

Kognitiv bias och bedömning av vetenskaplig kvalitet:
En analys av nationell klassificering av vetenskapliga tidskrifter i Finland

Julia Nummenmaa & Annika Tuomi

Pro gradu-avhandling i psykologi

Handledare: Matti Laine

Fakulteten för humaniora, psykologi och teologi

Åbo Akademi

Åbo 2019

**ÅBO AKADEMI - FAKULTETEN FÖR HUMANIORA, PSYKOLOGI OCH
TEOLOGI**

Ämne: Psykologi	
Skribenter: Julia Nummenmaa och Annika Tuomi	
Rubrik: Kognitiv bias och bedömning av vetenskaplig kvalitet: En analys av nationell klassificering av vetenskapliga tidskrifter i Finland	
Handledare: Matti Laine	
<p>Abstrakt: Inom vetenskaplig forskning används kollegial granskning eller <i>peer review</i> för att bedöma kvaliteten på bland annat tidskrifter, forskningsanslag samt vetenskapliga texter och ansökningar. Den kollegiala granskningsprocessen har blivit utsatt för kritik och hävdats vara otillräcklig. Tidigare forskning tyder på att granskare kan influeras av faktorer som påverkar värderingen av vetenskapligt material. Till dessa faktorer hör kognitiva biaser såsom ingrupsbias, som kan uppstå då en person värderar fenomen som hen har varit i kontakt med tidigare högre än obekanta fenomen.</p> <p>I Finland ansvarar det nationella Publikationsforumet (Julkaisufoorumi, JUFO) och dess kommittéer för den kollegiala granskningen av vetenskapligt material. Syftet med denna avhandling var därför att undersöka om JUFO-kommittémedlemmars egna forskningsprofiler påverkade deras klassificering av vetenskapliga tidskrifter, således om det förekom ingrupsbias under bedömningsprocessen. Detta mättes genom att undersöka om kommittémedlemmar som hade publicerat i en tidskrift tenderade att värdera tidskriften högre än andra. I avhandlingen undersöktes även eventuell variation i bias inom naturvetenskaper, bio- och hälsovetenskaper, samhällsvetenskaper och humanistiska vetenskaper. Detta undersöktes genom att såväl jämföra kommittéklassificeringen med ett objektiva jämförelsemått (Impact Factor) som genom ett subjektivt frågeformulär som JUFO kommittémedlemmarna ombads svara på.</p> <p>Resultaten från denna avhandling tydde på att ingrupsbias förekom i någon mån vid klassificeringen av tidskrifter i Finland. Samhällsvetenskaper använde sig mest av Impact Factors under bedömningsprocessen. Effekten av antalet kommittémedlemmar som hade publicerat i en tidskrift på JUFO-klassificering var däremot signifikant störst inom samhällsvetenskaper i jämförelse med de andra ämnesområdena. Kommittémedlemmarnas åsikter om utvärderingsprocessen var varierande och bestod av både positiva och negativa synpunkter om JUFO-systemet. Resultaten bör tolkas med försiktighet då samplen var begränsade och skillnaderna mellan grupperna var relativt små.</p>	
Nyckelord: kollegial granskning, bias, JUFO	
Datum: 07.03.2019	Antal sidor: 33
Nivå: Pro gradu-avhandling	

Tack

Vi vill tacka följande personer som gjort genomförandet av denna undersökning möjligt: ett stort tack till vår handledare, professor Matti Laine, för all vägledning och alla idéer du gett oss under arbetets gång. Ett stort tack till professor Pekka Santtila, som var den som ursprungligen kom med idén att börja forska i bias i JUFO-systemet, och den som handledde oss i början. Vi vill också rikta ett varmt tack till doktorand Jussi Jylkkä, som gett oss ovärderlig hjälp med att utföra de statistiska analyserna. Tack till universitetslektor Mira Karrasch och alla deltagare på pro gradu-seminariet i psykologi hösten 2018 för att ni hjälpt oss se vår text med nya ögon. Tack till den forskningsetiska nämnden vid ämnet psykologi vid Åbo Akademi för tillståndet att sätta upp projektet.

Projektet skulle inte ha gått att genomföra utan en del förarbete och anskaffande av förhandsmaterial. Tack till Eeva Savolainen och Janne Pölönen på JUFO-sekretariatet för att ni bidragit med svar på våra frågor och med oumbärligt bakgrunds- och forskningsmaterial. Tack till Elias Lassila och Emma Vihervaara för ert arbete i vårt gemensamma forskningsprojekt 2016, där de preliminära analyserna av en del av materialet gjordes.

Till sist, men inte minst, vill vi tacka varandra för alla idéer och allt stöd på vägen, och all uppmuntran då det känts tungt eller då någondera kört fast. Ingendera av oss kunde ha önskat en bättre medskribent att arbeta med. Vi vill också tacka alla personer som står oss nära för ert stöd och för att ni förgyller vardagen.

Innehållsförteckning

1 Inledning	1
1.1 Kognitiv bias i bedömningar och beslutsfattande	2
1.2 Publikationsforum	5
1.2.1 Klassificeringen av tidskrifter	7
1.2.2 Undervisnings- och kulturministeriets finansiering av universitetsverksamheten.....	9
1.3 Studiens syfte	10
2 Metod	11
2.1 Material	12
3 Resultat	12
3.1 Analys av JUFO-bedömningen 2015	13
3.1.1 Signifikanta resultat mellan ämnesområdena	14
3.1.2 Figurer och interaktionseffekter	15
3.2 Analys av panelisternas upplevelse av bedömningen	17
4 Diskussion	24
4.1 Bias inom JUFO-klassificeringen av tidskrifter	24
4.2 Enkätresultaten	26
4.3 Begränsningar.....	27
4.4 Sammanfattning	27
Referenser	29
Bilagor	

1 Inledning

Kognitiva biaser uppstår då personliga erfarenheter, åsikter och utvärderingar påverkar en persons omdömesförmåga, beslutsfattande och bedömning. Trots medvetna försök till objektiva beslut och värderingar, påverkas alla människor i någon mån av kognitiva biaser vid beslutsfattande. Det här kan anses problematisk i många situationer, inte minst inom vetenskapliga sammanhang. Inom forskning är det vanligt att använda sig av kollegial granskning eller *peer review* för att utvärdera vetenskapliga texter, ansökningar och större helheter som forskningsprojekt och tidskrifter. Kollegial granskning är en av de äldsta metoderna inom vetenskaplig forskning och består av forskare som utvärderar vetenskapliga material och har som uppgift att särskilja det material som är av högsta vetenskapliga kvalitet och identifiera tänkbara brister. Besluten ska vara objektiva och basera sig enbart på materialets korrekthet och relevans (Bornmann & Daniel, 2004).

Förespråkare av kollegial granskning anser att metoden är den mest funktionella utvärderingsmetoden i vetenskapliga sammanhang, medan motståndare menar att den kollegiala granskningsprocessen är för bristfällig för att den ska anses vara ändamålsenlig (Bornmann & Daniel, 2004). En av fördelarna med kollegial granskning är att granskningsprocessen ökar sannolikheten för att speciellt högkvalitativ forskning gynnas. Granskningsprocessen ser även till att forskning av hög kvalitet publiceras i tidskrifter som anses vara av likvärdig kvalitet (Ware & Mabe, 2015). Kritiken som riktas mot kollegial granskning innefattar att det finns begränsat med vetenskaplig evidens som påvisar effektiviteten med processen (Demicheli & Pietrantonj, 2007; Jefferson, Rudin, Brodney Folsø & Davidoff, 2007). Metoden har också anklagats för bristfällig reliabilitet samt hävdats vara långsam och dålig på att upptäcka felaktigheter. Metoden ger även utrymme för ökad risk för biaser och misstag gjorda av granskarna då arbetet är såväl arbets- som tidskrävande (Ware, 2011). En enkät utförd av Ware (2008) visade däremot att 93 procent av enkättagarna inte höll med om att kollegial granskning skulle vara onödigt, 85 procent ansåg att kollegial granskning främjar vetenskaplig kommunikation och 83 procent ansåg att det inte skulle finnas någon kontroll utan kollegial granskning. Frågan skulle enligt Smith (2006) därmed inte vara huruvida man borde sluta med kollegial granskning eller inte, utan hur den kollegiala granskningsprocessen borde förbättras.

I nuläget genomförs den kollegiala granskningsprocessen ofta av granskare, redaktörer och författare som inte får någon ersättning för sitt arbete, trots att granskandet är

arbetsdrygt och tidskrävande. Utvärderingen av en vetenskaplig text tar i regel flera timmar till en hel dags arbete, men i genomsnitt 9 timmar (Ware & Mabe, 2015). Granskare av vetenskapligt material har en betydande inverkan på spridningen och omfattningen av kunskap inom olika vetenskapsområden. De har även inflytande på samhället överlag, eftersom deras evalueringar av vetenskapligt material avgör vilken forskning som understöds, publiceras för allmänheten och får erkännande av forskare. En enda negativ evaluering av en vetenskaplig text kan påverka det slutgiltiga resultatet av granskningen (Armstrong, 1996). De som granskar vetenskapliga texter har därmed stort inflytande på publikations- och anslagsbeslut (Crane, 1967).

I Finland bedöms sedan 2010 tidskrifters vetenskapliga nivå av ett nationellt organ, Publikationsforum (i fortsättningen JUFO, från finskans Julkaisufoorumi), som är en del av Vetenskapliga samfundens delegation (VSD), en takorganisation för de vetenskapliga samfunden i Finland (Publikationsforum 2017). JUFO-klassificeringar används för att rikta statsfinansiering till universiteten. Arbetet i JUFO sker i 23 kommittéer, som representerar samtliga vetenskapsområden som det görs forskning inom i Finland (ibid).

I föreliggande avhandling undersöktes om JUFO-kommittémedlemmarnas klassificering av vetenskapliga tidskrifter påverkades av deras tidigare publikationer i tidskrifter de bedömde, dvs. om de tidigare publikationerna utgjorde en bias, antingen så att man förhöll sig mer positivt eller mer negativt till de tidskrifter man kände till än till andra. För att kunna kartlägga om något ämnesområde var mer eller mindre utsatt för biaser under bedömningsprocessen för tidskrifter, jämfördes ämnena naturvetenskaper, biovetenskaper, samhällsvetenskaper och humanistiska vetenskaper med varandra. Slutligen undersöktes även JUFO-kommittémedlemmarnas subjektiva upplevelse av bedömningsarbetet genom ett anonymt frågeformulär.

1.1 Kognitiv bias i bedömningar och beslutsfattande

Faktorer som kan påverka en granskares evaluering av vetenskapligt material har undersökts sedan många år tillbaka. Till dessa faktorer hör bland annat författarens identitet och granskares identitet samt kognitiva biaser såsom förankringseffekt, baktändning, blinda fläcken-bias, konfirmationsbias, falsk slutledning, haloeffekt samt ingrups- och utgrupsbias (Cho m.fl., 1998; Strauss, 1994; Justice m.fl., 1998; Peters & Ceci, 1982; McIntosh & Ross, 1987; Godlee & Martyn, 1998; Van Rooyen m.fl., 1999). En redogörelse för dessa biaser och exempel på hur de kan förekomma under granskningsprocessen av tidskrifter framkommer i

tabell 1. Dessa biaser kan uppkomma skilt från varandra eller i olika kombinationer, beroende på situationen (Kaatz m.fl., 2014; Caelleigh m.fl., 2003). Trots att granskare ämnar vara objektiva vid evalueringen av vetenskapligt material påverkas alla människor av olika former av bias vid beslutsfattande. Bedömningarna och besluten är inte längre objektiva om en person har påverkats av bias under utvärderingsprocessen; de är istället gjorda utgående från personens egna erfarenheter, värderingar eller förutfattade meningar (Bornmann & Daniel, 2004). Om bias förekommer under den kollegiala granskningsprocessen påverkas således utvärderingen av det vetenskapliga materialet av andra faktorer än materialets kvalitet, vilket är den enda faktorn som ska tas i beaktande under utvärderingsprocessen (Kaatz m.fl., 2014; Caelleigh m.fl., 2003).

