

ACTA  
REGIAE SOCIETATIS SCIENTIARUM ET LITTERARUM  
GOTHOBURGENSIS

---

Interdisciplinaria  
18

# Upprepbarhet och generaliserbarhet i forskningen

Redaktör: Bo Lindberg



---

Kungl. Vetenskaps- och Vitterhets-Samhället  
GÖTEBORG

© Författarna och Kungl. Vetenskaps- och Vitterhets- Samhället i Göteborg, 2019  
Bokomslag: Gunnar Dahlström

*Distribution:*

Kungl. Vetenskaps- och Vitterhets-Samhället i Göteborg  
Box 222, 405 30 Göteborg

ISBN 978-91-984647-2-6

ISSN 0347-4925

Rundqvists Boktryckeri AB, Göteborg, 2019

## INNEHÅLL

Förord .....	5
<i>Sven-Eric Liedman</i> : Vad går att upprepa och vad går inte? .....	7
<i>Peter Jagers</i> : Slumpvariationer i det annars upprepbara .....	15
<i>Anna Dreber och Magnus Johannesson</i> : Tillförlitlighet av forskningsresultat – vilka resultat går att upprepa? .....	23
<i>Lars Nyberg</i> : Aktivering av hjärnan och reproducerbarhet .....	37
<i>Jan-Erik Gustafsson</i> : Begrepps referensgeneralitet och datas inferensnivå som determinanter för möjligheter till generalisering ..	49
<i>Elisabet Engdahl</i> : Sökandet efter en generell språk teori .....	65
<i>Arne Jarrick</i> : Historiens händelser kan inte köras om som i en partikelaccelerator – vad gör man då med upprepbarhet och generalisering? .....	81
<i>Hans Ruin</i> : En delad blick – Husserl, Derrida och det teoretiska seendets kall .....	95
De medverkande .....	109



## Förord

Upprepning och kunskap hör ihop. *Repetitio est mater studiorum*, repetition är kunskapens moder, uttrycker en sliten men grundläggande pedagogisk erfarenhet. Men upprepningen är central också i vetenskapen, där det gäller att generalisera sina resultat. Att generalisera innebär enligt ordboken att ge en (mer) allmän giltighet åt en regel eller utsaga som endast gällt ett fåtal fall. Filosoferna talar om induktion, där man av iakttagna fenomen sluter sig till regelmässiga samband. I en svag mening generaliserar alla forskare: också humanister, som antas syssla med det unika, måste redovisa var de gjort sina iakttagelser och därigenom göra dem kontrollerbara. Men i stark mening förekommer generalisering i naturvetenskaperna och de s.k. beteendevetenskaperna, där man strävar efter att fastställa allmängiltiga samband. I de vetenskaperna söker man bekräfta generaliseringarna genom att upprepa observationer, experiment och jämförelser för att se om de ger samma resultat. För att upprepningen skall vara tillförlitlig måste omständigheterna vara desamma som i den ursprungliga undersökningen och man måste också ta hänsyn till att resultat kan bero på slumpen.

Den här boken handlar om detta upprepande, eller *upprepbarheten*, i forskningen. Ordet saknas i SAOL och är knappast etablerat i svenskan; från engelskan inlånade synonymer är reproducerbarhet, replicerbarhet och replikabilitet. Oavsett benämning behövs begreppet för att diskutera de metodiska, filosofiska och forskningsetiska problem som är förenade med bemödandena att generalisera forskningsresultaten. Så sker i bidragen i denna volym som är produkten av ett symposium om som hölls i Kungl. Vetenskaps- och Vitterhets- Samhället i Göteborg i oktober 2018. Symposiet planerades och utformades i enlighet med sedvänjan i Kungliga Samhället av ordföranden det året, Erland Hjelmquist.

Texterna behandlar ämnet från olika disciplinära synpunkter.

Sven-Eric Liedman ger det historiska perspektivet från Aristoteles och framåt; en poäng hos honom är att skillnaderna mellan vetenskaperna inte är så stora som man ofta gör gällande. Peter Jagers diskuterar från statistikerns

synpunkt slumpens underskattade roll i ansträngningarna att generalisera vetenskapliga resultat.

Anna Dreber och Magnus Johannesson belyser upprepbarhetens svårigheter och fällor i de kvantitativa samhällsvetenskaperna.

Lars Nyberg beskriver hur man i neurovetenskapen registrerar hjärnans aktivitet med magnetkamera och hur upprepbarhetens problem då kan bemästras.

Jan-Eric Gustafsson beskriver upprepbarhetens roll i pedagogisk forskning med inriktning på begreppet inferensnivå som ett sätt att hantera skillnaden mellan kvalitativa och kvantitativa undersökningar.

Elisabet Engdahl skriver om hur man i modern språkvetenskap sökt formulera en universell grammatik tillämplig på alla språk; här är det inte upprepbarhetens svårigheter som står i fokus utan snarare möjligheten att finna de generella likheterna mellan språken.

Också Arne Jarrick är från sin utsiktspunkt som historiker snarare intresserad av möjligheterna till generalisering än problemen med upprepbarheten: hans fråga är i vilken grad det är möjligt att formulera empiriskt grundade generaliseringar om det förflutna, som ju inte låter sig upprepas.

Hans Ruin slutligen ägnar sitt bidrag åt Edmund Husserls vetenskapsfilosofi, som var ett försök att övervinna naturalistisk objektivering och relativistisk historisering av vetandet och istället ge det ett etiskt sammanhang; upprepbarheten ligger här inte i repeterbara experiment utan i traderingen av etiskt meningsfull kunskap.

*Bo Lindberg*

*Sven-Eric Liedman*

## Vad går att upprepa och vad går inte?

En stor del av den moderna vetenskapen bygger på förutsättningen att iakttagelser och experiment kan upprepas. Vad den ene forskaren sett måste den andre också kunna se. Resultat som redovisas måste kunna prövas av vem som helst annan som är kompetent inom området, godtas eller förkastas.

Det är på det sättet vetenskapen fungerar och kan göra framsteg. Men innebär det att allt också kan generaliseras? Med experiment är det uppenbarligen så. Har man upptäckt ett mönster, så ska man kunna se det mönstret varje gång man upprepar experimentet.

Frågan är om det som gäller experiment också gäller all vetenskaplig kunskap. Jag kommer att behandla den i slutet av framställningen. Men för att göra den mer hanterlig kommer jag att belysa den historiska bakgrunden. Vi brukar göra det genom att gå tillbaka till de gamla grekerna. Denna artikel utgör inget undantag. Den tar sin utgångspunkt hos den store filosofen och forskaren Aristoteles (384-322 f. v. t.) och hans syn på den mänskliga kunskapen.

Aristoteles skiljer mellan vad han kallar den teoretiska, den praktiska och den tekniska kunskapen. Den teoretiska gäller förhållanden som vi inte kan påverka, bara iakttä (det grekiska verbet som ligger till grund för beteckningen betyder att noga iakttä, att skärskåda). Den praktiska motsvarar inte vad vi idag främst lägger in i ordet ”praktisk” utan har med människors handlande mot andra människor att göra, alltså helt enkelt den mellanmänskliga sfären. Teknisk kunskap är den kunskap med vars hjälp vi förmår förändra den icke-mänskliga verkligheten till exempel i jordbruk, bergsbruk och över huvud all verksamhet där vi använder redskap.

Aristoteles syn på kunskapen fick i synnerhet under hög- och senmedeltiden ett enormt inflytande i både islam och kristenheten. En distinkt världsbild tog form. I den kombinerades Aristoteles med element från astronomen och geografen Klaudios Ptolemaios (cirka 90 – cirka 170) lära om jordens, planeternas, solens och fixstjärnornas inbördes ordning.

Enligt denna världsbild, som fick sitt mest åskådliga och lysande uttryck i *Divina commedia*, Den gudomliga komedin, Dante Alighieris (1265-1321) stora diktverk, befinner sig jordklotet i vila i universums mitt. Där råder ständig förändring, de fyra elementen jord, vatten, luft och eld blandas oavslåttligt med varandra, och i jordens mitt finns helvetet. Vi brukar säga att världsbilden är geocentrisk, det vill säga med jorden i centrum, men mer exakt är det helvetet som intar den allra mest centrala positionen.

Runt jorden cirklar himlakropparna, från månen och uppåt, med solen som den mellersta och mest betydande planeten. Alla är fästade vid kristalliska sfärer. Längst bort befinner sig fixstjärnorna som egentligen är hål i den yttersta sfären, hål som ger oss en aning om härligheten i empyrén, de saligas boningar och därmed också Guds rike.

Alltifrån månens sfär är allting oföränderligt, formerna perfekta och himlasfärerna därmed helt cirkulära. Jordens fyra ständigt föränderliga element motsvaras här av ett femte, *quinta essentia* eller på ren svenska kvintessen.

Till denna världsbild hörde en mycket bestämd föreställning om vad som kunde upprepas.

Först och främst gällde det iakttagelser av himlakropparna och deras rörelser. Alltsedan de gamla babylonierna hade det utvecklats en avancerad himmelsk geometri som ytterligare förfinats av greker och inom den islamiska kulturkretsen.

Vidare gällde det iakttagelser och berättelser om olika typer av mellanmänskliga beteenden som utgjorde den empiriska basen för moralfilosofin.

Slutligen handlade det om det som utgör grunden för hantverk och jordbruk: en ständigt upprepad uppsättning verksamheter utförda med en lång rad olika verktyg. I antiken och långt fram i tiden ansågs denna hantering som en lägre kunskapsart. Förändringen kom först med renässansen och på sina håll än senare.

Avgörande för den världsbild som mer eller mindre direkt inspirerades av Aristoteles och som präglade det allra mesta av både kristen och muslimsk världsbild under århundraden var att kunskaper som gällde de celesta förhållandena, det vill säga allt från månens sfär och uppåt, inte kunde överföras på jordiska förlopp och vice versa. Det utfördes många sofistikerade experiment och många exakta iakttagelser om jordiska förhållanden, men de kunde enligt den auktoritativa uppfattningen inte säga någonting om de himmelska sfärerna där allt var geometriskt fulländat med perfekta cirklar och klot. Jor-



den med sina berg och dalar och andra oregelbundenheter hade ingen likhet med dessa upphöjda ting.

Mot denna världsbild fanns det hela tiden en mer eller mindre medveten opposition. Den blev påtaglig redan under 1300- och 1400-talen och nådde full kraft under 1500- och 1600-talen. Under det som ofta kallas den naturvetenskapliga revolutionen skedde en rad genombrott som kullkastade den härskande uppfattningen. Nicolaus Copernicus (1473-1543) placerade i sitt system solen i universums mitt och förklarade att jorden var en planet bland andra. Den stora förskräckelse som denna världsbild väckte, gällde inte som eftervärlden ofta trott att människan inte länge befann sig i mitten av allt. Detta centrum var som vi sett ofullständigt, ja syndigt, och i det absoluta centrum befann sig som vi sett helvetet – inte en position att eftersträva. Nej, den främsta skandalen bestod i att människan, denna syndfulla varelse, flyttade upp i himlarna.

Nästa grundskott mot den härskande ordningen var Johannes Keplers (1571-1630) teori som bland annat visade att planetbanorna inte var cirklar utan ellipser; även jorden rörde sig så. Men det definitiva genombrottet för en ny världsbild kom genom Galileo Galilei (1564-1642). Men hjälp av ett tämligen nytt instrument, kikaren, kunde Galileo visa att månen liksom den ofullständiga jorden hade berg, att planeten Venus liksom jorden genomgick faser och att planeten Jupiter liksom jorden hade måne, ja flera stycken.

Galileo spelade också stor roll genom sina experiment som numera upprepas i den elementäraste fysikundervisning, till exempel det sluttande planet.

Den ärevördiga föreställningen att jordiska experiment aldrig skulle kunna säga något om de celesta förhållandena fick sin definitiva vederläggning genom Isaac Newtons (1642-1727) gravitationsteori som visar att himlakropparnas rörelser kan beräknas på exakt samma sätt som projektilbanor just över jordens yta.

Experiment blev desto mer centrala för kunskapsutvecklingen som det skedde en snabb och genomgripande utveckling av olika redskap som mikroskop, exaktare klockor, kemiska instrument och så vidare. Teknik och forskning överlappade varandra alltmer. Experimentet kännetecknas av att det kan upprepas. Vad den ene kommit fram till kan bekräftas eller modifieras av den andre. I den vetenskapliga korrespondensen, som efter hand utvecklas till vad som är de vetenskapliga tidskrifterna, kunde forskare redogöra för sina resultat och väcka reaktioner av olika slag bland likasinnade i andra länder.

Här spelade naturligtvis tryckkonsten en avgörande roll för spridningen

av resultaten. Den mödosamma kopieringen av manuskript avlöstes av det mycket effektivare och säkrare mångfaldigande som trycket ger möjlighet till.

Denna utveckling var lika viktig för humanistisk forskning som för naturvetenskaplig. Noggranna editioner av klassiska verk, profana eller religiösa och försedda med kritiska kommentarer, ser dagen. Böcker och artiklar med fotnoter blir allt vanligare. På samma sätt som experimentet ger kolleger på andra orter möjlighet att med egna ögon kontrollera resultatet, på samma sätt kan fotnoterna med hänvisningar till den ena eller andra texten inbjuda andra att konstatera: ja, så står det verkligen. Grunden för den moderna källkritiken läggs.

Men humanistiska discipliner och naturvetenskapliga utvecklades inte helt och hållet i god harmoni. Ett av tvisteämnena blev just upprepbarheten och i förlängningen generaliseringen. Experiment kan upprepas och ligga till grund för generaliseringar och i förlängningen lagpåståenden och kanske hela teorier. Det finns förvisso också generaliseringar som kan göras om texter, om författare och konstnärer, om historiska och samhälleliga förhållanden. Men är de lika centrala där som när det gäller naturen?

Med den traditionella, mer eller mindre aristoteliska vetenskapsindelningen tycktes harmonin lätt att bevara. Å ena sidan hade man naturfilosofin där generella utsagor om naturen fogades samman till en fast byggnad. Newton såg sig som naturfilosof. Han presenterade sin gravitationsteori i *Philosophiae Naturalis Principia Mathematica* (Naturfilosofins matematiska principer, 1687). Ännu in på 1800-talet kallade sig naturvetare naturfilosofer; den store fysikern Michael Faraday (1791-1867) var en av de sista som ståndaktigt höll fast vid beteckningen.

Moralfilosofin utgjorde den självklara pendangen till naturfilosofin. Även den gjorde anspråk på att utlägga allmängiltiga förhållanden när det gällde mänskligt handlande. Det var visserligen ingen matematik som byggde upp slutsatserna. Men moralfilosoferna hävdade att de kunde fånga den mänskliga naturen.

Den kunskap som inte – eller ännu inte – kunde generaliseras fördes å ena sidan till naturalhistorien, å andra till moralhistorien. Carl von Linné (1707-78) var självklart naturalhistoriker. Det var naturens mångfald han kartlade, och de generaliseringar som han ibland dristade sig till hade vad han själv än menade en underordnad betydelse.

Hans motsvarighet på den humanistiska sidan var en moralhistoriker som

Sven Lagerbring (1707-87) som även han kunde pröva sina vingar med svepande sammanfattningar av det sant mänskliga men utan att göra anspråk på att vara moralfilosof.

Hela denna till synes så harmoniska uppdelning bryts sönder redan i början av 1800-talet. Naturalhistorien antikveras; naturforskningen splittras upp i allt fler specialiteter. Moralthistorien blir historia rätt och slätt eller finfördelas i politisk historia, konsthistoria, idéhistoria och så vidare.

Moralfilosofin lever förvisso vidare men uppfattar sig inte längre som en motsvarighet till naturfilosofin.

Vad gäller naturfilosofin stöter den i Storbritannien på ett häftigt motstånd. Ledare för denna attack heter William Whewell (1794-1866), en brittisk filosof och naturforskare. Han och hans åsiktsfränder, som skapade The British Association for the Advancement of Science, propagerade för att det gamla ordet "science" (latinets *scientia*) nu skulle förbehållas naturvetare som arbetade i Newtons tradition och därmed eftersträvade matematiskt exakta forskningsresultat. Whewell uppfann till och med ordet "scientist" för den som utövade denna typ av verksamhet.

Whewell var framgångsrik. I den anglosaxiska världen trängdes beteckningen "natural philosopher" snart ut eller blev beteckningen på en skum verksamhet präglad av naturspekulation.

Utvecklingen i det tyska språkområdet liksom i Norden gick en annan väg. Där blev i stället ordet "Wissenschaft" ("vetenskap", "videnskab" etc.) den favoriserade beteckningen, mycket tack vare universitetens allt högre status i samband med den reformverksamhet för vilken Wilhelm von Humboldt (1767-1835) blivit symbolen. En vetenskap blev helt enkelt det som hade en position i form av en lärostol vid ett universitet. Det gällde alltså allt från teologi till fysik och från juridik till fysiologi.

Denna brokiga skara av discipliner präglade inte av någon enhetlighet. Frågan blev hur man skulle karakterisera dem som forskningsdiscipliner. Vissa menade att alla borde sträva i riktning av den exakta naturvetenskapen. Andra hävdade att det fanns distinkt olika sorters vetenskaper med olika metoder och olika mål. En som delvis satte agendan för den senare gruppen var filosofen och idéhistorikern Wilhelm Dilthey (1833-1911). Han påpekade att gapet mellan naturvetenskaper och humaniora inte var så stort som det ibland gjordes gällande. Det fanns naturvetenskaper som syftade till att framställa (också) det unika, till exempel paleontologin. Även himlakropparna, däribland jorden hade en historia bortom människan.

Humaniora – till vilka samhällsvetenskaperna ännu räknades – syftade på motsvarande sätt stundtals till att nå fram till det generella. Dilthey nämner den tidiga motsvarigheten till dagens bibliometri, där målet är att komma fram till generella utsagor, kanske till och med någon sorts lagbundenheter när det gäller utgivningen av böcker och artiklar.

Men otvivelaktigt finns det enligt Dilthey en skillnad mellan de mest extrema naturvetenskaperna, som fysik, och de mest extrema humanvetenskaperna, som studiet av den enskilda men avgörande historiska händelsen, det enskilda storartade konstverket eller den intressanta unika personligheten. Den typiska naturvetenskapen syftade till lagpåstående infogade i väl uppbyggda teorier, den typiska humanistiska vetenskapen till det särpräglade som aldrig kunde upprepas.

Det fanns andra filosofer efter Dilthey som snarare sökte skillnaden mellan naturvetenskaper och humaniora i deras metoder, syftande till att fånga det som låter sig generaliseras respektive det som inte går att upprepa.

Det senare 1800-talets stora debatt om skillnader mellan olika vetenskapstyper är inte lika livaktig som den var då. Whewells ”science” har fått en vidare innebörd även om den exakta vetenskapen à la Newton (eller snarare Einstein och Planck) fortfarande framstår som mönstret. Numera använder man beteckningar som ”social sciences” och till och med ”human sciences”. Universitetet är inte längre den självklara hemorten för alla ”Wissenschaften” eller vetenskaper som de var i länder som Tyskland och Sverige ännu långt in i förra århundradet.

Men det finns stora vetenskapsområden där det fortfarande är kontroversiellt hur långt man kommer med generaliseringar. Hur är det med historieforskningen? Är inte intresset där självklart inriktat på det speciella förloppet, den unika händelsen, den fascinerande personligheten? Vägen för att komma dit är visserligen kantad av generaliseringar av vilka de flesta är på tämligen låg nivå (när det inte rör sig om rena trivialiteter). Men målet är att kunna förstå det särskilda.

Detta gäller även fenomen som är vitt spridda. Låt oss ta ett enda exempel: nationalismen. I sin klassiska form spelade den en avgörande roll från slutet av 1800-talet fram till och med andra världskriget. Den hade vissa mycket typiska särdrag som dyrkan av det egna landet och dess folk, språk, historia och så vidare. Samtidigt riktade den sitt ursinne mot vissa andra, likaledes nationalistiskt sinnade länder.

I vår samtid har en ny typ av nationalism vunnit allt större inflytande. Den är lika upptagen av det egna folket, språket och kanske även historien som den gamla. Men den formar nu ett slags mer eller mindre informell internationell där olika länders nationalister stöder och uppmuntrar varandra. De inre fienderna som bekämpas är delvis desamma för förra gången, nu preciserade som feminister, globalister eller kulturmarxister. Men framför allt är man upptagen av en motståndare som påstås på väg att ta över det egna landet och den egna kulturen: islam.

Detta är ett enkelt exempel på en jämförelse där likheter och skillnader mellan två komplexa företeelser, i detta fall nationalistiska rörelser, kastar ljus över båda. Bara genom jämförelserna får vi fram det särpräglade.

Vi skulle kunna fortsätta med andra mänskliga områden, till exempel konsten, musiken, litteraturen och filmen. Överallt möter vi samma mönster.

I samhällsvetenskaperna spelar normalt sätt generaliseringarna en större roll. Sociologerna syftar vanligen till att säga något allmängiltigt om någon aspekt av samhället. Men även där är en motspänstig empiri viktig, en empiri som naggar generaliseringarna i kanten. Människans värld är mångfacetterad. Noga sett är det oss själva som vi försöker lägga under luppen.

Nationalekonomin har anammat det matematiska modellbyggande som visat sig så framgångsrikt inom naturvetenskapen. Men man har inte lyckats nå fram till de exakta förutsägelser som karakteriserar de förebildliga vetenskaperna. Vid en rymdfärd beräknas rymdskeppens banor med häpnadsväckande exakthet. Den minsta meteorit har man lyckats styra fram till på det sättet.

Men aldrig så mycket matematik hjälper inte ekonomerna att förutsäga nästa stora ekonomiska kris. Det är en mängd mänskliga beslut som utlöser krisen, och dessa beslut kan inte fångas i någon kalkyl. En dos psykologi hjälper inte heller för att råda bot på problemet. Människan är helt enkelt en alltför oförutsägbar varelse.

Ämnet psykologi har till uppgift att fånga detta nyckfulla mänskliga psyke med vetenskaplig strängighet. Det har utvecklats en väldig mångfald av subdiscipliner där olika aspekter utvecklas med ackuratess. I långa stycken lyckas man genom experiment uppnå både exakthet och förutsägbarhet. Men så snart det är det enskilda psyket, eller snarare den enskilda människan med sin unika livshistoria och sina speciella problem som är i fokus ställs man inför nya problem där den medmänskliga förståelsen och erfarenheten är de bästa hjälpmedlen.

Kunskapens och därmed vetenskapens område är outtömligt. Det är inte beklagligt: i annat fall skulle vi eller våra efterkommande en dag stå arbetslösa.

I vårt kunskapssökande finns det ett mycket klokt råd att följa. Det gavs av den legendariske vetenskapsteoretikern Håkan Törnebohm, en gång medlem av detta Samhälle. Han uppmanade oss att skilja mellan vad som är vetbart och vad som är värt att veta. Det finns mycket som är vetbart men knappast värt att veta. Men det som är värt att veta men ännu inte vetbart bör vara det övergripande målet för vårt arbete, i dag och i morgon.

### Referenser

- Bolotin, David: *An Approach to Aristotle's Physics: With Particular Attention to the Role of His Manner of Writing* (Albany: SUNY Press, 2004).
- Broberg, Gunnar: *Carl Linnaeus* (Stockholm: Svenska institutet, 2006).
- Drake, Stillman: *Galileo at Work* (Chicago: University of Chicago Press, 1978).
- Eriksson, Nils: *Dalin – Botin – Lagerbring: historieforskning och historieskrivning i Sverige 1747–1787* (Göteborg: Institutionen för idé- och lärdomshistoria, 1976).
- Lagercrantz, Olof: *Från helvetet till paradiset: En bok om Dante och hans komedi* (Stockholm: Wahlström & Widstrand, 1964).
- Liedman, Sven-Eric, *Stenarna i själen: Form, materia och innehåll från antiken till våra dagar* (Stockholm: Bonniers, 2006).
- Morrell, Jack och Thackray, Arnold: *Gentlemen of Science: The Early Years of the Association of the Advancement of Science* (Oxford: Clarendon Press, 1981).
- Pasnau, Robert och Van Dyke, Christina (utg.): *The Cambridge History of Medieval Philosophy* (Cambridge, UK: Cambridge University Press, 2010).
- Törnebohm, Håkan: *Världar, vetande och värden* (Göteborg: Institutionen för vetenskapsteori, 1989).
- Westfall, Richard S., *Never at Rest: A Biography of Isaac Newton* (Cambridge UK: Cambridge University Press, 1983).

*Peter Jagers*

## Slumpvariationer i det annars upprepbara

Slumpmässiga fenomen är definitionsmässigt oförutsägbara, i den grundläggande meningen att utfallen inte kan upprepas på beställning. Omvänt kallar vi just de fenomen, vars orsakskedjor vi inte bemästrar, och som vi därför inte kan upprepa efter behag, för just slumpmässiga eller stokastiska. Samtidigt utgör resultatvariation en oundviklig ingrediens i noggrann och mer avancerad vetenskap och uppprepbarhet ett naturligt krav på vetenskapliga försök. Det är i denna konfliktpunkt sannolikhetssteori och statistisk vetenskap borde ha något att säga om ämnet för denna konferens om uppprepbarhet och generaliserbarhet i vetenskapen.

