det fortfarande relativt få observationsstudier som ger en direkt inblick i vad som faktiskt sker när en grupp sakkunniga bedömer, förhandlar och fattar beslut om vilka ansökningar som ska beviljas anslag. Även om policyfrågan förvisso är av stor vikt i sig så har mitt ursprungliga vetenskaps sociologiska intresse för peer review processen kretsat kring det akademiska omdömet och expertisens villkor i denna process. Hur går det till när man fattar ett gruppbeslut baserat på ett flertal bedömares individuella synsätt om vad som betraktas som nyskapande och värdefulla forskningsidéer?

Materialiet till min studie är ganska unikt i sitt slag. Jag hade turen att få tillträde till Vetenskapsrådets (VR) beredningsgrupper inom natur- och teknikvetenskaper för utlysningssåret 2013. Min första redogörelse av detta material publicerades 2015 i Scores rapportserie med titeln, "Akademiska omdömen under osäkerhet: en studie av kollektiva ankringseffekter i Vetenskapsrådets beredningsgrupper". Föreliggande rapport kan ses som en fortsättning på flera av tankegångarna som fanns med redan i den första rapporten. Den begreppsliga utgångspunkten har visserligen ändrats en aning, fokus likaså. Men trots detta är det övergripande problemkomplexet i grunden detsamma – frågan om förhållandet mellan konsensus och slump i peer review processen. För att bättre förstå studien som helhet skulle jag dock vilja rekommendera den som är intresserad av detta ämne att även läsa den första rapporten; dels därför att det breddar den teoretiska diskussionen, dels för att flera relevanta observationer finns med där som fördjupar bilden av vad peer review i grunden handlar om.

Jag vill härmed rikta ett stort tack till alla involverade forskare och administratörer på VR som så vänligen gick med på att låta sig observeras under sensommaren 2013. Jag vill också passa på att tacka Adrienne Sörbom och Staffan Furusten som utan att tveka erbjöd sig att hjälpa till med att göra observationer i några av beredningsgrupperna. Deras insats resulterade i att jag fick ett betydligt rikare empiriskt material som jag senare har haft stor glädje av i mitt analysarbete.

“We are all agreed that your theory is crazy. The question which divides us is whether it is crazy enough to have a chance of being correct. My own *feeling* is that it is not crazy enough.”

– Niels Bohr

## Introduktion

Ibland är det enbart känslan eller intuitionen som skiljer experter åt i deras uppfattningar om en ny forskningsidé – en känsla som kan få ganska olika effekter beroende på situationen. Bedömningen av forskningsansökningars kvalitet har vid upprepade tillfällen visat sig hemsökas av ett tillsynes ofrånkomligt mått av slump som är kopplat till variationerna och samspelet mellan olika expertomdömen (Cole et al. 1981; Chubin och Hackett 1990; Fogelholm et al. 2012; Graves et al. 2011; Luukkonen 2012; Roumbanis 2017). Inom den vetenskapssociologiska forskningen om peer review har man sedan länge uppmärksammat vad man kallat ”the luck of the reviewer draw” som innebär att olika uppsättningar av bedömare praktiskt taget alltid leder till substantiella skillnader i beslut. Med hjälp av experiment har man kunnat påvisa att vissa ansökningar som rankats högst av en grupp skulle blivit utan anslag ifall beslutet istället baserades på bedömningarna från en helt annan grupp (Cole et al. 1981; se även Mayo et al. 2006). Vad som identifieras och erkänns som högsta vetenskapliga kvalitet kan variera åtskilligt både inom och mellan olika beredningsgrupper (Brezis och Birukou 2020; Pier et al. 2018). Men att bedömare värderar kvaliteten i ansökningar på olika sätt är kanske inte så konstigt, ”When confronted with multidisciplinary issues involving scientific complexity and uncertainty, even competent, honest, and disinterested scientists may arrive at different conclusions” (Mumpower och Stewart 1996, s.194). Villkoren för att skapa konsensus i en beredningsgrupp tycks snarare alltid präglas av motstridiga omdömen och tvetydigheter som kräver kompromisser. Detta kan dock medföra att genuint nyskapande, interdisciplinära och/eller mer riskfyllda projekt elimineras till förmån för de ”solida” projekten som är enklare att enas om i ett initialt skede (Franssen et al. 2018; Hackett 1990; Linton 2016; Luukkonen 2012). Samtidigt är forskningsfinansierarnas uttalade målsättning att deras resurser primärt sett ska delas ut till de forskare/forskargrupper vars projekt just framstår som de mest nyskapande, originella och med störst chans om att kunna leda till viktiga upptäckter.

Oavsett om förekomsten av variationer och oenighet i bedömning är att betrakta som ett svårt reliabilitetsproblem eller tvärtom som en naturlig följd av den kognitiva mångfalden som finns inom de flesta forskningsfält, så går det inte att bortse från de praktiska konsekvenserna som fördelningen av resurser medför i det epistemiska landskapet över tid (Avin 2015; Roumbanis 2019). Det görs visserligen ständiga försök till att förbättra och säkerställa peer review processens tillförlitlighet genom olika insatser. Trots detta kvarstår det faktum att all kvalitetsbedömning i grunden handlar om mer eller mindre kvalificerade tolkningar som baserar sig på bedömarnas förkroppsligade expertis. För vad är egentligen vetenskaplig kvalitet sett i relation till sådana förhållandevis svårfångade kriterier som innovation och originalitet? Även en sådan sak som att värdera tidigare meriter och publikationer – en central kvalitetsindikator i peer review – är inte heller alltid givet ur en bedömnings synpunkt, utan kan variera åtskilligt både mellan och inom olika discipliner (Hammarfelt 2017). Att försöka förutse och att i grupp komma överens om vilka forskningsprojekt som har störst chans att

kunna bidra till de mest värdefulla resultaten i framtiden har fått vissa kritiska kommentatorer att tala om peer review som en ”higher form of nonsense” (Ziman 1983).<sup>1</sup>

Syftet med den här studien är att belysa förhandlingssituationer där bedömare på olika sätt måste *hantera oenighet* under relativt stor tidspress. Genom att fokusera på just oenighet – hur det uppstår och hanteras – är det min ambition att försöka bidra till en mer kvalitativt inriktad förståelse om det diskreta förhållandet mellan konsensus och slump i peer review. Och detta följer Cole et als. (1981, s.885) rekommendation: ”In order to clarify the way in which the luck of the reviewer draw works, we must look at,...reviewer disagreement”. Materialet till följande undersökning baserar sig på observationer gjorda på tio av Vetenskapsrådets (VR) beredningsgrupper inom ramen för natur- och teknikvetenskaper. Inledningsvis kommer jag att belysa begreppet *akademiskt omdöme* utifrån en bredare vetenskapssociologisk och epistemologisk kontext. Peer review processen genomsyras av en mångfald av akademiska omdömen, vilket också är anledningen till att det förtjänar en särskild teoretisk uppmärksamhet. En av studiens centrala utgångspunkter är att oenighet många gånger handlar om skillnader i epistemisk känsla och intuition, vilket ibland kan vara svårt att artikulera under förhandlingarna. Detta berör i sin tur den mer övergripande frågan om hur den enskildes expertkänsla ibland ställs i direkt motsättning till gruppens samlade expertis, något som kanske inte alltid medför de mest tillfredställande besluten i termer av innovation och risk.

## **Hur är akademiskt omdöme i peer review möjligt?**

Inom all grundforskning existerar det som flera tidigare vetenskapsteoretiker har kommit att tala om som den ”essentiella spänningen” mellan tradition och innovation (Kuhn 1977; Foster, Rzhetsky, Evans 2015; Hackett 2005; Paavola 2001). I peer review utgör detta inneboende epistemiska spänningsmoment något som på olika sätt kan påverka varje enskild bedömare som fått till uppgift att värdera graden av originalitet och korrekthet i ansökningar; – hur väl är författaren/na till en ansökan förtrogen med kunskapen vid forskningsfronten och vilka nya idéer presenteras som eventuellt kan leda till substantiella bidrag? Att bedömare i många fall utgår från någorlunda likartade regler och standarder när de fäller sina omdömen kan vanligtvis förklaras med att de ingår och interagerar i liknande epistemiska kulturer, nätverk och forskarmiljöer (Knorr Cetina 1999; Lamont 2009; Peterson 2017). Men dessa regler och standarder utmanas ständigt genom det vetenskapliga tänkandets kreativitet och öppenhet inför verklighetens oändliga mångfacetter, eller som Poincaré för länge sedan påpekade: ”Is not human intelligence, more specifically the intelligence of the scientists, susceptible of infinite variation?” (1907, s.12). Även om meningsutbyten mellan forskare i allmänhet bygger på att det existerar etablerade begrepp och synsätt, så kan dessa meningsutbyten vara ytterst komplexa och innehålla olika former av tvetydigheter, vaghet och polysemi (McMahan och Evans 2018; se även Abbott 1997; Levine 1989). Sociologen David

<sup>1</sup> Här kan också nämnas Daniel H. Osmonds (1983) dräpande kritik av peer review systemet och det stora fokus på konkurrensutsatta forskningsanslag som så länge dominerat i USA, ”the word ‘competition’ has less meaning than is commonly supposed. It smacks of the quest for excellence, but may militate against it. Those who conduct ‘competitions’ must be more humble and realistic about the validity of what they do. In most cases they are in fact deciding that one shade of blue is competitively superior to another shade of blue, which is, of course, nonsense.” (s.99). Osmonds poäng är att tävlingen om forskningsanslag är betydligt mer öppen för godtycklighet därför att det är genuint osäkert att ”mäta” vilken ansökan som är bäst, till skillnad från riktiga tävlingar som t.ex. friidrott, där det oftast är enklare att avgöra med objektiva mättekniker vem som faktiskt lyckas springa snabbast eller hoppa högst.