Forskning har genomförts för att undersöka om biaser förekommer och hur de i så fall kommer till uttryck under bedömningsprocesser av vetenskapligt material. Mahoney beskrev i en studie från 1977 hur konfirmationsbias påverkade granskningen av artiklar. I studien ombads 75 granskare utvärdera ett manuskript som rapporterade antingen positiva, negativa, blandade eller inga resultat. Studiens resultat visade att granskarna inte var eniga och att de dessutom var partiska gentemot manuskript som uppvisade motsägande resultat i förhållande till deras teoretiska perspektiv. I en senare engelsk studie kom Luty, Arokiadass, Easow och Anapreddy (2009) fram till att bias förekom i medicinska tidskrifters redaktioner gällande vilka artiklar som publicerades. Redaktörerna tenderade att favorisera bidrag från den egna redaktionens forskare på konkurrerande redaktioners bekostnad, även om det inte fanns skillnader i Impact Factor (IF) mellan tidskrifterna.

Nuförtiden har det blivit vanligare att använda andra metoder inom kollegial granskning för att minska på risken att bias ska förekomma under granskningsprocessen. En av dessa metoder är tillämpandet av dubbelblind-design (Godlee & Jefferson, 2003; Kassier & Campion, 1994). I en dubbelblind-design känner varken granskaren eller den granskade till personlig information om den andra personen, såsom namn eller kön. Dubbelblind-designer används däremot inte i all kollegial granskning, eftersom det finns bristande med forskning som bevisar nyttan med metoden (Ross m.fl., 2006). Det är även tidskrävande att till exempel dölja en författares namn genom en hel vetenskaplig text, vilket gör det i vissa fall praktiskt omöjligt för en kollegial granskningskommitté att använda dubbelblind-designer (Pitkin, 1995). Hojat, Gonnella & Caelleigh menade dessutom i en studie från 2003 att metoder som blindtester och dubbel-blind tester inte är tillräckligt effektiva för att undvika bias vid beslutsfattandet. De menade att granskarna istället måste bli mer medvetna, uppmärksamma

och ansvarstagande gällande vilka biaser de kan påverkas av under bedömningsprocessen för att öka sannolikheten för en rättvis kollegial granskningsprocess.

Tabell 1

Beskrivning av biaser som kan förekomma under kollegial granskning och exempel på hur en granskare kan påverkas av biaserna under utvärderingsprocessen för vetenskapliga tidskrifter.

Kognitiva biaser	Biaser under utvärderingsprocessen av tidskrifter
<p>Förankringseffekt/ankaret (<i>Anchoring bias</i>): Tendensen att i allt för hög grad lite på den första informationen man kommer i kontakt med vid beslutsfattande.</p>	<p>En granskare kan exempelvis ha haft bra/dåliga erfarenheter av tidskriften som bedöms, vilket påverkar granskarens bearbetning av ny information om tidskriften.</p>
<p>Baktändning (<i>Backfire effect</i>): Uppstår då information som motbevisar ens tidigare åsikt om någonting inte ändrar åsikten, utan istället förstärker den.</p>	<p>En granskare kan hålla fast vid sin tidigare åsikt eller givna klassificering, trots ny motsäggande information.</p>
<p>Blinda fläcken-bias (<i>Blind-spot bias</i>): Förmågan att se andra personers benägenhet att påverkas av olika biaser, men inte sin egen.</p>	<p>En granskare tror att i motsats till många kollegor fattar hen objektiva beslut och inte påverkas av någon bias.</p>
<p>Konfirmationsbias (<i>Confirmation bias</i>): Tendens att söka sig till information som stämmer överens med ett tidigare antagande eller intryck.</p>	<p>En granskare har overseende med misstag gjorda av en tidskrift om hen redan har gett en hög klassificering åt den.</p>
<p>Falsk slutledning/felslut (<i>Ecological fallacy</i>): Tendensen att göra en slutledning om en person baserat på en samling av data eller antaganden om en grupp människor.</p>	<p>En granskares klassificering av en tidskrift påverkas av hens antaganden om ämnesområdet tidskriften tillhör.</p>
<p>Halo-effekt (<i>Halo effect</i>): Om en person är kompetent/inkompetent inom ett område, antar man att hen är kompetent/inkompetent inom flera områden.</p>	<p>En granskare anser att en tidskrift förtjänar en låg/hög klassificering beroende på om tidskriften tidigare har ansetts vara av låg/hög kvalitet/status.</p>
<p>Ingrupps-/utgruppsbias (<i>In group/out group bias</i>): Det krävs mindre information för positiva omdömen om två personer har någon slags samhörighet än då de inte har det. Det krävs även mindre information för negativa omdömen om två personer inte har någon samhörighet.</p>	<p>En granskare som tidigare har varit i kontakt med en tidskrift tenderar att värdera tidskriften högre än en tidskrift som hen inte har varit i kontakt med. Alternativt, granskaren värderar en tidskrift som sämre om den omfattar ett ämnesområde som granskaren inte själv tillhör.</p>

1.2 Publikationsforum

Vetenskapliga tidskrifter kan klassificeras genom så kallade objektiva mått som baserar sig på hur många gånger artiklar i en tidskrift refererats till, och genom olika former av expertbedömningar, som tillämpas till exempel i Finland, Norge och Danmark (Auranen & Pölönen, 2012, 18). I Finland görs expertbedömningen av Publikationsforum, en enhet under Vetenskapliga samfundens delegation (VSD), en takorganisation för de vetenskapliga samfunden i Finland (Publikationsforum, 2017a). Ett välanvänt objektiva mått är å sin sida Thomson Reuters Journal Impact Factor (i fortsättningen IF), som speglar en tidskrifts inflytande genom att beräkna förhållandet mellan hur många av tidskriftens artiklar som citerats och hur många artiklar som publicerats i den under en viss tidsperiod (Clarivate Analytics 2018). IF gäller ofta två år åt gången, men de rapporteras årligen (ibid). Såväl olika former av expertbedömningar, till exempel JUFO-klassificering, och IF berättar om tidskriftens kvalitet och inflytande. Ju högre JUFO klassificering och IF en tidskrift har, desto större vikt anses den ha.

Styrelsen för Vetenskapliga samfundens delegation (VSD) utnämner JUFOs styrgrupp, som i sin tur väljer kommittémedlemmarna för de 23 kommittéer som bedömer tidskrifterna. De 23 kommittéerna representerar samtliga vetenskapsområden i Finland. I styrgruppen sitter representanter för de olika vetenskapssamfunden och andra aktörer inom utbildning i Finland, närmare bestämt VSD, Undervisnings- och kulturministeriet, Finlands Universitet UNIFI r.f., Finlands Akademi, Finlands universitetsbiblioteks råd, Nationalbiblioteket, Centret för Informationsteknik för vetenskap (CSC) och representanter för olika vetenskapsområden (Savolainen, 2018). Styrgruppens och också kommittéernas arbete förbereds av Publikationsforums sekretariat (ibid).

JUFO-systemet i Finland fick sin början år 2009, då Rådet för Finlands universitetsrektorer framförde ett förslag på ett klassificeringssystem för tidskrifter som skulle basera sig på den norska och danska modellen, där experter bedömer tidskrifters kvalitet (Auranen & Pölönen 2012, 3). Fördelarna med det här systemet skulle vara att finansieringen universiteten får för publikationer skulle basera sig på publikationernas kvalitet, inte mängd som tidigare var fallet (ibid). Systemet skulle också ge forskare i Finland information om tidskrifter (och vilka det lönar sig att publicera i) (ibid). IF ansågs dessutom inte vara det bästa måttet på publikationers kvalitet inom humaniora och teknik, eftersom forskare inom dessa områden oftare publicerar andra typer av publikationer än artiklar i

tidskrifter, till exempel böcker (ibid; Rantala 2012; Publikationsforum 2017b). Med JUFO-systemet ville man således ta olika vetenskapsområdets publikationstraditioner bättre i beaktande vid universitetsfinansieringen. Rantala (2012) nämner också i sitt blogginlägg att man genom systemet ville lyfta fram inhemska tidskrifter, som annars inte får stor synlighet då enbart IF används som mått.

På basen av Finlands universitetsrektorers förslag sattes ett JUFO-projekt upp av VSD år 2010, som pågick till 2012 (Auranen & Pölönen 2012, 3). Under projektet fastställdes att kommittéerna är 23 stycken till antalet och att tidskrifterna som ska bedömas begränsas till 20 000. Dessa 20 000 tidskrifter stämmer långt överens med de tidskrifter som bedömts i Norges, Danmarks och delvis Australiens system (Auranen & Pölönen 2012, 5, 6). I bedömningsförslaget som JUFO-sekretariatet gav åt kommittémedlemmarna användes Norges, Danmarks och Australiens bedömningar som referenspunkter, liksom också Journal Impact Factor (JIF), Source Normalized Impact per Paper och Scimago Journal impaktfaktorerna för naturvetenskaper, biovetenskaper, ekonomi samt beteendevetenskaper och pedagogik (kommittéerna 1-15 samt 16 och 18), och enbart Source Normalized Impact per Paper för humaniora och samhällsvetenskaper (kommittéerna 17 samt 19-23) (Auranen & Pölönen 2012, 6)¹. På basen av projektet fortsatte sedan de valda kommittémedlemmarnas arbete fram till 2013, och nya medlemmar valdes år 2014 (Auranen & Pölönen 2012, 7). Den nuvarande kommittésammansättningen är den tredje i ordningen, och deras verksamhetsperiod sträcker sig från 2018 till 2022 (Publikationsforum, 2017c).

Val av kommittémedlemmar sker via ansökan, antingen så att enskilda forskare söker själva till kommittén eller, vilket är mycket vanligare, via en organisation, såsom ett universitet, som stöder anställningen av forskaren i fråga (Publikationsforum, 2017c). I antagningen beaktas olika vetenskapsorganisationers rekommendationer, erfarenhet av forskning samt erfarenhet av tidskriftsbedömning (ibid). Man tar också i beaktande att det ska vara så jämn fördelning som möjligt mellan olika universitet och andra

¹ Journal Impact Factor rapporteras årligen av nyhets- och publikationsverksamhetsföretaget Thomson Reuters, och mäter en tidskrifts påverkan genom att räkna ut förhållandet mellan antalet citeringar av tidskriften och antalet publikationer i tidskriften under två år (Clarivate Analytics 2018).

Source Normalized Impact per Paper är ett påverkningsfaktormått som tar vetenskapsområdets publikationsvidd i beaktande i beräkandet av påverkan. Det räknas ut genom att räkna det genomsnittliga antalet referenser per publikation, samt hur mycket källor artikeln använder sig av (ju färre källor, desto större vikt anses publikationen som refereras till att ha) (Uleåborgs universitet, 2019).

Scimago Journal Impact Factor tar förutom det genomsnittliga antalet referenser per publikation också den citerande och den citerade journalens prestige och närhet till varandra i beaktande i beräkandet av påverkan (Guerrero-Bote & Moya-Anegón, 2012, 674).

vetenskapsorganisationer, kön, vilken etapp i forskarkarriären kommittémedlemmarna befinner sig i (nytt för 2018–2021), och vetenskapsområden.

De sittande kommittéerna består sammanlagt av 235 personer, och alla Finlands universitet samt några av statens forskningsinstitutioner finns också representerade bland dem (Publikationsforum, 2017c). De består till 24 procent av medlemmar från den förra perioden, 37 procent kvinnor, 63 procent män, samt 13 procent unga forskare (blivit utexaminerade 2010 eller senare). I antagningsprocessen vägde det tyngst om ett samfund rekommenderat personen, vilket var fallet vid största delen av ansökningarna, bara 22 procent ansökte på egen hand (ibid). Att JUFO strävar efter att ha en så demografiskt varierande uppsättning kommittémedlemmar som möjligt, och t.ex. inkluderandet av även unga forskare som inte ännu hunnit publicera så mycket visar på att de försöker undvika bias som beror på en alltför homogen bedömargrupp.