I själva verket är det just det slumpmässiga inslaget i försök eller observationer som konstituerar det problematiska i upprepbarheten, i vart fall om vi talar om hederliga försök, och om styrda sådana, snarare än observationer av komplicerade skeenden, som samhällseliga eller historiska, där vi inte kan arrangera förutsättningarna eller, förstås, om simpelt fusk eller omedvetna klavertramp.

Vid Kungl. Vetenskaps- och Vitterhets-Samhällets i Göteborg septembersammankomst 2018 erinrade Bengt Alexanderson, i ett annat sammanhang, om Sven Delblancs bok "Trampa vatten" (1972). I denna arbetar bokens jag på en vetenskaplig institution, men störs var tredje minut av ett vrål från vinden där ett ständigt uppreat försök pågår:

*"Nakenstedt är docent i teoretisk filosofi och genomför sedan många år tillbaka en experimentell undersökning. Apparaturen består av ett strykjärn, kasserat av fru Nakenstedt, samt ett protokoll uppdelat på kolumnerna A och B. Var tredje minut släpper Nakenstedt strykjärnet på sin stortå. Om, men bara om, strykjärnet därvid följer naturlagarna och faller genom luften med av mekaniken föreskriven hastighet för att drabba Nakenstedts stortå med smärtsamt eftertryck, då, men bara då, ritar Nakenstedt en kråka i kolumn A. Om däremot strykjärnet trotsar naturlagarna och förblir*

*svävande i luften som i en rymdkapsel, då, men bara då, ritar Nakenstedt en kråka i kolumn B... Det senare alternativet har emellertid aldrig inträffat, inte under de år jag varit här. .... Vi kan fastställa, att var gång vi släppt strykjärnet (a) har detta följts av (b) svullenhet, sveda, värk och avgrundsvrål, ...”*

I Delblancs berättelse nämndes att strykjärnet faller med ”av mekaniken förskrivna hastighet”. Det sägs dock inget om någon kontroll av den tid fallet tar. Hade Nakenstedt bekymrat sig om denna, hade han haft en mer avancerad fråga framför sig. Här hade startpunktens exakta position kommit in, liksom hans kroppshållning. Än svårare och mer varierad hade tidsmätningen blivit om det inte rört sig om strykjärnet utan om en mer tåvändig bomullstuss, som hade låtit sig påverkas av drag och andra banaliteter.

Naturlagens, det vill säga den mekaniska teorins, exakta tid, hade störts av en slumpmässig avvikelse. Fortfarande är strukturen dock tämligen enkel: Forskaren observerar tiden ”i sig” med denna avvikelse. Just det resultatet kan vi svårligen reproducera, men försöket, att släppa strykjärnet eller bomullstussen, kan Nakenstedt eller vi upprepa.

Men nu uppkommer frågan om Nakenstedts så utvidgade försök med en slutsats också om falltiden kan bekräftas eller motsägas vid en upprepning och vad vi i så fall ska mena med detta. Det är just här den statistiska slutledningen kommer in.

Och denna enkla struktur återkommer i en generaliserad form i en rad olika situationer, även sådana som inte självklart har formen att det som observeras är ett teoretiskt värde med en ”slumpavvikelse”, så som ofta är fallet i klassisk experimentell naturvetenskap och så som många samhälls- och beteendevetare vill föreställa sig sina vetenskaper. Vad vi observerar är utfall av stokastiska, alltså slumpmässiga, variabler. Vad vi kan hoppas på vid upprepning är inte likhet mellan utfallen utan mellan dessas fördelningar.

Det rör sig här om en klassisk, arketyrisk form som möter i snart sagt all vetenskap, att undersöka huruvida två situationer, grupper eller experimentbetingelser ger olika resultat. Detta är inte en fråga om huruvida utfallen skiljer sig åt, det gör de som sagt alltid, utan en fråga om huruvida fördelningarna för utfallen kan anses vara olika i de två situationerna. Och reproducerbarhet eller upprepbarhet av hela försöket blir en fråga om man får samma svar på denna fråga om fördelningarna, när hela proceduren görs om.

Inom natur- och samhällsvetenskap är detta helt centralt för kunskapsbildningen. Egentligen är det bara inom nationalekonomins huvudfåra som ett



annat förhållningssätt dominerar, även om mer beteendevetenskapliga ansatser ju också förekommer – såsom framgår av Anna Drebers och Magnus Johannessons bidrag i denna volym. Men i den dominerande skolan – den som brukar belönas med ännu ett Riksbankspris till Alfred Nobels minne – där gör man inga experiment eller planerade upprepbara observationer av välpreciserade skeenden. Snarare eftersträvar man ett mer matematiskt arbetssätt, med den teoretiska fysiken som förebild. Ur empiriska observationer extraherar man en förenklad modell, ur vilken man matematiskt-logiskt deducerar påståenden, som egentligen gäller modellen, men som man frejdigt applicerar på verkligheten, och så ger man råd om vilka åtgärder som bör vidtagas. Genom sin abstraktion förutsätts modellen ha nått en generell relevans, som gör att den inte behöver ifrågasättas eller prövas experimentellt. Därmed har jag också sagt att de matematiska vetenskapernas kunskapsbildning följer ett helt annat mönster än de upprepbara observationernas – liksom, på sin kant, många humanistiska ämnen.

Men detta var en randanmärkning. Åter till huvudproblematiken. Extra tydligt möter denna i läkemedelsprovningens hårt reglerade metodik. Här rör det sig i princip om ett nytt preparat eller en behandling, som skall jämföras med placebo, eller en traditionell behandling. Låt oss nu anta att man i ett försök har kunnat påvisa en skillnad. Vad betyder det då att detta resultat är upprepbart?

Samma mönster är mer än välkänt i snart sagt all vetenskap från klassisk experimentell naturvetenskap till socialmedicin eller psykologi, med den skillnaden att det här ofta inte är självklart vilka de två grupperna, X och Y, är som i läkemedelsprovningen. I allmänhet är det forskaren eller uppdragsgivaren som definierar dem, vilket innebär en frihet men också ibland en fallgrop, nämligen om undersökningens data tillåts påverka gruppindelningen – jag återkommer till detta klassiska missgrepp.

Den statistiska vetenskapens genombrott i början av förra seklet innebar att en rad standardprocedurer för detta och liknande problem etablerades. De byggde alla på samma grundläggande logik. Man anade eller hoppades, att en skillnad eller behandlingseffekt förelåg och testade sedan mothypotesen, som man kallade nollhypotesen, genom att undersöka till vilken grad det erhållna resultatet gick att förena med föreställningen att ingen skillnad, noll effekt, var för handen. Var detta svårt att tänka sig, förkastades nollhypotesen och man ansåg sig ha goda skäl att tro på en reell effekt.

Ett reproducerbart vetenskapligt försök av denna form innehåller då en rad komponenter.

1. Två grupper som skall jämföras. Dessa skall vara beskrivna i förväg, annars kan man inte bestämma sannolikheten för det erhållna och mer extrema utfall.
2. En försöksbeskrivning, inklusive mätmetoder, som skall vara noggrant angivna.
3. Det som idag kallas koder, program etc., d.v.s. de transformationer och uträkningar som görs i databehandlingen. De skall preciseras.
4. De statistiska fördelningar för observationsvärden, som detta resulterar i, under olika förutsättningar.
5. Vilket så kallat p-värde observationen ger, d.v.s. hur stor sannolikheten är att under nollhypotesen få ett resultat som minst lika mycket som det erhållna motsäger nollhypotesen.

Om allt detta är för handen kan man säga att försöket är reproducerbart eller upprepbart. Man kan göra om försöket och se huruvida p-värdet hamnar i samma härad som förra gången. Det kan man kalla att man har bekräftat försöket.

Litteraturen i dessa frågor tyngs av en i min mening ganska beskäftig diskussion om distinktioner mellan upprepbarhet, repeterbarhet, replikerbarhet och reproducerbarhet. Jag går inte in på denna.

Dessa idéer för signifikansprövning går tillbaka till Ronald Fisher vid förra seklets början. Enligt min uppfattning hör de till 1900-talets grundläggande vetenskapliga vinningar. De fick också snabbt ett starkt genomslag i olika vetenskaper, som dock tyvärr inte alltid förenades med grundlig förståelse. När min generation började studera, kunde den självsäkra användningen av schablonmässiga statistiska fördelningar stå en upp i halsen. Jag minns min ungdoms föreläsningar i ekonomi och sociologi. Allt var, som en av dåtidens göteborgska professorer stolt kunde utropa, ”t-testat och klart”.

Modellen var självklar, lite linjära samband och så normalfördelade fel. Kvalitativa data kvantifierades frejdigt i Likertskalor, särskilt inom sociologiämnet, som om avståndet mellan kärlek och likgiltighet var detsamma som mellan likgiltighet och hat.

Till detta kom mer sofistikerade vanföreställningar och missgrepp.

Logiken bakom signifikanstestet, och därmed användningen av p-värdet,

kan ju sammanfattas som ”Om ingen skillnad mellan grupperna föreligger, har något så konstigt inträffat, att jag vägrar att tro på att där inte finns någon skillnad.” I sannolikhetsterminologi kan man säga att sannolikheten för data som så starkt motsäger nollhypotesen, är mycket liten under densamma. Men nu gäller det att hålla tungan rätt i mun: att sannolikheten för A givet B är liten betyder inte att sannolikheten för B givet A är liten. Tvärtom är detta ett felslut som är så etablerat att det har hedrats med ett eget namn, ”transponerad betingning”. Men det är ett frestande misstag, som säkerligen föresvävat många aningslösa användare av p-värden. Själva namnet ”p-värde”, ”p” givetvis för probabilitet = sannolikhet är ju här förföriskt. Och inte blev situationen bättre av att många statistiker försökte använda ordet ”trolighet” – eller än värre ”likelihood” – i en sorts teknisk betydelse, skild från ”sannolikhet”, trots att orden vardagspråkligt uppenbart avser samma sak. Det var alltså lätt för den empiriske forskaren med ytliga statistikkunskaper att gå i den fälla som gillrades av den transponerade betingningen. Och det blev ännu lättare, när så kallade Bayesianska metoder blev moderna mot förra seklets slut. Dessa innebar ju en sorts legitimitet åt att tillskriva alla möjliga utsagor sannolikhetsvärden.

Men ytterligare en lömsk fara lurade på den eleganta Fisherska idén om hypotesprövning och p-värden. Genom IT-revolutionen blev allt större datamängder tillgängliga, allt fler gruppindelningar kunde göras. De lätt amoralska bland forskare kunde förledas att glutta bak förlåten och kanske ana att det kanske förelåg något samband någonstans. Och sedan glömma det hela och blåögt testa nollhypotesen om att just detta – redan observerade – samband inte förelåg, för att sedan triumferande förkasta den. Detta så kallade flersignifikansproblem har säkert lett till ett antal forskningsrapporter med falska signifikanser. Saken kompliceras ytterligare därutav att i ett stort, brett tillgängligt material olika forskare oberoende kan välja olika hypoteser att pröva. Och, som jag brukar säga, om tillräckligt många forskare prövar tillräckligt många nollhypoteser, så kommer alltid någon att hitta något som kan förkastas med besked. Sådana resultat går förstås inte att upprepa i annat än den triviala meningen att om exakt samma hypotes testas på samma material – och inget annat fel gjorts – så får man samma resultat.

En sorts oavsiktlig variant av denna flersignifikansproblematik utgörs av det mångomtalade problemet med publiceringsskevheter. Det är ju helt enkelt så att det är lättare att publicera belagda resultat än uppsatser som mynnar ut

i att inget resultat har kunnat fastläggas. Och helt utan ond vilja kan ju en forskare drabbas av ett slumpgenererat utfall, som är mycket osannolikt under nollhypotesen. 5% av alla resultat, där ingen skillnad mellan alternativen föreligger, kommer ju att uppvisa  $p < 0,05$ . Dessa 5% kommer sedan att publiceras i långt högre grad än de 95% av försöken som inte uppvisar något extremt resultat.

Summan av kardemumman – och naturligtvis av en rad andra omständigheter av utomstatistisk art, som regelrätt forskningsfusk, publikationspress, koncentration på citeringar av snarare än innehåll i artiklar – har lett till den replikeringskris som ligger bakom dagens debatt. Denna är förstås i hög grad motiverad. Observationers upprepbarhet utgör en kärnpunkt i vetenskapen. Inte desto mindre finns det anledning att varna för en del av de motmedel som föreskrivits. Ont ska inte med ont utan med gott fördrivas.

Ett mycket märkligt motmedel är det som har varit populärt i vissa psykologkretsar, framför allt i och kring en tidskrift som heter Basic and Applied Social Psychology. Denna gick 2015 så lång som att bannlysa all användning av p-värden och statistiska hypotestest från tidskriften. Grunden förefaller helt enkelt ha varit att redaktörerna aldrig hade förstått den grundläggande metodiska idén, utan krävde att en godtagbar metodik skulle explicit ange en sannolikhet för nollhypotesen, vilket helt enkelt är omöjligt ur observationsdata, om man nu inte ska begå det fundamentala felet med transponerad betingning, eller förutsätta att sannolikheter för olika hypoteser på något magiskt sätt har blivit oss från början givna.

Annan kritik, som har formulerats av bl. a. två ekonomer, Ziliak och McCloskey (2008), har pekat på att statistisk signifikans inte är samma sak som relevans. Detta är förvisso sant, och viktigt att betänka, men något vid sidan om diskussionen.

För min del tycker jag att denna typ av kritik förvisso kan motiveras, men framför allt av det schablonmässiga användandet inom kritikernas egna discipliner, mindre av begreppens egna brister. Återigen har vi fått exempel på hur forskare förleds att dra felaktiga slutsatser av vad som råkar vara dåliga vanor inom det egna facket. Lite bättre kunskap om diskussionen inom andra vetenskapsfält hade hjälpt även debatten om upprepbarhet i vissa discipliner.

Låt mig avsluta med två tack till ledamöter av KVVVS. Jag har redan nämnt Bengt Alexanderson. Därutöver rekommenderar jag uppsatser i ämnet av Olle Häggström, som är tillgängliga på hans hemsida.

## ***Referenser***

Cox, D.R & Hinkley, D.V., *Theoretical Statistics* (London:Chapman and Hall 1974).

Delblanc, Sven, *Trampa vatten* (Stockholm: Författarförlaget 1972).

Hägström, Olle, The need for nuance in the null hypothesis testing debate. *Educational and Psychological Measurement* 77 (2017)

Challenges in irreproducible research. Temanummer, *Nature* 515 (2014).

Statistical Inference in the 21st Century: A World Beyond  $p < 0.05$ . Temanummer, *The American Statistician* 73 (2019).



*Anna Dreber Almenberg och Magnus Johannesson*

## Tillförlitlighet av forskningsresultat – vilka resultat går att upprepa?

### 1. Introduktion

Vilka publicerade forskningsresultat inom kategorin kvantitativ samhällsvetenskap kan vi lita på? Ett kort svar är *få*. Varför få? Det är just det vi ska diskutera i det här kapitlet. Men få kan bli fler – det finns många sätt att öka tillförlitligheten av forskningsresultat. Vi har tillsammans med medförfattare inom olika samhällsvetenskapliga fält forskat om just vilka resultat som går att upprepa eller finna stöd för med nya data, vilka resultat som går att förutse och vilka lösningar som potentiellt kan förbättra tillförlitligheten av forskningsresultat. Det är främst denna forskning vi i det här kapitlet fokuserar på.

Ett något längre och mer informativt svar på den inledande frågan är att det inte är helt lätt att definiera vad 'lita på' betyder. Definitioner av upprepbarhet och generaliserbarhet spelar centrala roller och kan ta olika former.

Guldstandarden för att förstå kausala effekter är att göra dubbel-blinda randomiserade experiment där deltagare randomiseras till att ingå i en kontrollgrupp eller en eller flera behandlingsgrupper eller treatmentgrupper. Vår fokus på upprepbarhet i denna översikt är just på experimentella studier. Mycket ekonometri och statistik inom de kvantitativa samhällsvetenskaperna handlar om att få icke-experimentella data att likna experimentella data för att vi ska kunna dra just kausala slutsatser, men i det här kapitlet befinner vi oss alltså främst i den bästa av kausala världar.

Den första nivån av upprepbarhet är att se om vi med hjälp av samma data och analys som i ursprungsstudien kan upprepa resultaten. Denna upprepbarhet tar vi ofta som en självklarhet, men de systematiska försök inom nationalekonomi som har gjorts för att testa detta tyder på problem redan här (Dewald, Thursby, och Anderson 1986; McCullough och Vinod 2003; McCullough, McGeary, och Harrison 2006; Chang och Li 2015). Problemen orsakas av otillgängliga data och kod, men även när dessa är tillgängliga går

alla resultat inte att återskapa p.g.a. t.ex. fel i koden. Detta slags upprepning kallas ibland för ”ren replikation” (Hamermesh 2007). Ett färskt exempel kommer från Gertler, Galiani och Romero (2018) som gick igenom en månads publicerade artiklar i nio ledande tidskrifter inom nationalekonomi. För maj månad 2016 fann Gertler et al. 203 empiriska artiklar utan legala restriktioner på datatillgänglighet. Trots denna datatillgänglighet kunde Gertler et al. endast replikera resultaten i en minoritet av artiklarna. Blott 16% av artiklarna hade både rådata och användbar kod som fungerade. Den vanligaste filen som originalförfattarna hade laddat upp på tidskriftens sida var filen med estimeringskoden; med hjälp av denna kod kunde Gertler et al. producera originalartikelns slutgiltiga tabeller och figurer i endast 37% av fallen. För blott 14% av de 203 artiklarna kunde forskarna med hjälp av rådata nå fram till samma resultat som i originalartiklarna. Det finns således mycket utrymme för förbättring även på denna nivå där resultat borde vara uppenbart upprepbara.

Nästa nivå av upprepbarhet kan vara att se om resultaten håller vid insamling av nya data, där materialet i studien helst hålls konstant och där urvalet är nytt. Det är vad vi hädanefter kallar direkt replikation. I en direkt replikation ställer vi oss frågan om vi med samma material som i originalstudien kan upprepa resultaten i ett nytt urval där deltagarna randomiseras till kontroll eller ”behandling” (treatment). Flera olika kriterier har föreslagits för att avgöra om ett resultat replikerar eller inte. Det vanligaste sättet, och det vi kommer att fokusera på, är att i den direkta replikationen använda samma statistiska test som i ursprungsartikeln och testa om replikationsresultatet är i samma riktning som i ursprungsstudien med ett statistiskt signifikant p-värde ( $p < 0,05$  i ett två-sidigt test).

Nästa steg av upprepbarhet kan vara att göra en konceptuell replikation. Samma konceptuella hypotes testas, men med annan metod eller under något andra (kontrollerade) omständigheter. Konceptuella replikationer tar oss in på ursprungsresultatens generaliserbarhet. I vilken utsträckning finner vi stöd för hypotesen när vi varierar hur hypotesen testas, t.ex. när vi varierar designen, måtten och urvalet?

Det är inte alltid lätt att dra gränsen mellan direkta och konceptuella replikationer. Tillsammans med medförfattare genomförde vi en replikation av fenomenet ”power posing”. Ursprungsartikeln (Carney, Cuddy, och Yap 2010) presenterade resultat från en studie där 42 deltagare rapporterade ha randomiserats till att inta en expansiv maktfull kroppsposition eller en återhållen



mindre maktfull kroppsposition i någon minut. Högmaktspositionen hade dramatiska effekter – deltagarna i denna grupp fick i genomsnitt statistiskt signifikant högre testosteronnivåer, lägre kortisolnivåer, blev mer ekonomiskt risktagande och upplevde högre känsla av makt jämfört med deltagarna som randomiserats till lågmaktspositionen. Med sådana fantastiska effekter med små medel är det inte konstigt att artikeln ledde till ett TED-talk med mer än 52 miljoner visningar i april 2019. *What's not to like?*

Ett av problemen med ”power posing” uppstod när vi 2015 publicerade vår misslyckade replikation, där vi varken lyckades replikera resultaten för testosteron, kortisol eller risktagande (Ranehill et al. 2015). Däremot replikerade vi resultatet för känsla av makt men med mycket mindre effektstorlek – men det är ett resultat som senare studier i stort misslyckats med att replikera. En potentiell invändning mot vår replikation är att replikationspopulationen – i det här fallet schweiziska studenter – inte var identisk med originalpopulationen – i det här fallet amerikanska studenter på ett specifikt universitet.

När vi studerar behandlingseffekter finns det dock anledning att tro att studiepopulationen är mindre viktig än om vi studerar nivåeffekter. Om treatmenteffekten varierar mellan populationer leder det till ökad variation i resultat mellan originalstudier och replikationsstudier men det leder inte till någon systematisk bias i replikationernas genomsnittligt uppskattade effekttorlek. Några av de replikationsprojekt vi diskuterar nedan (de så kallade *Many Labs* projekten) finner inte heller stöd för att det skulle vara mycket systematisk variation i genomsnittliga behandlingseffekter från psykologistudier över olika urval (Klein et al. 2014; Ebersole et al. 2016).

## 2. Förklaringar till varför resultat inte går att upprepa

Det finns många olika anledningar till att falska resultat publiceras i den vetenskapliga litteraturen. En anledning är bedrägeri – att ursprungsstudiens data är påhittade och därmed falska. Ett flertal kända fall av akademiska bedragare inom samhällsvetenskaperna har uppmärksammats (se t.ex. Callaway (2018) för mer diskussion). Men även utan svindlare finns det många skäl till att tro att andelen falska resultat kan vara hög (Ioannidis 2005).

Inom de flesta kvantitativa samhällsvetenskaper har vi dragit gränsen vid  $p < 0,05$  för att kalla ett resultat för statistiskt signifikant. Ett sådant  $p$ -värde tolkas ofta som starkt stöd för den alternativa hypotesen, men  $p$ -värdet i sig

säger bara hur sannolikt vi skulle observera något lika eller mer extremt som det vi observerar om nollhypotesen vore sann. Ett lågt p-värde ger endast starkt stöd för att den alternativa hypotesen är sann om apriori-sannolikheten – den initiala sannolikheten att den testade hypotesen är sann – och den statistiska styrkan är hög. Apriori-sannolikheten är dock ofta subjektiv och otillgänglig vilket gör att den ofta ignoreras.

Statistisk styrka, eller sannolikheten att hitta ett statistiskt signifikant resultat givet att den alternativa hypotesen är sann, är också en nyckelfaktor i att förklara andelen falska resultat. Statistisk styrka är en funktion av studiens storlek (antal observationer), variansen i data samt den förväntade effektstorleken om den alternativa hypotesen är sann. Låg statistisk styrka ökar inte bara sannolikheten att statistiskt signifikanta resultat är falska positiva resultat och att rapporterade effektstorlekar för sanna positiva resultat är kraftigt överdrivna jämfört med de sanna effektstorlekarna, utan även att observerade statistiskt signifikanta resultat är i direkt motsatt riktning till den sanna effekten (Leamer 1983; Gelman och Carlin 2014). Det finns många anledningar att vara orolig för låg statistisk styrka och de problem som medföljer. Inom nationalekonomi har Ioannidis et al. (2018) visat att med deras mest generösa mått på statistisk styrka så är medianstyrkan endast 18%. För t.ex. neurovetenskap har den rapporterats vara 21% (Button et al. 2013).

Falska positiva resultat publiceras också p.g.a. publikationsbias, där forskare inte kommunicerar nollresultat de finner och där tidskrifter föredrar att publicera positiva resultat över nollresultat (Sterling 1959; Rosenthal 1979). Franco et al. (2014) hittar stöd för båda sorternas publikationsbias när de genom en specifik plattform kan följa studier från design till resultat och publikation och kan observera att starka resultat publiceras oftare och bättre än blandade resultat, medan sannolikheten är minst för att nollresultat skall publiceras eller ens skriva ihop till ett papper.

De kanske främsta anledningarna till att falska resultat publiceras är något som fram till relativt nyligen fått mindre uppmärksamhet än vad de förtjänar – de olika frihetsgraderna inom forskningen (*“researcher degrees of freedom”*). Forskarens frihetsgrader innebär att även för studier där forskaren tror sig testa en specifik hypotes finns det ofta frihetsgrader i hur hypotestestet ska göras och forskaren kan mer eller mindre medvetet lura sig själv och andra till att tro att ett specifikt statistiskt signifikant resultat är något slags ”sanning” med låg falsk positiv sannolikhet. P-hacking beskriver den process där forskaren mer eller mindre aktivt söker efter ett statistiskt signifikant re-

sultat. Och det finns många sätt att få fram ett statistiskt signifikant resultat: i ett experiment kan forskaren sluta samla in data så fort hen fått ett resultat med  $p < 0,05$ , eller forskaren kan analysera många mått men bara rapportera dem med  $p < 0,05$  resultat, eller forskaren kan samla in och analysera många behandlingar eller grupper men bara rapportera dem med  $p < 0,05$  resultat, eller forskaren kan använda kontrollvariabler i analysen för att få  $p < 0,05$  resultat, eller forskaren kan exkludera deltagare för att få  $p < 0,05$  resultat, eller forskaren kan transformera sina data för att få  $p < 0,05$  resultat (Simmons, Nelson, och Simonsohn 2011). Slutprodukten blir ett statistiskt signifikant resultat med hög sannolikhet för att vara ett falskt positivt resultat, vilket gör p-värdet meningslöst.