Peterson har till exempel gjort intressanta observationer från det molekylärbio-logiska forskningsfältet och noterat följande:

“Rather than a picture of high consensus and cognitive integration, the molecular biology lab was a scene of constant, sometimes chaotic, change. Where bench-building is a possibility, it is a necessity. [...] Bench-building is a type of Schumpeterian ‘creative destruction’ that transforms practice from within.” (Peterson 2015, s.1219)

Det intellektuella klimat som odlas vid olika institutioner, forskningscenter och laboratorier följer ofta naturligt med in i samtalen under beredningsgruppers möten. Och även sådant som idiosynkrasi och personlighetsdrag kan spela en inte oansenlig roll i hur en ny teori eller forskningsresultat tas emot under en viss period (Kuhn 1977). Ta till exempel fallet med materialvetaren Daniel Shechtman och hans epokgörande upptäckt av kvasi-kristaller. Redan tidigt kom han att utsättas för hårt motstånd inifrån det egna forskningsfältet, inte minst från den berömde Nobelpristagaren i kemi, Linus Pauling, som under en konferens på ett föraktfullt sätt sade, ”there are no quasi-crystals, just quasi-scientists.” I sinomtid skulle dock de upptäckter som Shechtman gjorde under tidigt 1980-tal komma att resultera i det som idag allmänt betraktas som ett nytt paradigm inom modern kemi (Nobelprize.org 2011). Orsaken till att jag nämner detta fall är, att rent hypotetiskt sett skulle Shechtman förmodligen aldrig fått ett anslag för att forska vidare om kvasi-kristaller ifall Pauling hade fått bestämma. Och anledningen skulle då vara att både idén och resultatet på ett fundamentalt sätt stred mot Paulings djuprotade intuition om hur en kristall överhuvudtaget kan vara beskaffad. Fallet illustrerar, på en kollektiv nivå, hur svårt det kan vara att vid vissa tidpunkter identifiera och uppskatta vetenskapliga innovationer; det sätter fingret på den epistemologiska frågan om, ”the interconnections between the problem of innovation, disagreement and consensus” (Laudan och Laudan 1989, s. 226). Även om exemplet med Shechtman kan uppfattas som extremt, så är det knappast unikt i vetenskapens historia. Fallet säger oss också något väsentligt om den makt som enskilda bedömare kan utöva mot projektidéer som de av någon anledning inte gillar (Travis och Collins, 1991; van den Besselaar et al. 2018). Att få avslag flera år i rad kan vara förödande för en frisinnad forskare, inte minst därför att det kan göra det ytterst svårt att genomföra forskningen ifråga. Flera av de mest nyskapande och originella ansökningarna befinner sig påfallande ofta någonstans i den bedömningsmässiga gråzonen, detta på grund av att dess idéinnehåll lätt kan ge upphov till delade meningar om hur man ska tolka och rangordna dem. Här blir det relevant att även understryka vad Luukkonen (2012, s.57) har framhållit:

“What will tip such proposals to one or the other side of the boundary may depend on contingent factors. It should also be noted that the customary rule of deferring to expertise implies that the fate of a proposal can depend on the views of only one or two experts.”

En enskild bedömare kan med andra ord få ett avgörande inflytande, på ett positivt eller negativt sätt, över hur vissa ansökningar slutligen värderas. Men ofta är det likväl de aggregerade medelvärdena som styr gruppens allmänna överenskommelser och som medför att olika fraktioner av majoriteten vinner. Den ”evaluative culture” som Lamont (2009) har analyserat och som exempelvis innebär att bedömare förväntas ta hänsyn till varandras expertis, bygger samtidigt på att bedömare själva lägger en heder i att försöka vara kritiska på ett så opartiskt sätt som möjligt. Men vad betyder det i praktiken? Grundidén med peer review består i att man vill undvika att enskilda bedömare ska få oproportionerligt mycket inflytande; istället vill man ge utrymme för gruppbeslut baserat på *konsensus*. En beredningsgrupp är tänkt att genomsyras av forskarsamfundets grundvärderingar, vilket innebär att de både bör slå vakt om det som anses vara beprövade sanningar och objektiva fakta, samtidigt som det ska värdesätta originalitet, självständighet, integritet och genuint



nytänkande (Polanyi 1962). Och detta ska på ett eller annat sätt avspegla sig i de beslut som gruppen fattar gemensamt. Vad som emellertid sker i ett initialt skede av bedömningsprocessen är att avvikande expertomdömen lätt tenderar att reduceras i relation till gruppens aggregerade medelvärden och rangordningen. Ett positivt omdöme har ofta svårt att hävda sig mot flera negativa, men ett enda negativt omdöme kan lätt rasera chansen för en ansökan där två andra varit positiva. Men som nämndes ovan kan det ibland uppstå situationer där vissa omdömen får en slags vågmästarfunktion, och de är i dessa många olika brytpunkter som slumpen på ett diskret sätt tycks uppstå och integreras i den konsensus som gruppen gemensamt lyckas konstruera.

I en studie genomförd av Pier et al. (2018) replikerades och analyserades samtliga bedömningsaspekter i amerikanska NIH peer review process. Resultatet visade tydligt att bedömarna värderat kvaliteten i ansökningarna på väldigt olika sätt, såväl i ord som i siffror. Den centrala slutsatsen man kom fram till i studien var att trots att alla bedömare fick exakt samma instruktioner om hur de skulle betygsätta ansökningarna och hur de skulle skriva sina utlåtanden, så fann man "no agreement in how reviewers 'translated' a given number of strengths and weaknesses into a numerical rating" (Pier et al 2018, s.2952). Även om bedömare vanligtvis inte är oeniga om exakt allt, så bidrar studien ändå med viktig evidens om hur olika vissa bedömare kan tolka kvaliteten i ansökningar. Peer review handlar aldrig enbart om att bedöma kvaliteten i ansökningar, utan i lika hög grad om att bedöma rimligheten i kollegornas omdömen, och att därtill omvärdera sitt eget omdöme i relation till den nya information som framkommer under samtalen (Roumbanis 2017).

### *Konsten att bedöma forskningsidéer: kognitiv partikularism och det akademiska omdömet estetisk*

Omdömesakten kan, vetenskapsfilosofiskt betraktat, förstås som en vital länk mellan tänkande och handlande; det står för ett kreativt urskilningsmoment och en förmåga att skapa en kritisk distans till det som ska bedömas. Michel de Certeau (1984, s.72) har skrivit: "acknowledging an *art* at the root of thought, makes judgment a 'middle term' ...between theory and praxis." Att värdera forskningsansökningars kvalitet kan ses som en särskild form av tolkningskonst som drar upp olika gränser. Travis och Collins (1991, s.327) har påpekat att både likhet/olikhet i bedömning har att göra med det man kallar för *kognitiv partikularism*, det vill säga, "the existence of cognitive boundaries within and between scientific specialities and disciplines." Vad detta begrepp särskilt fokuserar på är att bedömare av naturliga skäl föredrar sådan forskning som ligger dem nära intellektuellt och kunskapsmässigt. På ett plan kan detta synas ofrånkomligt och nästintill tautologt. För varför skulle en bedömare överhuvudtaget ha anledning att förorda en annan ansökan framför den ansökan som direkt tilltalar honom/henne mest – vore det inte psykologiskt motsägelsefullt? Här kan också det epistemologiska krusket med "confirmation bias" nämnas, alltså att forskare generellt sett tenderar att bedöma forskning som överensstämmer med deras egna befästa uppfattningar, standarder och förväntningar mer positivt än sådan forskning som inte gör det (Hergovish et al. 2010; Lee et al. 2013; Mahoney 1977). Likväl är själva frågan om kognitiv partikularism långt ifrån trivial, tvärtom frammanar det tanken om att sambandet mellan graden av intellektuell distans och bedömningen av vetenskaplig kvalitet faktiskt aldrig är helt given i peer review. Den bedömaren som står intellektuellt närmast ett forskningsämne kan ofta vara den allra hårdaste domaren och avfärda projektidéer som en annan bedömare med större distans skulle uppfatta som spännande och nyskapande (Boudreau et al. 2016). Fast i vissa fall kan den som står närmast vara den som lättast kan se subtila och spännande aspekter i ett projekt som en annan bedömare med större distans inte direkt kan urskilja. Vad som attraherar och engagerar vissa

bedömare intellektuellt men inte andra, kan många gånger vara väldigt svårt att exakt peka ut. De flesta bedömare kan förvisso gå med på kollegornas åsikter till en viss gräns, eller som en av Lamonts intervjupersoner uttryckte sig: "I'm not sure that I could articulate what that point is, but I'm pretty confident that I'd know it when I see it. You develop a little bit of a nose for it" (Lamont 2009, s.1).