1.2.1 Klassificeringen av tidskrifter enligt JUFO

Kommittémedlemmarna har tillgång till en handbok med instruktioner för bedömning, som också finns allmänt tillgänglig på Internet (Savolainen, 2018). I handboken gås kriterierna för olika nivåer igenom, liksom vad besluten ska grunda sig på. Nivåerna är 0, 1, 2 och 3. 80 procent av tidskrifterna ligger på nivå 1 (Auranen & Pölönen 2012, 24). För denna nivå krävs att tidskriften uppfyller vetenskapliga krav, det vill säga publicerar vetenskapliga resultat, har en redaktion som utgörs av experter inom vetenskap, och att manuskripten går i sin helhet igenom kollegial granskning (Savolainen, 2018). Till nivå 1 väljs i regel inte tidskrifter som annars uppfyller kraven på god vetenskaplig praxis, men som är ”lokala”, det vill säga över hälften av antingen publikationerna eller redaktionen kommer från samma organisation, om organisationen också är förläggaren (ibid). I handboken nämns också att tidskrifter vars ”vetenskapliga nivå eller relevans kan ifrågasättas” (”joiden tieteellinen taso tai relevanssi on kyseenalainen”) (Savolainen, 2018) inte heller når nivå 1. Det här kan tänkas vara en grogrund för bias, eftersom det inte listas några kriterier för hur man bestämmer vad vetenskaplig nivå eller relevans är, utan det är upp till kommittémedlemmarna att bestämma.

Nivå 2 och 3, där 20 procent av tidskrifterna ligger, representerar mycket högklassiga tidskrifter, som publicerar internationellt ledande forskning inom området (Auranen & Pölönen 2012, 25). Nivå 3 är den ledande delen av nivå 2, och 25 procent av tidskrifterna på nivå 2 uppnår nivå 3; alla tidskrifter som uppfyller kriterierna för nivå 3 måste alltså också uppfylla kriterierna för nivå 2 (ibid). Kriterierna som listas för nivå 3 är bland annat att

tidskriften måste representera den högsta nivån av forskning inom sitt område, impaktfaktormått (till exempel IF) måste vara mycket höga, tidskriften måste täcka sitt vetenskapsområde i sin helhet (alltså inte bara fokusera på en enskild del av fältet) och tidskriftens redaktörer måste vara toppforskare av internationella mått inom sitt område (ibid). Kommittéerna får själva besluta om vilka tidskrifter som ska placeras på nivå 1 och 2, förutom för inhemska tidskrifter på nivå 2 inom humaniora, där ordförandena för ifrågavarande paneler fattar det slutgiltiga beslutet, medan ordförandena från alla kommittéer tillsammans beslutar om vilka tidskrifter som ska placeras på nivå 3, på basen av ifrågavarande kommittés förslag (Publikationsforum, 2017d). Det finns också kvoter för hur stora publikationsvolymerna (det vill säga sammanlagda antalet artiklar) som får kategoriseras som nivå 2 och 3. Sedan 2015 har det varit max 20 procent av den sammanlagda publikationsvolymen kommittén kan lista på nivå 2, respektive 25 procent av nivå 2 på nivå 3 (ibid). Kvoterna bestäms enligt artikelvolym, och inte enligt antalet tidskrifter från och med 2015, eftersom tidskrifternas storlek varierar mycket, särskilt inom naturvetenskaper, där det finns väldigt stora och ofta utkommande tidskrifter (ibid). Då kvoterna bestäms enligt artikelvolym jämnas den här skillnaden ut mellan kommittéerna.

Kommittémedlemmarna instrueras i handboken att använda sig av bedömningarna i Norge och Danmark, olika impaktfaktormått, råd från vetenskapssamfund, samt sin egen expertis i klassificeringen (Savolainen, 2018). Vad egen expertis är specificeras inte närmare, och det här kan också tänkas vara en potentiell källa till bias. I föreliggande arbete används kommittémedlemmarnas egen publiceringshistoria som ett biasmått vars effekt på rangordning av tidskrifterna undersöktes.

Vidare motiveringar för klassificeringen krävs om bedömningen av en tidskrift avviker mycket från hur den bedömts i Norge och Danmark, liksom om den får ett högre betyg än en annan tidskrift från samma område, som har betydligt högre IF eller andra impaktfaktormått (Savolainen, 2018). Dessutom kan panelerna bli tvungna att förklara allmänt vilka kriterier de använt för att bestämma vad som inte uppfyller nivå 1, liksom vad som uppfyller kraven för nivå 2 och 3 (ibid).

Kommittémedlemmarna är skyldiga att meddela sekretariatet om olika förbindelser till publikationskanalerna de kan tänkas ha som kan påverka bedömningen. I praktiken innebär detta att de är skyldiga att meddela om förbindelser om deras panel behandlar en publikationskanal som de publicerat i mer än en gång inom 5 år, eller fungerat som redaktörer eller medlemmar i redaktionen för (utan tidsbegränsning). De får ändå delta i beslutfattningsprocessen, men kommittén kan som helhet ta förbindelsen i beaktande, och de

blir meddelade om den av sekretariatet (Savolainen, 2018). Man har således tagit det föreliggande biasmålet i beaktande i JUFO-systemet, men hur detta har lyckats är öppet, även därför att kommittéerna själva har fått besluta hur detta behandlas.

1.2.2 JUFO-klassificeringens betydelse för universitetens finansiering

Undervisnings- och kulturministeriet bidrar med finansiering åt universiteten i Finland baserat på olika kriterier (Opetus- ja kulttuuriministeriön asetus yliopistojen perusrahoituksen laskentakriteereistä 331/2016). Av dessa pengar som delas ut årligen baserar sig 13 procent på universitetens publikationer (Opetus- ja kulttuuriministeriön asetus yliopistojen perusrahoituksen laskentakriteereistä 331/2016, 5). Alla publikationer beaktas, men beroende på JUFO-klassificeringen och publikationstypen ges varje publikation en koefficient som bestämmer publikationens ”värde”. I tabell 2 beskrivs koefficienterna för kollegialgranskade artiklar, som undersöks i denna avhandling.

Tabell 2

Koefficienter för kollegialgranskade artiklar och redigerade vetenskapliga verk (Opetus- ja kulttuuriministeriön asetus yliopistojen perusrahoituksen laskentakriteereistä 331/2016, 5)

JUFO-poäng	Koefficient
0	0,1
1	1
2	3
3	4

Redan en JUFO-klassificering på 0 ger universitetet vars anställda publicerat i tidskriften pengar, medan en klassificering på 3 ger fyra gånger mer pengar än klassificering på 1. Eftersom JUFO-klassificeringen således direkt påverkar den statliga finansieringen av

universitetens verksamhet är det av yttersta vikt att bedömningsprocessen är transparent och konstant utvärderas till exempel med tanke på potentiell bias.

1.3 Studiens syfte

Syftet med denna avhandling var att undersöka om JUFO kommittémedlemmarnas egen publikationshistoria påverkade deras klassificering av vetenskapliga tidskrifter, således om det förekommit bias under bedömningsprocessen. I avhandlingen undersöktes även eventuell variation i denna bias inom naturvetenskaper, bio- och hälsovetenskaper, samhällsvetenskaper och humanistiska vetenskaper. Ämnena delades in i dessa fyra huvudämnesgrupper enligt den nya indelningen av Finlands Akademi (2018). Denna indelning motiveras även med tanke på kritik som riktats mot JUFO-systemet främst från humanistiskt håll, eftersom humanistisk forskning inte lika ofta publiceras i tidskrifter (åtminstone inte stora, internationella sådana) som till exempel naturvetenskaplig forskning (Rantala, 2012). Detta undersöktes genom såväl objektiva jämförelsemått som genom subjektiva frågeformulär som JUFO kommittémedlemmarna ombads svara på.

Forskningsfrågan för föreliggande avhandling var således om det förekommer bias i forskares bedömning av tidskrifters kvalitet och om det förekommer några skillnader i detta hänseende mellan olika vetenskapsområden i Finland. Bias som mättes här gällde ingrupsbias, det vill säga huruvida forskare var mer positivt inställda till tidskrifter de själva publicerat i än andra. Förutom arkivdata användes en enkät för att erhålla information om hur JUFO-kommittémedlemmarna uppfattade bedömningsprocessen och hur de upplevde att olika faktorer påverkade deras bedömningar. Om eventuella kopplingar till en tidskrift, såsom tidigare publikationer i tidskriften, påverkade deras bedömning kan JUFO-klassificeringens reliabilitet kritiseras.

2 Metod

Forskningsfrågan för föreliggande undersökning var huruvida det förekommer skillnader i mängden bias mellan de fyra vetenskapsområdena naturvetenskaper, biovetenskaper, samhällsvetenskaper och humanistiska vetenskaper, samt hur JUFO-kommittémedlemmarna uppfattar klassificeringsprocessen av tidskrifter. Bias operationaliserades som att forskarna gav ett högre vitsord åt en tidskrift dom själv publicerat i, vilket gör att vitsordet inte enbart kunde förklaras av att tidskriften verkligen objektivt sett var högklassig, vilket mättes med Impact Factor.

I en forskningsövning med samma material som Annika Tuomi, Julia Nummenmaa, Elias Lassila och Emma Vihervaara utförde år 2016 var syftet att få reda på huruvida det överhuvudtaget fanns bias i kommittémedlemmarnas klassificeringar, oberoende av vetenskapsområde. Då gjordes en linjär regressionsanalys, där JUFO-vitsordet var den beroende variabeln, medan Impact Factor-poängen, antalet kommittémedlemmar som publicerat i tidskriften och det sammanlagda antalet publikationer för kommittén var de oberoende variabler. Eftersom denna undersökning visade att Impact Factor och antalet kommittémedlemmar som publicerat tillsammans (*adjusted* $R^2 = 0,096$, $F[1, 970] = 14,190$, $p < 0,001$) förklarade mera än enbart Impact Factor (*adjusted* $R^2 = 0,083$, $F[1, 971] = 89,251$, $p < 0,001$), medan effekten av sammanlagda antalet publikationer inte var signifikant (*adjusted* $R^2 = 0,097$, $F[1, 969] = 2,406$, $p = 0,121$), användes de två signifikanta oberoende variablerna, Impact Factor och antalet kommittémedlemmar som publicerat i en viss tidskrift, i föreliggande avhandling.

Med denna forskningsövning som grund väcktes frågan om vad biasen berodde på, och var i samplet den förekom. Materialet i denna avhandling delades därför in enligt vetenskapsområde och analyser utfördes för att undersöka hur väl regressionsmodellerna passade för varje vetenskapsområde. Detta gjordes för att kunna få en inblick i möjliga skillnader i bias mellan de olika områdena. För att undersöka hur eventuell bias kom till uttryck under klassificeringsprocessen skickades ett frågeformulär ut till JUFO-kommittémedlemmarna, se bilaga 2. Syftet med frågeformuläret var att undersöka hur medlemmarna själva upplevt bedömningsprocessen och om de trodde att biasen kan ha påverkat deras bedömning. Frågeformuläret utgjorde således ett komplement till den huvudsakliga analysen.

2.1 Material

Två olika material analyserades för att besvara forskningsfrågan, det vill säga huruvida det förekom bias i olika grad mellan naturvetenskaper, biovetenskaper, samhällsvetenskaper och humaniora. För att besvara frågan huruvida det förekom variationer i bias mellan de fyra vetenskapsområdena användes ett slumpmässigt sampel på 973 av 12 027 JUFO-bedömda tidskrifter (ursprungligen 1000, varav vissa föll bort bland annat på grund av avsaknad av Impact Factor), deras JUFO-bedömningar och Impact Factor-poäng från år 2015. JUFO-poängen för tidskrifterna erhöles i en datafil från sekretariatet. Impact Factor-poängen för samplet erhöles genom webbsidan ISI Web of Knowledge, där samtliga tidskrifters Impact Factor-poäng fanns samlade. Den andra oberoende variabeln, det vill säga antalet kommittémedlemmar som publicerat i tidskriften, erhöles genom en systematisk genomgång av kommittémedlemmarnas publikationer före år 2015 i Google Scholar. Detta gjordes genom att skriva in en specifik tidskrift i sökfältet tillsammans med namn på alla kommittémedlemmar för kommittén som hade bedömt tidskriften. Resultaten gick sedan manuellt igenom för att undersöka om någon av kommittémedlemmarna hade publicerat i tidskriften före år 2015. Slutligen delades kommittéerna in i respektive vetenskapsområden, se bilaga 3 för mer information.