Den kanske vanligaste frihetsgraden är samtidigt den mest förrädiska i termer av hur forskaren kan tro att hen hittat något. *The garden of forking paths* (Gelman och Loken 2013) är när forskaren avser att testa en specifik hypotes på ett dataset men låter data avgöra hur den ska analyseras (inklusive hur variabler ska definieras). När den exakta analysen inte specificerats i förväg uppstår ett multipelt testningsproblem där universumet av tester inte är tydligt eftersom de genomförda testerna hade blivit annorlunda om data sett annorlunda ut. Det blir lätt att hitta ett statistiskt signifikant resultat i slutet av någon specifik väg forskaren vandrat ner för i data, men precis som för p-hacking blir p-värdet meningslöst. Det räcker med att göra ett enda test och det är fortfarande *forking* eftersom, om data sett lite annorlunda ut, fler tester hade gjorts eller variabler definierats annorlunda.

### 3. Replikationsprojekt

Vi definierar att en studie replikerar om vi i replikationen finner en effekt i samma riktning som i ursprungsstudien med ett p-värde  $< 0,05$  i ett två-sidigt test. Detta är vårt huvudmått men vi diskuterar även resultaten med andra mått.

Ett flertal stora replikationsprojekt har haft syftet att med direkta replikationer testa upprepbarheten av publicerade resultat. I termer av antalet studier (och replikationsförfattare!) är *Reproducibility Project: Psychology* (RPP) (Open Science Collaboration 2015) det största. 100 studier från en årgång (2008) av tre tidsskrifter inom psykologi (*Psychological Science*, *Journal of Personality and Social Psychology*, och *Journal of Experimental Psychology: Learning, Memory, and Cognition*) valdes av medlemmarna av ett kollektiv

på 270 forskare ut för att studeras. Replikationerna hade i genomsnitt 92% statistisk styrka till att hitta 100% av originalstudiens effektstorlek på en 5% signifikansnivå. 97 av 100 originalstudier rapporterade ett statistiskt signifikant resultat men för dessa 97 studier var det endast 35 av replikationerna som rapporterade en effekt i samma riktning som i originalstudien med  $p$ -värde  $< 0,05$  i ett två-sidigt test.

I ett replikationsprojekt inom experimentell nationalekonomi (*Experimental Economics Replication Project*; EERP) (Camerer et al. 2016) genomförde vi tillsammans med våra medförfattare 18 replikationer av studier publicerade inom två nationalekonomiska topptidsskrifter (*American Economic Review* och *Quarterly Journal of Economics*) under perioden 2011-2014. Medan RPP inkluderade studier som testade interaktionseffekter och även trippelinteraktioner valde vi i EERP att endast fokusera på huvudeffekter i mellan-deltagarstudier. Våra replikationer hade liknande statistisk styrka som i RPP; i genomsnitt 92% statistisk styrka till att hitta 100% av originalstudiens effektstorlek på en 5% signifikansnivå. Av de 18 originalresultaten fann vi stöd för 11 resultat.

I vårt senaste publicerade replikationsprojekt inom samhällsvetenskaplig forskning (*Social Science Replication Project*; SSRP) (Camerer et al. 2018) genomförde vi 21 replikationer av studier publicerade inom de två allmänvetenskapliga topptidsskrifterna *Nature* och *Science* under perioden 2010-2015. Dessa replikationer hade betydligt högre statistisk styrka än de två tidigare projekten, för att ta hänsyn till att även sanna positiva resultat i genomsnitt har överskattade effektstorlekar i originalstudierna. Vi genomförde replikationerna i två steg. I steg 1 av datainsamlingen hade vi 90% statistisk styrka att hitta 75% av originalstudiens effektstorlek. Om vi då inte lyckades replikera resultatet gick studien vidare till steg 2 där datainsamlingen i det poolade urvalet av steg 1 och steg 2 hade 90% statistisk styrka till att hitta 50% av originalstudiens effektstorlek. För studier som fortsatte till steg 2 hade vi således cirka fyra gånger så stora urval som vi hade haft om vi hade nöjt oss med 90% styrka för att hitta 100% av originalstudiens effektstorlek, och jämfört med originalstudierna är urvalen ca sex gånger så stora. Av de 21 ursprungsresultaten replikerade 13 resultat. Vi lär oss också något av att ha hög statistisk styrka – två resultat som inte replikerade i steg 1 replikerade i steg 2.

Gemensamt för alla ovannämnda projekt är de replikationsrapporter som skrevs och kommunicerades till originalförfattarna. I dem beskrev replika-

tionsförfattarna vilken specifik delstudie och vilket statistiskt test som replikationen fokuserade på (många originalartiklar innehöll flera studier; i RPP valdes ofta den sista och i SSRP valdes istället den första) och replikationen jämfördes i detalj med originalstudien (där t.ex. skillnader i urval beskrevs). Dessa rapporter skickades till originalförfattarna som i en majoritet av fallen godkände rapporten innan replikationen satte igång.

Andra projekt med hög statistisk styrka som vi inte har varit inblandade i är *Many Labs*-projekten. I dessa projekt har forskarna valt att replikera både nya och klassiska effekter inom psykologi för att studera hur robusta effekterna är i olika sammanhang. *Many Labs* projekten har hög statistik styrka, där många lab (d.v.s. av många grupper) replikerar varje studie. I *Many Labs 1* replikerade 10 av 13 resultat (Klein et al. 2014), i *Many Labs 2* replikerade 15 av 28 resultat (Klein et al. 2018), och i *Many Labs 3* replikerade 3 av 10 resultat (Ebersole et al. 2016). Generellt hittar de olika *Many Labs* projekten inte starkt stöd för att kontext skulle spela stor roll för huruvida ett resultat replikerar. Fler *Many Labs* projekt är på väg.

Det finns även många andra intressanta replikationsprojekt inom psykologi. Ett exempel är s.k. *Registered Replication Reports* (RRRs) där en studie väljs ut för replikation, replikations- och originalförfattarna utarbetar replikationsprotokollet, och studien replikeras sedan i många labb. En av oss (Magnus) var med på en RRR kring huruvida tidspress påverkar samarbete i fångarnas dilemma – resultaten från studien med 21 grupper och 3596 experimentdeltagare fann inte stöd för detta när analysen gjordes korrekt (vilket inte var fallet i originalstudien) (Bouwmeester et al. 2017).

Inom experimentell filosofi har det också genomförts ett replikationsprojekt på 40 studier publicerade 2003-2015 (Cova et al. 2018). Av de 40 originalresultaten rapporterade 37 studier statistiskt signifikanta resultat – 29 av dessa replikerade.

Det finns många komplement till det binära replikationsmåttet vi rapporterar ovan. Ett sådant är den relativa effektstorleken, där effektstorlekar standardiseras så att originalstudier och replikationer kan jämföras. Detta blir ett kontinuerligt mått på graden av replikation. Utan systematisk bias i publicerade resultat borde relativa effektstorlekar i genomsnitt vara 100%. I RPP var de i genomsnitt ca 50% och i EERP ca 60%. I SSRP var de nära 50% (74% för studierna som replikerade och 0% för dem som inte replikerade).

Andra mått som föreslagits är huruvida replikationens 95 procentiga konfidensintervall inkluderar originaleffekten eller inte; det 95 procentiga pre-

diktionsintervallet som föreslår att man estimerar ett 95 procentigt prediktionsintervall för originaleffekten och sen räknar hur många replikerade effekter som ligger inom detta intervall (vilket innebär att testa om originalresultatet och replikationsresultatet skiljer sig åt statistiskt signifikant) (Patil, Peng, och Leek 2016); och det ”lilla teleskop”-mättet som innebär att testa om replikationens effektstorlek är signifikant mindre ( $p < 0,05$  i ett en-sidigt test) än en ”liten effekt” i originalstudien, där liten effekt definieras som den effektstorlek originalstudien skulle ha haft 33% statistisk styrka till att hitta (Simonsohn 2015). Alla dessa mått har en del problem som gör att vi inte anser att de är att föredra över det binära mått vi presenterat ovan eller den relativa effektstorleken. Se t.ex. Camerer et al. (2019) för mer diskussion. Även Bayes-faktorn har blivit populärare att rapportera. Den representerar styrkan av bevis till förmån för den alternativa hypotesen jämfört med den för nollhypotesen (Wagenmakers, Verhagen, and Ly 2016; Marsman et al. 2017). För både RPP och EERP ledde ovannämnda mått till olika slutsatser, men för SSRP med dess höga statistiska styrka konvergerade måtten mer.

#### 4. Förutsägbarhet

I vilken utsträckning kan forskare förutse vilka resultat som replikerar? För RPP, EERP, SSRP och *Many Labs 2* bjöd vi in forskare till att försöka förutse replikationsresultaten med hjälp av en enkät och en prognosmarknad (Dreber et al. 2015; Camerer et al. 2016, 2018; Forsell et al. 2018). I enkäten frågade vi forskarna hur sannolikt de bedömde det vara att studien skulle replikera, och på prognosmarknaden lät vi deltagarna handla kontrakt som representerade replikationerna. Om en studie replikerade var kontraktet värt ett visst antal poäng, där poängen motsvarade pengar vi betalade ut, och om studien inte replikerade var kontraktet värt noll poäng. Med vissa brasklappar (se t.ex. Manski (2006) kan man tolka priset på sådana kontrakt som ett mått på hur sannolik marknaden bedömer möjligheten av ett utfall – i vårt fall att studien replikerar. Prognosmarknader föreslogs för användning i forskning av Robin Hanson (Hanson 1995) och testades sen i labbet i en forskningskontext av Almenberg, Kittlitz, och Pfeiffer (2009). Vi inkluderade inte alla studier i RPP och *Many Labs 2* utan vårt poolade urval just nu består av 104 studier från prognosmarknader och 103 studier från enkäterna. Våra resultat tyder på att det finns något slags ”*wisdom of crowds*” där forskare kan förutse vilka resultat som håller. I den enklaste analysen där vi säger att enkätsvar eller

priser över 50% tolkas som att forskarna tror att studien kommer att replikera, finner vi en 73% (76/104 studier) korrekt prognosandel för prognosmarknaderna (Forsell et al. 2018). För enkäten är motsvarande siffra 66% (68/103 studier). Vi genomför nu fler studier på det här temat.

## 5. Framåt

Våra replikationsprojekt tyder på att vi har problem med upprepbarhet av många publicerade resultat bland de kvantitativa samhällsvetenskaperna. Våra studier med prognosmarknader och enkäter tyder också på att det finns något systematiskt i vilka resultat som inte replikerar. Varför publiceras då dessa studier i toptidskrifter? Vi har inget tydligt svar på den frågan.

De relativt låga replikationssiffrorna kan lätt göra en nedstämd. Men om vi ser på utvecklingen inom psykologi finns det många anledningar till att vara positiv. En lösning på problemet med de olika frihetsgraderna i forskningen är att förregistrerade analysplaner som beskriver exakt hur analysen ska göras. I ”power posing”-fallet framgick det att originalförfattarna hade varit inblandade i flera olika sorters p-hacking utan att förstå problemen de ledde till (Carney 2016). Med en förregistrerad analysplan hade detta kunnat undvikas (och ”power posing” hade nog aldrig blivit ett fenomen). Förregistrerade analysplaner har blivit vanliga inom psykologi och många andra fält (Nosek et al. 2018). Inom nationalekonomi har de främst fått genomslag inom utvecklingsekonomi (Casey, Glennerster, och Miguel 2012; Olken 2015; Christensen och Miguel 2018), men till vår besvikelse har de ännu inte blivit vanliga inom beteendekonomi. *Registered reports* (RRs) är som förregistrerade analysplaner där en RR genomgår referentgranskning innan studien genomförs och den vetenskapliga tidskriften bestämmer innan data samlas in om studien ska publiceras. Även den utvecklingen är mycket positiv, och leder till att även publikationsbias undviks.

Diskussioner kring vilken p-värdeströskel som ska användas för statistisk signifikans, om någon alls, är också i full gång vilket vi tycker är positivt. Vi tillhör dem som anser att p-värdegränsen för statistik signifikans borde sänkas till 0,005 (Benjamin et al. 2018). Med denna nya gräns minskar andelen falska positiva resultat och urvalsstorlekar blir större för en viss given statistisk styrka. Trots att  $p < 0,005$  är en ny arbiträr gräns så är den mer meningsfull än  $p < 0,05$  i termer av falsk positiv sannolikhet för olika apriori-sannolikheter för de testade hypoteserna. Med hjälp av prognosmarknaden för RPP upp-



skattar vi att medianen för sannolikheter a priori inom psykologi är cirka 1:10 (Dreber et al. 2015), och även med denna låga apriori-sannolikhet hålls sannolikheten för ett falskt positivt resultat nere för en stor spännvidd av statistisk styrka med  $p < 0,005$ .

I det här kapitlet har vi främst diskuterat experimentell forskning. Vi tror dock att problemen är större för andra slags empirisk kvantitativ forskning eftersom forskarens frihetsgrader där är betydligt större och även om utmaningarna där är större hoppas vi att t.ex. förregistrering blir vanligare så att vi kan få mer tillförlitliga resultat

### Referenser

- Almenberg, Johan, Ken Kittlitz, och Thomas Pfeiffer. 2009. "An Experiment on Prediction Markets in Science." *PLOS ONE* 4 (12): e8500.
- Benjamin, Daniel J., James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, et al. 2018. "Redefine Statistical Significance." *Nature Human Behaviour* 2 (1): 6–10.
- Bouwmeester, S., P. P. J. L. Verkoeijen, B. Aczel, F. Barbosa, L. Bègue, P. Brañas-Garza, T. G. H. Chmura, et al. 2017. "Registered Replication Report: Rand, Greene, and Nowak (2012)." *Perspectives on Psychological Science* 12 (3): 527–42.
- Button, Katherine S., John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, och Marcus R. Munafò. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews Neuroscience* 14 (April): 365.
- Callaway, Ewen. 2018. "Report Finds Massive Fraud at Dutch Universities : Nature News." June 18, 2018. <https://www.nature.com/news/2011/111101/full/479015a.html>.
- Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2016. "Evaluating Replicability of Laboratory Experiments in Economics." *Science* 351 (6280): 1433–36.
- Camerer, Colin F., Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2018. "Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015." *Nature Human Behaviour* 2 (9): 637–44.



- Camerer, Colin F., Anna Dreber, och Magnus Johannesson. 2019. "Replication and Other Practices for Improving Scientific Quality in Experimental Economics." In *Handbook of Research Methods and Applications in Experimental Economics*.
- Carney, Dana R. 2016. "My Position on 'Power Poses.'"
- Carney, Dana R., Amy J.C. Cuddy, och Andy J. Yap. 2010. "Power Posing: Brief Non-verbal Displays Affect Neuroendocrine Levels and Risk Tolerance." *Psychological Science* 21 (10): 1363–68.
- Casey, Katherine, Rachel Glennerster, och Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan\*." *The Quarterly Journal of Economics* 127 (4): 1755–1812.
- Chang, Andrew C., och Phillip Li. 2015. "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say 'Usually Not.'" Finance and Economics Discussion Series. Board of Governors of the Federal Reserve System (U.S.). <https://ideas.repec.org/p/fip/fedgfe/2015-83.html>.
- Christensen, Garret, och Edward Miguel. 2018. "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature* 56 (3): 920–80.
- Clemens, Michael A. 2017. "The Meaning of Failed Replications: A Review and Proposal." *Journal of Economic Surveys* 31 (1): 326–42.
- Cova, Florian, Brent Strickland, Angela Abatista, Aurélien Allard, James Andow, Mario Attie, James Beebe, et al. 2018. "Estimating the Reproducibility of Experimental Philosophy." *Review of Philosophy and Psychology*, June.
- Dewald, William G., Jerry G. Thursby, och Richard G. Anderson. 1986. "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project." *The American Economic Review* 76 (4): 587–603.
- Dreber, Anna, och Magnus Johannesson. 2019. "Statistical Significance and the Replication Crisis in the Social Sciences." In *Oxford Research Encyclopedia of Economics and Finance*.
- Dreber, Anna, och Magnus Johannesson. 2018. "Vilka Forskningsresultat Kan vi Lita På?" *Ekonomisk Debatt* 46 (2): 17–28.
- Dreber, Anna, Thomas Pfeiffer, Johan Almenberg, Siri Isaksson, Brad Wilson, Yiling Chen, Brian A. Nosek, och Magnus Johannesson. 2015. "Using Prediction Markets to Estimate the Reproducibility of Scientific Research." *Proceedings of the National Academy of Sciences* 112 (50): 15343.
- Ebersole, Charles R., Olivia E. Atherton, Aimee L. Belanger, Hayley M. Skulborstad, Jill M. Allen, Jonathan B. Banks, Erica Baranski, et al. 2016. "Many Labs 3: Evaluating Participant Pool Quality across the Academic Semester via Replication." *Journal of Experimental Social Psychology*, Special Issue: Confirmatory, 67 (November): 68–82.
- Forsell, Eskil, Domenico Viganola, Thomas Pfeiffer, Johan Almenberg, Brad Wilson, Yiling Chen, Brian A. Nosek, Magnus Johannesson, och Anna Dreber. 2018. "Predicting Replication Outcomes in the Many Labs 2 Study." *Journal of Economic Psychology*, October.

- Franco, Annie, Neil Malhotra, och Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345 (6203): 1502.
- Gelman, Andrew, och John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science* 9 (6): 641–51.
- Gelman, Andrew, och Eric Loken. 2013. "The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No 'Fishing Expedition' or 'p-Hacking' and the Research Hypothesis Was Posited Ahead of Time." Working Paper.
- Gertler, Paul, Sebastian Galiani, och Mauricio Romero. 2018. "How to Make Replication the Norm." *Nature* 554: 417–19.
- Hamermesh, Daniel S. 2007. "Viewpoint: Replication in Economics." *Canadian Journal of Economics/Revue Canadienne d'économique* 40 (3): 715–33.
- Hanson, Robin. 1995. "Could Gambling Save Science? Encouraging an Honest Consensus." *Social Epistemology* 9 (1): 3–33.
- Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLOS Medicine* 2 (8): e124.
- Ioannidis, John P. A., T. D. Stanley, och Hristos Doucouliagos. 2018. "The Power of Bias in Economics Research." *The Economic Journal* 127 (605): F236–65.
- Klein, Richard A., Kate A. Ratliff, Michelangelo Vianello, Reginald B. Adams Jr., Štěpán Bahník, Michael J. Bernstein, Konrad Bocian, et al. 2014. "Investigating Variation in Replicability: A 'Many Labs' Replication Project." *Social Psychology* 45 (3): 142–52.
- Klein, Richard A., Michelangelo Vianello, Fred Hasselman, Byron G. Adams, Reginald B. Adams, Sinan Alper, Mark Aveyard, et al. 2018. "Many Labs 2: Investigating Variation in Replicability Across Samples and Settings." *Advances in Methods and Practices in Psychological Science* 1 (4): 443–90.
- Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *The American Economic Review* 73 (1): 31–43.
- Manski, Charles F. 2006. "Interpreting the Predictions of Prediction Markets." *Economics Letters* 91 (3): 425–29.
- Marsman, Maarten, Felix D. Schönbrodt, Richard D. Morey, Yuling Yao, Andrew Gelman, och Eric-Jan Wagenmakers. 2017. "A Bayesian Bird's Eye View of 'Replications of Important Results in Social Psychology.'" *Royal Society Open Science* 4 (1): 160426.
- McCullough, B. D., Kerry Anne McGeary, och Teresa D. Harrison. 2006. "Lessons from the JMCB Archive." *Journal of Money, Credit and Banking* 38 (4): 1093–1107.
- McCullough, Bruce D., och H. D. Vinod. 2003. "Verifying the Solution from a Nonlinear Solver: A Case Study." *American Economic Review* 93 (3): 873–92.
- Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, och David T. Mellor. 2018. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences*, March.
- Olken, Benjamin A. 2015. "Promises and Perils of Pre-Analysis Plans." *Journal of Economic Perspectives* 29 (3): 61–80.

- Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251).
- Patil, Prasad, Roger D. Peng, och Jeffrey T. Leek. 2016. "What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science." *Perspectives on Psychological Science* 11 (4): 539–44.
- Ranehill, Eva, Anna Dreber, Magnus Johannesson, Susanne Leiberg, Sunhae Sul, och Roberto A. Weber. 2015. "Assessing the Robustness of Power Posing: No Effect on Hormones and Risk Tolerance in a Large Sample of Men and Women." *Psychological Science* 26 (5): 653–56.
- Rosenthal, Robert. 1979. "The File Drawer Problem and Tolerance for Null Results." *Psychological Bulletin* 86 (3): 638–41.
- Simmons, Joseph P., Leif D. Nelson, och Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22 (11): 1359–66.
- Simonsohn, Uri. 2015. "Small Telescopes: Detectability and the Evaluation of Replication Results." *Psychological Science* 26 (5): 559–69.
- Sterling, Theodore D. 1959. "Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance--Or Vice Versa." *Journal of the American Statistical Association* 54 (285): 30–34.
- Wagenmakers, Eric-Jan, Josine Verhagen, och Alexander Ly. 2016. "How to Quantify the Evidence for the Absence of a Correlation." *Behavior Research Methods* 48 (2): 413–26.



Lars Nyberg

## Aktivering av hjärnan – reproducerbarhet och generaliserbarhet

### En ny hjärnavbildningsteknik växer fram

Kan tanken fångas på bild? Var i hjärnan kan medvetandet lokaliseras, och var sitter minnet? Kan romantisk kärlek eller empati kopplas till särskilda hjärncentra? Har hjärnan redan gjort vårt val då vi upplever att vi fattar det?

Detta är några exempel på frågeställningar som formulerats i olika *funktionella hjärnavbildningsstudier*, vilka som namnet antyder söker koppla olika funktioner till hjärnans områden eller konstellationer av områden (s.k. nätverk). Förvisso har hjärnavbildningsfältet utvecklats i rasande takt de senaste decennierna, men fortfarande återstår ett stort antal metodologiska utmaningar innan detta slags frågor låter sig definitivt besvaras.

Den idag dominerande metoden är att använda funktionell magnetkamera. Jag kommer fortsättningsvis i kapitlet att använda den engelska förkortningen *fMRI*, som står för *functional magnetic resonance imaging*. Den teoretiska basen för MRI-tekniken publicerades av Paul Lauterbur i början av 1970-talet<sup>1</sup>, och han skulle sedermera år 2003 dela Nobelpriset i medicin eller fysiologi med Peter Mansfield för deras upptäckter kring just MRI. Tillämpningen av fMRI för att studera hjärnans funktioner introducerades i början av 1990-talet<sup>2</sup>. Då genomfördes den typen av studier på endast ett fåtal platser i världen och enbart ett litet antal studier publicerades varje år. Resten är som det brukar heta historia; fMRI-fältet har internationellt fullkomligt exploderat med flera tusen publikationer per år. Förklaringen till den snabba framväxten är till stor del att magnetkameror finns tillgängliga på sjukhus och institut världen över, medan andra hjärnavbildningstekniker (t.ex. positronemissionstomografi; PET) är mer sällsynta.

Naturligt nog präglades de inledande åren av fMRI-forskning av en iver att tillämpa metoden på nya spännande frågeställningar, vilket rikligen belönades med frekventa publikationer i topprankade tidskrifter som *Science* och *Nature*. Ett annat utmärkande drag för de tidiga studierna är att resultaten

vanligen baserades på undersökningar av små stickprov. Dessa karakteristika av tidiga fMRI-studier leder naturligtvis till frågetecken kring hur reproducerbara och generaliserbara de påvisade fynden kan tänkas vara. Dessa frågor är högaktuella även inom dagens hjärnavbildningsforskning, och jag kommer i resten av kapitlet att diskutera dem ur några olika perspektiv.

## Det är svårt att registrera hjärnaktivitet

Det ska sägas direkt, att fMRI endast erbjuder ett indirekt mått på hjärnans aktivitet<sup>3</sup> genom att kartlägga blodflödesförändringar och variation i syrehalten i olika hjärnområden [*Blood Oxygen Level Dependent (BOLD) signal*]. Andra tekniker, som då elektroder placeras i eller i nära anslutning till hjärnans nervceller, erbjuder mer direkt registrering av hjärnaktivitet. Den typen av registrering är emellertid i humanförsök vanligen begränsad till situationer då skallbenet tas bort i samband med olika kliniska behandlingar, och de erbjuder därför endast mätning i en begränsad del av hjärnan, medan fMRI-signalen ofta avspeglar hela hjärnans aktivitet. Jämförande undersökningar indikerar att fMRI-mätningar överensstämmer tämligen väl med mer direkta registreringar, men med begränsad spatial och temporal upplösning. Detta innebär att fMRI registreringen influeras av signaler från ett mycket stort antal nervceller och därmed har en begränsning vad gäller hur specifikt den kan sägas avspegla vad som händer i en viss del av hjärnan, samt att tidsskalan för registreringen är på sekund- snarare än millisekundnivå. Dessa teknikaliteter problematiserar fMRI-teknikens validitet vad gäller mätning av faktisk neural aktivitet, men de tangerar även frågor kring reproducerbarhet. Till exempel så skulle en bristande grad av replikering av ett fynd från ett stickprov till ett annat kunna påverkas av skillnader kopplade till hjärnans blodflöde (exempelvis orsakade av åldersskillnader mellan de två stickproven).