Men inte sällan uppstår det också situationer då två eller flera bedömare med ungefär samma intellektuella distans tolkar innehållet i en ansökan på helt olika sätt. Då är det ofta en fråga om det som Travis och Collins beskrev som gränsskillnader inom samma vetenskapliga specialitet. Det är i dessa fall som omdömet subjektiva aspekter kan få en ganska framträdande roll i tolkningen av olika kvalitetskriterier och den mening som läggs vid det som sägs i en ansökan. Även om en ansökan visserligen är relativt begränsad till sitt format så kan det meningsinnehåll som författaren försöker kommunicera ändå öppna upp för ganska olika visioner om vad det skulle kunna resultera i. Bedömare kan ibland t.o.m. läsa in saker i ett forskningsprojekt som inte finns direkt utskrivet i själva ansökan. Det handlar därför inte enbart om gränserna, utan även om variationer i *kognitivt djup*, något som i grunden är omöjligt att mäta därför att det rör den vetenskapliga kreativitetens förbluffande öppenhet inför det okända.

Mumpower och Stewart (1996) har påpekat att oenighet ofta har sitt ursprung i att två eller flera experter kombinerar relevant information på olika sätt eller att de lägger olika vikt vid olika delar av informationen i en komplex vetenskaplig bedömningsituation. De idéer och påståenden som görs i en ansökan kan till exempel ge upphov till ganska olika typer av associationer och insikter, och i flera tidigare experimentstudier har det visat sig att det är, "impossible to account for all the variance in individuals' judgements solely in terms of cues, at least when the judgements have an intuitive component" (s.197). I peer review använder sig bedömarna ständigt av sin intuition som är kopplad till deras förkroppsligade expertis; detta sker både när de enskilt granskar ansökningar och när de förhandlar i grupp. Akademiskt omdöme genereras både genom analytiskt regel-baserat tänkande och genom det kognitiva djup som intuitionen erbjuder. Vad som dock är viktigt att anmärka är att man inte bör uppfatta dessa två – det analytiska och det intuitiva – som helt åtskilda system, utan snarare som en betydligt mer komplex kognitiv samverkan som tillsammans grundar sig på bedömarens tidigare erfarenheter och expertkunskaper (se t.ex. Glöckner och Wittman 2010; Craciun 2018; Loureiro och Garcia-Marques 2018; Leschziner och Brett 2019). Det finns många olika typer av epistemiska kvaliteter som kan fylla en bedömare med en särskild känsla av intellektuell entusiasm och välbehag inför en viss forskningsidé. O'Loughlin och McCallum (2019, s.333) har påpekat att, "criteria such as coherence, fruitfulness, non-ad-hoc-ness, and internal consistency as *genuinely aesthetic criteria*...are also sought in theory selection and development in science." Både de estetiska och intuitiva aspekterna i det akademiska omdömet är på det hela taget ytterst viktiga för att förstå hur variationer och oenighet kan uppstå i peer review processen.

## **Mötesetnografiska observationer**

Materialet som ligger till grund för den här studien samlades in runt månadsskiftet augusti-september 2013, då tio av Vetenskapsrådets (VR) sammanlagt nitton beredningsgrupper inom ramen för natur- och teknikvetenskaper observerades under deras tvådagarsmöten. Förmånen att få observera dessa tio grupper har resulterat i ett relativt omfattande material. Valet av grupperna gjordes dels med syfte att få en ämnesmässig spridning och dels utifrån den varierande graden av internationella och inhemska bedömare. I några av beredningsgrupperna

var alla bedömare från Norden, vilket gjorde att man då valde att tala på ”skandinaviska” istället för på engelska. VR:s regler gör gällande att det måste finnas minst 20 % utländska bedömare representerade i varje beredningsgrupp. I den gruppen som hade högsta antalet utländska bedömare uppgick den siffran till 75%, men i merparten av grupperna dominerade bedömare från svenska universitet och högskolor. Då flera av gruppernas möten var förlagda parallellt under den aktuella perioden beslöt jag mig för att be två kollegor om hjälp med att göra observationer i några av grupperna. På så sätt fick jag ett mer omfattande material med fler grupper att jämföra. En ytterligare fördel med att vi var tre forskare som gjorde observationer var att jag kunde jämföra och diskutera mina egna iakttagelser med mina två kollegor efter mötenas slut. Därigenom hade jag en bredare ingång till det analysarbete som jag sedan på egen hand hade att genomföra.

### *Mötet som händelsernas centrum i peer review processen*

Utgångspunkten för denna studie var redan från första början att med ett så öppet sinne som möjligt beskriva vad som utspelade sig under beredningsgruppernas möten. Att studera formella möten har visat sig vara en fruktbar metod som kan bidra till närgående inblickar i hur samtal, beslutsfattande och konsensusskapande går till i praktiken (Haug 2015; Sandler och Thedvall 2017). Det som äger rum under möten ger ofta en god bild av vad som står på spel i en organisation och de meningsutbyten som kan få konsekvenser för många fler än de som närvarar under själva mötet. Socialantropologen Helene Schwartzman har identifierat vad det är som gör formella möten till en så central skådeplats vid studiet av organisatoriska processer:

“...a meeting is defined as a communicative event involving three or more people who agree to assemble for a purpose ostensibly related to the functioning of an organization or group, for example, to exchange ideas or opinions, to solve a problem, to make a decision or negotiate an agreement.” (Schwarzman, 1989, s.7)

Att studera peer review genom att observera beredningsgruppers möten ger ett unikt tillfälle att få kunskap om en av de allra viktigaste urvals- och kontrollmekanismerna inom dagens forskarsamfund. Det bidrar till en förståelse för hur man inom en relativt starkt begränsad tidsram måste jämföra sina olika omdömen mot varandra och skapa kompromisser för att nå ett gemensamt beslut. Mötena erbjuder, kort sagt, ”an ideal setting for observing competing academic definitions of excellence.” (Lamont 2009, s.3).

Den allmänna strategin som jag och mina kollegor enades om innan observationerna påbörjades var att försöka skriva ner så mycket som möjligt av det vi kunde se och uppfatta när vi var på plats. I linje med denna ambition försökte vi så obemärkt som möjligt utföra vårt arbete utan att på något nämnvärt sätt störa processen. Även om det alltid finns en potentiell risk med att vissa personer som observeras omedvetet anpassar sitt beteende, så var detta inget vi kunde se skedde direkt med anledning av vår närvaro. Överlag var bedömarna allt för upptagna med sitt eget arbete och verkade knappt berörda av att vi iakttog dem under två dagar. Några av bedömarna ställde nyfikna frågor till oss under pauserna om vad vi kunde observera, men på det hela taget verkade de flest ha en ganska avslappnad attityd till att vi genomförde studien.

Det som sades under mötena, då bedömningsarbetet var i full gång, utgör det mest väsentliga materialet till denna studie. Likväl tillkom även mycket annan intressant information under luncher och kafferasterna då bedömarna samtalade med varandra. Dessa samtal anknöt ibland

till sådant som rörde bedömningsarbetet, men även till helt andra typer av spörsmål. Under dessa pauser fick även vi som observatörer chansen att delta i flera samtal med bedömare och administratörer från Vetenskapsrådet. Dessa samtal utgör i sig en rik källa till kunskap om hur personerna ifråga tyckte och tänkte kring sitt arbete, och har både direkt och indirekt bidragit till den bredare förståelsen om peer review som socialt fenomen. Materialet till föreliggande rapport bygger på de samlade iakttagelserna i hur man skapade konsensus i de tio beredningsgrupperna, dock har jag här valt att särskilt fokusera på hanteringen av oenighet. Dessa gränsfall utgör i själva verket viktiga bedömningsmässiga brytpunkter under förhandlingarna som på ett bra sätt illustrerar de motsättningar som varje konsensusbeslut i efterhand alltid döljer.