Det andra materialet, som utgjorde ett komplement till det första, bestod av en enkät med frågor om hur kommittémedlemmarna upplevt JUFO-beslutfattningsprocessen. I enkäten uppmanades deltagarna att ange endast sitt vetenskapsområde och kön, så att ingen personlig information eller information om en specifik panel kunde framkomma. Enkäten skickades ut till samtliga av år 2014–2017 kommittémedlemmar via JUFO:s sekretariat. De utgjorde en population på 201 personer, varav 126 var män och 75 kvinnor. Sextiotvå representerade naturvetenskaper, 60 bio- och hälsovetenskaper, 41 samhälls- och beteendevetenskaper, och 38 humaniora. Sammanlagt 99 svar mottogs och efter uteslutning av fall som hoppat av under enkäten bestod samplet av 85 personer, 53 män och 32 kvinnor.

I enkäten ingick nio frågor om beslutfattningsprocessen i Publikationsforum, se bilaga 1. Den första frågan var indelad i sex delar och berörde hur mycket personen upplevde att olika faktorer påverkade hans beslut. Den sista frågan var en öppen fråga där respondenterna allmänt fick uttrycka sina åsikter om beslutfattningsprocessen. Den frågan analyserades kvalitativt, medan det i alla andra frågor fanns en Likertskala med 10 nivåer. Enkäten genomgick också etikprövning i den forskningsetiska nämnden för psykologi och logopedi vid Åbo Akademi.

3 Resultat

3.1 Analys av JUFO-bedömningen 2015

En linjär regressionsanalys (lm-funktion i R) utfördes på hela JUFO-datan från 2015, där JUFO-poängen var den beroende variabeln och Impact Factor och antal kommittémedlemmar som publicerat i en tidskrift var oberoende variabler. Resultaten av analysen visas i tabell 3.

Tabell 3

Estimat för modell med Impact Factors och antal kommittémedlemmar som publicerat i en tidskrift som prediktorer för variationen i JUFO-poäng

	Estimat	Standardfel	t-värde
(Intercept)	0,05413	0,02907	1,862
Impact Factor	0,26142	0,02894	9,034***
Kommittémedlemmar	0,07819	0,03593	2,176*

Modellen var signifikant med en adjusted R^2 på 0,09708 ($F(3,970) = 35,87$, $p < ,001$), och av prediktorerna bidrog Impact Factor med mest förklaring i modellen (Estimat = 0,26142, $p < 0,001$), men även antalet kommittémedlemmar bidrog signifikant till förklaringen (Estimat = 0,07819, $p < 0,05$).

Tabell 4 visar en variansanalys av hur effekten av de olika prediktorerna skiljde sig mellan grupperna, det vill säga naturvetenskap, bio- och hälsovetenskaper, samhällsvetenskaper och humanistiska vetenskaper. Variansanalysen baserade sig på modellen: $JUFO \sim \text{grupp} * (\text{IF} + \text{kommittémedlemmar})$, vilket var en linjärmodell. Resultaten visade att det fanns en effekt av grupp på hur Impact Factor påverkade JUFO-poängen ($F(3,970) = 30,7731$, $p < 0,001$).

Tabell 4

Variansanalys av prediktorernas effekter mellan vetenskapsområdena

	df	Sum Sq	Mean Sq	F-värde
Grupp	4	64,60	16,150	24,6151 ***
Impact Factor	1	104,21	104,214	158,8381 ***
Kommittémedlemmar	1	17,90	17,896	27,2760 ***
Grupp * Impact Factor	3	60,49	20,164	30,7731 ***
Grupp * Kommittémedlemmar	3	2,50	0,832	1,2688

Till näst analyserades om det förekom skillnader i hur väl prediktorerna förutsåg JUFO-poäng mellan vetenskapsområdena, vilket gjordes genom att undersöka skilda estimat i lm-modellen mot olika referensnivåer. Detta gjordes genom att varje vetenskapsområde i tur och ordning var referensgrupp och de övriga grupperna jämfördes med referensgruppen. Nedan rapporteras de signifikanta resultaten för vetenskapsområdena med hänvisning till bilaga 1.

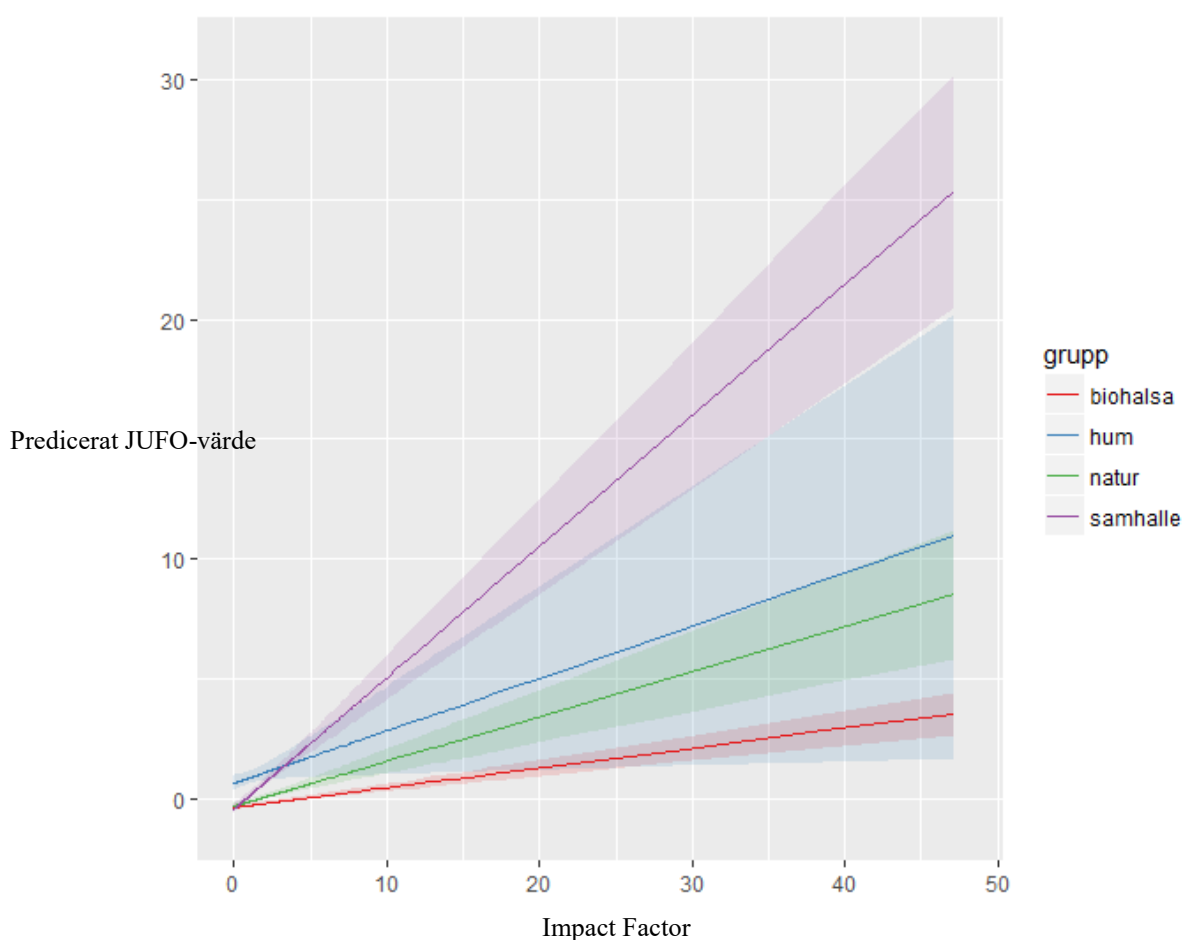
3.1.1 Faktorer som predicerar JUFO-klassificering inom de olika ämnesområdena

Effekten av Impact Factor på JUFO-klassificering var signifikant lägre inom bio- och hälsovetenskaper än inom naturvetenskaper och samhällsvetenskaper ($p < 0,001$). Se figur 1 för interaktionseffekter mellan ämnesområde och Impact Factor på JUFO-poäng. Effekten av antalet kommittémedlemmar som hade publicerat i en tidskrift på JUFO-klassificering var signifikant mindre inom bio- och hälsovetenskaper än inom samhällsvetenskaper ($p < 0,01$). Se figur 2 för interaktionseffekter mellan ämnesområde och variabeln Kommittémedlemmar på JUFO-poäng. Det här betyder att bio- och hälsovetenskaper i lägre grad än naturvetenskaper och samhällsvetenskaper följde Impact Factors under JUFO-klassificeringen. JUFO-klassificeringarna utförda av bio- och hälsovetenskaper påverkades mindre av antalet kommittémedlemmar som hade publicerat i den utvärderade tidskriften i jämförelse med samhällsvetenskaper.

Effekten av Impact Factor var mindre inom naturvetenskaper i jämförelse med samhällsvetenskaper ($p < 0,001$). Även effekten av antalet kommittémedlemmar som hade publicerat i en tidskrift på JUFO-klassificering var signifikant mindre inom naturvetenskaper

än inom samhällsvetenskaper ($p < 0,05$). Resultaten tyder på att klassificeringarna gjorda av naturvetenskaperna påverkades i lägre grad av Impact Factor och av kommittémedlemmarnas egna publikationer i tidskriften de bedömde i jämförelse med samhällsvetenskaper.

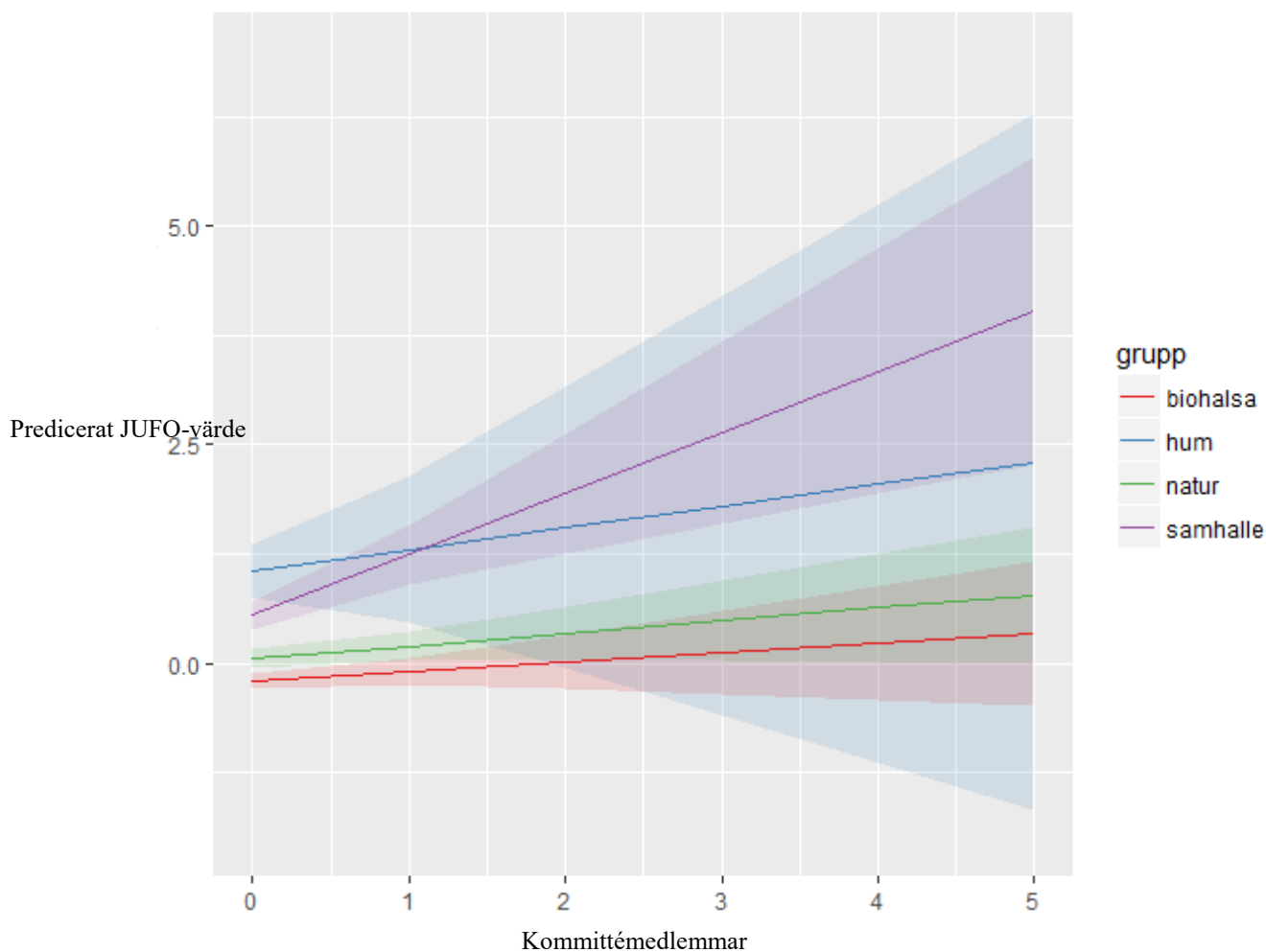
Effekten av Impact Factor på JUFO-klassificering var signifikant större inom samhällsvetenskaper än inom alla övriga ämnesområden (inom bio- och hälsovetenskaper och naturvetenskaper $p < 0,001$, inom humanistiska vetenskaper $p < 0,01$). Även effekten av antalet kommittémedlemmar som hade publicerat i en tidskrift på JUFO-klassificering var signifikant större inom samhällsvetenskaper än inom bio- och hälsovetenskaper ($p < 0,01$) och naturvetenskaper ($p < 0,05$).