Magnetkameran har som sin viktigaste komponent just en stark magnet. Styrkan på denna kan variera och idag är 3 Tesla (T) vanligt, men 1,5T, 4T och till och med 7T förekommer. En annan viktig komponent är den specifika spole som huvudet placeras i under hjärnmätningar, eftersom prestandan av olika spolar varierar. Således, om inte resultaten från ett experiment utfört i ett visst lab kan upprepas i ett annat, så kan det till viss del förklaras med olika mätegenskaper hos respektive utrustning – även om båda utrustningarna har lika stark magnet. En starkt framväxande trend är att i olika storskaliga internationella konsortier (t.ex. ENIGMA) slå samman fMRI-registreringar

från olika fMRI-kameror och relatera dessa till vissa variabler (t.ex. genetisk variation). Där kommer en stor del av variansen att utgöras av metodvariation, men förhoppningen är att storskaligheten i ansatsen med 1000-tals individer kommer att uppväga detta. Försök görs för att kompensera för variation i mätningar från olika lab. Exempelvis har vi i ett pågående projekt, *Lifebrain*, låtit en grupp individer resa runt för att scannas vid samtliga ingående lab. Detta ger en uppfattning av hur stor påverkan på fynden som variation i utrustning faktiskt har, och kan erbjuda sätt att kompensera för detta.

Andra mättekniska aspekter som kan påverka hur väl fMRI-fynd reproduceras inkluderar hantering av deltagarnas huvudrörelser under scanning. Det är av stor vikt att deltagarna ligger stilla under hela försöket då rörelse kan negativt påverka lokalisering samt introducera artefakter i signalen. Minimering av huvudrörelse kan uppnås på olika sätt, såsom att i huvudspolen infoga mjuka kuddar så att huvudet ”fixeras”, samt försäkra sig om att deltagarna lägger sig bekvämt och slappnar av. I vissa lab sköts detta av professionell MR-personal, medan det i andra fall utförs av doktorander eller annan forskande personal. Min erfarenhet är att ett professionellt team bidrar till att göra deltagarna lugna och avslappnade, vilket sannolikt reducerar rörelser under fMRI-mätningar.

## Konsten att designa hjärnabbildningsstudier

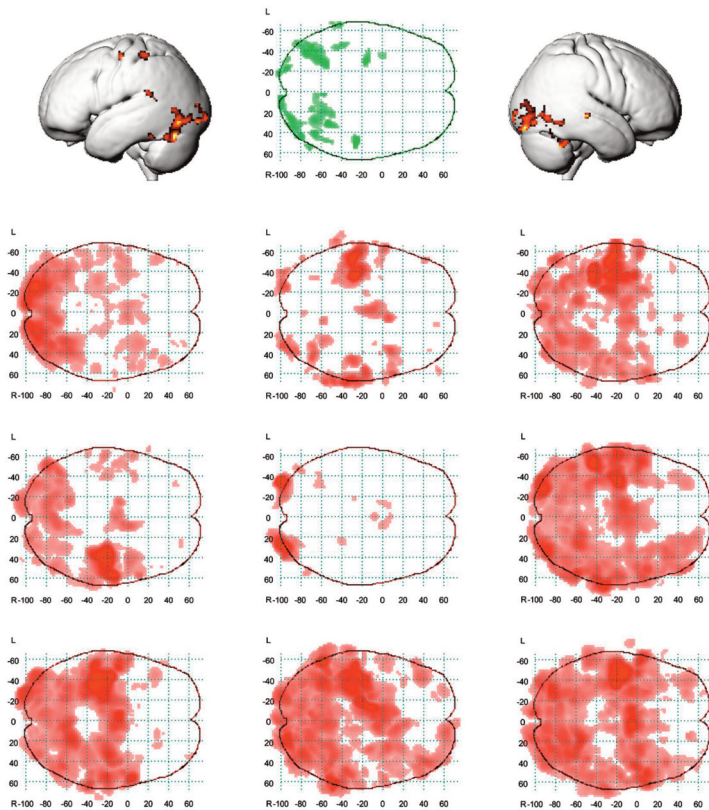
Det finns likheter mellan fMRI och mätning av temperatur i den mening att det inte finns någon absolut nollpunkt. Den dominerande trenden i fMRI-studier har därför kommit att bli att jämföra fMRI-signalen som uppmäts i en experimentell betingelse med mätningar under en kontrollbetingelse. Utifrån denna ”subtraktionslogik” är förhoppningen att signal som kopplas till den intressant variabeln, och som därmed skiljer sig mellan betingelserna, kan isoleras. Med tiden har mer avancerade designer och statistiska metoder kommit att introduceras, men grunden är att fånga den relativa signalförändringen mellan två eller flera betingelser. I tidiga fMRI-studier kunde detta handla om att i vissa delar av försöket, d.v.s. under vissa segment av tidsserien, exponera deltagarna för ett blinkande ljus. Under andra segment så förekom ingen yttre visuell stimulering. Detta upplägg kan påvisa en förändring av fMRI-signalen i delar av hjärnan som bearbetar syninformation. På likartat sätt kan inkludering av motorik eller beröring i vissa men inte andra betingelser identifiera delar av de motoriska och sensoriska systemen. Den här typen



av kontrast mellan en sensorisk eller motorisk betingelse med en vilobetingelse genererar vanligen robusta aktiveringar såväl inom individer (tämmligen god reliabilitet) som mellan individer (se Figur 1).

**Figur 1.**

Genomsnittlig och individualspecifik fMRI-aktivering. Figuren visar nio individers hjärnaktivering (rött) sett från ovan i en "glashjärna". Resultaten är från en fMRI-studie där en kontrollbetingelse med ingen visuell stimulering alternerade med en experimentell betingelse som bestod av att ett mönster med blinkande rutor presenterades och där deltagarna tryckte med höger pekfinger varje gång ett mönster visades. Den experimentella betingelsen förväntades medföra ökad signal i synbarken samt i kontralaterala (vänstra) motorcortex. Som framgår av varje deltagares hjärnaktivitet så observerades genomgående aktivering i syncortex baktill i hjärnan. Trots att varje individuell aktivitetskarta visas på samma statistiska tröskel ( $p < 0,05$  efter FDR-korrigerings för multipla jämförelser) så ses även individuella skillnader, och aktivering i vänster motorcortex kunde bara skönjas för vissa av deltagarna. Den gröna glashjärnan högst upp visar hjärnaktivitet för samma kontrast mellan experimentell betingelse och kontrollbetingelse sammanslaget över de nio individerna ( $p < 0,0001$  utan korrektion för multipla jämförelser). I denna genomsnittskarta framträder som förväntat robust bilateral aktivering i synbarken och även i vänstra motorbarken. Detta mönster illustreras även på en "mallhjärna" för att bättre illustrera aktiveringarna i relation till hjärnans anatomi. Deltagarna var samtliga cirka 55 år och deltog tillsammans med många andra deltagare i en av våra fMRI-studier inom Betula-projektet (se referens 4).



Tack till Micael Andersson för hjälp med figuren.



Efter att inledningsvis ha varit fullt nöjda med att påvisa förändrad signal i synområden för visuell stimulering kontra ingen stimulering kom forskare snabbt in på att ställa mer förfinade frågor. Ett exempel är huruvida olika delar av synsystemet är särskilt benägna att aktiveras av vissa typer av stimuli. Ett prominent exempel är att se på ansikten kontra andra visuella stimuli (bilar, byggnader, musikinstrument, m.m.). Inom perceptionspsykologin har det sedan länge funnits ett stort intresse för just ansikten, och ibland brukar det debatteras om vi har en särskild ”modul” för ansiktigenkänning eller om denna förmåga snabbt utvecklas då ansiktsperception redan hos spädbarn är en viktig förmåga. Från den typen av teoretiserande var steget kort till att söka identifiera ett ”ansiktsområde” i människans hjärna, och det har visat sig att delar av den nedre (ventrala) synbanan uppvisar särskilt stark signalförändring då just ansikten visas (*fusiform face area*, FFA). Detta har studerats vidare i ett stort antal studier. Vi har inom det Umeå-baserade Betula-projektet visat att under både inkodning och framplöckning av ansikten (och namn) så är den starkaste fMRI-signalen lokaliserad i trakterna av FFA-regionen<sup>4</sup>. Den analysen baserades på nära 300 deltagare, vilket är ett anmärkningsvärt stort stickprov i fMRI-studier, och ger därmed starkt stöd för god överensstämmelse i fMRI-signalen över ett stort antal individer.

Den fortsatta utvecklingen har i rask takt gått mot alltmer komplexa design och frågeställningar. I sammanhanget reproducerbarhet så är en observation att uppgifter som har fler frihetsgrader vad gäller hur de kan lösas kommer att riskera ha en mindre sannolikhet att generera samstämmiga fMRI-fynd över individer och över upprepade experiment (dvs lägre reliabilitet). Detta är förstås ingen brist vad gäller fMRI-metoden i sig, utan en avspiegling av att mer uppgifter kan lösas på flera sätt och därmed åtminstone till viss del aktivera olika nätverk i hjärnan. Om detta skall ses som ett problem eller en intressant aspekt som i sig är väl värd att studera beror på forskningsfrågan. Om en forskare vill undersöka huruvida aktiveringsgraden i region A predicerar hur väl uppgift B löses, då bör upplägget på studien försöka minimera frihetsgrader så att deltagarna tar sig an uppgiften på ett likartat sätt (”använda en specifik strategi som taxerar region A”). I andra sammanhang kan variation i strategieval vara av intresse i sig, och då kan fMRI-metoden erbjuda ett fönster för att fånga just den dimensionen – vilket inte skulle vara möjligt enbart utifrån registrering av huruvida rätt eller fel svar genererades<sup>5</sup>. Således kan variationer i fMRI-utfallet i det senare exemplet liknas med talesättet att det finns många vägar till Rom!

## Den genomsnittliga kontra den individuella och föränderliga hjärnan

Att variation i strategi för att lösa en uppgift avspeglar sig i fMRI-signalen för oss in på en dominerande trend i fältet de senaste åren, nämligen *individuella skillnader*. De inledande studierna analyserade resultaten på individnivå och kunde sedan rapportera huruvida samtliga av (de fåtaliga) deltagarna aktiverade en viss region<sup>2</sup>. Snart följde en intensiv utveckling mot att skapa mallhjärnor för sammanslagning av resultat från flera deltagare<sup>6</sup>. Detta behov var sprunget ur det faktum att olika individer har olika stora hjärnor, och hjärnor kan även variera till betydande del vad gäller anatomiska särdrag. För att kunna säga att deltagarna i genomsnitt aktiverade exempelvis amygdala, den lilla mandelformade kärna i temporalloben som förknippas med emotion, ja då gällde det att den hopslagna signalen kunde anses avspegla en signal från just amygdala hos samtliga individer. De mallar som utvecklades kan likställas med ett tredimensionellt koordinatsystem, så att specifika områden kan beskrivas i termer av x,y,z-koordinater, Brodmann-regioner, eller andra anatomiska termer.

Att kunna slå samman resultat från enskilda individer till gruppdata innebär helt klart en revolution inom fältet. Det gjorde det möjligt att kommunicera fynd i ett enhetligt ramverk och styrkan på aktiveringar kunde anges i konventionella statistiska termer. Tillsammans med min kollega Roberto Cabeza summerade jag genomsnittsaktivering för flera kognitiva funktioner på basis av ett stort antal enskilda studier<sup>7</sup>. Resultaten uttryckte vi i summeringstabeller som sammanfattade de mest typiska aktiveringsmönstren för respektive funktion. Denna ansats rönt stort intresse och artikeln ligger enligt en färsk analys<sup>8</sup> på plats 13 av världens 100 mest citerade arbeten inom *neuroimaging*. Jag tror att den starkaste anledningen till att vår artikel rönt så stort intresse är just att den gav läsaren en snabb översikt av de mest typiska aktiveringarna för olika funktioner. Detta perspektiv fortsätter att vara av intresse, men nu har pendeln åter svängt mot att analysera resultat på individnivå för att fånga den rika variation som finns i hur olika individers hjärnor aktiveras<sup>9</sup> – en variation som går förlorad i gruppanalyserna. Vilken som är den ”rätta” analysnivån avgörs från fall till fall utifrån den specifika forskningsfrågan.

Variation mellan individer attraherar således mer och mer intresse. En relaterad, men mindre uppmärksam, variationskälla handlar om variation inom individer. Ett stort antal dynamiska förlopp kan influera fMRI-signalen.

Ett exempel handlar om uppgiftens svårighetsgrad. Resultat från meta-analyser som integrerar fynd från flera olika delstudier visar på en stark trend att hjärnaktivitet i vissa delar av hjärnan reduceras om ett test upprepas<sup>10-11</sup>. Tolkningen av detta utfall är att upprepningen gör att uppgiften blir mindre krävande, vilket återspeglas i en reducerad fMRI-signal. Detta gör att det i träningsstudier är viktigt att inkludera kontrollgrupper för att uppskatta graden av test-retest effekter inom individer. En generell implikation när det gäller reproducerbarhet är att då likheter och skillnader i utfall mellan olika studier av samma fenomen utvärderas så är det viktigt att beakta faktorer som grad av förträning (t.ex. i en ”låtsas”-fMRI-scanner i ett beteendelab), huruvida stimuli visas upprepade gånger, *state*-effekter (som trötthet) och liknande faktorer som kan påverka variation i hjärnaktiviteten inom och mellan individer.

## Analysmetoder och statistisk power

I ett kapitel om reproducerbarhet inom fMRI-området kan jag inte undgå att kommentera bearbetningen av fMRI-signalen, vilket i sig är en faktor som påverkar utfall och därmed reproducerbarhet. Ett flertal olika program har genom åren utvecklats för så kallad pre-processande av signalen och följande statistisk utvärdering. Ett exempel är SPM (*Statistical Parametric Mapping*) som utvecklades i London i början av 1990-talet, inledningsvis med fokus på analys av resultat från PET-experiment men sedan med ett ökande fokus på fMRI-data. Själv har jag sedan början av min tid inom hjärnbildningsfältet använt denna mjukvara<sup>12</sup>, och vi fortsätter än idag att använda den i våra studier. För såväl SPM som andra mjukvaror gäller att det finns flera olika val som kan påverka slutresultatet. Utan att i detalj diskutera dessa så bör det i sammanhanget noteras att det gradvis växer fram en förståelse för denna potentiella felkälla, vilket bland annat tar sig i uttryck i att krav ställs på en detaljerad beskrivning av genomförda analyssteg. Idealt bör valen göras och dokumenteras (exempelvis i offentliga arkiv) före den faktiska analysen, men om vägval görs under analysarbetet så ska dessa redovisas och eventuellt beaktas i den statistiska utvärderingen<sup>13</sup>. Nyckelordet är transparens och öppenhet vad gäller att dela information, data, kod för analys och annat som kan vara informativt vid utvärdering och försök till att replikera.

Den faktiska statistiska utvärderingen är precis som de olika preprocess-stegen förknippad med olika valmöjligheter. Den sannolikt mest diskuterade aspekten handlar om hur man på bästa sätt kan gardera sig mot falska positiva

fynd. Även i sin enklaste form, med en experimentell betingelse och en kontrollbetingelse (som beskrevs ovan), kommer den statistiska analysen innebära att ett mycket stort antal jämförelser görs (tusentals). Den gängse hållningen i dylika fall är att korrigera för antalet test, exempelvis med något som kan liknas vid Bonferroni-metoden. Tillämpningen av denna metod på fMRI-data kompliceras emellertid av att närliggande delar (voxlar) i hjärnan tenderar att vara inbördes korrelerade då de utgör en funktionell enhet (t.ex. primära synbarken). Detta försvårar appliceringen av korrektionsmetoder, och frågan om hur risken för falska positiva resultat bäst hanteras har utvecklats till en ”vetenskap inom vetenskapen” – och den fortsätter att generera diskussion och debatt<sup>14</sup>. En relaterad problematik, som även den orsakas av icke-beroende statistiska analyser, handlar om att först göra en övergripande analys och på basis av denna fokusera efterföljande analyser på en specifik delmängd av den rika dataskörden från ett fMRI-experiment<sup>15</sup>. Att undvika denna typ av ”cirkularitet” i analysen är angeläget då den leder till bias som negativt kommer påverka reproducerbarheten.

Som noterades inledningsvis så har fMRI-resultat tenderat att baseras på små stickprov. Detta inverkar förstas negativt på den statistiska kraft (*power*) för att detektera sanna effekter som föreligger i ett försök. Trenden går så sakteliga bort från de 10-tal deltagare per studie som var typfallet i de tidiga fMRI-studierna mot ett 30-tal deltagare per studie idag<sup>16</sup>. Därmed minskar förstas risken för att missa sanna effekter. En annan positiv effekt av större stickprov, som ofta röner mindre uppmärksamhet, är att chansen ökar för att påvisade effekter faktiskt är sanna effekter som sannolikt kommer att reproduceras. I sammanhanget bör också betonas att ett experiments totala *power* inte bara påverkas av antalet deltagare utan även själva försöksupplägget, där tumregeln är att det alltid är bra att få med fler mätningar per individ. Naturligtvis måste detta vägas mot andra faktorer som negativt kan inverka, som trötthet och tappad fokus på uppgiften. Hur uppgiften i sig är konstruerad vad gäller längd på registreringar (*blocks*), antal stimuli, samt fördelningen mellan dessa (”jitter”) påverkar försökets totala *power*.

### **Generaliserbarhet – vilkas hjärnaktivitet är det som undersöks?**

En aspekt av den problematik som har diskuterats i detta kapitel handlar om hur generaliserbara resultat från fMRI-studier är. I likhet med vad som är fallet inom psykologisk forskning så rekryteras många deltagare till fMRI-stu-

dier från universitetspopulationen. Detta kan inverka på forskningsresultatens generaliserbarhet då denna student-subpopulation sannolikt skiljer sig på avgörande sätt från den större populationen som de ska representera (t.ex. med avseende på utbildningsnivå). I mina egna studier som är inriktade på den åldrande hjärnan kan en likartad problematik skönjas. Det är inte ovanligt att de äldre deltagare som inkluderas i fMRI-studier är hårt selekterade vad gäller hälsostatus (t.ex. genom att inte inkludera äldre med förhöjt blodtryck vilket direkt medför en selektionseffekt), men ändå görs ofta generaliseringar till ”den åldrande populationen”.

En relaterad problematik i studier av hur åldrande påverkar hjärnaktiviteten är att slutsatser vanligtvis baseras på tvärsnittjämförelser mellan yngre och äldre individer – inte på longitudinella studier av hur hjärnaktivitet förändras inom personer över tid då de åldras. Vi har visat att longitudinella observationer avsevärt kan avvika från tvärsnittsestimat, exempelvis genom att de senare indikerar att frontala regioners hjärnaktivitet är högre hos de äldre än de yngre individerna. Longitudinella fynd tyder på att denna gruppskillnad avspeglar en selektion av högfungerande äldre deltagare som sannolikt alltid har uppvisat stark hjärnaktivering – men longitudinellt reduceras aktiviteten även för dessa<sup>17</sup>.

I sammanhanget longitudinella studier kan även noteras att det i fMRI-studier finns utvecklingspotential vad gäller de statistiska modellerna, vilka vanligen begränsas till de deltagare som återkommer till förnyad testning. Denna praxis bygger, explicit eller implicit, på antagandet att bortfall från studier är slumpmässigt. I realiteten är det däremot ofta så att bortfallet är selektivt. Statistiska metoder för att kunna hantera bortfall i fMRI-analyser är under utveckling, och det har börjat komma studier som beaktar denna viktiga dimension<sup>18</sup>. Detta kommer högst sannolikt ha positiva konsekvenser för reproducerbarhet och generaliserbarhet inom fMRI-området.

## Uppsummering och en blick framåt

I kapitlet har jag diskuterat några av de många faktorer som kan påverka graden av reproducerbarhet i fMRI-studier. De olika faktorerna kan sägas påverka olika *dimensioner* av reproducerbarhet<sup>13</sup>. *Metod*-reproducerbarhet utvärderas i fall då ett dataset från en publicerad studie rekvideras av en oberoende forskargrupp som söker återskapa de ursprungliga fynden genom att tillämpa de analysmetoder som beskrivs i den publicerade artikeln. Det

kan tyckas som en självklarhet att metodreproducerbarhet då uppnås, men om inte de olika analysstegen och valen som diskuterades ovan är beskrivna på ett uttömmande sätt så är risken stor för att fullständig metodreproducerbarhet inte uppnås. *Resultat*-reproducerbarhet handlar om att söka upprepa en studie, inklusive genomförande av en ny datainsamling som självklart ska efterlikna den ursprungliga i alla väsentligheter. Flera av ovan diskuterade faktorer kan påverka denna dimension, såsom stickprovets komposition och storlek, de funktionella uppgifternas reliabilitet och komplexitet, samt variationer i deltagarnas tillstånd (states) samt i mätapparaturen. En tredje dimension, *inferens*-reproducerbarhet är kanske den mest komplexa och handlar om konsensus vad gäller tolkningen av vad ett visst aktiveringsmönster avspeglar. En faktor som är relevant för denna dimension är den indirekta mätning av hjärnaktivitet som fMRI baseras på. Det medför exempelvis att där en forskargrupp tolkar en skillnad i fMRI-signal mellan två grupper som att de aktiverar olika kognitiva strategier, skulle en annan grupp kunna föreslå att den påvisade skillnaden är ett tecken på vasculära olikheter. Mer generellt belyser frågan om *inferens*-reproducerbarhet den stora utmaningen att förstå vad hjärnaktiveringar ”betyder”. Denna svårighet diskuteras ibland i relation till problemet med ”*reverse inference*”, som kan förklaras med exemplet att, även om rörelse på ena handens fingrar utlöser fMRI-aktivitet i vissa motoriska delar av hjärnan, så gäller inte nödvändigtvis den motsatta slutsatsen (observation av fMRI-aktivering i motoriska delar *kan* avspegla handrörelse men andra tolkningar är också möjliga såsom inre föreställning av rörelse utan faktisk rörelse).

Jag tycker att framtiden ser ljus(are) ut vad gäller de frågor som har diskuterats i kapitlet och i hela den här volymen. För det första är det viktigt i sig att frågorna lyfts upp och diskuteras i olika forum. Inom kognitiv neurovetenskap skickades nyligen viktiga signaler genom att redaktörer för viktiga tidskrifter, *Nature Neuroscience* och *Nature Communications*, på ett konstruktivt sätt dryftat dessa frågor<sup>19</sup>. Detta handlar om att synliggöra, att reliabiliteten för fMRI-fynd i många fall är tämligen modest<sup>20</sup>, men även om att betona, att många fynd faktiskt är reproducerbara och i linje med observationer från mer direkta mätningar av hjärnaktivitet<sup>19</sup>. Detta är bra nyheter för forskare, men även för försöken att praktiskt använda fMRI-tekniken, exempelvis för diagnos eller utvärdering av kliniska interventioner. Vidare har den aktuella debatten en framåtsyftande karaktär, och betonar hur forskare, granskare, redaktörer och andra inblandade ytterligare kan bidra till en positiv ut-

veckling vad gäller reproducerbarhet genom att ställa krav på transparent redovisning av metodologiska och statistiska detaljer – inklusive uppgifter om test-retest reliabilitet i den specifika studien. En annan positiv trend är att betona vinsterna med att göra data tillgängliga för andra, vilket möjliggör insyn och försök till reproducering, samt även aggregering av flera småskaliga dataset för att öka den statistiska styrkan. Genom att aktivt beakta de faktorer som har diskuterats i detta kapitel och fortsätta att lära mer om den fortfarande ”ungdomliga” fMRI-tekniken (den har ännu inte passerat 30 år!) så kommer fMRI även framgent ge oss unika – och reproducerbara – insikter i den komplexa relationen mellan mentala förmågor och hjärnans funktionella organisation.

### Referenser

1. Lauterbur, P.C. (1973). Image formation by induced local interactions: Examples employing nuclear magnetic resonance imaging. *Nature*, 242, 190-191.
2. Kwong, K.K., Belliveau, J.W., Chesler, D.A., et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences, USA*, 89, 5675-5679.
3. Huettel, S.A., Song, A.W., & McCarthy, G. (2004). *Functional Magnetic Resonance Imaging*. Sunderland, MA: Sinauer.
4. Salami, A., Eriksson, J., & Nyberg, L. (2012). Opposing effects of aging on large-scale brain systems for memory encoding and cognitive control. *Journal of Neuroscience*, 23, 10749-757.
5. Sohn, M-H., Goode, A., Koedinger, K.R. et al. (2004). Behavioral equivalence, but not neural equivalence – neural evidence of alterantive strategies in mathematical thinking. *Nature Neuroscience*, 7, 1193-1194.
6. Evans, A.C., Janke, A.L., Collins, D.L., & Baillet, S. (2012). Brain templates and atlases. *NeuroImage*, 62, 911-922.
7. Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, 12, 1-47.
8. Kim, H. J., Yoon, D.Y., Kim, E.S., et al. (2016). The 100 most-cited articles in neuroimaging: A bibliometric analysis. *NeuroImage*, 139, 149-156.
9. Braga, R.M. & Buckner, R.L. (2017). Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron*, 95, 457471.
10. Chein, J.M. & Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, 25, 607-623.