### *Organiseringen av peer review: några grundläggande regler och villkor*

Peer review processen omgärdas av en mängd olika regler och riktlinjer. De viktigaste punkterna finns nedskrivet i en handbok som VR delar ut till alla bedömare och som är tänkt att ligga till grund för beredningsgruppernas arbete (Swedish Research Council 2019). Jag har inte för avsikt att i detalj gå igenom allt som står där, men jag kommer beröra några av de mer betydande reglerna som genomsyrar hela bedömningsarbetet. En av dessa regler är den som berör jäv och jävssituationer, något som VR tar på mycket stort allvar och gör allt för att motverka. Den som är jävig måste lämna rummet och ska absolut inte bedöma en ansökan om det råder någon som helst partiskhet. Huruvida man är partisk eller inte är dock ytterst sett en samvetsfråga som bedömarna själva bör redogöra för, och under de möten som vi observerade hände det vid ett antal tillfällen att jävsfrågor kom på tal när det rörde oklarheter. I några enstaka situationer uppstod till och med förhandlingar om huruvida någon i gruppen kunde anses vara jävig eller inte. En amerikansk studie (Gallo et al. 2016) har visat att många bedömare förefaller underskatta flera potentiella risker för jäv och därför också missar att inrapportera detta till forskningsfinansiärens administratörer. Man kan alltså förmoda att jäv ändå kan slinka in i systemet, trots de många försiktighetsåtgärder som organisationen vidtar samt bedömarnas egna goda uppsåt. När det finns ansökningar där man av någon anledning behöver kompletterande expertis, på grund av jäv eller brist på specialistkunskaper, rekryterar man ofta en eller ett par externa bedömare. Det är emellertid alltid upp till beredningsgruppen att ta ställning till hur dessa externa bedömares utlåtanden ska vägas in under överläggningarna.

Två villkor som särskilt påverkar bedömningsarbetet har dels att göra med hur *bedömningsmetoderna* är utformade och dels på *budgetens* storlek. Under observationerna som gjordes i den här studien föreföll det vara en allmän uppfattning bland bedömarna i samtliga tio grupper att budgeten till projektmedel var alldeles för snålt tilltagen. Detta berodde inte på att det fanns mindre pengar i budgeten än vanligt, utan därför att VR hade beslutat om att göra vissa större specialsatsningar detta utlysningssår. Flera av bedömarna uttryckte sitt missnöje över hur lite pengar man hade att fördela i relation till det omfattande arbetet som man tillsammans utförde. Budgetens storlek styr alltid hur många anslag som ytterst sett kan fördelas.

När det gäller bedömningsmetodernas utformning (betygsskalor, rankingförfaranden, granskningsansvar) kan det få en ganska påtaglig inverkan på utsikterna för projekt som till exempel uppfattas som säkra respektive riskfyllda. Vissa bedömningsmetoder kan resultera i att just säkrare och mer ”solida” ansökningar har större chanser att få anslag än riskprojekt. I en studie har Langfeldt (2006) uppmärksammat bedömningsmetodernas inverkan på bedömarnas handlingsmöjligheter. Anslagsgivare skiljer sig ofta åt en aning i hur de utformat

sina arbetsmetoder. Större statliga forskningsråd i Norden och flera andra länder runtom i världen använder sig dock ofta av likartade metoder inspirerade av amerikanska statliga forskningsfinansiärer som t.ex. National Institute of Health (Sandström 2009). Om man till exempel använder sig av grovranking och ett öppet beslutsfattande tenderar det, enligt Langfeldt, att kunna ge ett större utrymme för vetenskaplig mångfald och nyskapande projekt. Entusiastiska bedömare kan då i lite större utsträckning förändra sina kollegors åsikter om ansökningar som de först ansett vara för riskabla, perifera eller omogna. Langfeldt påpekar vidare, att ett större antal bedömare per grupp som sätter betyg på alla ansökningar i kombination med en fingradig betygsskala och medelranking, tenderar att bidra till ett mera grundligt och förutsägbart slutresultat. Det minskar risken för att godtycklighet och slumpen påverkar beslut. Men frågan om hur många bedömare som är mest optimalt är långt ifrån givet och beror delvis på vilken typ av peer review det gäller; utgångspunkten är att höja reliabiliteten och kvaliteten i beslut utan att det ska kosta för mycket resurser (Snell 2015; se även Mayo et al. 2006).

På VR använder man sig av en relativt öppen diskussion mot bakgrund av en medelranking som baserar sig på betygen från tre bedömare, och ibland även av en eller ett par externa bedömare. Man använder sig av en mer grovskalig betygsskala (1-7), men i flera av de observerade grupperna använde man sig provisoriskt av decimaler för att kunna sortera och rangordna ansökningarna. De kriterier som man betygsätter är: i) *Vetenskaplig kvalitet*, ii) *Nytänkande och originalitet*, iii) *Sökandes kompetens*, iv) *Genomförbarhet*, och v) *Sammanfattande betyg*. Kriteriet genomförbarhet betygsätts endast med en tregradig betygsskala. Sifferbetygen översätts till ord enligt följande: 7="Enastående", 6="Utmärkt", 5="Mycket bra till utmärkt", 4="Mycket bra", 3="Bra", 2="Svag", och 1="Dålig". Bedömarna har även gjort sina egna individuella rangordningar av ansökningarna, något som ofta används som argument under själva förhandlingarna ("Detta är min nummer 2 av 21" eller "Den har jag rankat näst sist"). Varje ansökan presenteras av den bedömare som har fått till uppgift att fungera som s.k. *förslagsställare*, och som dessutom måste skriva ett yttrande till den/de som skrivit ansökan. Ett yttrande skrivs i regel på ett standardiserat sätt om ansökan inte var särskilt nära att få anslag, men mer detaljerat och specificerat om det var på gränsen till att få anslag. Presentation av varje ansökan går vanligtvis till på så vis att förslagsställaren kort summera ansökan, meriter, och vilka för- och/eller nackdelar som finns med projektbeskrivningen i sin helhet. Detta följs av att förslagsställaren nämner sina betyg och vilken plats den aktuella ansökan får i dennes individuella rangordning. Därefter får de två övriga bedömarna berätta hur de värderat och betygsatt ansökan, vilket sedan följs av överläggningar och justering av betyg.

## Resultat och analys

### *Tidspress och tvetydigheter i bedömningsarbetet*

För att bättre förstå oenighetens kontextuella villkor i peer review bör ljuset även riktas mot några faktorer som på grundläggande sätt präglar bedömningsarbetet. En av de första sakerna man direkt blir varse när man observerar en beredningsgrupp är den ständiga tidspress bedömarna arbetar under. Enligt VR:s regler måste grupperna bli klara inom utsatt tid. Oftast fortlöpper diskussionerna utan några större hinder, men inte alltid. I vissa situationer uppstår det långdragna och heta meningsutbyten. Varje ordförande har en viktig roll som samtalsledare och som en slags ordningens väktare under förhandlingarna. Till sin hjälp har gruppen den preliminära medelrankingen – den ger en överblick och fungerar samtidigt som

en tidsbesparande gallringsmekanism. Denna metod medför dock att många ansökningar behandlas relativt snabbt och rutinmässigt, vilket en bedömare kommenterade på följande sätt under mötet: ”Det känns inte riktigt bra att behöva avverka ansökningar som på löpande band” (NT-6). Denna åsikt kom till uttryck i flera av grupperna. Många ansökningar får inte någon ordentlig chans till omvärdering – de förefaller snarare dömda på förhand genom det aggregerat av individuella betyg som ligger till grund för gruppens medelranking. Mest tid disponeras istället på de ansökningar som uppfattas ha en någorlunda rimlig chans att kunna få anslag; det rör sig då oftast om ansökningar som fått femmor och sexor i betyg. Dessa ansökningar är många gånger svårast att separera, eller som en bedömaren påpekade om några ansökningar som placerat sig intill varandra i rangordningen: ”They are really close, so in some sense it is impossible to chose” (NT-4). Denna form av tvetydighet är inte alls ovanlig, tvärtom uppstår det nästan alltid kritiska ögonblick i bedömningsprocessen då man måste välja mellan jämnvärdiga ansökningar. Under den andra mötesdagen blev en bedömare märkbart frustrerad över situationen och ställde då kritiskt frågan till sina kollegor om de tyckte att man hade lyckats åstadkomma särskilt mycket under den föregående dagen:

“We spent eight hours discussing this. Our views might have changed slightly. The top proposals has overall sevens, and then there is a group of proposals that have sixes. The next row of proposals in the ranking have fives. Are we happy with this? Should the sixes really be sixes?” (NT-4)

Detta följdes av en ganska lång tystnad. Kommentaren ger oss en bild av hur svårt det kan vara att dra upp gränsvärdena även med avseende på de allra bästa ansökningarna. Det hela skapar en slags andra ordningens tvetydigheter som delvis skiljer sig åt från de tvetydigheter som kan uppstå vid själva tolkningen av vad som kommuniceras i en ansökan. Gruppens ordförande hade precis innan påpekat att: ”Of course, it is a delicate thing to draw a line here” apropå alla de ansökningar som låg på +5 och som i likhet med de ansökningar som hade fått 6:or mycket väl förtjänade att få anslag. För hur bestämmer man var gränsen går mellan ”Utmärkt” (=6) och ”Mycket bra till utmärkt”(=5), måste det inte ligga en bedömningsmässig tvetydighet i hur man väljer att tillämpa betygen på de olika ansökningarna?