Figur 1. Interaktionseffekter mellan ämnesområde och Impact Factor på JUFO-poäng.

I figur 1 är det möjligt att avläsa att samhällsvetenskaper var den gruppen som i högsta grad påverkades av eller använde sig av tidskrifters Impact Factor vid bedömningen av tidskrifternas kvalitet. Bio- och hälsovetenskaper var gruppen som minst följde tidskrifters Impact Factors, vilket ökar risken för bias under bedömningsprocessen. Humanistiska

vetenskaper och naturvetenskaper följde Impact Factors i lägre grad än samhällsvetenskaper, men inte i samma utsträckning som bio- och hälsovetenskaper. Det förekom ingen signifikant skillnad i påverkan av Impact Factors mellan humanistiska vetenskaper och naturvetenskaper.



Figur 2. Interaktionseffekter mellan ämnesområde och variabeln Kommittémedlemmar på JUFO-poäng.

I figur 2 framkommer interaktionseffekterna mellan ämnesgrupperna och variabeln Kommittémedlemmar och dess effekt på den beroende variabeln, det vill säga JUFO-poäng. Effekten av antalet kommittémedlemmar som hade publicerat i en tidskrift på JUFO-klassificering var störst bland samhällsvetenskaper, vilket betyder att ju fler som tidigare hade publicerat i tidskriften desto högre JUFO-poäng tilldelades tidskriften. Denna statistiskt signifikanta ingrupsbias förekom enbart inom samhällsvetenskaperna, inte inom de övriga ämnesområdena.

För att få en bild av om JUFO-systemet tillför något till bedömningen av en tidskrifts kvalitet som IF inte redan ger analyserades korrelationen mellan IF och JUFO-poäng. Om den här korrelationen var stark (nära 1) är det möjligt att ifrågasätta om JUFO-systemet alls behövs, då den ligger så nära IF, men om korrelationen var svag (under 0,5) kan man ställa frågan ifall JUFO-systemet är utsatt även för bias, och om det tillräckligt objektivt mäter en tidskrifts kvalitet. Korrelationen mellan IF och JUFO-poängen visade sig vara relativt svag, men statistiskt signifikant (Pearson korrelation = 0,293**).

Det är tänkbart att den positiva påverkan av antalet kommittémedlemmar som publicerat i tidskriften på JUFO-poängen berodde på att forskarna bara råkat publicera i synnerligen högklassiga tidskrifter. Därför analyserades ytterligare korrelationen mellan antalet kommittémedlemmar som publicerat och IF, och även här fanns ett svagt positivt samband (Pearson korrelation 0,102**), men den var inte tillräckligt stor för att kunna förklara variationen i JUFO-poäng. De olika vetenskapsområdena analyserades skilt för att undersöka var korrelationen förekom. En statistiskt signifikant korrelation påträffades mellan antalet kommittémedlemmar och IF enbart inom bio- och hälsovetenskaperna (Pearson korrelation 0,100*), och en nästan statistiskt signifikant korrelation mellan samma variabler inom naturvetenskaper (Pearson korrelation 0,112, $p=0,054$).

Sammanfattningsvis verkade samhällsvetenskaper mest ta i beaktande tidskrifters Impact Factors vid bedömningen av tidskrifternas kvalitet. Samtidigt var även effekten av antalet kommittémedlemmar som hade publicerat i en tidskrift på JUFO-klassificering störst bland samhällsvetenskaper.

3.2 Analys av panelisternas upplevelse av bedömningen

Ett Chi-kvadrattest utfördes på könsvariabeln och vetenskapsområdesvariabeln i enkäten för att undersöka om såväl könsfördelningen som fördelningen mellan vetenskapsområden i samplet motsvarade respektive fördelning i populationen (alla kommittémedlemmar), vilket de gjorde, då samplet bestod av 32 kvinnor och 53 män, medan den förväntade könsfrekvensen var 31,7 kvinnor respektive 53,3 män (Chi-kvadrat [1] = 0,004, $p = 0,949$). Tjugofem av respondenterna var verksamma inom naturvetenskaper (förväntat $N = 26,2$), 22 inom bio- och hälsovetenskaper (förväntat $N = 25,4$), 19 inom samhälls- och beteendevetenskaper (förväntat $N = 17,3$), och 19 inom humanistiska vetenskaper (förväntat $N = 16,1$) (Chi-kvadrat [3] = 1,199, $p = 0,753$).

Nedan visas medelvärden och standardavvikelser för alla enkätfrågor i hela samplet (tabell 6). Ur tabellen framgår att respondenterna uppgav att deras bedömning mest påverkades av den egna uppfattningen av tidskriften (medelvärde 8,26) och dess Impact Factor (7,55), medan andra kollegors uppfattning påverkade bedömningen minst (medelvärde 6,40). Det respondenterna uppgav att överlag påverkade dem minst var deras egen koppling till publikationskanalen, till exempel tidigare publikationer (medelvärde 4,78). Detta var intressant, eftersom det kunde tänkas betyda att effekten av föreliggande biasmåtten i vår analys av JUFO-bedömningen inte borde vara signifikant, vilket den var.

Tabell 6

Medelvärden och standardavvikelser för enkätfrågorna

Fråga	N	Medelvärde	Standardavvikelse
<i>Hur mycket påverkade följande faktorer bedömningen</i>			
Impact Factor	85	7,55	2,32
Motsvarande bedömning i Norge och Danmark	85	7,12	2,03
Egen uppfattning	85	8,26	1,68
Andra kommittémedlemmars uppfattning	85	7,19	1,97
Andra kollegers uppfattning	85	6,40	2,33
Annat	85	7,42	3,80
<i>Övriga frågor</i>			
Hur väl kände man personligen	85	6,46	2,09
Hur svårt att bedöma	85	5,13	2,80
Hur lätt att komma överens i kommittéen	85	6,81	2,32
Andras åsikts påverkan på mig	85	5,89	2,25
Min åsikts påverkan på andra	85	6,25	1,93
Betydelse av koppling till publikationskanalen	85	4,78	2,96
Hur väl togs alla i beaktande	85	7,35	1,94

En envägs-ANOVA utfördes för att undersöka om det fanns skillnader mellan vetenskapsområdena på de olika variablerna. I tabell 7 framkommer medelvärde och standardavvikelse för alla de enkätfrågor där det framkom skillnader mellan vetenskapsområdena. Det förekom signifikanta skillnader mellan grupperna på hur mycket respondenterna upplevde att Impact Factor påverkade bedömningen (fråga 1a) ($F [3, 84] = 9,948, p < 0,001$), hur mycket de upplevde att andra kommittémedlemmars uppfattningar påverkade den egna bedömningen (fråga 1d) ($F [3, 84] = 2,953, p = 0,037$), hur väl kommittémedlemmarna upplevde att de personligen kände till tidskrifterna (fråga 2) ($F [3, 84] = 4,705, p = 0,004$), och hur mycket de upplevde att andras åsikter påverkade deras åsikt om tidskrifterna (fråga 5) ($F [3, 84] = 3,125, p < 0,030$). Det förekom inga signifikanta skillnader mellan grupperna på de resterande frågorna.

Gällande Impact Factor och andras åsikters påverkan på den egna bedömningen kunde antagandet om homogenitet i varianserna förkastas, eftersom Levenes test var signifikant för dessa två variabler: för IF = $F(3, 81) = 4,133; p = 0,009$, för andras åsikter = $F(3, 81) = 2,797; p = 0,045$. På grund av att grupperna var ungefär lika stora borde detta dock inte utgöra något större problem (Field 2013, 445) ($n = 25$ för naturvetenskaper, 22 för bio- och hälsovetenskaper, 19 för samhällsvetenskaper respektive humaniora). Post hoc-testen visade att humaniora var den enda gruppen som skilde sig signifikant från alla andra grupper, se tabell 7.

Respondenterna inom humaniora använde sig i lägre grad av Impact Factor än alla andra grupper. Både inom naturvetenskaper och inom biovetenskaper upplevde respondenterna att de kände bättre till tidskrifterna än respondenterna inom humaniora. Ingen skillnad förekom på den här punkten mellan naturvetenskaper och biovetenskaper (båda hade ett medelvärde på 7 på den frågan, då skalan var 0–10).

En skillnad i upplevelse av hur andras åsikter påverkade den egna bedömningen framkom mellan biovetenskaper och humaniora; personer inom humaniora upplevde att de påverkades signifikant mer av andra än vad personer inom biovetenskaper upplevde. En skillnad i hur mycket kommittémedlemmarna använde sig av andra kommittémedlemmars åsikt i sin bedömning framkom mellan samhällsvetenskaper och humaniora; personer inom humaniora använde sig signifikant mer av andra kommittémedlemmars åsikter i bedömningsprocessen än vad personer inom samhällsvetenskaper gjorde.

Tabell 7

Medelvärde, standardavvikelse och skillnad mellan vetenskapsområden från envägs-ANOVA och post hoc-test på fråga 1a, 1d, 2 och 5.

Fråga	NV	BI	SA	HU	Post hoc-analys
1a. "How much do you think the following factors influenced your ratings of the publication channels while working in a JUFO committee? Please indicate your response with a number, 1 being no influence and 10 being very high influence: The publication channel's impact factors"	8,12 (1,88)	8,55 (1,14)	7,84 (1,98)	5,37 (2,87)	HU<NV*** HU<BI*** HU<SA**
1d. "The other committee members' perception of the quality of the publication channel" (1 = no influence, 10 = very high influence)	7,16 (1,75)	6,82 (2,24)	6,58 (1,92)	8,26 (1,63)	HU>SA*
2. "How well do you think you personally knew the publication channels you rated?" (1 = not at all, 10 = very well)	7,00 (1,56)	7,00 (2,12)	6,58 (2,36)	5,00 (1,83)	HU<NV** HU<BI**
5. "To what degree do you think the other committee members' opinion of the publication channels affected your own opinion of the publication channels?" (1 = not at all, 10 = to a very high degree)	5,76 (2,19)	4,91 (2,69)	6,16 (2,17)	6,95 (1,31)	HU>BI*

*** p< 0,001, ** p< 0,01, * p< 0,05. NV= Naturvetenskaper, BI = Bio- och hälsovetenskaper, SA= Samhälls- och beteendevetenskaper, HU = Humanistiska vetenskaper. 1= mycket låg påverkan, 10= mycket hög påverkan.