11. Salmi, J., Nyberg, L., & Laine, M. (2018). Working memory training mostly engages general-purpose large-scale networks for learning. *Neuroscience and Biobehavioral Reviews*, 93, 108-122.
12. Nyberg, L., Tulving, E., Habib, R., et al. (1995). Functional brain maps of retrieval mode and recovery of episodic information. *NeuroReport*, 7, 249-252.
13. Gilmore, R.O., Diaz, M.T., Wyble, B.A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. *Ann. N.Y. Acad. Sci.*, 1396, 5-18.
14. Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences, USA*, 113, 7900-7905.
15. Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., & Baker, C.I. (2009). Circular analysis in systems neuroscience – the dangers of double dipping. *Nature Neuroscience*, 12, 535-540.
16. Poldrack, R.A., Baker, C.I., Durnez, J. et al. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, 18, 115-126.
17. Nyberg, L., Salami, A., Andersson, M., et al. (2010). Longitudinal evidence for diminished frontal-cortex function in aging. *Proceedings of the National Academy of Sciences, USA*, 107, 22682-22686.
18. Nyberg, L., Pudas, S. & Lundquist, A. (2017). Longitudinal structural and functional brain imaging. In *Cognitive Neuroscience of Aging* (2nd ed.): by R Cabeza, L. Nyberg, D. Park (Eds.), Oxford University Press.
19. Editorial (2017). Fostering reproducible fMRI research. *Nature Neuroscience*, 20, 298.
20. Bennett, C.M. & Miller, M.B. (2010). How reliable are the results from functional magnetic resonance imaging? *Ann. N.Y. Acad. Sci.*, 1191, 133-155.



*Jan-Erik Gustafsson*

## Begrepps referensgeneralitet och datas inferensnivå som determinanter för möjligheter till generalisering

### **Inledning**

Den pedagogiska forskningen har såväl nationellt som internationellt rötter i psykologin och då i synnerhet i utvecklings-, inlärnings- och differentialpsykologi. Fram till mitten av 1950-talet var pedagogik och psykologi i Sverige ett gemensamt ämne, men kom då att delas och separata professurer inrättades i pedagogik och psykologi.

Fram till 1970-talet var kvantitativa forskningsansatser dominerande inom det pedagogiska forskningsfältet i Sverige. Forskningen ställde ofta kausala frågor, t.ex. om lämplig tidpunkt för organisatorisk differentiering av utbildningen, och i synnerhet i Göteborg var mycket forskning inriktad på generaliserbara beskrivningar baserade på stora datamaterial (Härnqvist 1987). Då, liksom nu, fanns det dock hot mot möjligheten att nå generaliserbara resultat.

Cronbach (1957) pekade i en "Presidential address" till American Educational Research Association på att forskningen i större utsträckning bör uppmärksamma att olika interventioner och behandlingar har olika effekt på olika individer, genom att systematiskt undersöka interaktionseffekter mellan individ- och behandlingsvariabler. Denna uppmaning hade viss effekt, och inte minst startade Cronbach och hans medarbetare systematiska forskningsprogram med sådan inriktning. Cronbach (1975) kunde presentera en rad intressanta fynd, men huvudslutsatsen var ändå negativ. Forskningen hade nämligen också visat på begränsningar i möjligheten att generalisera resultaten på grund av att de intressanta interaktionseffekterna i sin tur var involverade i interaktioner av högre ordning. Cronbach (1975) drog därför slutsatsen att samhälls- och beteendevetenskaplig forskning inte skulle kunna göra generaliseringar som är hållbara utanför laboratoriemiljöer, och han menade också att intresset i mindre grad borde riktas mot hypotesprövning och i större utsträckning fokuseras på lokal, kontextuell, tolkning av forskningsresultat.

Dessa slutsatser kan ses som uttryck för en tidig generaliserbarhetskris, men de kom inte att få något större genomslag. En anledning till detta kan vara att ungefär samtidigt som Cronbach (1975) lade fram sina pessimistiska slutsatser presenterade Glass (1976) meta-analys som en teknik för att göra kvantitativa generaliseringar baserade på ett stort antal undersökningar, och där flera konkreta exempel övertygade om att detta var en fruktbar ansats.

Med början under 1970-talet kom den svenska pedagogiska forskningen att genomgå dramatiska förändringar när det gällde val av metod och frågeställningar. Kvalitativa ansatser kom successivt att bli i det närmaste allena-rådande. Samma tendens fanns inom andra samhälls- och beteendevetenskapliga ämnen såväl i Sverige som i andra länder, men det torde finnas få andra exempel på en lika dramatisk förändring som den som den svenska pedagogiska forskningen genomgått. Det finns säkert många förklaringar till detta men jag dristar mig till att som en partiell förklaring framföra att ämnets inomvetenskapliga stadga var svag, dels som en följd av behovet att frigöra sig från psykologins individorienterade metodarsenal (Lundgren 2018), dels som en följd av att mycket av verksamheten kom att fokuseras på utrednings-, utvecklings- och implementeringsarbete i anslutning till grundskolereformen och andra förändringar på utbildningsområdet (Härnqvist 1987). Finansieringen av denna verksamhet kom i huvudsak från statliga verk och myndigheter, medan finansieringen av grundforskning från forskningsråd och fakultetsanslag var blygsam. En följd av detta var att pedagogers deltagande i den internationella vetenskapliga diskursen var av blygsam omfattning och att de problem som sysselsatte såväl seniora som juniora forskare inte ställde krav på forskningsmetoder som utvecklats för att stödja generaliseringar; istället kunde de med fördel angripas med mjukare metoder. Den kompetens som tidigare byggts upp vad gäller kvantitativa metoder vittrade därför delvis bort, och nyrekryteringen till området kom att begränsas.

Samtidigt finns det anledning att understryka att distinktionen mellan kvalitativa och kvantitativa forskningsmetoder ofta är oklar (Åsberg 2001). En avgörande skillnad är dock att medan den kvantitativa forskningsmetodiken utvecklats med huvudsyftet att stödja möjligheterna att göra generaliseringar, gäller detta i mindre utsträckning för de kvalitativa ansatserna.

Syftet med detta kapitel är att peka på några grundläggande principer som har relevans för möjligheten att göra generaliseringar. De exempel och frågor jag tar upp till diskussion emanerar i stor utsträckning från egna forsknings-erfarenheter, och har formen av personliga reflexioner.

## Ett exempel: kunskapsbedömning i kvalitativa och kvantitativa termer

Jag tar utgångspunkt i ett konkret exempel. Schoultz, Säljö & Wyndham (2001) ifrågasatte korrektheten i den internationella undersökningen Third International Mathematics and Science Study (TIMSS 1995) (Beaton et al. 1996), bland annat därför att man där undersöker kunskaper inom matematik och naturvetenskap med papper-och-penna uppgifter. De argumenterade för ett synsätt som innebär att prestation skapas i en kommunikativ praktik, snarare än att prestation är ett uttryck för elevernas förmågor och kunskaper.

Schoultz et al. (2001) valde ut två uppgifter från TIMSS 1995 för studium i en intervjuundersökning som omfattade 25 svenska elever i åk 7. En av uppgifterna var en optikuppgift. I uppgiften visas två ficklampor, en med reflektor och en utan, och uppgiften var att ange vilken av ficklamporna som ger mest ljus på en vägg 5 meter bort. Ett öppet svar krävdes, och för att svaret skulle bedömas som korrekt måste det innehålla en förklaring som framhöll att reflektorn fokuserar ljuset på väggen.

I den svenska TIMSS-undersökningen var det bara 39 % av eleverna som gav det korrekta svaret men i intervjuundersökningen gav 66 % av eleverna det rätta svaret. Även om intervjuundersökningen endast omfattade 25 icke slumpmässigt utvalda elever var denna skillnad så stor att den inte torde ha uppkommit av en slump. Enligt den tolkning Schoultz et al. (2001) gjorde på grundval av intervjumaterialet berodde det bättre resultatet i den kvalitativa undersökningen huvudsakligen på det stöd som intervjuaren gav i en sokratisk dialog. Schoultz et al. (2001, s. 234) drog slutsatsen att: "Knowing is in context and relative to circumstance. This would seem an important premise to keep in mind when discussing the outcomes of psychometric exercises."

Detta framstår måhända som en förödande kritik, inte endast av TIMSS, utan också av de resultat som papper-och-penna uppgifter i allmänhet kan ge. Det är dock tveksamt om man kan betrakta denna undersökning som en studie av validiteten i TIMSS. Schoultz et al. (2001) tolkar resultaten på så sätt att prestationsskillnader mellan olika kontexter är absoluta, och att de bättre prestationerna i intervjusituationen visar på en högre kunskapsnivå och en högre grad av insikt. I TIMSS ser man i stället prestationsskillnaderna som relativa, därför att de observerade prestationsskillnaderna ses som bestämda inte endast av elevens förmåga, utan också av uppgiftens svårighetsgrad. Enligt detta betraktelsesätt kan man förklara skillnaden i resultat mellan de två situationerna med hänvisning till att uppgiften är lättare i intervjusituationen

än i papper-och-penna situationen. Med hjälp av så kallad ”item-response theory” (se t.ex. Embretson & Reise 2013) är det också med utgångspunkt i stora datamaterial, som t.ex. TIMSS, möjligt att fastställa den relativa svårighetsgraden hos olika uppgifter, liksom olika elevers förmåga. Även om man skattar uppgifternas relativa svårighetsgrad från datamaterial där eleverna har olika prestationsnivåer erhålls samma resultat, inom ramar bestämda av slumpen, och på samma sätt är det möjligt att erhålla samma skattningar av individers förmåga från svar på olika uppsättningar uppgifter. En styrka med den kvantitativa, storskaliga, ansatsen är sålunda att den ger möjlighet till generalisering över såväl urval av uppgifter som urval av elever. Det faktum att det är möjligt att i en intervjusituation påvisa att dialogen kan påverka elevens tänkande och svarande utgör inte något argument mot användning av en psykometrisk ansats.

### **Kvalitativ och kvantitativ forskning**

Ercikan och Roth (2006) gör ett försök att fördjupa analysen av innebörden i tudelningen mellan kvantitativa och kvalitativa metoder. För det första pekar de på att alla fenomen och all kunskap samtidigt har både kvantitativa och kvalitativa aspekter. Exempelvis baseras det som brukar betecknas som kvantitativ forskning normalt på kvalitativa distinktioner vad gäller konceptualisering, utveckling av instrument för datainsamling, och kodning av observationer och elevsvar. För det andra menar de att den distinktion mellan subjektivitet och objektivitet som ofta associeras med kvalitativ respektive kvantitativ forskning varken är korrekt eller användbar. För det tredje pekar Ercikan och Roth (2006 s. 19) på att inte endast kvantitativ forskning utan även en del kvalitativ forskning syftar till generaliseringar. Som ett exempel på detta nämner de fenomenologiska ansatser för att förstå perceptionspsykologiska mekanismer baserade på intensivstudier av en person.

Ercikan och Roth (2006) föreslog att den dikotoma distinktionen mellan kvalitativ och kvantitativ forskning bör ersättas med en kontinuerlig skala där olika former av forskning kan placeras och där den ena ändpunkten representeras av forskning som utgår från och försöker beskriva och förstå människors levda erfarenhet, medan den andra ändpunkten representerar forskning som söker identifiera abstraherade och idealiserade mönster för mänsklig erfarenhet. Ercikan och Roth betecknade den första ändpunkten som forskning på låg inferensnivå, och den andra ändpunkten som forskning på hög

inferensnivå. Undersökningar på låg inferensnivå karakteriseras av fokus på det speciella och oförutsedda, de påverkas av sammanhanget och kännetecknas av konkretisering och närvaro. Forskning på hög inferensnivå däremot karakteriseras av standardisering, universalitet, avstånd och abstraktion. Forskning på låg inferensnivå kan använda både tekniker som traditionellt betecknas som kvalitativa (t.ex. observation, intervju) och kvantitativa (t.ex. deskriptiv statistik), vilket också gäller höginferensforskningen. Ercikan och Roth (2006) betonar också att forskning kan befinna sig längs hela skalan från låg till hög inferens liksom att forskningsfrågan skall styra valet av metod.

### **En metafor för låg- och höginferensforskning: väder och klimat**

I syfte att ge en mer konkret innebörd åt skalans ändpunkter presenteras nedan en metafor. Vädret påverkar våra dagliga liv i hög grad, genom att ge förutsättningar för aktiviteter både i arbetsliv och på fritid. Dock kan vi inte göra något åt vädret, utan vi får anpassa oss till det, bland annat genom val av aktiviteter och klädsel. I dessa val har vi glädje av att det på kort sikt är möjligt att förutsäga vädret, men över längre tid än ungefär en vecka är vädret oförutsägbart.

Klimat är generaliserat väder över en längre tid. Forskare definierar klimat som aggregerade aspekter av väder, med hjälp av indikatorer som genomsnittlig temperatur och genomsnittlig nederbörd. Medan väder är oförutsägbart på längre sikt, är klimat och klimatförändringar förhållandevis stabila fenomen, som det är meningsfullt att försöka förstå teoretiskt och för vilka empiriskt baserade modeller som syftar till att förutsäga den långsiktiga utvecklingen kan konstrueras. I termer av denna metafor handlar höginferensforskning om klimat, medan låginferensforskning handlar om väder.

### **Aggregering som en grund för höginferensforskning**

Klimatforskningen bygger på en högt utvecklad teknik för att generera och samla in data, och på överenskomna definitioner och analysmodeller. Men den grundläggande idén är att aggregera multipla observationer av olika väderaspekter. På samma sätt bygger utbildningsforskning som syftar till hög inferens på aggregering av observationer av olika aspekter av undervisning och lärande.

Vi kan skilja mellan två typer av aggregering. Den första innebär att aggregering sker över observationsenheter, som elever, klasser, skolor, kommuner och skolsystem. Detta är vad som brukar kallas statistisk aggregering. Den andra typen av aggregering sker över olika observationer för samma enhet, till exempel när svar på flera provuppgifter kombineras till en total poäng. En annan term för detta är mätning.

Det finns delade meningar om aggregering är bra eller dåligt. Yanchar and Williams (2006) hävdade att:

*... data aggregation and accompanying statistical tests often hide qualitative patterns and lead to excessively abstract or artificial conclusions ... ; statistical indices are often used as facile substitutes for careful interpretation and human judgment ... patterns in aggregate data are erroneously used to make inferences about the structure of psychological processes in individuals ... (s. 6).*

Men det är också möjligt att vända på resonemanget, och man kan hävda att för att de generella aspekterna ska kunna framträda är det nödvändigt att bli av med det som är specifikt och unikt, och aggregering av observationer är ett sätt att åstadkomma detta. Det går därför att hävda att metoder som döljer kontextberoende variation har styrkor snarare än svagheter när syftet är att undersöka generella mönster och relationer. Aggregering kan således vara både bra och dåligt, beroende på syftet med forskningen.

## Mätning

Inom samhälls- och beteendevetenskap finns det stora grupper av forskare som avvisar tanken att det är möjligt att på ett meningsfullt sätt mäta abstrakta begrepp som läsförmåga, intelligens, introversion och undervisningskvalitet. Här finns dock stora skillnader mellan olika samhälls- och beteendevetenskapliga ämnen. Medan forskare inom psykologi utan att tveka använder sig av olika mätmetoder, utövar de flesta forskare inom pedagogikfältet aktivt motstånd mot försök att kvantifiera kunskaper, färdigheter och elevgenskaper som begåvning.

Beteendevetenskapliga mätningar bygger på principen att sammanläggning av flera observationer som är påverkade av olika slumpfaktorer leder till stabilare och tillförlitligare mätvärden. Denna princip uttrycktes i en formel redan i början av förra seklet, då Spearman (1910) och Brown (1910) oberoende av varandra formulerade vad som senare skulle komma att betecknas

som ”The Spearman-Brown prophecy formula” vilken uttrycker hur mycket mer tillförlitligt ett mätinstrument kan förväntas bli då det förlängs ett visst antal gånger. Tillförlitlighet (eller reliabilitet) är dock endast en aspekt av mätningars kvalitet. En annan aspekt är att instrumentet faktiskt fångar det vi har för avsikt att mäta, eller att dess validitet är god. Om mätinstrumentet påverkas av irrelevanta faktorer försämrar detta validiteten. Exempelvis kan ett hörförståelseprov i engelska där eleverna skall producera omfattande skriftliga svar förväntas mäta en blandning av skrivförmåga och engelsk hörförståelse, där systematiskt inflytandet av skrivförmåga försämrar provets validitet som ett mått på hörförståelse. Ett annat vanligt förekommande validitetshot är att mätinstrumentet endast täcker delar av det begrepp det egentligen är avsett att mäta. En vanlig orsak till detta är att provet huvudsakligen innehåller uppgifter som fångar det som är lätt mätbart men i mindre utsträckning omfattar de mer svårfångade aspekterna av begreppet.

Ytterligare en aspekt som påverkar instrumentets reliabilitet och validitet är karaktären på det begrepp som vi önskar mäta och då framför allt med avseende på generalitetsgrad (”referent generality”, Coan 1964). Vissa begrepp omfattar ett brett spektrum av fenomen (t.ex. allmän kognitiv förmåga, politiskt deltagande), medan andra begrepp omfattar en mer avgränsad domän av fenomen (t.ex. spatial förmåga, deltagande i val).

Frågan om referensgeneralitet skapar både teoretiska och empiriska utmaningar eftersom denna egenskap hos begrepp är svårfångad och i hög grad kontextuellt betingad. Ett konkret exempel kan tydliggöra detta. Det finns en omfattande forskning kring elevers självbedömda kompetens, och hur denna ömsesidigt relaterar till skolprestation. Här finns flera forskningslinjer där en huvudskillnad går mellan två olika sätt att beskriva självuppfattad kompetens. Den ena använder sig av begreppet ”academic self-concept” (akademisk självuppfattning) (Shavelson, Hubner & Stanton 1976) och som avser mer allmänna aspekter, som hur man uppfattar graden av duktighet i olika skolämnen som svenska, engelska och matematik. Den andra använder sig av begreppet ”self-efficacy” (Bandura 1997), vilket möjligen kan översättas med ”självförtroende” och som avser hur man bedömer den egna förmågan att lösa olika konkreta uppgifter, som att skriva ett brev på engelska eller genomföra addition. Akademisk självuppfattning avser sålunda mer generella prestationsaspekter, och involverar ofta implicita eller explicita jämförelser med andra, medan uppgifter som mäter självförtroende är domänspecifika och avser förmåga att klara av specifika uppgifter. Begreppet akademisk självuppfatt-



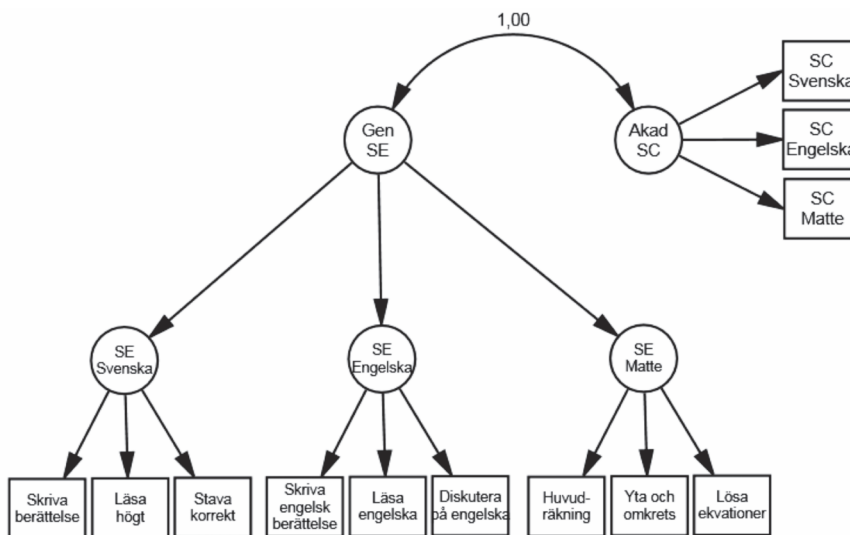
ning har sålunda högre referensgeneralitet än begreppet självtilltro, och denna skillnad har betydande teoretiska implikationer vad avser uppkomstdeterminanter och konsekvenser av skillnader i dessa två former av självbedömd kompetens (Bong & Skaalvik 2003).

Samtidigt förhåller det sig så att mått på "self-efficacy" inom olika domäner (t. ex. matematik, svenska och engelska) i allmänhet har ganska betydande inbördes positiva samband, vilket pekar på att det finns faktorer utöver den domänspecifika variationen som påverkar elevernas bedömningar. Med hjälp av statistiska metoder för multivariat analys (s.k. faktoranalys, se Jöreskog 1969) är det möjligt att formulera modeller som förklarar samvariation mellan variabler i termer av icke observerbara latent variabler. Om vi exempelvis har mätt självtilltro inom tre ämnesområden med ett antal frågor inom varje område kommer vi att kunna identifiera tre korrelerade faktorer, som svarar mot de tre områdena och som förklarar samvariationen mellan frågorna. Men vi kan också föra analysen ett steg vidare genom att analysera samvariationen mellan de tre faktorerna för att identifiera en s.k. andra-ordningens faktor, som representerar det som är gemensamt mellan de tre faktorerna av första ordningen. Faktorn av andra ordningen ger uttryck för vad som kan betecknas som en domänöverskridande eller generell självtilltro, vilket enligt Banduras (1997) teori om "self-efficacy" knappast är en möjlig konstruktion. Men om vi relaterar en akademisk självuppfattningsfaktor mätt med frågor om självbedömd kompetens inom ämnena svenska, engelska och matematik, till faktorn som uttrycker generell självtilltro kommer vi att finna att dessa har ett i det närmaste perfekt samband. Detta har sin grund i att begreppet generell akademisk självtilltro har samma höga referensgeneralitet som begreppet akademisk självuppfattning och att båda refererar till i huvudsak samma domän.

Den modell med tre faktorer av första ordningen och en faktor av andra ordningen som vi skisserat för akademisk självtilltro (se Figur 1) är ett exempel på en s.k. hierarkisk mätmodell, som kännetecknas av att de inkluderar både breda dimensioner med hög referensgeneralitet och smala dimensioner med låg referensgeneralitet. Denna typ av modeller har under senare decennier kommit att få allt större betydelse inom flera olika forskningsområden som personlighetspsykologi och forskning kring kognitiva förmågor.

Under 1970-talet kom personlighetsforskningens fokus på stabila personlighetsegenskaper att ifrågasättas därför att kritiker (t.ex. Endler & Magnusson 1976) menade att grundantagandet om beteendekonsistens över situatio-





Figur 1. Figuren presenterar en hypotetisk hierarkisk modell över "self-efficacy" (SE). På den lägsta nivån finns tre domänspecifika faktorer av första ordningen (SE för svenska, engelska och matematik) som mäts med självbedömning av förmåga att lösa olika uppgifter.

På nivån ovanför återfinns en generell akademisk SE dimension, som representeras av en andra ordningens faktor, med de tre domänspecifika faktorerna som indikatorer. I modellen finns också en faktor som representerar akademisk självuppfattning (Akad SC), vilken mäts med självbedömningar av grad av duktighet i olika ämnen.

I modellen antas både den generella akademiska SE faktorn och den generella akademiska SC faktorn ha ungefär samma höga referensgeneralitet, med en förväntad hög korrelation mellan faktorerna. De domänspecifika SE faktorerna, liksom bedömningarna av akademisk självuppfattning har lägre referensgeneralitet.

ner inte var uppfyllt. I stället förespråkades en interaktionell modell där beteendet ses som bestämt av en kontinuerlig och ömsesidig interaktion mellan person- och situationsvariabler. Individens grad av beteendekonsistens över olika situationer är dock olika för begrepp av olika generalitetsgrad. Observationer som avser begrepp med hög generalitetsgrad tenderar att uppvisa hög grad av stabilitet över olika mätinstrument och mättillfällen, medan observationer som avser begrepp med låg referensgeneralitet tenderar att vara mindre stabila. Under de senaste decennierna har därför mycket av forskningen tenderat att fokusera på mätning av begrepp med hög referensgene-

ralitet. Detta har i allmänhet skett inom ramen för teoretiska modeller som postulerar begrepp med olika generalitetsgrad, från mycket generella och inkluderande till mycket specifika och avgränsade. Den nu dominerande personlighetsmodellen är en hierarkisk modell med fem breda personlighetsdimensioner högst upp, under vilka det sedan finns smala ”facetter” som avser mer specifika personlighetsdimensioner (se t.ex. Digman 1990).

Forskningen kring kognitiva förmågor dominerades under de första decennierna av 1900-talet av Spearman's (1904) g-faktor modell, enligt vilken en enda förmågedimension är tillräcklig för att förklara individuella differenser i kognitiv förmåga. Med publiceringen av en alternativ modell som omfattade sju primära mentala förmågor (Thurstone 1938) kom Spearman's modell dock att förlora sin ledande ställning. Utvecklingen under de följande decennierna kännetecknades av en alltmer differentierad beskrivning av smala förmågedimensioner, vilka dock inte gav några större bidrag till vare sig teoriutvecklingen eller förbättrade praktiska tillämpningar. Bland annat genom en mycket omfattande reanalys och meta-analys av i stort sett all tidigare forskning kring de kognitiva förmågornas struktur (Carroll 1993) har dock en förhållandevis bred konsensus etablerats kring en modell med tre nivåer. På den lägsta nivån återfinns ett stort antal smala kognitiva förmågor; på den andra nivån ett dussintal breda förmågor, varav flera identifierats tidigare (Horn & Cattell 1996); och på den tredje nivån en enda generell faktor i linje med Spearman's g-faktor. Denna modell har bidragit till att fokus i forskningen har förskjutits mot de högre nivåerna i modellen där förmågebegreppen har högre referensgeneralitet (Gustafsson 1984, 2001).