Medelrankingen uppdragar också en annan form av tvetydighet som anknyter till det som belystes ovan, nämligen till frågan om hur bedömarna individuellt har översatt de olika kvalitetskriterierna i sifferbetyg och vad eventuella skillnader inom gruppen återspeglar. Detta är en central fråga som varje grupp måste förhålla sig till när man ska justera sina betyg för att nå konsensus. Vad är en rimlig och rättvis justering? En av svårigheterna med detta är kopplad till vad flera av bedömarna påpekade under mötena, nämligen att betygsskalan i praktiken förefaller tillämpas både relativt och absolut, något som bidrar till att själva medelrankingen i sig blir ett något tvetydigt bedömningsinstrument. Överläggningarna kan ofta upplevas som starkt begränsande just på grund av den preliminära medelrankingen, men små skillnader i betyg kan få ganska stora effekter. Eller som bedömaren i en beredningsgrupp kritiskt påpekade: ”We can not follow the ranking,...if we do that, then we can all go home” (NT-4). Och ändå bygger hela systemet på att man faktiskt måste utgå ifrån medelrankingen, annars skulle en hel del av meningen med allt förarbete som gjorts innan mötet gå förlorat, något som en av administratörerna var snabb med att påpeka. Gallringsproceduren skulle tappa en av sina grundläggande funktioner och risken för godtycke skulle öka. Ändå går det inte att bortse från att de aggregerade medelvärdena lätt ger en slags illusion om en slags given kvalitetsordning. Två bedömare påtalade under en kaffepaus att det alltid finns ”tur” med i peer review, men att man ändå fann det intressant att man utifrån den preliminära medelrankingen tycktes vara eniga om de bästa och de sämsta ansökningarna i gruppen (NT-5). Men en annan bedömare i samma grupp uttryckte tvärtom sitt missnöje med att man inom gruppen hade värderat ansökningarna så fundamentalt olika

att vissa ”toppsökningar” hade hamnat i botten. En hel del handlar förstås om från vilket perspektiv man väljer att betrakta de kluster och variationer i betyg som sammanställs mekaniskt genom konstruktionen av rangordningen. Den kritiska frågan som flera bedömare ställde sig kan uttryckas så här: varför ska man rätta sig efter de preliminära medelvärdena som i vissa fall bygger på omdömen som inte alls motsvarar det som man upplever är en korrekt tolkning av den vetenskapliga kvaliteten i vissa ansökningar? Detta sätter fingret på en öm punkt i värderingsarbetet: hur kompromissvillig kan man tillåta sig vara, och hur ska man förhålla sig till majoritetens åsikter som utifrån ett aggregerat perspektiv känns helt orubbliga eller ibland fullständigt godtyckliga?

Ytterligare en sak som också komplicerar bedömningsarbetet är jävsregeln, som innebär att bedömare då och då tvingas lämna mötesrummet. Detta får konsekvenser såtillvida att de enskilda bedömare förlorar en del av inblicken under överläggningarna. I en situation då det uppstått en smärre dispyt så påtalade en av bedömare sina svårigheter med att värdera en av ansökningarna som låg på gränsen: ”it was difficult for me, because I didn't see many of the surrounding applications because of conflict of interests” (NT-4).

I samtliga grupper kämpade man hårt tillsammans för att försöka nå ett så rättvist och välbalanserat beslut som möjligt. Inflationen i betyg var dock ett återkommande dilemma som kom på tal. I en grupp skämtade en av bedömare om hur svårt det var att göra avvägningar baserade på jämförelser av innovation och framtidspotential i ansökningar som alla uppvisade hög kvalitet: ”Almost everyone is between 5,5 and 6,0. So we haven't done anything” (NT-10). En stor del av det kollektiva grubberiet kretsar kring att försöka vara så konsekventa som möjligt i hur man jämför kriterierna i olika ansökningar. Allt från hur man bedömer idéers originalitet och genomförbarhet, till publikationens värde eller värdet av externa bedömares utlåtanden står ständigt i fokus. En ganska stor del av tiden under mötena disponeras till att diskutera ansökningar som har relativt höga betyg och som alla ligger på gränsen till att få anslag. Denna del av sorteringsarbetet utgör en särskilt vital del i beslutsprocessen – små ändringar i poängsättning kan förändra hela rangordningen. Därför krävs stor noggrannhet. En annan grupp av ansökningar som man också lägger mer tid på är ansökningar där man gjort helt olika bedömningar för att närmare undersöka om det finns utrymme för en omvärdering.

#### *Oenighet som utslagningsmekanism: vad är ett ensamt positivt omdöme värt?*

På den stora projektorskärmen som är uppsatt på väggen inne i mötesrummet går det att få en överblick hur variationerna i betyg ser ut, men det är först under förhandlingarna som de verkliga meningsskiljaktigheterna kommer till ytan. I en av grupperna påpekade ordförande redan i början av första mötesdagen att betygen spretade rejält och att man därför skulle behöva ta flera tuffa diskussioner för att nå konsensus. Några av de hårdaste förhandlingarna ägde rum i just denna grupp. För att illustrera detta ska jag ta upp ett exempel med två ansökningar som följde på varandra under mötet. Jag vill på så vis försöka belysa oenighet såsom en del i ett större spel om den nya rangordningen som gruppen kämpar för att skapa. Vad jag också vill beröra med dessa exempel är att oenighet många gånger leder till att man slår ut varandras favoritansökningar som råkat få väldigt blandade betyg.

I den aktuella situationen som jag här ska jag ge en inblick i hade en av bedömare satt höga betyg på en ansökan som han ansåg verkligen förtjänade att få anslag. Detta var, enligt honom, en mycket innovativ och djärv ansökan. Men denna åsikt var han ensam om, för de två andra bedömare hade, av olika skäl, satt betydligt lägre betyg.

- Bedömare 1: ”Jag satte honom högt förra året också. Han kommer ursprungligen från USA men är nu verksam vid universitet X. Han har en självständig position där och jag har *en känsla* av att han verkligen försöker göra något eget. Han har egna idéer. Han har dock inte producerat så många artiklar, men jag är förtjust i att han gör aktiva val. Detta är ett riktigt bra projekt!”
- Bedömare 2: ”Den här ansökan är näst sist i min ranking. Visserligen är det nytt. Men jag bedömde sannolikheten för att det ska flyga som låg. Han har inte övertygat mig alls. Jag kanske är för kritisk, men...”
- Bedömare 1: ”Visst, det är ett högriskprojekt, och jag vågar inte lova att det kommer att flyga.”
- Bedömare 3: ”Jag tycker det verkar vara ett roligt projekt, spännande. Men jag kan inte så mycket om det här området.”
- Bedömare 1: ”Det är lite tokiga grejer, men det är roligt.”
- Ordförande: ”Det kanske inte är bra att ge pengar till för mycket högrisk.” (NT-5)

Detta följdes sedan av en lång diskussion om risk och hur man värderar detta i förhållande till graden av innovation. Hur ska man tänka kring innovationspotential kontra risken att projektet misslyckas? Någon i gruppen påpekade att vissa av ansökningarna som låg i toppen av rangordningen också innehöll ganska betydande mått av risk och att det kanske alltid är ofrånkomligt med projekt som försöker bryta ny mark. Bedömare 1 försökte på olika sätt övertyga sina kollegor om att detta projekt inte bara var en kopia av redan etablerad forskning, utan tvärtom representerade vetenskapligt nytänkande och stort mod. Detta var den huvudsakliga anledningen, påpekade han högt inför alla i gruppen, till varför han ville ge anslag till just denna ansökan. Trots detta fick han inget gehör från kollegorna och man beslutade att justera betygen så att ansökan till slut hamnade under strecket för anslag. Eftersom ingen ansökan bedöms isolerat, utan alltid jämförs i relation med andra ansökningar, blir det nu intressant att titta närmare på situationen strax därefter för att åskådliggöra hur oenigheten här tycktes ge ringar på vattnet. Nu var det helt ombytta roller och Bedömaren 1 fick uttryckte starka invändningar (med stöd från en annan tredje bedömare) mot den ansökan som Bedömare 2 nu istället höjde till skyarna:

- Bedömare 2: ”Det här är en ansökan som är väldigt intressant. Projektet är originellt och innovativt. Hon är välmeriterad och jag har tolkat hennes publikationslista som stark.
- Bedömare 1: ”Det här tyckte jag inte alls kändes särskilt nytt. Vad är det som är nytt?”
- Bedömare 3: ”Jag reagerade på att hon inte pratar om hur hon ska bygga upp alltsammans. Det är den triviala delen hon måste beskriva bättre,...men visst, hon har publicerat bra grejer. Men hon borde ha formulerat en modell i kinetiska termer.”
- (Kort diskussion i hela gruppen: Bedömare 1 uttrycker bl.a. sin oro över bristen på verkligt nytänkande och djärighet i flera av ansökningarna han läst.)*
- Ordförande: ”Det verkar svårt att höja henne?”
- Bedömare 2: ”Jag talar utifrån *min känsla* att hon kommer lyckas. Så tolkar jag det.” (NT-5)

I diskussionerna om de två ansökningarna som jag här tagit upp hänvisar både Bedömare 1 och Bedömare 2 till sina känslor gällande de respektive ansökningarnas höga kvalitet och grad av innovation. Det kan tolkas som en hänvisning till den egna djupare vetenskapliga intuitionen såsom en slags garanti om projektens dolda potential. Likväl kom ingen av ansökningarna att få anslag. Flera kriterier fanns med i bilden. Samtalen handlade stundtals om hur man skulle värdera olika meriter eller externa bedömares utlåtanden. Ibland kunde någon uttrycka att en viss publikation i en viss tidskrift inte alls var märkvärdig, och detta kunde då ge upphov till livlig debatt. De externa bedömarens utlåtanden kunde också användas strategiskt i argumentationen för eller emot en ansökan. Men när det väl kom till den centrala delen av bedömningen – tolkningen av det rent vetenskapliga innehållet – tycktes det uppstå en osynlig gräns. Det gick inte att komma längre än att man enades om att man var oeniga. Allt skedde i en professionell anda och med ett pragmatiskt sinnelag. Fast det gick samtidigt att ana en lite ödesmättad och något uppgiven stämning bland bedömarena som fick



se sina favoritansökningar elimineras. Det hela blir en oundviklig del av peer review arbetet; att systematiskt gallra bort även de ansökningar som vissa i gruppen med hjälp av sitt kognitiva djup och expertis finner vara ytterst intressanta och löftesrika. Budgeten sätter med nödvändighet gränser, men på ett allmänt plan handlar det också om att en enskild bedömares positiva känsla aldrig riktigt räcker till för ett ge anslag när konkurrensen är för stor. Oenighet fungerar därvidlag som en social utslagningsmekanism. I den beredningsgrupp som jag här hämtat exemplen ifrån gick det tydligt att märka hur oenigheten ofta låg kvar i luften även när det gällde flera andra intilliggande ansökningar. Inga av bedömarna var direkt ohövlige, men man eliminerade metodiskt några av ansökningarna där man tolkat kvaliteten helt olika och där synen på vissa av kvalitetskriterierna ibland kunde skilja sig åt markant. Vid ett tillfälle lade en av bedömarna i gruppen stor möda i att kritisera och underminera en ansökan som han uppenbarligen inte alls gillade, varpå en annan bedömare svarade: ”Jag tycker precis likadant som du, fast tvärtom.” (NT-5).

### *Oenighetens olika upplösningar: ”a simple twist of fate”?*

Situationen såg lite annorlunda ut i en annan beredningsgrupp. Gruppens ordförande var djupt kritisk till innehållet i en ansökan som han hade granskat. Samtidigt var de två andra bedömarna positiva till ansökan, vilket gjorde att förhandlingarna ganska fort hårdnade. Till en början fick man känslan att denna ansökan inte hade en chans. Ordförande var mycket tydlig med vad han tyckte om projektet: ”It doesn’t make sense! So my grading is based on disappointment. Is it really possible to do this?” (NT-16). De övriga två bedömarna höll dock fast vid att detta var ett ytterst spännande projekt, och efter en stund hade man kört fast ordentligt i diskussionerna. Meriterna var det inget fel på, inte heller graden av innovation. Så långt var alla tre eniga. Men tiden gick och man var tvungna att komma vidare. Hur skulle man nå konsensus? Var det inte enklast att låta det negativa omdömet bli utslagsgivande, precis som med flera andra ansökningar som man gått igenom innan? Detta kom dock inte att ske, istället vek sig tillslut gruppens ordförande, fast inte utan protest: ”The merits is not 7, it is more like 5. Now I have given up all!...Either it [the research idea] is completely wrong or really, really, new. If it is correct it might be a contribution” (NT-16). Ordförande bedömde att det fanns för mycket oklarheter och risk förknippat med projektet i sin helhet. Likväl enades man om att sätta betygen 7, 6, 5, 3, och med det sammanfattande betyget 6, vilket ledde till att denna ansökan (som först såg ut att var utdömd) kom att beviljas anslag. Men vad var det egentligen som fick omvärderingen av innovation och risk att ta en annan vändning när det gällde just denna ansökan ifall man jämför med fallen från den tidigare gruppen ovan? Var det så enkelt att ”två positiva mot en negativ” fick styra, tillskillnad från ”en positiv mot två negativa”? Utifrån de befintliga observationerna är det svårt att exakt peka ut de avgörande orsakerna till varför det faktiskt blev som det blev. Det finns dock en hel del evidens från tidigare studier som visat att ett enda negativt omdöme i princip kan sätta stopp för en ansökan.<sup>2</sup> Men ordförande i den här gruppen hade vid några tidigare tillfällen under förhandlingarna varit ganska pådrivande i att lyfta fram vissa ansökningar och sänka andra. Oenigheten hade byggts upp på ett sådant sätt att han rent förhandlingsmässigt låg i skuld och behövde visa på god kompromissvilja, detta trots att det tog emot ur en rent intellektuell och vetenskaplig synvinkel. Men varför denna ordförande bestämde sig för att släppa fram just denna ansökan och inte någon av de föregående ansökningarna som han bidragit till att fälla vid målnöret, framstår delvis som ett resultat av slumpens diskreta och mångfacetterade

<sup>2</sup> Travis och Collins (1991, s.335) uppmärksammade detta för många år sedan när man talade om: ”a blackball system wherby one poor grade can damn a proposal.” Se även, Van den Besselaar et al. (2018) om hur negativa omdömen kan förstöra chanserna för högriskprojekt.

inverkan i processen. Hade den aktuella ansökan behandlats lite tidigare i ordningen hade den kanske blivit utan anslag.

I en annan beredningsgrupp hade förslagsställaren satt höga betyg på en tvärvetenskaplig ansökan som han bedömde vara i absolut världsklass, vilket motsvarade sexor och sjuor på de olika kvalitetskriterierna. Han framhöll att denna ansökan var den nästbästa i hans egen rangordning. De övriga två bedömarna var mer återhållsamma i sina betyg och var överlag inte fullt så entusiastiska. Visserligen kunde ingen förneka att huvudförfattaren till ansökan ägde ett gott anseende inom sitt forskningsfält, med välciterade artiklar i *Science*, *PNAS* och flera högt rankade specialisttidskrifter. Efter att förslagsställaren lyckats övertala de två andra kollegorna om den stora potentialen med detta projekt var det en hel del som tydde på att den skulle ha goda chanser att få anslag. Men plötsligt blandade sig en fjärde bedömare in i diskussionerna. Denne bedömare hade inte läst ansökan, men uttryckte tydligt att dess ämne låg inom hans expertis. Då det fortfarande fanns tveksamheter och viss osäkerhet kvar i luften från diskussionerna erbjöd han sig att läsa ansökan under kvällen. När den andra mötesdagen kom och denna ansökan skulle diskuteras igen så framförde den fjärde bedömaren en förödande kritik. Hans argumentation gick bland annat ut på att den tvärvetenskapliga dimensionen i ansökan helt klart uppvisade en avsaknad på förståelse för utvecklingen vid den aktuella forskningsfronten inom den ena disciplinen och att man utgick från ett allt för reduktionistiskt synsätt när det gällde en avgörande teoretisk aspekt. Bedömare 1 som i det här fallet var förslagsställaren kommenterade den nya informationen genom att säga: ”Jag tar till mig dina synpunkter, men anser fortfarande att ansökan är mycket bra. Men jag kan gå med på att flytta ner ansökan en liten bit på listan” (NT-1). Här var det tydligt att man hade helt olika teoretiska perspektiv, att skillnaderna nästan låg på ett vetenskapsfilosofiskt plan. Man beslutade sig för att se närmare på de intilliggande ansökningarna för att se om man kunde stuva om en aning i rangordningen. Men då uppstår nya problem. Man är även oenig när det gäller värderingen av andra ansökningar, vilket gör att man väljer att ta en större allmän diskussion i gruppen. Men då måste man förhålla sig till olika jävssituationer som praktiskt innebär att vissa av bedömarna måste lämna rummet i tur och ordning beroende på vilka ansökningar som diskuteras; då missar några av ansökningarna som har chans att få anslag sina främsta ”beskyddare” inne i rummet under delar av omförhandlingarna. Till sist sänks emellertid betygen på den tvärvetenskapliga ansökan som extrabedömts av en fjärde gruppmedlem, men då betydligt mer än vad förslagsställaren själv hade gått med på tidigare. Utan att riktigt sätta hårt mot hårt får han se en av sina favoritansökningar hamna precis under sträcket för anslag, och detta delvis på grund av de många andra kompromisser som han deltagit i att förhandla fram då oenigheten var stor. Sammantaget hade förhandlingsutrymmet efter ett tag krympt avsevärt; men oenigheten var inte borta, bara borträngd i termer av konsensus till förmån för en annan bra ansökan vars författare råkade ha lite av turen på sin sida denna dag.