På den frivilliga öppna frågan i slutet av enkäten där respondenterna ombads fritt kommentera sin upplevelse av arbetet i JUFO-kommittén erhöles 48 svar. Svaren analyserades kvalitativt så att återkommande teman identifierades i svaren, för att undersöka om det förekom tydliga skillnader mellan vetenskapsområdena i vilka teman som togs upp. Denna analys utfördes först gemensamt, så att båda skribenterna läste igenom svaren

tillsammans och diskuterade analysen, och sedan sammanfattades analysen skriftligt av en av skribenterna.

Tre återkommande, övergripande teman identifierades: för det första att det var omöjligt att känna till alla de tidskrifter som bedömdes, och att kommittémedlemmarna därför använde sig av olika strategier för att bedöma dem. För det andra identifierades åsikter om samarbetet i kommittén som ett återkommande tema. Det tredje och sista återkommande temat var olika åsikter om JUFO-systemet, eller kritik mot vissa aspekter av det.

Många uttryckte att det inte går att personligen känna till alla de tidskrifter som en panel bedömer, eftersom det kan handla om tusentals tidskrifter. Då använde kommittémedlemmarna sig av olika hjälpmedel för att kunna göra bedömningen, såsom att läsa om tidskrifterna på deras webbsidor eller låta de kommittémedlemmar som var bekanta med tidskrifterna eller ämnesområdet i fråga styra bedömningen. I vissa fall upplevdes den här arbetsfördelningen som positiv, i vissa fall som negativ. Sex svar indikerade positiva upplevelser av en dylik arbetsfördelning, såsom att det var naturligt att lita på kommittémedlemmar som har mest erfarenhet inom ett ämnesområde och att tilliten var ömsesidig. I tre fall antydde en negativ åsikt om den här arbetsfördelningen, till exempel att det kunde uppstå bias hos vissa kommittémedlemmar, så att de ville gynna tidskrifter inom sitt eget ämnesområde. I några fall uttryckte respondenterna också att det var svårt att svara på frågan om hur väl de personligen kände tidskrifterna de bedömde, eftersom de bara kände till tidskrifterna inom deras egna ämnesområden. Det kunde därför inte ta ställning till hur väl de generellt kände till tidskrifterna.

Impact Factor nämndes i sex svar, och ordet ”impact” i ett svar. I tre av dessa svar, som getts av två respondenter inom samhälls- och beteendevetenskaper och en inom bio- och hälsovetenskaper, uttrycktes frustration över att Impact Factor eller allmänt tidskriftens ”impact” användes för lite som referens för bedömningen. En respondent inom bio- och hälsovetenskaper nämnde att de bedömde många journaler med hög IF, och en annan inom samma vetenskapsområde menade att bedömningarna gjordes i linje med IF och andra internationella mått. En respondent inom humaniora menade att IF inte är så relevant för de ämnesområdena hans kommitté bedömde.

Temat om samarbetet i panelen hade visst överlapp med det föregående temat, men det togs också upp andra aspekter om panelsamarbetet, som gör det motiverat att konstruera det som ett eget tema: 27 svar behandlade frågan om samarbete i panelen på något sätt. Fördelningen mellan positiva och negativa upplevelser var väldigt jämn: 13 svar antydde att samarbetet var positivt, till exempel rättvist, och att alla fick sin åsikt hörd. 14 svar

behandlade negativa upplevelser av samarbetet: i sex fall handlade det om att andra kommittémedlemmar inte var så aktivt med i processen, i åtta fall om att det fanns politiska intressen, främst i att gynna det egna ämnesrådets tidskrifter, eller, som indikerades i två fall, att det fanns bias gällande tidskrifter man själv hade en koppling till. Det var alltså bara två respondenter som tog upp den bias som undersöktes i föreliggande studie. I ett av svaren menade respondenten att det fanns en negativ påverkan på nivå 2 och 3 av JUFO-poängen, genom att bedömningen var strängare om personen hade en koppling till publikationskanalen.

Det tredje temat handlade om åsikter om själva JUFO-systemet. Här fanns det två respondenter, en från humaniora och en från samhälls- och beteendevetenskaper, som uttryckte att det är ett bra system, medan det fanns 15 svar, jämnt fördelade mellan vetenskapsområdena, där kritik av någon aspekt av systemet eller systemet i sin helhet uttrycktes. Enbart tre respondenter, två inom biovetenskaper och en inom samhälls- och beteendevetenskaper, uttryckte att systemet i sin helhet är dåligt eller att det borde avskaffas. Kritiken handlade för det mesta om att panelindelningen (i ämnesområden) är för bred eller annars inte så ändamålsenlig, och att JUFO-poängen en tidskrift får påverkas av vilken panel den råkar bedömas av, eftersom olika paneler har olika mängder och storleks tidskrifter att bedöma, medan kvoterna är samma för alla. Volymbegränsningarna och kvoterna kritiserades i flera fall, eftersom det gör bedömningen svår, och inte kvalitetsbaserad. Flera respondenter uttryckte också att klassificeringen av tidskrifter på nivå 2 och 3 var särskilt svår, eller att skalan 1–3 inte var optimal.

Inga tydliga skillnader förekom mellan vetenskapsområdena i vilka teman som togs upp. Fyra positiva svar om arbetsfördelningen i kommittéerna, det vill säga arbetsfördelningen som följde av faktumet att den enskilde kommittémedlemmen inte kunde känna till alla tidskrifter och övrigt samarbete, förekom inom bio- och hälsovetenskaper, naturvetenskaper och samhällsvetenskaper. Inom humanistiska vetenskaper förekom tre positiva svar om kommittéernas arbetsfördelning. Inom naturvetenskaper och bio- och hälsovetenskaper framkom tre negativa svar, medan det framkom två negativa svar inom samhällsvetenskaper och humanistiska vetenskaper. Gällande positiva tankar om JUFO-systemet fanns det bara två svar, ett från samhälls- och beteendevetenskaper och ett från humaniora. Fördelningen av vetenskapsområdena på de 15 negativa svaren var jämn: fyra för såväl naturvetenskaper, bio- och hälsovetenskaper som samhälls- och beteendevetenskaper, och tre för humaniora.

Slutsatsen som drogs av analysen av de öppna frågorna var att det var vanligare att rikta kritik mot JUFO-systemet, eller någon aspekt av det, än att förhålla sig positivt till det.

Svaren tydde på att kommittémedlemmarna inte kände till alla tidskrifter, vilket gjorde att de ofta erhöll information om dem från webbsidor eller litade på de andra kommittémedlemmarnas bedömningar. De flesta svaren tydde dock på att detta ansågs vara ett fungerande tillvägagångssätt, framom att det skulle vara någonting negativt eller något som skulle ha hotat neutraliteten i bedömningen. Det förekom en jämn fördelning bland svaren mellan att uppleva samarbetet i panelen som positivt och att uppleva det som negativt: egen bias togs upp bara i ett fall, medan det i de negativa svaren handlade om att andra påverkades av bias, vägrade ändra sin åsikt, styrdes av politiska intressen eller inte var så aktivt med i processen.

4 Diskussion

I denna avhandling undersöktes om JUFO-kommittémedlemmars forskningsprofiler hade en inverkan på deras bedömning av vetenskapliga tidskrifter och om de därmed påverkades av denna typ av bias under klassificeringsprocessen. I avhandlingen undersöktes även om det förekom variation i bias mellan vetenskapsområdena naturvetenskaper, bio- och hälsovetenskaper, samhällsvetenskaper och humaniora. Kommittéernas klassificeringar av tidskrifter jämfördes med tidskrifternas Impact Factors för att avgöra om diskrepans mellan bedömningsmåttet förekom. Som biasmått undersöktes effekten av antalet kommittémedlemmar som hade publicerat i en tidskrift på JUFO-klassificering. En enkät skickades även ut till JUFO-kommittémedlemmarna där de ombads svara på frågor om bedömningsprocessen.

4.1 Bias inom JUFO-klassificeringen av tidskrifter

Resultaten från denna avhandling tydde på att en viss mängd bias förekom under bedömningen av tidskrifter i Finland. Samhällsvetenskaper var ämnesområdet som mest tog i beaktande tidskrifters Impact Factors under klassificeringsprocessen.

Samhällsvetenskapernas JUFO-klassificeringar påverkades däremot signifikant mest av antalet kommittémedlemmar som hade publicerat i tidskriften, genom att en tidskrift erhöll en högre JUFO-klassificering i takt med att antalet kommittémedlemmar som hade publicerat i tidskriften ökade. Det här resultatet kan tolkas som ett tecken på att mer ingruppsbias förekom inom samhällsvetenskaper genom att ju fler av kommittémedlemmarna som hade

publicerat i en tidskrift, desto bättre klassifikation fick tidskriften. Bio- och hälsovetenskaper och humanistiska vetenskaper var ämnesområdena som i lägst grad tog i beaktande tidskrifters Impact Factors under bedömningsprocessen. Mindre beaktande av Impact Factors under utvärderingsprocessen skulle ha kunnat ge utrymme för mer bias. Om en lägre effekt av Impact Factors skulle reflektera ökad bias, borde den potentiella biasen ha härstammat från andra, okända källor. Det är däremot även möjligt att bio- och hälsovetenskaper och humanistiska vetenskaper relativt sätt litade mer på andra, kanske icke-biaserade, informationskällor än Impact Factors.

De humanistiska vetenskaperna särskilde sig som ett vetenskapsområde där panelisterna främst litade mycket på andra kollegors och/eller andra kommittémedlemmars åsikter under bedömningsprocessen, inte lika mycket på Impact Factors. Det är möjligt att det här berodde på att flera av de tidskrifter som bedömdes (dock inte dom som ingick i föreliggande samplet) inte hade någon Impact Factor, alternativt att humanistiska vetenskaper i allmänhet publicerar i lägre mån i tidskrifter än andra vetenskapsområden. Humanistiska vetenskaper var även ämnesområdet vars kommittémedlemmar upplevde att de sämst kände till tidskrifterna de skulle bedöma. Detta kan däremot inte klassas som en bias, eftersom problemet löstes genom att lägga stor vikt vid gruppdiskussioner inom kommittéerna, vilket kan tolkas som kollegial granskning av bedömningsprocessen framom förlitande på de egna åsikterna.

Med hänvisning till biaserna som presenterades i tabell 1, var det komplicerat att avgöra vilken eller vilka biaser som kan ha förekommit under klassificeringsprocessen av tidskrifter. Utgående från resultaten från såväl analysen av JUFO-poäng mot Impact Factor-poäng och analysen av enkäten var det möjligt att samtliga biaser kan ha förekommit, det vill säga förankringseffekt, baktändning, blinda fläcken-bias, konfirmationsbias, falsk slutledning, haloeffekt samt ingrups- och utgrupsbias. Den enda predicerade biasfaktorn som bevisligen förekom under bedömningsprocessen var effekten av antalet kommittémedlemmar som hade publicerat i en tidskrift på JUFO-poäng. Denna bias klassificerades som en ingrupsbias, genom att tidskrifter som kommittémedlemmarna själva hade publicerat i favoriserades framom tidskrifter som de inte hade publicerat i.

Det var möjligt att tendensen att ge högre poäng åt en tidskrift som personen själv publicerat i kunde förklaras med att forskarna verkligen publicerat i mer högklassiga tidskrifter än genomsnittet av tidskrifter. En jämförelse mellan hela samplets Impact Factors och Impact Factors för de tidskrifter kommittémedlemmarna publicerat i gjordes. Det visade sig att Impact Factors var något högre för de tidskrifter kommittémedlemmarna publicerat i

än för hela samplet, vilket delvis kunde förklara de högre JUFO-poängen för dessa tidskrifter, eftersom de objektivt sett var något högklassigare än andra. Effekten av antalet kommittémedlemmar som publicerat på JUFO-poängen var också låg, även om den var signifikant. Dock var korrelationen mellan de tidskrifter kommittémedlemmarna publicerat i och IF tämligen låg. Effekten av antalet kommittémedlemmar som publicerat på JUFO-poängen borde ändå inte ha varit signifikant, om vi antar att ingen bias fanns, eftersom det också sägs i instruktionerna i handboken som kommittéerna får att man ska rapportera vilka kopplingar man har till tidskriften, och därmed medvetandegöra att det kan utgöra en risk för bias. Denna typ av bias representerar högst antagligen ingrupsbias, att man har en positivare inställning till något man haft kontakt med än något man inte haft.