### **Aggregering av data**

År 1958 bildades “The International Association for the Evaluation of Educational Achievement” (IEA) som en sammanslutning av forskare intresserade av att undersöka möjligheterna att göra jämförande studier av elevers kunskaper och färdigheter inom olika områden. Den första studien omfattade ett 20-tal länder och undersökte matematikkunskaper på tre olika stadier (Husén 1967). Undersökningen var en framgång i det att den visade att det faktiskt går att praktiskt genomföra komplexa empiriska undersökningar i många länder med meningsfulla resultat. Under de följande decennierna genomfördes flera undersökningar inom olika ämnesområden, men det stora genomslaget för IEA-undersökningarna kom först med ”Trends in Interna-

tional Mathematics and Science Study” (TIMSS) år 1995 med ett stort antal deltagande länder och förbättrade metoder för urval, mätning, analys och rapportering. Inte minst var denna studie betydelsefull genom att den var utgångspunkt för regelbundna mätningar vart fjärde år, med möjlighet att undersöka förändringar i kunskapsnivåer över tid. Det riktigt stora genomslaget för internationella kunskapsmätningar kom dock med Programme for International Student Assessment (PISA) vars första undersökning genomfördes 2001. Ansvaret för PISA ligger på OECD, som med milt tvång fått samtliga medlemsländer att delta, och även lockat många icke-OECD länder att medverka. Metodmässigt finns stora likheter mellan TIMSS och PISA, men en viktig skillnad är att i PISA är avsikten att mäta kunskaper och färdigheter inom matematik, naturvetenskap och läsning som är viktiga efter avslutad skolgång, medan TIMSS fokuserar på de kunskaper och färdigheter som anges i de medverkande ländernas läroplaner.

I de internationella studierna av kunskaper och färdigheter aggregeras resultaten för de deltagande ländernas urval av elever till nationell nivå och undersökningarna har visat på mycket stora skillnader mellan olika länders kunskapsresultat. De bästa prestationerna återfinns oftast i de östasiatiska länderna, (t.ex. Korea, Japan, Hongkong, Singapore och Taiwan), även om det finns undantag (t.ex. Finland), vilket rest frågor vilka faktorer som kan förklara detta. Den pedagogisk-psykologiska forskningen har pekat på att positiva attityder till lärande och hög uppfattning av den egna kompetensen påverkar motivationen positivt, vilket i sin tur leder till förbättrad prestation (t.ex. Bandura 1997; Marsh & Hau 2004). Mer ingående analyser av data från de internationella studierna ger dock inte stöd för sådana förklaringar av skillnaderna mellan länderna. De länder som har högst medelvärden på frågor till eleverna som avser akademisk självuppfattning, motivation och intresse för studier tenderar nämligen att ha låga medelvärden på de internationella kunskapsproven (se t.ex. Min, Cortina & Miller 2016). Om i stället sambandet mellan resultat på kunskapsproven och elevernas svar på frågor om självuppfattning, motivation och intresse beräknas inom varje land för sig framträder dock de förväntade positiva sambanden (Min et al. 2016). Vi står sålunda inför paradoxen att sambandet mellan kunskapsresultat och självbedömningar av förmåga, motivation och intresse är negativt då det beräknas på landnivå, och positivt då det beräknas på individnivå. Detta empiriska fynd har benämnts ”the attitude-achievement” paradox.

Denna paradox innebär att de faktorer och mekanismer som förklarar skill-

nader i resultat mellan länder delvis skiljer sig från de faktorer och mekanismer som förklarar skillnader i resultat mellan elever, klasser och skolor inom länder. Det finns flera tänkbara anledningar till detta. En är att vissa faktorer som varierar mellan länder är konstanta inom länder. Ålder för skolstart är ett exempel på detta, vilket har sin grund i att många länder har regelverk för skolstart som bygger på födelseår, med en variation mellan fyra år och sju år.

Undersökningsprocedurernas kvalitet påverkar också resultaten för de olika länderna, och eftersom det finns många potentiella hot mot korrekt genomförande av de internationella studierna kan de resultat som observeras avvika från de faktiska resultaten. Språkliga och kulturella skillnader mellan länder är andra faktorer som kan orsaka skillnader i svarsmönster. Det är också i denna senare kategori av faktorer som man framförallt sökt finna förklaringar till den ovan beskrivna paradoxen.

I de av konfucianismen präglade östasiatiska kulturerna finns normer om personlig återhållsamhet när det gäller att presentera den egna personen (Chen, Lee, & Stevenson 1995). Detta skapar en tendens hos asiatiska studenter att undvika extrema kategorier i den positiva änden av skalan, och att i stället välja kategorier närmare skalans mitt. Denna hypotes om ”modesty bias” har fått stöd i forskningen som åtminstone en av flera förklaringsfaktorer (Min et al. 2016). Intressant nog finns det resultat som pekar på att även elever i Finland, som också de uppvisar mycket goda prestationer i de internationella undersökningarna, tenderar att undvika de mest extrema svarskategorierna (Scherer & Gustafsson 2015). En annan förklaring till ”attitude-achievement” paradoxen fokuserar på det höga prestationstryck som kännetecknar de asiatiska utbildningssystemen, vilka tenderar att ha mer krävande läroplaner, högre satta akademiska normer och mer kännbara konsekvenser av såväl studieframgång som studiemisslyckande (Leung 2002, s. 106). Ytterligare en förklaring tar fasta på sociala jämförelser mellan elever och det många gånger replikerade resultatet att elever i en högpresterande klass eller skola tenderar att skatta den egna förmågan lägre än då klass- eller skolkamraterna uppvisar mindre goda akademiska resultat. Denna så kallade ”Big-Fish-in-Little-Pond” effekt (Marsh & Hau, 2004) är väl belagd då skillnaderna avser klasser och skolor, men det framstår inte som självklart att det går att överföra resonemanget till att gälla även skillnader mellan länder.

Det har länge funnits konsensus om att det är nödvändigt att ta hänsyn till att data från utbildningsområdet har en hierarkisk struktur, där elever finns

inom klasser, som finns inom skolor, som finns inom kommuner, som finns inom länder. Denna hierarkiska struktur skapar beroenden mellan observationer som hör till samma enhet; t.ex. tenderar elever som går i samma klass att vara mer lika varandra i prestationshänseende än ett slumpmässigt urval av elever. I statistiska analyser är det nödvändigt att ta hänsyn till sådana beroenden för att erhålla korrekta resultat, och många programvaror gör detta mer eller mindre automatiskt. Men i många fall är det inte tillräckligt med sådana lösningar, utan det krävs teorier och metoder som medger att observationer på olika aggregationsnivåer mäter olika begrepp. Exemplet ovan visade sålunda att elevers självuppfattning av förmåga har positiva samband med elevresultat då analysen görs på elevnivå, men då analysen görs på data som aggregerats till klass-, skol- eller landsnivå påverkas innebörden i mätningarna av kulturellt bestämda svarsmönster, och kontextuella och sammanställningsmässiga förhållanden som påverkar sociala jämförelser.

## Diskussion och slutsatser

Syftet med detta kapitel är att peka på några grundläggande principer som har relevans för möjligheten att göra generaliseringar från empiriska undersökningar. En utgångspunkt ges av den kontinuerliga skala med forskning som ligger nära människors levda erfarenhet som en ändpunkt (låginferensforskning) och forskning som söker identifiera abstraherade mönster som den andra ändpunkten (höginferensforskning). Denna skala sammanfaller inte med distinktionen mellan kvalitativ och kvantitativ forskning, men det är dock uppenbart att det finns en tendens att låginferensforskningen i större utsträckning använder sig kvalitativa metoder, medan höginferensforskningen i större utsträckning förlitar sig på kvantitativa metoder. Anledningen till detta är att mycket av den kvantitativa metodarsenalen utvecklats i syfte att stödja generaliseringar, t.ex. i form av statistisk inferens från stickprov till population, slutsatser om kausala effekter av interventioner, eller mätegenskaper hos psykologiska instrument. Det framstår dock som angeläget att den framtida metodutvecklingen också fokuserar på de behov som forskning på lägre inferensnivå har, där utmaningen ofta är att hantera en stor mängd information om ett begränsat antal enheter.

I anslutning till metaforen väder och klimat försöker jag också konkretisera skillnaden mellan låg- och höginferensforskning. Även om klimat och klimatförändringar är högst reella och påtagliga fenomen bygger vår förstå-

else av dessa fenomen på teoretiska konstruktioner som prövas mot observationer som många gånger är både felbemängda och begränsade i såväl tid som rum. I detta liknar klimatforskningens utmaningar de problem som den empiriska samhälls- och beteendevetenskapliga forskningen ställs inför, och lösningarna torde också vara likartade, nämligen införskaffande av mer och bättre data och utveckling av bättre metoder för att ta tillvara informationen i data.

I kapitlet tar jag också upp begreppet referensgeneralitet, vilket avser omfattningen av den domän som ett begrepp refererar till. De fenomen vi undersöker är abstraktioner som exempelvis avser personliga egenskaper eller förmågor, och referensgeneralitet avser då skillnader i grad av specificitet eller generalitet mellan olika begrepp. Såväl teoretiskt som metodmässigt innebär detta stora utmaningar, men en lösning som i praktiken har gett goda resultat är att samtidigt använda mer generella och mer specifika begrepp inom ramen för hierarkiska modeller. Inom detta område finns dock en lång rad teoretiska och metodmässiga problem som återstår att reda ut (se Gustafsson 2001), men det skulle föra för långt att diskutera dessa här.

Robinson (1950) visade att samband mellan variabler som beräknats för data som aggregerats till delstatsnivå inte kan tolkas som uttryck för samband på individuell nivå, och på grundval av dessa resultat myntades uttrycket "ecological fallacy" för felaktiga slutsatser från analys av aggregerade data till individuella data. Insikten om denna typ av felslut har lett till att utbildningsforskare har varit försiktiga med att använda aggregerade data i sina analyser. Som påpekats av Lubinski och Humphreys (1996) ger dock analys av aggregerade data den relevanta informationen när det gäller utvärdering av policyförändringar, vilket pekar på att forskarna varit alltför restriktiva i sin användning av aggregerade data.

Utvecklingen av statistiska tekniker särskilt avsedda för analys av flernivådata (t.ex. Muthén 2004) i kombination med den dramatiska ökningen av tillgång till komparativa data från undersökningar som omfattar stora antal länder, har gett rika möjligheter att undersöka i vilken utsträckning relationer mellan variabler är invarianta över olika aggregationsnivåer. Detta har gett upphov till en rad intressanta fynd, som det inte ens varit möjligt att formulera hypoteser kring utan tillgång till detta breddade empiriska underlag, vilket också öppnar en lång rad intressanta möjligheter för fortsatt forskning.

## Referenser

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.
- Beaton, A.E., Martin, M.O. Mullis, I.V.S., Gonzalez, E.J., Smith, T.A. & Kelly, D.L. (1996). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1-40.
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296-322.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Chen, C., Lee, S-Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6(3), 170-175.
- Coan, R. W. (1964). Facts, factors, and artifacts: The quest for psychological meaning. *Psychological Review*, 71(2), 123.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12(11), 671-684.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30(2), 116-127.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41(1), 417-440.
- Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.
- Endler, N. S., & Magnusson, D. (1976). Toward an interactional psychology of personality. *Psychological Bulletin*, 83(5), 956.
- Ercikan, K., & Roth, W. M. (2006). What good is polarizing research into qualitative and quantitative? *Educational Researcher*, 35(5), 14-23.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8.
- Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8(3), 179-203.
- Gustafsson, J. E. (2001). Measurement from a hierarchical point of view. In H. L. Braun, D. N. Jackson, & D. E. Wiley (Eds.) *The role of constructs in psychological and educational measurement* (pp. 87-111). Routledge.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57(5), 253.
- Husén, T. (1967). *International Study of Achievement in Mathematics: A Comparison of Twelve Countries, Vol I & II*. Stockholm: Almqvist & Wiksell.
- Härnqvist, K. (1987) En SOU-forskarens hågkomster. *Forskning om Utbildning*, 14(1), 20-29.



- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183-202.
- Leung, F.K.S. (2002). Behind the high achievement of East Asian students. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 8(1), 87-108.
- Lubinski, D., & Humphreys, L. G. (1996). Seeing the forest from the trees: When predicting the behavior or status of groups, correlate means. *Psychology, Public Policy, and Law*, 2(2), 363-376.
- Lundgren, U. P. (2018). Pedagogisk forskning vid Göteborgs universitet 1963 – 1975. En personlig betraktelse. *Pedagogisk Forskning i Sverige*, 23(5), 99-117.
- Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, 96(1), 56.
- Min, I., Cortina, K. S., & Miller, K. F. (2016). Modesty bias and the Attitude-Achievement paradox across nations: A Reanalysis of TIMSS. *Learning and Individual Differences*, 51, 359-366.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22(3), 376-398.
- Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, 15(3): 351-357.
- Scherer, R., & Gustafsson, J. E. (2015). Student assessment of teaching as a source of information about aspects of teaching quality in multiple subject domains: An application of multilevel bifactor structural equation modeling. *Frontiers in Psychology*, 6, 1550.
- Schultz, J., Säljö, R., & Wyndhamn, J. (2001). Conceptual knowledge in talk and text: What does it take to understand a science question?. *Instructional Science*, 29(3), 213-236.
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46(3), 407-441.
- Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-292.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*.
- Yanchar, S. C., & Williams, D. D. (2006). Reconsidering the compatibility thesis and eclecticism: Five proposed guidelines for method use. *Educational Researcher*, 35(9), 3-12.
- Åsberg, R. (2001). Det finns inga kvalitativa metoder–och inga kvantitativa heller för den delen: Det kvalitativa-kvantitativa argumentets missvisande retorik. *Pedagogisk Forskning i Sverige*, 6(4), 270-270.



*Elisabet Engdahl*

## Sökandet efter en generell språk teori\*

Detta sökande har pågått länge. Så snart människor började reflektera över hur de talar, började de även fundera på vad som utmärker språket. Redan Aristoteles formulerade tankar kring vilka språkliga kategorier som behövs,<sup>1</sup> men jag ska inte börja med klassikerna utan med ett empiriskt faktum. Ett nyfött barn kan lära sig vilket språk som helst. Om barnet föds i Korea och omgivningen talar koreanska och interagerar med barnet på koreanska, så lär det sig det koreanska ljudsystemet, lexikonet och satsbildningsmönstren. Om barnet i stället föds i Sverige, eller tidigt flyttar till en svensktalande familj, så lär det sig svenska. Fort går det också.

Givet detta faktum har språkvetare under lång tid försökt komma fram till en teori för mänskligt språk, närmare bestämt en teori som är så generell att den kan förklara både de gemensamma drag som alla språk uppvisar och den variation som finns mellan olika språkfamiljer. I denna artikel tar jag först upp vilka gemensamma drag som utmärker naturliga språk. Därefter beskriver jag hur språktypologer arbetar för att identifiera generella ordföljdmönster och samband mellan dessa. Detta att ett barn mycket snabbt lär sig det språk som omgivningen använder tolkar vissa språkvetare som att barn har ett medfött språkanlag, en universell grammatik. Jag redogör för några försök att konkretisera och testa denna universella grammatik innan jag slutligen kort tar upp frågan om det går att härleda sådana generella egenskaper hos språket ur andra kognitiva processer.

### Gemensamma drag

Låt oss börja med att identifiera några egenskaper som alla naturliga språk uppvisar. De använder en liten uppsättning ljud, i svenska t.ex. *a*, *o*, *b*, *m* och *r*. Dessa ljud, som i sig inte har någon betydelse, kan kombineras till ett stort antal ord med betydelse, t.ex. *bar*, *bro*, *bor*, *om*, *mor* och *ram*, som i sin tur kan kombineras till ett stort antal fraser: *en stor bro*, *i baren*, *bor i Mora* etc. Sådana fraser kan sedan kombineras till ett obegränsat antal yttranden. Detta

att vi, givet ett begränsat antal språkljud, ändå kan uttrycka alla de olika meddelanden och nyanser vi vill förmedla, gör språket mycket effektivt.<sup>2</sup>

Om vi nu tittar närmare på vad dessa yttranden förmedlar kan vi urskilja fem olika typer av *språkhandlingar*, alltså saker som vi gör genom att säga, eller skriva, en viss sorts yttrande. Om vi vill berätta eller tala om något, använder vi ett *påstående* och om vi undrar över något kan vi ställa en *fråga*. Med hjälp av en *uppmaning* kan vi få någon att göra något. Vi kan även reagera på något, antingen genom ett värderande *utrop* eller en kort *interjektion*. I (1) ges exempel från svenska på dessa språkhandlingar.

- (1)
- a. *Påstående*: Min bror har byggt ett hus.
  - b. *Fråga*: Har din bror byggt ett hus?
  - c. *Uppmaning*: Bygg ett hus!
  - d. *Utrop*: Vilket fint hus han har byggt!
  - e. *Interjektion*: Oj då! Aj!

Alla mänskliga språk kan uttrycka dessa språkhandlingar och en generell språk teori bör följaktligen ha något att säga om hur de är relaterade, både inom ett språk, och tvärspråkligt.

Hur byggs då de språkliga yttrandena upp? Som redan nämnts är byggstenarna inte enskilda ord utan fraser, med intern struktur. Detta betyder att lyssnaren inte behöver hålla ett antal separata ord i minnet utan kan gruppera ihop dem och tolka frasen som en helhet. Även ur talarens perspektiv – produktionsperspektivet – är det effektivare att planera och uttala fraser än serier av separata ord. För att illustrera detta kan vi återvända till exempel (1a) där det finns två nominalfraser, d.v.s. fraser som har ett substantiv som huvudord, nämligen *min bror* och *ett hus*. I stället för *min bror* hade jag kunnat välja någon av de kursiverade nominalfraserna i (2) eller någon annan nominalfras.

- (2)
- a. *Erik* har byggt ett hus.
  - b. *Den rikaste mannen i stan* har byggt ett hus.
  - c. *En erfaren snickare som kommunen rekommenderade* har byggt ett hus.

Så länge yttrandet inleds med en nominalfras, oavsett längd, och denna följs av ett verb, här *har*, kommer yttrandet att tolkas som ett påstående på svenska. Den andra nominalfrasen, *ett hus*, kan förstås också varieras i längd och komplexitet.

Även om *min bror* och *ett hus* båda är nominalfraser så har de olika *grammatisk funktion* i satsen.<sup>3</sup> *Min bror* identifierar vem som bygger och *ett hus* talar om vad som byggs. Det är vanligt att benämna dessa funktioner *subjekt* respektive *objekt*. Dessa termer är relationella; en fras är alltid subjekt eller objekt i förhållande till ett huvudord, vanligtvis ett verb, här *bygga*. Väljer vi huvudordet *bygga* behöver vi ett subjekt och ett objekt för att satsen ska bli komplett. Väljer vi i stället *bo*, behöver vi ett rumsadverbial.

- (3) a. Erik har *byggt* ett hus.  
       SUBJ   V     OBJ  
       b. Erik har *bott* här.  
       SUBJ   V     ADVL

Tanken att huvudordet i satsen bestämmer vilka övriga led som måste ingå återkommer i någon form i de flesta grammatiska teorier. Det är vanligt att tala om detta som *valens*, med en term lånad från kemin, d.v.s att verbet binder ett antal argument till sig.<sup>4</sup> För enkelhets skull har jag här illustrerat de gemensamma dragen med exempel från svenska. Men oavsett vilket språk jag hade utgått ifrån, hade det knappast varit möjligt att beskriva språket utan att identifiera de ljud, ordbildningsmönster, frasmönster och satsmönster som används. Dessa är alltså viktiga delar av en generell språk teori.<sup>5</sup>

## Språktypologi

För att kunna hitta likheter och skillnader mellan olika språk behöver vi kunna identifiera element med samma funktion, t.ex. vad som är subjekt eller objekt i satser i olika språk. Detta är något som främst språktypologer undersöker med hjälp av kriterier som är baserade på frekvens, distribution och böjning (se t.ex. Croft 1990). Givet att man kan identifiera de grammatiska funktionerna, har typologer kunnat studera ordföljds-mönster på en abstraktare nivå i en stor mängd språk. Om vi jämför de svenska exemplen i (3) med motsvarande tyska i (4), ser vi att huvud verbet i svenska kommer före objektet och adverbialiet, medan det i tyska kommer efter.

- (4) a. Erik hat ein Haus *gebaut*.  
       SUBJ   OBJ     V  
       b. Erik hat hier *gewohnt*.  
       SUBJ   ADVL   V

Om vi lägger till subjektets placering kan vi säga att svenska är ett SVO-språk, d.v.s. ett subjekt-verb-objekt-språk, medan tyska är ett SOV-språk. Den här typen av generalisering utgår från att vi kan etablera ett språks grundläggande ordföljd och att denna i sin tur förutsäger andra ordföljdmönster. I det banbrytande arbetet *Universals of Human Language* (1963) formulerade Joseph Greenberg vid Stanford ett stort antal *implikationella* universalen; ett exempel visas i (5).

- (5) a. Om ett språk är VO så har det prepositioner.  
b. Om ett språk är OV så har det postpositioner.

(5a) stämmer bra för modern svenska som ju placerar verbet före objektet och har så gott som enbart prepositioner, t.ex. *från X*, *på X* och *till X*. Några postpositioner förekommer visserligen, men bara i fasta uttryck som *inte mig emot*. För att uttrycka riktning placerar vi *emot* framför nominalfrasen: *bilen kom emot mig*. (5b) stämmer mindre bra på tyska som har fler prepositioner än postpositioner, men är tydligt för andra OV-språk som japanska och turkiska.<sup>6</sup>

Trots att Greenbergs undersökning var baserad på ett begränsat urval språk, trettio stycken, har de samband som han kunde se i stora drag bekräftats i undersökningar baserade på betydligt fler språk. Den största databasen över ordföljdmönster heter *World Atlas of Language Structure* (WALS) och är sedan 2011 tillgänglig på internet.<sup>7</sup> Databasen innehåller data om cirka 2700 språk och bygger på tryckta och otryckta språkbeskrivningar samt intervjuer med infödda talare. Sambandet mellan verbets placering och pre- eller postpositioner har undersökts i 1142 språk med följande resultat.

(6)	OV PostP	472
	VO PreP	456
	VO Postp	42
	OV PreP	14
	<i>övriga</i>	58
	Summa	1142

Vi ser att 928, drygt 80 %, av språken följer Greenbergs implikationella universaler. Som *övriga* räknas språk utan klara adpositioner, t.ex. de som har circumpositioner. Tittar man närmare på kartan eller listan med under-

sökta språk, ser man att tyska är ett av de enbart 14 OV-språken med prepositioner.<sup>8</sup>

WALS är ett utmärkt verktyg för språkvetare som vill undersöka eventuella samband mellan olika ordföljdmönster. Databasen är dock beroende av att det finns tillförlitliga grammatiker eller beskrivningar för språken som tas upp, vilket inte alltid är fallet. Ett initiativ som söker komma förbi detta beroende av språkbeskrivningar är projektet *Universal Dependencies*.<sup>9</sup> Även detta har sitt ursprung i den typologiska forskningen vid Stanford, men använder språkteknologiska metoder och utgår från autentiska texter på språket i fråga. Texterna analyseras med automatiska metoder som skapar s.k. trädbanker, alltså syntaktiskt annoterade strukturer. I dessa hierarkiska strukturer, kallade dependensträd, kan man läsa av vilken grammatisk funktion en fras har till sitt huvudord.<sup>10</sup>

Genom att analysera dessa annoterade texter kan forskarna utveckla såväl parsnings- och översättningsprogram som grammatiker för de ingående språken. Precis som för det typologiska arbetet som ligger till grund för valet av annoteringskategorier, krävs det att instruktionerna är klara och entydiga så att de kan användas av talare av olika språk och även av programmerare som kanske saknar djupare språkvetenskaplig träning. Projektet involverar cirka 200 forskare runt om i världen och leds av professor Joakim Nivre, Uppsala. För närvarande finns mer än 100 trädbanker för ett sjuttiotal språk tillgängliga på projektets hemsida.

En annan språkteknologisk tillämpning som baseras på generella drag i mänskliga språk är *Grammatical Framework* (GF), en programmeringsmiljö för flerspråkiga tillämpningar som skapats och koordineras av professor Aarne Ranta, Göteborg.<sup>11</sup> I denna miljö utvecklas system för generering av text och tal i ett fyrtiotal språk, översättning mellan språken, parsning och flerspråkiga tillämpningar. Som exempel på en sådan kan nämnas en app för mobiltelefoner som en svenskspråkig barnmorska kan använda för att kommunicera med en gravid kvinna som talar arabiska eller kurdiska.

Som nav för generering och översättning i GF används en logisk representation, kallad *abstrakt syntax*. Detta är en abstraktion av återkommande drag i de språk som analyserats, uttryckt i en komputationellt effektiv formalism. Idén om en gemensam abstrakt syntax kommer nära en generell språk teori och programmeringsmiljön har visat sig kunna hantera språk från skilda språkfamiljer som finska, mongoliska och urdu. Aarne Ranta antar, så vitt jag vet, inte att denna abstrakta syntax skulle finnas i hjärnan på språk-

användarna, men det är däremot vad Noam Chomsky gör med sin teori om universell grammatik.