## **Avslutande diskussion**

Peer review handlar om att värdera nya forskningsidéer, något som inte sällan öppnar upp för tvetydigheter och variationer av olika slag. Det förefaller delvis vara ofrånkomligt och får med nödvändighet konsekvenser för vilka ansökningar som beviljas anslag. Genom dessa tvetydigheter och variationer uppvisar slumpen sin diskreta makt. Det har hävdats att djupt originella och utmanande ansökningar oftare får avslag därför att de tenderar att kunna uppfattas som svårbegripliga, konstiga, riskfyllda eller omöjliga att genomföra (Lamont 2009; Langfeldt 2006; Luukkonen 2012). Det kan många gånger vara ytterst besvärligt att enas om

vad som är att betrakta som nyskapande och mest lovande, inte minst inom naturvetenskapens olika forskningsfält, där det okända utgör en ständigt närvarande faktor som kräver idel nya djärva förslag. Eller som fysikern Freeman J. Dyson på ett underhållande sätt har beskrivit forskningen om elementarpartiklar:

”When the great innovation appears, it will almost certainly be in a muddled, incomplete and confusing form. To the discoverer it will be only half-understood; to everybody else it will be a mystery. For any speculation which does not at first glance look crazy, there is no hope.” (Dyson 1958, s.80)

Även om denna utsaga är formulerad på ett lite tillskruvat sätt så säger det oss ändå något om hur svårt det kan vara att bedöma nya forskningsidéer vid ett givet ögonblick. Utsagan säger oss dessutom något om den ”essentiella spänningen” som finns inom all grundvetenskap när det gäller förhållandet mellan tradition och nyskapande. I peer review processen ställs detta på sin spets genom att det handlar om projekt som ännu inte genomförts och där själva värderingen av vetenskaplig kvalitet endast kan vara en slags proxy för vad som syns lovande inför framtiden. Den projektidé som presenteras i en ansökan kan antingen leda till banbrytande upptäckter eller en riktig flopp, men de allra flesta projekt bidrar ändå vanligtvis på något sätt till vetenskapens ackumulation.

Vem som bedömer en specifik ansökan får nästan alltid en viss betydelse. En viktig pusselbit för att förstå hur de konkreta effekterna av ”the luck of the reviewer draw” fungerar i peer review processen går via frågan om hur bedömarna tillsammans hanterar oenighet. Detta har också varit det centrala ämnet för denna studie. Utifrån ett vetenskapsteoretiskt perspektiv har frågan om oenighet ofta uppfattats som både förbryllande och djupt problematiskt därför att det leder till en slags ofrånkomlig relativisering av begreppet vetenskaplig kvalitet. Vad betyder det att en ansökan samtidigt kan uppfattas både som briljant och medioker, beroende på vem som granskar dess innehåll? En del ser oenighet som ett dilemma som ytterst sett rör hela peer review systemets legitimitet och tillförlitlighet. Om vi dock bara för ett ögonblick vänder på perspektivet så kanske det inte är så konstigt att bedömning av nya forskningsidéer lätt kan ge upphov till oenighet. Christensen (2007, s.208) har exempelvis hävdat att: ”The fact that disagreement by epistemic peers is *possible* is a constant and inevitable consequence of our being nonideal thinkers”. Trots att en sakkunnig bedömare förutsätts ha en särskild kompetens när det gäller att kunna urskilja och värdera vetenskaplig kvalitet, så kan han/hon likväl göra misstag eller överskatta sitt eget omdöme. Och det är förstås här som peer review kommer in i bilden såsom ett socialt korrigerandum som medför att överlappande expertiser kan skapa ett mer tillförlitligt slutresultat genom dialogens hjälp; det hela fungerar som ett sätt att reducera osäkerhet och för att rättfärdiga beslut (Aspers 2018; Lamont 2009; Polanyi 1962).

Men trots detta kvarstår grundfrågan: vad händer i praktiken när man inte är eniga, men likväl måste fatta beslut? Nya osäkerhetsmoment uppstår på gruppnivå. Vad jag har försökt påvisa i denna studie är att oenighet i många fall leder till att man systematiskt tvingas eliminera en del av varandras favoritansökningar. De går helt enkelt inte att förhandla upp ansökningar där bara en bedömare är positiv, så är systemet utformat; strategin går ut på att gallra bort den stora merparten av ansökningar som inte ska få anslag. I andra situationer kan dock oenighet lösas på annat sätt, då exempelvis två positiva bedömare lyckas stå emot en kollegas negativa omdöme. Ibland kan det handla om subtila former av makt och status, men inte alltid. Ofta handlar det om att man måste vara pragmatisk och kompromissa på ett sätt som man i en annan situation aldrig hade gått med på. I andra situationer kan det röra sig om en bedömare som raserar allt genom stark polemik eller att tidigare fall av oenighet sprider sig till andra tilliggande ansökningar där man också är oeniga. Makt och ett pragmatiskt sinnelag är integrerat i peer review. Det problem som allt detta emellertid aktualiserar är att den enskilde

bedömarens förmåga att faktiskt kunna upptäcka det som andra i gruppen kanske inte direkt är förmögna att upptäcka, dessvärre ofta tenderar att gå om intet till förmån för en slags bredare gruppkonsensus – en konsensus som likväl ändå är fylld av tvetydigheter om vilka projekt som verkligen borde få anslag. Detta har onekligen sitt pris. Att inte bli hörsammad för den djupare känslan man har inom sig om att ett visst projekt skulle kunna resultera i något riktigt värdefullt, leder inte sällan till viss frustration och uppgivenhet, något som gick att observera i princip alla beredningsgrupper. Flera tidigare studier har visat att peer review systemet fungerar på ett sådant sätt att det nästintill automatiskt leder till en bortgallring av ansökningar som avviker för mycket från de gängse epistemiska standarderna eller som framstår som lite för riskabla (Boudreau et al. 2016; Brezis och Birukou 2020; Luukkonen 2012). I den här studien har jag emellertid försökt framhålla att risk också noga övervägs i bedömningsprocessen; risk är inget som negligeras, tvärtom utgör det ett naturligt samtalsämne under mötena. Däremot är det en mycket fin balansakt att urskilja rimligheten och värdet i att satsa på ett projekt som dras med stor osäkerhet och som mycket väl kan komma att misslyckas. Men för att verkligen få med de riktigt våghalsiga och ”galna” projekten krävs att flera kan se samma potential, vilket är själva stöttestenen i peer review. Att flera tycker likadant i en grupp kan i bästa fall vara en garanti för god vetenskaplig kvalitet, men det kan också vara ett utslag för konventionell smak som inte leder till särskilt mycket nyskande forskning i framtiden. Små variationer eller händelser spelar en avgörande roll i processen; saker som sägs eller inte sägs, kan påverka vilka ansökningar som beviljas anslag eller inte. Slumpen är således alltid diskret integrerad i den konsensus man når under förhandlingarna, men rör också en djupare kontrafaktisk dimension i all form av peer review. För vad hade hänt om andra sakkunniga hade gjort bedömningarna istället?