4.2 Enkätresultaten

Till de mest relevanta frågorna i enkäten som skickades ut till JUFO-kommittémedlemmarna hörde hur mycket panelisterna upplevde att Impact Factors, den egna uppfattningen om tidskrifter och andras uppfattningar om tidskrifter påverkade bedömningen av tidskrifter. Utöver dessa frågor var även frågor om hur väl panelisterna upplevde att de kände till tidskrifterna de bedömde, hur mycket andra panelisters åsikter påverkade den egna åsikten, hur mycket den egna åsikten påverkade de andra panelisternas åsikter och hur väl alla åsikter togs i beaktande relevanta.

Resultaten från enkäten tydde på att den egna åsikten mest påverkade bedömningen (medeltal 8,55), medan andra kommittémedlemmars uppfattningar hade ett medeltal på 6,55. Impact Factors verkade även påverka bedömningen på en hög nivå, med ett medeltal på 7,55. Det här resultatet stämmer överens med resultatet från den linjära modellen där JUFO-bedömningen från 2015 fungerade som beroende variabel; Impact Factors förklarade bäst variationen i JUFO-poäng. Impact Factors nämndes däremot bara sex gånger i de öppna svaren av enkäten, men det här motstridiga resultatet kan ha berott på att kommittémedlemmarna ansåg att det var självklart att följa Impact Factors och det därför inte var nödvändigt att diskutera användningen av Impact Factors desto mer i de öppna frågorna.

Några av svaren på de öppna frågorna tydde på att en del av kommittémedlemmarna inte ansåg att enkätfrågorna var ändamålsenliga, eftersom det uppfattades vara självklart att den egna uppfattningen påverkade andras åsikter och den slutliga bedömningen då personen kände väl till tidskrifter eller ämnesområden, medan andra kommittémedlemmars åsikter påverkade mer den egna bedömningen då tidskriften föll utanför ens personliga

expertisområde. Det förekom däremot delade uppfattningar om huruvida det var ändamålsenligt att dela upp kommittéarbetet genom att främst bedöma tidskrifter inom det egna ämnesområdet och att lita på andras expertis inom andra ämnesområden. Dessa delade uppfattningar kom även fram i medelvärdena för hur mycket panelisterna uppfattade att andras åsikter påverkade den egna åsikten och hur mycket den egna åsikten påverkade andras åsikter, med medelvärden på 5,89 respektive 6,25, det vill säga värden som förekom i mitten av skalan. Medelvärdet för hur väl allas åsikter togs i beaktande var däremot 7,35, vilket tydde på att uppfattningen var att allas åsikter i regel behandlades rättvist oberoende av den personliga uppfattningen om arbetsfördelningen.

4.3 Begränsningar

En av avhandlingens analyser baserade sig på ett slumpmässigt sampel på 973 JUFO-bedömda tidskrifter samt deras JUFO-klassificeringar och Impact Factors från år 2015. Samtliga ämnesområden, det vill säga naturvetenskaper, bio- och hälsovetenskaper, samhällsvetenskaper och humaniora, var representerade, men de slutliga samplen var av olika storlek. Tidskrifter som representerade humanistiska vetenskaper var lägst till antalet, vilket kan ha påverkat resultaten då antalet tidskrifter inom till exempel naturvetenskaper var fyra gånger fler. Ämnesfördelningen var däremot jämnare i avhandlingens andra analys som baserade sig på enkätsvar av JUFO-kommittémedlemmar, inget ämnesområde var signifikant mer representerat än något annat.

Enkäten som användes i avhandlingen var konstruerad för att få ut så mycket information som möjligt om JUFO-kommittémedlemmarnas personliga åsikter om bedömningsprocessen av tidskrifter. Det är däremot möjligt att fler eller på ett annat sätt konstruerade frågor skulle ha behövts för att erhålla ytterligare information om bedömningsprocessen, till exempel hur bedömningen påverkades av kvoterna och ämnesområdesindelningen som JUFO-kommittémedlemmarna var tvungna att förhålla sig till.

4.4 Slutsatser

Tidigare utförd forskning tyder på att det finns såväl fördelar som nackdelar med kollegial granskning och liknande klassifikationssystem, vilket även kom fram genom enkäten som skickades ut till JUFO-kommittémedlemmarna. Flera upplevde att klassificeringsprocessen

fungerar och utgör ett centralt verktyg som tillför någonting utöver Impact Factor-poäng, medan andra upplevde att systemet är partiskt och ohållbart. Upprätthållandet av JUFO kräver mycket resurser; 20 000 tidskrifter ska klassificeras av 23 paneler som består av över 200 människor sammanlagt, vilket gör det möjligt att ifrågasätta om ett så pass resurskrävande system är nödvändigt och tillför någonting som exempelvis Impact Factor inte gör. Korrelationen mellan IF och JUFO-poängen visade sig vara relativt svag, men signifikant. Det här betyder att även om JUFO-poängen hade en koppling till IF fanns det ändå en hel del andra källor för varians i poängen, men det tycktes inte i så hög grad vara kommittémedlemmarnas egna publikationer.

Det är möjligt att konkludera att det förekommer en risk för bias i JUFO-systemet. Biasen baserar sig antagligen däremot inte på avhandlingens tidigare upplagda hypotes, det vill säga att forskarnas tidigare kopplingar till tidskrifter påverkar deras bedömning, utan snarare på brister inom systemet, med hänvisning till de öppna svaren i enkäten. Till dessa brister hör att bedömningarna inte baserar sig på enbart tidskrifters kvalitet, utan långt på kvoter om hur många tidskrifter som kan tilldelas respektive JUFO-klassificering. Utöver det verkade partiskhet förekomma inom respektive ämne, genom att forskare ville gynna tidskrifter inom det egna vetenskapsområdet.

Det är möjligt att diskutera vad bias och objektivitet egentligen är, och om objektivitet ens är möjligt att erhålla i vetenskap, eller om positionering och ett visst mått av (tydligt dokumenterad) subjektivitet egentligen är det närmaste vi kan komma sann objektivitet, vilket till exempel flera feministiska forskare hävdar (se t.ex. Harding, 1993; Alcoff, 1991). Impact Factor har bland annat kritiserats för att inte ta olika vetenskapsgrenars publiceringstraditioner i beaktande (Guerrero-Bote & Moya-Anegón, 2012), och har därför inte ansetts vara lämpligt som enda mått på högklassig vetenskap. För att råda bot på det här problemet har dock nya påverkningsfaktorer utvecklats (ibid). Var gränsen mellan bias och kollektiv bedömning går, såsom mellan JUFO-klassificering som ett alternativ till ”felbar” matematiskt uträknad objektivitet (IF), är alltså svår att dra.

Referenser

- Alcoff, Linda (1991) The Problem of Speaking for Others. *Cultural Critique 1991/1992* (20), 5–32.
- Armstrong, J. S. (1996). We need to rethink the editorial role of peer reviewers. *Chronicle of Higher Education*, 43(9), B3-B4.
- Auranen, Otto & Janne Pölönen (2012). Tieteellisten julkaisukanavien tasoluokitus. Julkaisufoorumi-hankkeen (2010-2012) loppuraportti. *Tieteellisten seurain valtuuskunnan verkkojulkaisuja* 1/2012. Tillgängligt på http://www.ctors-council.helsinki.fi/raportit_ja_julkaisut/Julkaisujen_laadunarviointi.pdf [Hämtat 27.9.2018].
- Caellegh, A. S., Hojat, M., Steinecke, A., & Gonnella, J. S. (2003). Effects of reviewers' gender on assessments of a gender-related standardized manuscript. *Teaching and Learning in Medicine*, 15(3), 163-167.
- Cho, M. K., Justice, A. C., Winker, M. A., Berlin, J. A., Waeckerle, J. F., Callahan, M. L., ... & peer investigators. (1998). Masking author identity in peer review: What factors influence masking success? *JAMA*, 280(3), 243–245.
- Clarivate Analytics (2018). Journal Impact Factor. Tillgängligt på <http://ipscience-help.thomsonreuters.com/inCites2Live/indicatorsGroup/aboutHandbook/usingCitationIndicatorsWisely/jif.html>. [Hämtat 29.10.2018]
- Crane, D. (1967). The gatekeepers of science: Some factors affecting the selection of articles for scientific journals. *The American Sociologist*, 195-201.
- Bornmann, L., & Daniel, H. D. (2004). Reliability, fairness and predictive validity of committee peer review. *BIF Futura*, 19, 7-19.

- Demicheli, V., & Di Pietrantonj, C. (2007). Peer review for improving the quality of grant applications. *Cochrane Database of Systematic Reviews*, utgåva 2. Art. No.: MR000003. DOI: 10.1002/14651858.MR000003.pub2.
- Finlands Akademi (2018). Tillgängligt på <http://www.aka.fi/sv/>. [Hämtat 18.10.2018]
- Godlee, F., Gale, C. R., & Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *JAMA*, 280(3), 237 – 240.
- Godlee, F., & Jefferson, T. (2003). *Peer Review in Health Sciences, 2nd Edition*, (91 – 117). London: BMJ Books.
- Guerrero-Bote, Vicente P. & Moya-Anegón, Félix. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics* 6 (2012), 674–688. <http://dx.doi.org/10.1016/j.joi.2012.07.001>. [Hämtat 7.2.2019].
- Harding, Sandra (1993) Rethinking standpoint epistemology: what is 'strong objectivity'? I Alcoff, Linda & Potter, Elizabeth (red.) *Feminist Epistemologies*. New York: Routledge, 49-82.
- Hojat, M., Gonnella, J. S., & Caelleigh, A. S. (2003). Impartial judgment by the “gatekeepers” of science: fallibility and accountability in the peer review process. *Advances in Health Sciences Education*, 8(1), 75 – 96.
- Jayasinghe, U. W., Marsh, H. W., & Bond, N. (2006). A new reader trial approach to peer review in funding research grants: An Australian experiment. *Scientometrics*, 69(3), 591-606.
- Jefferson, T., Rudin, M., Brodney Folse, S., & Davidoff, F. (2007). Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database of Systematic Reviews*, (2), MR000016. doi:10.1002/14651858.MR000016.pub3.

- Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A., Rennie, D., & Peer Investigators. (1998). Does masking author identity improve peer review quality? A randomized controlled trial. *JAMA*, 280(3), 240-242.
- Kaatz, A., Gutierrez, B., & Carnes, M. (2014). Threats to objectivity in peer review: the case of gender. *Trends in Pharmacological Sciences*, 35(8), 371-373.
- Kassirer, J. P., & Campion, E. W. (1994). Peer review: crude and understudied, but indispensable. *JAMA*, 272(2), 96-97.
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2-17.
- Luty, J., Arokiadass, S. M. R., Easow, J. M., & Anapreddy, J. R. (2009). Preferential publication of editorial board members in medical specialty journals. *Journal of Medical Ethics*, 35(3), 200-202.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2), 161-175.
- Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. *American Psychologist*, 63(3), 160.
- McIntosh, E. G., & Ross, S. (1987). Peer review in psychology: Institutional ranking as a factor. *Psychological Reports*, 60(3), 1049-1050.
- Opetus- ja kulttuuriministeriön asetus yliopistojen perusrahoituksen laskentakriteereistä 331/2016. Helsingfors 3.5.2016: Finlands Författningssamling.
- Peters, D. P., Ceci S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behaviour and Brain Sciences*, 95(5), 187-255.