## Universell grammatik

Det förhållande som jag tog upp i inledningen, att ett nyfött barn kan lära sig vilket mänskligt språk som helst, har av en del lingvister tolkats som att barn föds med en predisposition att analysera och strukturera språklig input på ett särskilt sätt. Noam Chomsky kallar detta medfödda språkanlag för den *universella grammatiken* (UG) och mycket forskning under de senaste 50 åren har gått ut på att identifiera egenskaper hos denna grammatik. Chomskys metod har varit att utveckla ett formellt system som kan *generera*<sup>12</sup> i princip alla de strukturer som förekommer i naturliga språk (låt oss kalla dem A), men som inte kan generera sådana som inte förekommer i något språk (B). Genom att jämföra A och B försöker man identifiera systematiska begränsningar på språkliga strukturer. Dessa antas reflektera UG och följaktligen vara medfödda.

Chomsky och hans efterföljare är särskilt intresserade av sådana restriktioner som inte verkar ha någon direkt förklaring i termer av komplexitet eller processbarhet. Ett intressant exempel är den så kallade *bindningsteorin*, först utvecklad i Chomsky (1981). Denna antas komplettera den ovan nämnda valenstanken – att huvudordet bestämmer vilka argument som kan eller måste finnas – genom att bestämma var i satsen olika sorters nominalfraser kan stå och hur de påverkar varandras tolkning. Till grund för bindningsteorin ligger ett antagande om att det finns tre olika sorters nominalfraser i mänskliga språk. Dessa illustreras i (7) med exempel från svenska.

- (7)
- a. Anaforer: reflexiva pronomen *sig, sin*
  - b. Pronomen: *han, hon, de, honom, henne, ...*
  - c. R-uttryck: egennamn, nominalfraser *Eva, pojken, en hund*

Anaforer utmärks av att de är beroende av en annan referent för sin tolkning. Pronomen kan vara beroende, men kan också syfta självständigt medan R-uttryck, referentiella uttryck, aldrig är beroende.

Bindningsteorien består av tre principer, eller villkor, som reglerar hur dessa nominalfraser kan tolkas och var i satsen de kan stå

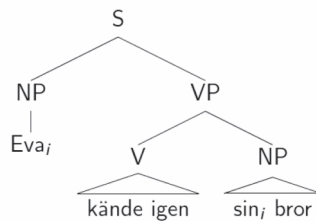
- (8) *Princip A*: En anafor måste vara *bunden* i sin bindningsdomän.

Principen illustreras bäst med några exempel. Satsen *Eva kände igen sin bror* tolkas av svensktalande som att en person, Eva, kände igen en annan person som är bror till Eva. För att uttrycka att det rör sig om samma Eva, annoteras satsen med så kallade referentiella index *i* i form av subskript. Om två nominalfraser (hädanefter NP) har index *i*, antas de referera till samma individ. Om de har olika index, är koreferens utesluten.

- (9) a.  $Eva_i$  kände igen  $sin_i$  bror.  
 b. \* $Eva_i$  kände igen  $sin_j$  bror.

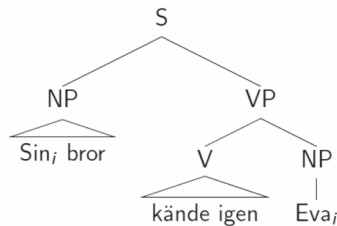
Den annoterade versionen i (9a) återspeglar alltså det sätt som vi tolkar satsen, medan (9b), med olika index, inte är en möjlig tolkning av denna sats, vilket anges med en asterisk \* framför satsen. Dessutom måste antecedenten, *i* det här fallet  $Eva_i$ , sitta högre upp i satsen för att kunna binda anaforen.<sup>13</sup> Detta visas med den något förenklade hierarkiska strukturen i (10).

(10)



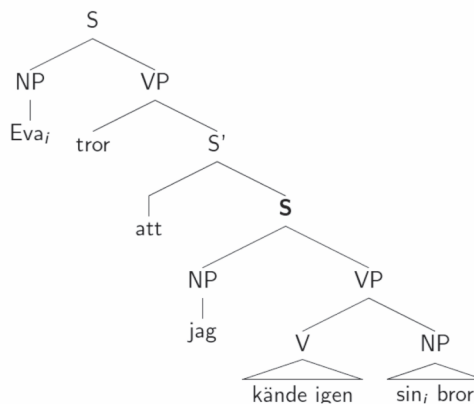
I (10) sitter NP  $Eva_i$  högre upp än NP  $sin_i$  bror. Satsen uppfyller därmed villkoret i Princip A att anaforen *sin* ska vara bunden av en antecedent. (11a) är däremot inte en möjlig sats på svenska eftersom antecedenten *Eva* sitter lägre ner, och följaktligen inte kan binda anaforen *sin*.

- (11) a. \* $Sin_i$  bror kände igen  $Eva_i$ .  
 b.



Det att exemplet i (12a) inte heller är möjligt motiverar ytterligare en precisering.

- (12) a. \*Eva<sub>i</sub> tror att jag kände igen sin<sub>i</sub> bror.  
b.



Här uppfylls visserligen villkoret att antecedenten ska sitta högre upp i strukturen än anaforen. Problemet är att antecedenten och anaforen separeras av en satsgräns, här markerad med fetstil, **S**. I svenska är alltså bindningsdomänen för anaforer den finita satsen.

Efter denna genomgång av Princip A, är det rätt enkelt att förklara Princip B.

- (13) *Princip B*: Ett pronomen måste vara *fritt* i sin bindningsdomän.

Kravet på pronomen är alltså motsatsen till kravet på anaforer; pronomen ska vara fria, vilket betyder att de inte får vara samindicerade med en antecedent högre upp i samma sats.<sup>14</sup> Vi får följande paradigm:

- (14) a. Eva<sub>i</sub> kände igen henne<sub>k</sub>.  
b. \*Eva<sub>i</sub> kände igen henne<sub>i</sub>.  
c. Eva<sub>i</sub> tror att jag kände igen henne<sub>i</sub>.  
d. Eva<sub>i</sub> tror att hon<sub>i</sub> vinner.

(14a) med olika index tolkas som att Eva kände igen en annan kvinnlig person. Tolkningen i (14b) med samma index är däremot inte möjlig. Om talaren avsett den tolkningen skulle hen sagt *Eva kände igen sig*. (14c,d) visar att



samindicerung med ett pronomen är möjlig om det inte står i samma sats som antecedenten.

(15) *Princip C*: Ett R-uttryck måste vara *fritt*.

Ett R-uttryck, som alltså har egen referens, kan inte vara bundet alls. Det får följande konsekvenser för tolkningen.

- (16) a. Hon tror att Eva vinner.  
 b.  $Hon_i$  tror att  $Eva_j$  vinner.  
 c.  $*Hon_i$  tror att  $Eva_i$  vinner.

Satsen i (16a) tolkas normalt som i (16b) där olika index anger att *hon* och *Eva* syftar på olika personer. (16c), med koreferens, är däremot inte en möjlig tolkning. Likaså får upprepade egennamn som konsekvens att det handlar om olika personer.

- (17) a. Eva tror att Eva vinner.  
 b.  $Eva_i$  tror att  $Eva_j$  vinner.  
 c.  $*Eva_i$  tror att  $Eva_i$  vinner.

(17a) kan betyda att Eva Andersson tror att Eva Bengtsson vinner, alltså med olika index som i (17b). Den samindicerade tolkningen i (17c) är däremot inte möjlig. Om talaren vet att det handlar om samma person väljer hen snarare (17d).

Sammantagna begränsar alltså de tre principerna var i satsen olika nominalfraser kan uppträda och hur de kan tolkas. Dessa principer har bekräftats i ett stort antal språk, trots att varken kategorierna eller de ganska abstrakta villkoren lärs ut. Det är detta som ligger bakom antagandet att bindningsteorin återspeglar den medfödda universella grammatiken. För att komma fram till huruvida principerna är medfödda eller inte, är det följaktligen viktigt att undersöka hur små barn tolkar språkliga beskrivningar. När det handlar om vuxna informanter går det att ställa frågor om hur de tolkar satser, men detta är sällan möjligt med riktigt små barn. I stället kan forskaren t.ex. låta barnen se hur några dockor eller leksaksdjur utför en handling. Sedan beskriver forskaren handlingen med en mening. Tanken är att om meningen stämmer med barnens upplevelse av handlingen, så instämmer de; annars protesterar de.

Metoden fungerar väl och har använts på olika språk.

I en serie experiment testade Crain & McKee (1985) hur amerikanska barn tolkade satser med egennamn som aktualiserade Princip C. I ett scenario visade de hur en av figurerna, Smurf, gick in i ett hus och åt en hamburgare. Sen fick barnen ta ställning till en av följande beskrivningar.

- (18) a. Smurf ate the hamburger when he was inside the house.  
 b. When he was inside the house, Smurf ate the hamburger.  
 c. When Smurf was inside the house, he ate the hamburger.  
 d. He ate the hamburger when Smurf was inside the house.  
 e. \*He<sub>i</sub> ate the hamburger when Smurf<sub>i</sub> was inside the house.

Crain & McKee fann att 80 % av de yngsta barnen (medelålder 3 år 1 månad) protesterade mot beskrivningen i (18d) men att de accepterade beskrivningarna i (18a–c). Författarna tolkar detta som att barnen reagerade mot att de återgav händelseförloppet som i (18d) eftersom detta tolkningsmönster bryter mot Princip C; R-uttrycket *Smurf* är bundet av det koreferentiella *he*, (visas i (18e)).<sup>15</sup> Crain & McKees studie har replikerats på många andra språk (se Crain 1991) och slutsatsen som många forskare drar är att bindningsteorins principer i någon form ingår i den universella grammatiken. Och om principerna är medfödda, borde de gälla för alla mänskliga språk. Här blir thailändska relevant.

### Thailändska

Ganska snart efter det att bindningsteorin blivit allmänt känd bland generativa lingvister på 1980-talet, började det komma rapporter om att man på thailändska kunde upprepa egennamn utan att få effekten att det handlar om olika personer. En thailändsk motsvarighet till (17c), som i (19a), betraktas som grammatisk, även när de två egennamnen refererar till samma person.<sup>16</sup>

- (19) a. nòi<sub>i</sub> khít wâa nòi<sub>i</sub> cà chaná.  
 Noi tror att Noi ska vinna  
 b. Noi<sub>i</sub> tror att hon<sub>i</sub> ska vinna.

På svenska, och de flesta andra språk, måste man uttrycka den betydelsen med ett bundet pronomen i bisatsen, som i (19b), men i thailändska fungerar

(19a) bra, trots att den bryter mot Princip C (se t.ex. Hoonchamlong 1991).<sup>17</sup> Många teoretiska studier har publicerats som försöker revidera bindningsteorin så att den även håller för thailändska, men jag går inte in på dem här. I stället vill jag ta upp en intressant studie av Deen & Timyam (2018) som har undersökt hur thailändska barn tolkar exempel som (19a).

Deen & Timyams hypotes var att bindningsteorins Princip C är medfödd och att thailändska barn inte skulle acceptera upprepade egennamn, till skillnad från vuxna thailändare. För att testa detta designade de ett experiment där försökspersonerna får titta på en animerad film på en surfplatta. I filmen uppträder fyra djur, en stor hund, en liten hund, en sköldpadda och en apa. Den stora hunden föreslår att djuren ska tävla om vem som kan springa fortast och priset är en fin apelsin. Filmen visar hur apan, den lilla hunden och sköldpaddan springer i tur och ordning. Sen ska domaren, den stora hunden, avgöra vem som har vunnit. Hon säger: Apan vann. Då protesterar den lilla hunden och säger: Nej, jag sprang fortast så jag vann. Filmen tar slut och försöksledaren tar fram en kasperdocka som sammanfattar filmen och avslutar med att yttra (20), där ordet för hund, *mǎa*, upprepas.

(20) mǎa phūut wǎa mǎa chaná kaan-khæng-khǎn  
hund säger att hund vinner tävling

När Deen & Timyam testade vuxna thailändare så instämde de och accepterade beskrivningen i (20). Detta betyder att de tolkade de två förekomsterna av *mǎa* som refererande till samma hund. De gjorde alltså precis den tolkning som utesluts av Princip C, men som har visats förekomma i thailändska.

Majoriteten av det sjuttiofå thailändska barn i ålder 4–6 år som testades reagerade däremot genom att protestera mot beskrivningen i (20). De svarade t.ex. som i (21).

(21) may-chay mǎa phūut wǎa khǎw chaná  
nej hund säger att hon vinner

Barnens reaktion tyder på att de tolkade de två förekomsterna av *mǎa* i (20) som att de refererade till olika hundar, helt i enlighet med Princip C. I stället producerade några en mening med ett pronomen, *khǎw*, som kan tolkas som koreferent med det första *mǎa* enligt Princip B.

Deen & Timyam tolkar resultaten som stöd för hypotesen att Princip C är

medfödd. När de testade lite äldre thailändska barn (9–12 år), svarade dessa på samma sätt som de vuxna. De accepterade att ett refererande uttryck kan vara bundet och koreferent med ett uttryck högre upp i satsen. Någon gång i 6–8-årsåldern börjar alltså thailändska barn anpassa sig till språkbruket runt omkring och lär sig att upprepade egennamn och substantiv snarare betar sig som pronomen.<sup>18</sup>

## Språk och kognition

Tanken att de tolkningsrestriktioner som beskrivs i bindningsteorin skulle reflektera en medfödd universell grammatik accepteras dock inte av alla språkvetare. Andra forskare menar att bindningsteorins effekter följer av allmänna kognitiva förmågor eller pragmatiska strategier. Den kognitiva lingvisten Karen van Hoek (1995) har t.ex. föreslagit att Princip C följer av att vi tenderar att välja nominalfraser efter hur tillgänglig referenten är i samtalsituationen.<sup>19</sup> För att introducera en ny referent använder vi vanligen egennamn eller substantiv. När vi väl har introducerat referenten, och den blivit så att säga *tillgänglig* för lyssnaren, använder vi i stället ett pronomen. På så sätt menar hon sig kunna förklara varför exemplen i (22) tolkas som de gör.

- (22) a. \*Eva<sub>i</sub> tror att Eva<sub>i</sub> vinner.  
b. Eva<sub>i</sub> tror att Eva<sub>j</sub> vinner.

När en talare yttrar det första *Eva* i (22a) förutsätter hen att lyssnaren förstår vilken person hen syftar på.<sup>20</sup> Denna person blir då en tillgänglig referent i talsituationen; vi kan kalla den *ref1*. Om talaren nu fortsätter att säga något om *ref1*, bör hen signalera att det handlar om samma *ref1* genom att använda ett pronomen, i det här fallet *hon*. Om talaren i stället väljer ett egennamn, uppmanas lyssnaren att söka efter en ny referent med namnet *Eva*. Detta leder till att vi tolkar satsen som i (22b), med två olika referenter, *ref1* och *ref2*, som båda heter *Eva*.

På liknande sätt förklarar van Hoek varför (23a) är ett missvisande sätt att uttrycka samma innebörd som i (23b) eller (23c).

- (23) a. \*Hon<sub>i</sub> tror att Eva<sub>i</sub> vinner.  
b. Hon<sub>i</sub> tror att hon<sub>i</sub> vinner.  
c. Eva<sub>i</sub> tror att hon<sub>i</sub> vinner.

Om talaren börjar med ett pronomen, *hon*, som i (23a), uppfattar lyssnaren det som en instruktion att söka efter en referent som redan är tillgänglig i samtalssituationen, kanske någon som nyss nämnts eller någon som står i närheten. När talaren sedan yttrar egennamnet *Eva* utgår lyssnaren från att det måste hänvisa till en ny referent. Hade det varit samma referent, skulle ju talaren använt ett pronomen, som i (23b). Om inte förutsättningarna för att inleda yttrandet med *hon* är uppfyllda, d.v.s. om inte talaren kan räkna med att lyssnaren förstår vem det är hen åsyftar med *hon*, gör talaren klokt i att i stället välja ett uttryck som pekar ut referenten, som i (23c).

Den här sortens resonemang om vilka slutledningar en lyssnare kan dra av talarens val av nominalfras förutsätter en ganska välutvecklad förmåga att föreställa sig hur andra människor tänker, det som brukar kallas *theory of mind* (Wellman 2018, Hjelmquist 2019). Det är inte något som barn kan automatiskt utan något de lär sig under flera år i samspel med vuxna talare.<sup>21</sup>

## Slutord

Sökandet efter en generell språk teori har fört språkvetenskapen framåt på flera sätt. Genom att studera många olika språk har man kunnat identifiera vissa egenskaper som alla mänskliga språk uppvisar. Till dem hör den effektiva användningen av en liten mängd betydelselösa ljud som kombineras till betydelsebärande ord och vidare till fraser och satser. En annan karakteristisk egenskap är att satsens, eller frasens, huvudord bestämmer vilka argument som kan eller måste ingå för att satsen ska kunna användas som en språkhandling. Därav följer också att satser har en hierarkisk struktur. Huruvida dessa egenskaper är karakteristiska för mänskligt språk eller är något som också utmärker andra kognitiva processer är en fråga som har varit aktuell sedan 1950-talet. Den har t.o.m. fått ökad aktualitet under de senaste decennierna när hjärnabbildningsstudier i realtid gjort det möjligt att undersöka var i hjärnan språk processas. Det är nu möjligt att jämföra vad som händer vid perception och processande av icke-verbala stimuli med processandet av språkliga stimuli. Denna forskning kommer säkert att kasta nytt ljus över den långvariga debatten mellan nativister och kognitivister. Nativister söker förklaringen till de gemensamma dragen i mänskliga språk i ett särskilt medfött anlag för språkinlärning medan kognitivister snarare ser de gemensamma dragen som konsekvenser av hur hjärnan utvecklats för att vi ska kunna kommunicera med varandra.

## Noter

- \* Jag tackar de ledamöter som ställde intressanta frågor i samband med symposiet samt Robin Cooper och Kerstin Norén som läst och kommenterat en preliminär version av denna artikel.
- 1 Liknande tankegångar finns även hos sanskritgrammatikern Panini, verksam i Indien på 300-talet f.Kr.
  - 2 Martinet (1961) talade om ”la double articulation du langage” d.v.s att betydelselösa men betydelseskiljande ljud (*fonem*) fogas samman till betydelsebärande ord (*morfem*). Jag menar att även den tredje och fjärde fasen, när ord fogas samman till fraser och fraser till yttranden, bör tas med som något karaktéristiskt för mänskliga språk.
  - 3 Andra beteckningar på grammatisk funktion är *satsdel* eller *satsled*.
  - 4 Valensmetaforen etablerades på 1950-talet av den franske språkvetaren Lucien Tesnière inom den s.k. dependensgrammatiken men introducerades enligt Przepiórkowski (2018) redan 1897 av Charles Sanders Peirce. Liknande tankegångar återfinns inom kategorialgrammatik, X-bar-teorin, Lexical-Functional Grammar, Head-driven Phrase Structure Grammar och även inom den senaste versionen av transformationsgrammatik, det s.k. Minimalistprogrammet där huvudordet genereras med en uppsättning abstrakta särdrag som måste satisfieras. Se Platzack (2010) för en introduktion till denna teori.
  - 5 En utmärkt översikt över dessa delar finns i Inledningen till *Svenska Akademiens grammatik* (Teleman et al. 1999).
  - 6 Senare forskning har visat att sambandet i (5) följer av en ännu generellare tendens, nämligen att språk tenderar att vara antingen huvudinitiala, som svenska, eller huvudfinala, som japanska och turkiska.
  - 7 <https://wals.info>
  - 8 <https://wals.info/feature/95A#2/14.9/152.9>
  - 9 <http://universaldependencies.org>
  - 10 Exempel på sådana dependensträd finns på <http://universaldependencies.org/introduction.html>. Dependensträden liknar den uppmärkning som visas i Språkbankens sökgränssnitt *Korp* när man väljer funktionen ’visa dependensträd’, se <https://sprakbanken.gu.se/korp>.
  - 11 <https://www.grammaticalframework.org>. Se även Ranta (2011).
  - 12 Därav benämningen generativ grammatik.
  - 13 Den tekniska termen är att antecedenten måste *c-kommandera* anaforen. Detta innebär att den första förgrenande noden ovanför antecedenten, också dominerar anaforen.
  - 14 I bindningsteorin används termerna *fri* och *bunden* på liknande sätt som fria och bundna variabler i logik.
  - 15 (18d) är det enda exemplet som bryter mot Princip C. Där står *he* högre upp och skulle binda *Smurf*, om de var samindicerade. I (18b) står *he* inuti en bisats och kan följaktligen inte binda *Smurf*. Även i (18c) står bisatsen högre upp än subjektet *he*, som därför inte kan binda *Smurf*.

- 16 De thailändska exemplen är hämtade från Deen & Timyam (2018) och återges här i en förenklad transkription.
- 17 Versionen med ett bundet pronomen fungerar också bra på thailändska.
- (i) nòi<sub>i</sub> khít wâa kháw<sub>i</sub> cà cháa.  
Noi<sub>i</sub> tror att hon<sub>i</sub> ska vinna
- 18 Jag har här inte gått in på andra faktorer som påverkar tolkningsmöjligheterna såsom förekomsten av adjektiv och klassificerare. Se Deen & Timyam (2018) för en fullständigare redogörelse. I onlineversionen av artikeln kan man även se stillbilder från filmen.
- 19 van Hoek bygger här på Ariels (1990) kända *accessibility hierarchy*.
- 20 Annars skulle talaren introducera referenten, t.ex. genom att säga *Jag träffade en kvinna som heter Eva. Hon ...*
- 21 Se t.ex. Karmiloff-Smith (1979) och senare studier.

### Referenser

- Ariel, Mira (1990) *Accessing Noun-phrase Antecedents*. London: Routledge.
- Chomsky, Noam (1981) *Lectures on Government and Binding*. Dordrecht: Foris.
- Crain, Stephen (1991) Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 4:597–612.
- Crain, Stephen & Cecile McKee (1985) The acquisition of structural restrictions on anaphora. *North East Linguistic Society (NELS)* 16:94–110.
- Croft, William (1990) *Typology and Universals*. Cambridge: Cambridge University Press.
- Deen, Kamil Ud & Napasri Timyam (2018) Condition C in Adult and Child Thai. *Language* 94:157–190.
- Greenberg, Joseph H. 1963. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In Greenberg, Joseph H. (ed.), *Universals of Human Language*, 73–113. Cambridge, Mass: MIT Press.
- Hoonchamlong, Yuphaphann (1991) Some issues in Thai anaphora: A government and binding approach. Madison: University of Wisconsin Madison dissertation.
- Hjelmquist, Erland (2019) Barnets tolkning av världen, med och utan språk. Föreläsning vid KVVVS högtidssammankomst den 24 januari 2019. Utkommer i KVVVS Årsbok 2019.
- Karmiloff-Smith, Annette (1979) *A Functional Approach to Child Language: A Study of Determiners and Reference*. Cambridge: Cambridge University Press.
- Martinet, André (1961) *Éléments de linguistique générale*. Colin.
- Platzack, Christer (2010) *Det fantastiska grammatiken. En minimalistisk beskrivning av svenskan*. Stockholm: Norstedts.
- Przepiórkowski, Adam (2018) The origin of the valency metaphor in linguistics. *Linguisticae Investigationes* 41:152–159.

- Ranta, Arne (2011) *Grammatical Framework: Programming with Multilingual Grammars*. Stanford: CSLI.
- Teleman, Ulf, Staffan Hellberg & Erik Andersson (1999) *Svenska Akademiens grammatik*. Stockholm: Norstedts Ordbok. [tillgänglig som pdf på <https://svenska.se/grammatik/>]
- Van Hoek, Karen (1995) Conceptual reference points: A cognitive grammar account of pronominal anaphora constraints. *Language* 71:310–40.
- Wellman, Henry M. (2014) *Making minds: How Theory of Mind Develops*. Oxford: Oxford University Press



*Arne Jarrick*

## Historiens händelser kan inte köras om, som i en partikelaccelerator – vad gör man då med upprepbarhet och generalisering?

### Det generella i det speciella

Världen som vi känner den är omåttligt rik på variation. Det gäller för den samhällseliga värld människor har byggt för att tillgodose sina behov, men också för den naturvärld som utgör dess byggstenar – inuti såväl som utanför oss. Det krävs ingen stor ansträngning för att se och inse detta. Människor förmedlar ofta sin förundran över all denna osannolika rikedom. Om det länge var frågan om en förundran på eller över gränsen till det religiösa, har denna allteftersom omvandlats till en alltmer världslig nyfikenhet på tillvarons överflödande och stundtals förvirrande mångfald.

Man kunde säga att det är här vetenskapen tar sin början – i den kunskaps-sökande människans strävan efter att städa upp i den aldrig upphörande raden av skiftande sinneserfarenheter, för att därefter så troget och systematiskt som möjligt registrera och rapportera om dem. Man kunde tillägga att detta också är början till nedrivningen av alla de varianter av generaliserande myter som människor i alla tider och på alla platser utvecklats om världen och en föreställd hinsides värld, för att förstå den och för att känna sig tryggare i den.

Men detta är verkligen bara början. Att stanna där vore som att avbilda världen i skala 1:1.