## Bibliografi

- Abbott, Andrew 1997, "Seven types of ambiguity." *Theory and Society* 26(2-3), 357-399.
- Aspers, Patrik 2018, "Forms of uncertainty reduction: decision, valuation, and contest." *Theory and Society* 47, 133-149.
- Avin, Shahar 2015, *Breaking the grant cycle: On the rational allocation of public resources to scientific research projects*. PhD thesis. University of Cambridge.
- Boudreau, Kevin J., Eva C. Guinan, Karim R. Lakhani, Christoph Riedl 2016, "Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science." *Management Science* 62(10), 2765-2783.
- Brezis, Elise och Birukou, Aleksander 2020, "Arbitrariness in peer review". *Scientometrics* <https://doi.org/10.1007/s11192-020-03348-1>.
- de Certeau, Michel 1984, *The Practice of Everyday Life*, Berkeley: University of California Press.
- Christensen, David 2007, "Epistemology of Disagreement: The Good News." *The Philosophical Review*, 116(2), 187-217.
- Chubin, Daryl E. and Hackett, Edward J. 1990, *Peerless Science: Peer Review and U.S. Science Policy*. Albany: State University of New York Press.
- Cole, Stephen, Cole, Jonathan, and Simon, Gary A. 1981, "Chance and Consensus in Peer Review." *Science* 214, 881-886.
- Craciun, Mariana 2018, "Emotions and Knowledge in Expert Work: A Comparison of Two Psychotherapies." *American Journal of Sociology* 123(4), 959-1003.
- Dyson, Freeman J. 1958, "Innovation in Physics." *Scientific American* 199(3), 74-83.
- Franssen, Thomas, Scholten, Wout, Hessels, Laurens K., and de Rijcke, Sarah 2018, "The Drawbacks of Project Funding for Epistemic Innovation: Comparing Institutional Affordances and Constraints of Different Types of Research Funding." *Minerva* 56(1), 11-33.
- Fogelholm, Mikael, Saara Leppinen, Anssi Auvinen, Jani Raitanen, Anu Nuutinen, and Kalervo Väänänen 2012, "Panel discussion does not improve reliability of peer review for medical research grant proposals." *Journal of Clinical Epidemiology* 65, 47-52.
- Foster, Jacob G., Rzhetsky, Andrey, and James A. Evans 2015, "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review* 80(5), 875-908.
- Gallo, Stephen A., Lemaster, Michael, and Glisson, Scott R. 2016, "Frequency and Type of Conflicts of Interest in the Peer Review of Basic Biomedical Research Funding Applications: Self-Reporting versus Manual Detection." *Science and Engineering Ethics* 22, 189-197.
- Graves, Nicholas, Barnett, Adrian G., and Clarke Philip 2011, "Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel." *BMJ* 34.
- Glöckner, Andreas and Witteman, Cilia 2010, "Beyond dual-process models: A categorisation of processes underlying intuitive judgment and decision making." *Thinking & Reasoning* 16(1), 1-25.
- Hackett, Edward 1990, "Science as a vocation in the 1990s: The changing organizational culture of academic science." *The Journal of Higher Education* 61(3), 241-279.
- Hackett, Edward 2005, "Essential Tensions: Identity, Control, and Risk in Research." *Social*

- Studies of Science* 35(5), 787-826.
- Hammarfelt, Björn 2017, "Recognition and reward in the academy: Valuing publication oeuvres in biomedicine, economics, and history." *Aslib Journal of Information Management* 69(5), 607-623.
- Haug, Christoph 2015, "What is consensus and how is it achieved in meetings? Four practices of consensus decision-making." In: J.A. Allen, N. Lehmann-Willenbrock, S.G. Rogelberg (Eds.): *The Cambridge handbook of meeting science*. NY: Cambridge University Press.
- Hergovich, Andreas, Schott Reinhard, and Burger Christoph 2010, "Biased Evaluation of Abstracts Depending on Topic and Conclusion: Further Evidence of a Confirmation Bias Within Scientific Psychology." *Current Psychology* 29, 188-209.
- Knorr Cetina, Karin 1999, *Epistemic Cultures*. Cambridge: Harvard University Press.
- Kuhn, Thomas 1977, "Objectivity, Value Judgment, and Theory Choice." In: *The Essential Tension: Selected Studies in the Scientific Tradition and Change*. Chicago: University of Chicago Press.
- Lamont, Michèle 2009, *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.
- Langfeldt, Liv 2001, "The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome." *Social Studies of Science*, 31(6), 820-841.
- Langfeldt, Liv 2006, "The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments." *Research Evaluation*, 15(1), 31-41.
- Laudan, Rachel och Laudan, Larry 1989, "Dominance and the Disunity of Method: Solving the Problem of Innovation and Consensus." *Philosophy of Science* 56(2), 221-237.
- Lee, Carole J., Sugimoto, Cassidy R., Zhang, Guo, and Cronin Blaise 2013, "Bias in Peer Review." *Journal of the American Society for Information Science and Technology* 64(1), 2-17.
- Leschziner, Vanina och Brett Gordon. 2019. "Beyond Two Minds: Cognitive, Embodied, and Evaluative Processes in Creativity." *Social Psychology Quarterly* 82(4), 340-366.
- Levine, Donald 1988, *The Flight from Ambiguity: Essays in Social and Cultural Theory*. Chicago: University of Chicago Press.
- Linton, Jonathan 2016, "Improving the Peer review process: Capturing more information and enabling high-risk high-return research." *Research Policy* 45, 1936-1938.
- Loureiro, Filipe och Garcia-Marques, Teresa 2018, "I know the rule, but I'll just go with my gut: is there a rational use of intuition?" *Thinking & Reasoning* 24(4), 469-497.
- Luukkonen, Tertu 2012, "Conservatism and risk-taking in peer review: Emerging ERC practices." *Research Evaluation* 21(2), 48-60.
- Mayo, Nancy E., James Brophy, Mark S. Goldberg, Marina B. Klein, Sydney Miller, Robert W. Platt, Judith Ritchie 2006, "Peering at peer review revealed high degree of chance associated with funding of grant applications." *Journal of Clinical Epidemiology* 59, 842-848.
- Mahoney, Michael J. 1977, "Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System." *Cognitive Therapy and Research* 1(2), 161-175.
- McMahan, Peter och Evans, James 2018, "Ambiguity and Engagement." *American Journal of Sociology* 124(3), 860-912.
- Merton, Robert. 1973. *The Sociology of Science*. Chicago: Chicago University Press.

- Mumpower, Jeryl L. och Stewart, Thomas R. 1996, "Expert Judgment and Expert Disagreement." *Thinking & Reasoning* 2(2-3), 191-211.
- Nobel Prize. 2011. <https://www.nobelprize.org/prizes/chemistry/2011/summary/>.
- O'Loughlin, Ian och McCallum, Kate 2019, "The Aesthetics of Theory Selection and the Logics of Art." *Philosophy of Science* 86, 325-343.
- Osmond, Daniel H. 1983, "Malice's Wonderland: Research Funding and Peer Review." *Journal of Neurobiology* 14(2), 95-112.
- Paavola, Sami 2001, "Essential Tensions in Scientific Discovery". In: Petri Ylikoski och Mika Kiikeri (Eds.) *Explanatory Connections. Electronic essays dedicated to Matti Sintonen*.
- Peterson, David 2015, "All That Is Solid: Bench-Building at the Frontiers of Two Experimental Sciences." *American Sociological Review* 80(6), 1201-1225.
- Peterson, David 2017, "The depth of fields: Managing focus in the epistemic subcultures of mind and brain science." *Social Studies of Science* 47(1), 53-74.
- Pier, Elizabeth L., Markus Brauer, Amarette Filut, Anna Kaatz, Joshua Raclaw, Mitchell J. Nathan, Cecilia E. Ford, and Molly Carnes, M. 2018, "Low agreement among reviewers evaluating the same NIH grant applications." *Proceedings of the National Academy of Sciences of the United States of America*. 115(12), 2952-2957.
- Poincaré, Henri 1907, *The Value of Science*. New York: The Science Press.
- Polanyi, Michael 1962, "The Republic of Science." *Minerva* 1(1), 54-73.
- Roumbanis, Lambros 2017, "Academic judgments under uncertainty: A study of collective anchoring effects in Swedish Research Council panel groups." *Social Studies of Science* 47(1), 95-116.
- Roumbanis, Lambros 2019, "Peer review or Lottery? A critical analysis of two different forms of decision-making mechanisms for allocation of research grants." *Science, Technology, & Human Values* 44(6), 994-1019.
- Sandler, Jen och Thedvall, Renita 2017, *Meeting Ethnography. Meetings as Key Technologies of Contemporary Governance, Development, and Resistance*. New York: Routledge.
- Sandström, Ulf 2009, "Cognitive Bias in Peer Review." *12th International Conference on Scientometrics and Informetrics (ISSI)*, Rio de Janeiro.
- Schwartzman, Helene 1989, *The Meeting: Gatherings in Organizations and Communities*. New York: Springer Science.
- Snell, Richard R. 2015, "Menage a Quoi? Optimal Number of Peer Reviewers." *PloS ONE* 10(4), e0120838
- Swedish Research Council. 2019. *Peer Review Handbook. Research project grant and starting grant in natural and engineering sciences 2019*.  
<https://www.vr.se/download/18.6969eb1a16a5bec8b59ac7/1557913247633/Peer%20review%20handbook%20NT%202019.pdf>
- Travis, GDL and Collins, HM. 1991, "New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System." *Science, Technology, & Human Values*, 16(3), 322-341.
- Van den Besselaar, Peter, Sandström, Ulf, Schiffbaenker, Hélène 2018, "Studying grant decision-making: a linguistic analysis of review reports" *Scientometrics* 117, 313-329.
- Ziman, John 1983, "The collectivization of science." *Proceedings of the Royal Society*, J.D. Bernal Lecture, B219, 1-19.