- Pitkin, R. M. (1995). Blinded manuscript review: an idea whose time has come? *Obstetrics & Gynecology*, 85(5), 781–782.
- Publikationsforum (2017a). Julkaisufoorumi. Tillgängligt på <http://www.julkaisufoorumi.fi/fi/julkaisufoorumi>. [Hämtat 28.10.2018]
- Publikationsforum (2017b). Usein kysytyt kysymykset. Tillgängligt på https://www.tsv.fi/julkaisufoorumi/usein_kysytyä.php. [Hämtat 3.10.2017]
- Publikationsforum (2017c). Uusien panelistien valintaperusteet. Tillgängligt på <http://www.julkaisufoorumi.fi/fi/arkisto/tiedotteet/uusien-panelistien-valintaperusteet>. [Hämtat 28.10.2018]
- Publikationsforum (2017d). Luokitteluperusteet. Tillgängligt på <http://www.julkaisufoorumi.fi/fi/arvioinnit/luokitteluperusteet>. [Hämtat 28.10.2018]
- Rantala, Jukka (2012). Julkaisufoorumi-hanke huolestuttaa humanisteja ja yhteiskuntatieteilijöitä. *Kasvatus ja aika* 1/2012. Tillgängligt på http://www.kasvatus-ja-aika.fi/site/?lan=1&page_id=448. [Hämtat 15.9.2018]
- Ross, J. S., Gross, C. P., Desai, M. M., Hong, Y., Grant, A. O., Daniels, S. R., ... & Krumholz, H. M. (2006). Effect of blinded peer review on abstract acceptance. *JAMA*, 295(14), 1675-1680.
- Savolainen, Eeva (2018). Jufo-käsikirja paneeleille. Tillgängligt på <https://wiki.eduuni.fi/pages/viewpage.action?pageId=66327793>. [Hämtat 15.9.2018].
- Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99(4), 178-182.
- Strauss, B. (1994). The effects of blinding on acceptance of research papers by peer review. *JAMA*, 272(2), 143-146.

- Uleåborgs universitet (2019). Tieteellisiin julkaisuihin pohjautuva arviointi: Source normalized impact per paper, SNIP. Tillgängligt på <http://libguides oulu.fi/c.php?g=4184&p=718944>. [Hämtat 7.2.2019].
- Van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *British Medical Journal*, 318(7175), 23-27.
- Ware, M. (2008). Peer review in scholarly journals: Perspective of the scholarly community—Results from an international study. *Information Services & Use*, 28(2), 109-112.
- Ware, M. (2011). Peer review: Recent experience and future directions. *New Review of Information Networking*, 16(1), 23–53.
- Ware, M., & Mabe, M. (2015). *The STM Report: An Overview of Scientific and Scholarly Journal Publishing*. Haag: International Association of Scientific, Technical and Medical Publishers.

Bilagor

Bilaga 1

Översikt på signifikanta resultat med jämförelser mellan vetenskapsområdena

Grupp	Estimat	Standardfel	t-värde	Pr(> t)
<i>Bio- och hälsovetenskaper som referensgrupp</i>				
(Ref = biohälsa)hum	1,31203	0,17731	7,400	2,98e-13 ***
(Ref = biohälsa)natur	0,26561	0,06298	4,217	2,71e-05 ***
(Ref = biohälsa)samhälle	0,83615	0,07585	11,024	< 2e-16 ***
(Ref = biohälsa)natur:Impact Factor	0,30458	0,09206	3,308	0,000973 ***
(Ref = biohälsa)samhälle:Impact Factor	1,28737	0,13991	9,201	< 2e-16***
(Ref = biohälsa)samhälle:Kommittémedlemmar	0,29990	0,10917	2,747	0,006126 **
<i>Naturvetenskaper som referensgrupp</i>				
(Ref = natur)hum	1,04049	0,17949	5,797	9,17e-09 ***
(Ref = natur)samhälle	0,57054	0,08099	7,045	3,54e-12 ***
(Ref = natur)samhälle:Impact Factor	0,98279	0,16253	6,047	2,12e-09 ***
(Ref = natur)samhälle:Kommittémedlemmar	0,27744	0,10781	2,573	0,010222 *
<i>Samhällsvetenskaper som referensgrupp</i>				
(Ref = samhälle)hum	0,42602	0,18563	2,295	0,021950 *
(Ref = samhälle)hum:Impact Factor	-0,88587	0,32834	-2,698	0,007098 **

Signifikansnivåer: 0 '****' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1. Residualernas standardfel: 0,81 på 957 frihetsgrader (29 observationer raderade på grund av att värdena saknades). Multipel R²: 0,2883, Adjusted R²: 0,2772, F-statistik: 25,85 på 15 och 957 frihetsgrader, p-värde: < 2,2e-16.

Bilaga 2

Enkät utskickad till alla JUFO-kommittémedlemmar 2014–2017

Questionnaire on the decision-making process in the JUFO committees for the period 2014–2017

Dear recipient,

We are conducting a study for our master's thesis on the process of rating publication channels in the Julkaisufoorumi committees. A publication channel is defined here as a journal, publishing house, journal series or conference. With this questionnaire, we are hoping to get a picture of the committee members' own experience of this decision-making process. The answers will be collected anonymously, and no specific person or committee can be recognized from the results. We would only like you to specify your field of research and gender. Filling out the questionnaire should not take longer than 5-10 minutes.

Julia Nummenmaa, Department of Psychology, Åbo Akademi University
julia.nummenmaa@abo.fi

Annika Tuomi, Department of Psychology, Åbo Akademi University
annika.tuomi@abo.fi

Matti Laine (supervisor), Professor of Psychology, Åbo Akademi University
matti.laine@abo.fi

I accept that by answering and returning this questionnaire my anonymous responses will be used in the thesis mentioned above, and in a possible future publication.

Yes

No

Please specify your field of research according to one of the four categories listed below.

Natural sciences (e.g. maths, information technology, physics, chemistry, geosciences, automation sciences and engineering)

Biosciences (e.g. biology, medicine, health sciences and agricultural sciences)

Political and behavioural sciences (e.g. economics, social sciences, political science and psychology)

Humanities (e.g. languages, philosophy, cultural sciences, theology, history and archaeology)

Please specify your gender

Female

Male

Other

1. How much do you think the following factors influenced your ratings of the publication channels while working in a JUFO committee? Please indicate your response with a number, 1 being no influence and 10 being very high influence.

- a) The publication channel's impact factors
1 2 3 4 5 6 7 8 9 10
 - b) The rating the publication channel has received in similar committee rating systems as JUFO, such as the Norwegian and Danish counterparts
1 2 3 4 5 6 7 8 9 10
 - c) Your own perception of the quality of the publication channel
1 2 3 4 5 6 7 8 9 10
 - d) The other committee members' perception of the quality of the publication channel
1 2 3 4 5 6 7 8 9 10
 - e) Other colleagues' perception of the quality of the publication channel
1 2 3 4 5 6 7 8 9 10
 - f) Other, please specify
1 2 3 4 5 6 7 8 9 10
-

- 2. How well do you think you personally knew the publication channels you rated? Please indicate your response with a number, 1 being not at all and 10 being very well.
1 2 3 4 5 6 7 8 9 10
- 3. How difficult do you think rating the publication channels was for you? Please indicate your response with a number, 1 being not at all difficult and 10 being very difficult.
1 2 3 4 5 6 7 8 9 10
- 4. How easy do you think agreeing on a rating for a publication channel was in your committee? Please indicate your response with a number, 1 being very difficult and 10 being very easy.
1 2 3 4 5 6 7 8 9 10
- 5. To what degree do you think the other committee members' opinion of the publication channels affected your own opinion of the publication channels? Please indicate your response with a number, 1 being not at all and 10 being to a very high degree.
1 2 3 4 5 6 7 8 9 10
- 6. To what degree do you think your opinion of the publication channels affected the other committee members' opinions of the publication channels? Please indicate your response with a number, 1 being not at all and 10 being to a very high degree.
1 2 3 4 5 6 7 8 9 10
- 7. To what extent do you think your own connections to a publication channel (e.g. having previously published in it or been a member of the editorial board) affected

your rating of the publication channel in question? Please indicate your response with a number, 1 being not at all and 10 being to a very high extent.

1 2 3 4 5 6 7 8 9 10

8. To what extent do you think each committee member's opinion of the publication channel was taken into account in the rating process? Please indicate your response with a number, 1 being not at all and 10 being to a very high extent.

1 2 3 4 5 6 7 8 9 10

9. Optional: please describe your experience of the committee work and the rating process in general

Thank you very much for answering and helping us with our study!

Bilaga 3

Indelning av JUFO-kommittéerna i vetenskapsområden

Publikationsforums bedömningskommittéer 2018–2021

Naturvetenskaper:

- 1 Matematik och statistik
- 2 Informationsbehandling och informationsvetenskaper
- 3 Fysik, rymdforskning och astronomi
- 4 Kemi
- 5 Geologi och miljövetenskaper
- 8 Byggnads- och samhällsteknik, maskin- och produktionsteknik
- 9 El- och automationsteknik, informations- och kommunikationsteknik
- 10 Kemiteknik, materialteknik och miljöteknik

Bio- och hälsovetenskaper:

- 6 Biovetenskaper I
- 7 Biovetenskaper II
- 11 Medicinteknik, bioteknologi och allmänmedicin
- 12 Klinisk medicin I
- 13 Klinisk medicin II och odontologi
- 14 Hälsovetenskaper och övriga medicinska vetenskaper
- 15 Jord- och skogsbruksvetenskaper

Samhälls- och beteendevetenskaper:

- 16 Ekonomi
- 17 Socialvetenskaper, medievetenskaper och kommunikation samt övriga samhällsvetenskaper
- 18 Psykologi och pedagogik
- 19 Statsvetenskap, förvaltningslära och rättsvetenskaper

Humanistiska vetenskaper:

20 Filosofi och teologi

21 Språkvetenskaper

22 Litteraturvetenskap, konstvetenskap och arkitektur

23 Historia, arkeologi och kulturforskning

Pressmeddelande

JUFO-bedömningssystemet av vetenskapliga tidskrifter i Finland potentiellt utsatt för bias

Resultaten från en pro gradu-avhandling i psykologi vid Åbo Akademi tyder på att finländska forskare tenderar att värdera tidskrifter de själva publicerat i högre än andra tidskrifter. I avhandlingen analyserades klassificeringar av vetenskapliga tidskrifter utförda av Publikationsforum (JUFO), som är ett klassificeringssystem vars uppgift är att stöda den vetenskapliga publikationsverksamhetens kvalitetsbedömning. Forskarnas preferens för tidskrifter de själva hade publicerat i framkom tydligast inom samhällsvetenskaper, framom naturvetenskaper, bio- och hälsovetenskaper samt humanistiska vetenskaper. Samhällsvetenskaper var även det ämnesområde som trognast följde tidskrifters Impact Factors, vilket är ett matematiskt mått på en tidskrifts inflytande, i sina bedömningar. En enkät skickades ut till alla JUFO-panelister för att erhålla information om deras åsikter om bedömningsprocessen. I enkätsvaren framkom en del kritik riktad mot bedömningssystemet, bland annat genom att flera panelister upplevde att de strikta kvoter som måste följas i bedömningarna kan utgöra en grogrund för bias, vilket gör processen politisk snarare än en avspegling av god vetenskaplig nivå. Vetenskapliga tidskrifter bedöms enligt skalan 1-3 av paneler bestående av forskare. Klassificeringssystemets reliabilitet och validitet kan ifrågasättas om partiskhet eller bias gentemot tidskrifter som forskarna själva har publicerat i förekommer.

Biaser uppstår då personliga åsikter, utvärderingar och erfarenheter påverkas en persons omdömesförmåga, bedömning och beslutfattande. Alla människor påverkas i viss mån av biaser vid beslutfattande, trots medvetna försök till objektivitet. Resultaten från denna avhandling tydde på att ingrupsbias förekom vid klassificeringen av tidskrifter i Finland, genom att bedömningarna i viss mån påverkades av kommittémedlemmarnas egen publiceringsaktivitet i de tidskrifter som bedömdes.

Föreliggande avhandlings sampel bestod av 973 JUFO-bedömda tidskrifter samt deras JUFO-klassificeringar och så kallade Impact Factors från år 2015. Enkäten besvarades av 85 av sammanlagt 201 panelister, 53 män och 32 kvinnor. Tjugofem av respondenterna var verksamma inom naturvetenskaper, 22 inom bio- och hälsovetenskaper, 19 inom samhällsvetenskaper och 19 inom humanistiska vetenskaper. Alla panelister var verksamma inom JUFO år 2015.

Ytterligare information fås av avhandlingens författare:

Julia Nummenmaa/Annika Tuomi

Tel. 044-2501177/040-5415644

E-post: julia.nummenmaa@abo.fi/annika.tuomi@abo.fi