Den utvecklade vetenskapens överordnade syfte är ett annat: det är att vara mönstersökande. Och vad är det? Det är att söka ordningen i det som synes sakna ordning, att identifiera upprepningen i det som synes variera utan gräns. Det är att överskrida det partikulära i försöken att formulera generella sanningar om tillvaron, men nu på stadigt empirisk grund och inte längre som ett utflöde av den fria tanken.

Att urskilja mönster, det vill säga att se bort från det som skiljer för att se till det som förenar, är onekligen att förenkla det som ter sig komplicerat. Får man göra så? Man inte bara får, man bör. Det är i själva verket en av veten-

skapens huvuduppgifter. Det betyder inte att det är den enda uppgiften, och det är inte liktydigt med att förneka den allestädes närvarande variationen. Tvärtom. Identifikationen av det som avviker från mönstret är ett viktigt korrektiv till alltför långtgående generaliseringar och förenklingar. Det som tycktes gälla alla människor, visar sig ibland bara gälla vissa. Det har visat sig att hjärtfel skiljer sig åt och yttrar sig på olika sätt för kvinnor och män, till exempel.<sup>1</sup> Också detta är dock en generell inferens, om än mindre generell än den som korrigerats. Ännu viktigare är dock iakttagelsen att det generella banar sig väg genom det speciella, genom variation. Det naturliga urvalet är ett skolexempel på det – det har variationen inom arterna som nödvändig förutsättning. Detsamma torde gälla det mänskliga samhällets utvecklingshistoria, som gett upphov till en unik rikedom av kulturella former till följd av människans lika stora som generella förmåga att anpassa sig till skiftande omständigheter. På motsvarande sätt vore det vilseledande att säga att väder är en sak för sig och klimat är en annan. Långsiktiga temperaturförändringar (klimat) manifesterar sig genom kortsiktiga växlingar i temperatur, vind och nederbörd (väder). På något annat sätt kan det inte ske. Att hävda motsatsen är en flykt undan insikt och ansvar.

Forskare förfogar över en rad metoder för att söka och finna mer och mindre generella mönster, samt för att skilja skenbara mönster från verkliga. Men allmänt sett kan de inte slå sig till ro med sina slutledningar förrän andra forskare har gjort försök att upprepa dem i nya studier av samma eller samma slags fenomen. Detta är ett av baskraven på vetenskapen, närmare bestämt på den empiriska vetenskapen, matematiken, den teoretiska filosofin och andra teoretiska ämnen undantagna. Hur det står till med efterlevnaden av baskravet i den verkliga vetenskapsvärlden är temat för det här symposiet. Här ska jag successivt närma mig hur historieforskningen hanterar det, men göra det via en allmän diskussion om både gemensamma och varierande förutsättningar – variation, var ordet – för olika typer av forskning. Jag börjar dock med det som är grundförutsättningen för alltsammans: människan, den unikt kunskapssökande kulturvarelse.

## **Människan – den unikt kunskapssökande och kunskapsundvikande arten**

Som ingen annan art på jorden söker och finner vi människor kunskap. Att vi håller på med det där sökandet är lika uppenbart som att det har gett enorma

resultat, manifesterat i nästan allt vi ser och tar i. Det betyder inte att vi alltid är lika pigga på att veta. Ibland söker vi istället undvika kunskap, när vi inte aktivt motsätter oss bättre vetande. Men den kognitiva nettoeffekten är otvivelaktigt positiv, vilket inte utesluter att det i sällsynta fall vore positivt för vårt öde att vara negativ till viss kunskap (kunskap om hur man tillverkar kärnvapen, till exempel). Jag skulle ändå säga att den kognitivt positiva nettoeffekten i huvudsak också har varit mänskligt positiv.

Det går att urskilja ett antal omständigheter som ser ut att betinga och driva den unikt mänskliga, oavåtligt pågående kunskapsutveckling som försiggår utan att kräva någon genetisk förändring av kunskapens bärare (som ändå långsamt genomgår en sådan förändring). Dynamiken ligger i föreningen av en sällsynt stor påhittighet med en lika sällsynt benägenhet och förmåga att såväl lära ut som lära in. Och det är verkligen frågan om en förening, ett ständigt pågående växelspel mellan kreativitet och imitation. Att ta till sig och tillämpa en innovation som någon försökt föra vidare kräver anpassning till en situation som sällan eller aldrig är exakt lik den där den ursprungliga innovationen gjordes. Därför inbegriper anpassningen – inte bara den som lyckas – ett mått av kreativitet. Själlös kopiering utan kreativ anpassning skulle i längden innebära att den kulturellt överförda innovationen gick förlorad.<sup>2</sup> Men så ser det inte ut att vara, för trots att kunskap oavåtligt också läcker ut ur kunskapsförrådet är tillflödena större än läckagen.<sup>3</sup> Så går det på, varv på varv, generation för generation. En grönländsk kajak, till exempel, är inte ett och detsamma, utan varierar på många sätt beroende på behoven den ska tillgodose (jakt eller transport, m.m.) och omständigheterna den ska verka i (sötvatten eller öppet hav).

Vi söker kunskap. Och vi gör det på många olika sätt, med tiden alltmer vetenskapligt. Vi gör det av nödvändighet, strävan och lust, men aldrig bara på grund av brist på vetande. Mängden okända men potentiellt vetbara fenomen är nämligen oändligt stor, varför något annat än bristen på vetande måste till som grund för beslutet att söka kunskap om det ena istället för det andra. Detta gäller också den vetenskapligt verksamma forskaren, vars kunskapsmotiv, i enlighet med denna enkla logik, ytterst måste vara utomvetenskapligt i likhet med hur det är för alla andra människor.

Den principiella likheten i motiv ska dock inte skymma olikheten i hur man i och utanför vetenskapen kan och bör förhålla sig till sina motiv. I den vetenskapliga verksamheten gäller kunskapsmotivets åtskillnad från kunskapssökandet; där krävs att forskaren distanserar sig från sina motiv för att

söka sanningen, vart den än leder. Den vetenskapligt verksamma forskarens huvudsakliga uppgift är att utveckla så pålitlig kunskap som möjligt, oavsett vilka personliga motiv vederbörande har att göra det. Några sådana krav kan inte ställas på medborgaren som verkar utanför akademien, trots att också hen skulle tjäna på att göra åtskillnad mellan sina motiv och de verkliga resultaten (om de inte råkar sammanfalla). Forskaren möter också högre krav på systematik än den allmänna medborgaren, som bara alltför ofta (i varje fall oftare än forskaren) leds vilse av sin ofullkomliga statistiska intuition.<sup>4</sup>

Jag vet att forskare, som andra människor, är svaga käril, och att de inte alltid lever som de lär. Men även om de gjorde det skulle distanseringen från utomvetenskapliga motiv inte betyda att dessa vore utslocknade för gott. De fortsätter att driva på därinne, kanske slumrande, och kan återaktiveras när resultaten är på plats. Kanske vill forskaren verka för att resultaten sätts i verket för något angeläget ändamål i samhället. Eller kanske inte. Kanske är det tillräckligt att se dem publicerade i en tidskrift med högt vetenskapligt anseende. Men vilka den enskilda forskarens motiv och intressen än är – de får vara vilka som helst – har de ansvariga på systemnivå ett uppdrag att säkerställa kommunikationen av den vunna kunskapen till alla de medborgare som med sina skattemedel har gjort kunskapen möjlig. Den akademiska världen får gärna befolkas av en och annan forskare som mest är till för sig själv och sin personliga karriär, men vetenskapen som system får inte vara självtillräckligt – den är en samhällsangelägenhet. Kring huvudsaken förenas dock båda, den att producera ny och pålitlig kunskap. Och väl att märka: också den kunskap som blivit säkerställd genom att reproduceras är att betrakta som ny.

## Vetenskap och generaliseringar i allmänhet

När kan man då tala om att kunskapsprocessen har karaktär av vetenskap?

Det mest minimalistiska kravet på vetenskapen är att dess företrädare framlägger sina rön i form av påståenden som kan testas av andra, egentligen av vem som helst som förfogar över de intellektuella instrumenten att göra det. I detta ligger att påståendena ska vara falsifierbara. Däremot krävs inte att de ska hänga samman – de får vara hur partikulära som helst. Redan detta blygsamma krav skulle dock göra det svårt för en del akademiska aktiviteter att passera, till exempel den konstnärliga forskning där gestaltning och undersökning inte skiljs åt, där inga kriterier har utformats med vars hjälp fors-

karen kan ta ställning till vad som är sant och falskt. Ett mer långtgående krav är att undersökningen är systematiskt upplagd på ett sådant sätt att de falsifierbara påståenden som framkommer relaterar till en sammanhängande idé eller hypotes. Om också idén syftar till att undersökningarna ska ge generaliserbara resultat, har kraven skärpts ytterligare. Och ju större de generaliserande anspråken är, desto viktigare är det att också kunna visa att de relevanta resultaten kan upprepas. På så sätt ryms de minimalistiska kraven inom de maximalistiska.

Ambitionerna varierar, som bekant. Medan den ena nöjer sig med att säkerställa enskilda observationer, arbetar den andra på systematisk kartläggning av ett fält samtidigt som en tredje – den med maximalistiska anspråk – söker generella förklaringar. De har alla hemorts rätt i den vetenskapliga världen, ja de behövs för varandra, trots att vetenskapen sedd som system syftar till generell kunskap. I samverkan borgar de för en kumulativ kunskapsutveckling. I början av 1600-talet identifierade William Harvey blodomloppet som ett i viss mening slutet system, men han visste varken vilken funktion blodets eviga cirkulation har eller hur förbindelsen ser ut mellan artärer och vener. Det återstod för andra att upptäcka, bland andra Marcello Malpighi som med mikroskopets hjälp fick syn på kapillärerna. Ändå bidrog Harvey med sina ytterst noggranna vivisektioner av grodor och andra djur till att riva ned den gamla humoralpatologins allmänna lära om kroppsvätskorna. Det var en av förutsättningarna för att en helt annan allmän lära om kroppens gasutbyte med tiden skulle komma i dess ställe.<sup>5</sup> Detta är ett av oräkneliga exempel på hur partikulär och generell kunskap flätas samman i en gemensam rörelse mot ökat vetande.

För att röra sig från en trogen representation av ett fenomen till generaliseringar över en klass eller flera klasser av fenomen, krävs klassifikatoriskt konsekventa jämförelser av så likvärdiga fall som möjligt, såvida exakt upprepning inte är möjlig. Det är lättare sagt än gjort, eftersom också de fenomen som utan närmare eftertanke beskrivs under ett begrepp vid närmare eftertanke ofta verkar ha alltför lite gemensamt – där är variationen igen – samtidigt som de kan vara svåra att avgränsa från fenomen som faller utanför begreppet. Men det går, vilket jag tror mig ha visat i en storskalig studie av lagarnas världshistoria, utförd tillsammans med historikern Maria Wallenberg Bondesson.<sup>6</sup> Att närma sig generella slutsatser går dock bara om man är beredd att också räkna efter, att kvantifiera sina iakttagelser, oavsett svårigheterna.

Att generaliseringar kräver statistik för att kunna göras kan tyckas alltför

självkklart för att behöva påpekas. Det är det inte, särskilt inte för de humanister som hävdar att deras särpräglade angreppssätt är den kvalitativa metoden samtidigt som de framhåller att den humanistiska forskningens lika särpräglade bidrag till vetenskapen är att den skapar insikt om vad det är att vara människa – ett påstående med generella sanningsanspråk. Ett sådant anspråk kan förverkligas bara av den som är beredd att gå utöver fallstudien.

Kanske vore det bäst att utmönstra distinktionen mellan kvalitativa och kvantitativa data helt och hållet. Alla data måste ha kvalitet, det vill säga möta särskilda kvalitetskrav för att kunna användas. De måste vara äkta, reliabla och valida i förhållande till den fråga som ställts. Men till exempel ska observationen av ett enstaka korntal, uttryckt i siffror, inte betraktas som vare sig mer eller mindre kvantitativ än observationen av ett enskilt religiöst uttryck i en andlig självbiografi. Eller rättare sagt: i inget fall har kvantifieringen påbörjats. Det sker först när den ena datapunkten läggs till den andra och så vidare, och då sker det på principiellt samma sätt i båda fallen. Att jämförbarheten då det gäller religiösa uttryck kan vara svårare att uppnå ändrar inte på detta, men det ställer större krav på distinkt operationalisering än studiet av korntalen gör.

## Hur förutsättningarna varierar mellan vetenskapsområden

Eftersom förutsättningarna är likartade mellan olika vetenskapsområden bör också kraven vara det. Det är ”sensmoralen” så här långt. Men allt är sig inte likt från område till område, varför också anspråken och kraven måste skifta på basis av en för all vetenskap gemensam epistemologi. Nu är det inte skillnaderna mellan humaniora och naturvetenskap jag har i åtanke, utan skillnader som skär rakt igenom de konventionsbundna institutionella indelningarna. Det handlar om hur forskaren vinner tillträde till den verklighet som ska studeras. Historikern skulle kanske tala om skillnader i källornas beskaffenhet, filosofen kanske om skilda ontologiska villkor för att söka kunskap. Jag ska vara specifik, likgiltigt vilken beteckning som skulle passa bäst.

Som historikern söker astronomen och evolutionsbiologen kunskap om det förflutna i hög grad via spåren i det nuvarande. Historikern kan inte studera passerade händelser direkt, utan bara indirekt genom nedtecknade vittnesbörd som i bästa fall är samtida med händelserna som återges. Ingen historiker har bevittnat avrättningen av Ludvig XVI, men många anser sig ändå kunna säkerställa att den ägde rum – och hur. På motsvarande sätt studerar

astronomen dramatiska händelser i ett kosmiskt förflutet genom de ”svallvågor” som når oss idag. Likadant med många evolutionsbiologer som sluter sig till en långsam artutveckling utifrån förhistoriska lämningar i nutiden, fossil till exempel. Allt är naturligtvis inte gemensamt mellan dessa discipliner, men detta är.

Att inte kunna observera de fenomen man vill ha kunskap om direkt är en otvetydig begränsning. Fördelen är att det forskande subjektet därmed inte heller kan påverka det beforskade fenomenet. Det ”korrumpas” inte, eftersom det ju observeras bara via spåren i nuet. Vi kan visserligen ta miste om Ludvig XVI:s avrättning, men han kan inte ha några synpunkter på våra missgärningar och inte resa sig ur graven för att korrigera oss.

Det senare kan inte sägas vara en gemensam nämnare för alla de tre nämnda områdena, men det gäller i varje fall för historieforskningen (och kanske i viss mån för evolutionsbiologin). Föreställ er en historiker som fick möjlighet att göra en tidsresa bakåt för att bli ett direkt vittne till rättegången mot Galilei Galileo i 1630-talets Italien. I samma stund skulle historikern förvandlas till en deltagande observatör och därmed å ena sidan njuta favören av omedelbar tillgång till studieobjektet, men å andra sidan riskera att ändra på skeendet redan genom sin för alla parter synliga närvaro där. Sådana ontologiska villkor skulle den föreställda tidsresenären dela med antropologer, etnologer och så kallade gröna biologer med flera forskare, som i sin egen tid ägnar sig åt direkt observation av förlopp som i princip fortgår vare sig de är där eller ej. Men medan till exempel svanforskaren kan gömma sig i en konstgjord svan för att inte påverka det svanliv som ska studeras (jag har sett foton på det), är det svårare för antropologen att förbli helt obemärkt bland de grupper av människor hen vill få fördjupad kännedom om. Det naturliga experimentet kan därmed ofta vara äkta naturligt, medan det kulturella experimentet inte riktigt kan spegla den ”äkta” kulturen, det vill säga kulturen sådan den var före antropologens ankomst dit.

Här är alltså två typer av forskning på ömse sidor om fakultetsgränserna, vilka i det ena fallet förenas av ontologiska för- och nackdelar som i det andra fallet är omvända. I jämförelse med dessa finns forskningsområden, åter på varsin sida om hävdvunna gränser, som i fråga om källornas beskaffenhet har obestridliga fördelar utan att belastas av nackdelarna. Kärnfysiken, den kliniska medicinen och den experimentella psykologin är exempel på det. Något som förenar dem är att de på ett kontrollerat sätt iscensätter en verklighet som de sedan studerar i realtid. På så sätt kan de också bestämma vad som ska



hållas konstant och vad som ska tillåtas eller fås att variera, vare sig det gäller elementarpartiklar, kemiska preparat eller psykiska omständigheter. I likhet med den gröna biologen och antropologen, gör fysikern, medicinaren och psykologen direkta observationer, men till skillnad från dessa kan de arrangera sina experiment så att de tar bort sin egen inverkan på objektet (där den alls finns) eller på sin tolkning av experimentet (t.ex. genom dubbelblinda tester), om de inte använder sig av sin inverkan på ett systematiskt sätt (som t.ex. Stanley Milgram en gång gjorde<sup>7</sup>).

Ur dessa skillnader mellan olika discipliner i villkoren för vetenskaplig verklighetskontakt framgår nu ytterligare en besvärande skillnad, den att vissa men inte alla kan leva upp till det basala vetenskapliga replikationskravet.

Trots att det inte verkar möjligt att återskapa jätteödlorna, kan en del evolutionära processer snabbspolas i realtid, till exempel med hjälp av bananflugornas snabba omsättning. På motsvarande sätt kan psykologiska och fysiska experiment replikeras, liksom fåglarna och deras flyttmönster kan studeras gång på gång i naturliga experiment, då de inte utsätts för regelrätta laboratorieförsök. Samtidigt är det uppenbart att historieforskningen inte kan leva upp till replikationskravet. Det går inte att upprepa avrättningen av Ludvig XVI, varken kontrollerat eller okontrollerat. Men det går inte heller inte med Big Bang. Detta har dock varken hindrat historiker eller astronomer från att framföra generella teorier om samhällets långa och världsalltets ännu längre historia. Ska då varken historieforskningen eller astronomin räknas till vetenskapen? Eller kan något göras för att kompensera bristen? Jag kan bara svara för mitt eget skrå. Det gör jag nedan.

## **Historieforskningen – vad har gjorts och vad kan göras?**

I årtusenden har tänkare av olika slag presenterat teorier om den mänskliga historiens generella karaktär och riktning. Längre var teorierna cykliska och mestadels kretsade de kring förändringar i människors attityder i allmänhet och deras moral i synnerhet. Dessutom var världsförklaringarna länge en intellektuell brygd av intuitivt och logiskt tänkande utan särskilt många empiriska ingredienser.

Den kinesiske historikern Ssu Ma-Ch'iens över 2 000 år gamla idé om dygdernas cykliska rörelse upp och ner är ett exempel på en sådan teori, Augustinus nästan lika gamla teori om syndafall och mänsklig strävan ett annat. Långt senare, på 1300-talet, formulerade den tunisiska historikern Ibn Khal-



dun en teori om växelverkan mellan kulturell uppgång och kulturellt förfall, byggd på idén om Asabiya, ett slags pionjäranda dömd att försvagas och småningom förfalla generation för generation, till dess nya grupper tar över på basis av ny pionjäranda, och så vidare i all oändlighet. Med sådana här teorier fortsatte det, till exempel med den allvetande filosofen och historikern Giambattista Vico som på 1700-talet framförde ännu en generell cyklisk teori om uppgångs- och förfallsfaser ifråga om människors förnuft och moral.<sup>8</sup>

Från mitten av 1700-talet förändrades det generaliserande historiska tänkandet på flera sätt. För det första övergick historien från att vara cyklisk till att tilldelas en riktning uppåt eller framåt. För det andra började teorierna alltmer handla om den materiella eller sociala utvecklingen och allt mindre om andens väg genom tidevarven, med Friedrich Hegel som idealistiskt undantag. Adam Smith, Karl Marx och Auguste Comte är nog de mest kända exponenterna för denna intellektuella förskjutning, trots att alla tre till viss del fortsatte att operera med psykologiska och kulturella kategorier.<sup>9</sup> Emile Durkheim och Max Weber kvalificerar sig nog också som representanter för den teoretiska strömkantringen. För det tredje började det nu göras mer ambitiösa försök att förankra teorierna i empiriska iakttagelser. Men statistiken var outvecklad och de empiriska ansträngningarna fortsatte länge än att vara ofullgångna.<sup>10</sup> Den ända in i vår tid hyllade Max Weber, till exempel, hade visserligen en gigantisk beläsenhet och enorma kunskaper om världen och dess historia, och han refererade till alla möjliga konkreta iakttagelser som grund för sina generella slutledningar. Ändå var dessa slutledningar i hög grad impressionistiska och byggde sällan på systematiska kvantifieringar.

Så, inte nog med att historieforskningen per definition är oförmögen att efterleva de grundläggande replikations- och reproduktionskraven – det senare följer i bästa fall av det förra. Många av de försök som har gjorts är desutom alltför lösliga konstruktioner, empiriskt såväl som teoretiskt. Med tiden har det dock blivit betydligt bättre, och i vår tid arbetar många historiker seriöst på att formulera och systematiskt testa sina teorier eller empiriska generaliseringar.

Ett prominent exempel på det är antikhistorikern Walter Scheidel som nyligen, stödd på gini-index, har publicerat en omfattande studie av den materiella resursfördelningens långsiktiga utveckling, särskilt under de senaste 2 000 åren. Hans generella slutsats är att utvecklingen huvudsakligen går i riktning mot ökad ojämlikhet. I synnerhet sker detta när samhällen befinner sig i stabila omständigheter. Omvänt är en sammanpressning av resursskill-

naderna mer sannolik under instabila perioder kännetecknade av krig, epidemier, statens sönderfall eller revolutioner.<sup>11</sup> Thomas Pikettys omtalade arbete om kapital- och inkomstfördelningen i det tjugoförsta århundradet faller väl in i Scheidels mångseklära empiriska generalisering. Också denne tycker sig ha identifierat ökad ojämlikhet som den spontant underliggande trenden, bara tillfälligt avbruten av den kapitalförstörelse som de två världskrigen medförde.<sup>12</sup>

Andra forskare har mot bakgrund av samma grundläggande iakttagelse försökt urskilja de mekanismer som principiellt verkar i riktning mot ökade klyftor. Per Molander är en av dem. Peter Turchin är en annan. Han är också en nutida forskare som arbetat med massiva mängder av data för att studera långsiktiga samhällsliga makroprocesser, bland annat samhällenas rörelse mot ökad komplexitet. Om man kombinerar Turchins rön med historikern Manuel Eisners omfattande data över det dödliga våldets långsiktiga nedgång går det dessutom att urskilja ett samband mellan det senare och tillväxten av samhällslig komplexitet.<sup>14</sup> Hur dessa två processer hänger samman, återstår dock att komma underfund med.

Scheidels och Pikettys analyser, som innebär en kritik av Simon Kuznets omvända teori om resursfördelningens utveckling, har inte fått stå oemotsagda, oaktat att nästan alla erkänner det gedigna i deras studier. Bland dem kan nämnas Branko Milanovic som har trätt till Kuznets försvar, också han på empirisk grund men enligt min mening utan att riktigt lyckas.<sup>15</sup>

Exemplen ovan visar att det går att komma en bra bit på väg mot generella inferenser också inom historieforskningen, trots att den är i avsaknad av replikationsstudien som korrektiv. Vissa långsiktiga makroprocesser är dessutom så uppenbara att replikationer vore helt överflödiga, för att inte säga meningslösa. Etableringen av bofasthet i och med den neolitiska revolutionen är exempel på det, liksom den successiva övergången till allt större dominans för ”immateriella näringar” eller den demografiska transitionen som allteftersom verkar slå igenom överallt, med mera.

Så också historieforskningen har lämnat viktiga bidrag till mänsklighetens kunskapsförråd, trots att den verkar under en särskild sorts begränsande ontologiska villkor. Mycket har också gjorts och mycket kan ännu göras för att kompensera för begränsningarna och öka möjligheten till generaliseringar, men också för att i ett särskilt avseende systematiskt upprepa tidigare studier.

För det första borde historikerna i högre rad än hittills bygga sina arbeten

på långa tidsserier av jämförbara data, eller göra ansträngningar för att arrangera sina undersökningar så att data blir jämförbara. Jämförbarheten bör läggas till grund för systematiskt komparativa analyser som ytterst syftar till att förklara historiska förlopp, det vill säga till att urskilja likheter i det som ter sig olikt. Det är något utöver att registrera kulturell variation bland mänskliga samhällen, oavsett att också detta behöver göras. Det görs dock redan i övermått.

Det är svårt att arrangera jämförelser över lång tid på ett sådant sätt så att jämförbarheten bevaras från första till sista observationen. Det komparativa systemet måste vara på en gång stabilt och flexibelt. Stabilt så att kategoriernas betydelser inte börjar glida, och flexibelt så att förutsett oförutsedda observationer kan införlivas i det komparativa systemet. I en relativt storskalig studie av lagarnas världshistoria har Maria Wallenberg Bondesson och jag försökt oss på just detta. Metoden har varit att notera vad vi finner i lagarna på tre olika abstraktions- eller aggregationsnivåer, vilket har gjort det möjligt att ändra indelningarna på en nivå utan att behöva rubba något på de andra nivåerna. Kategorier har också kunnat läggas till. Systemet är hierarkiskt och har formen av en trädstruktur.<sup>16</sup>

För det andra måste historikerna göra sina grunddata och sina metoder helt synliga, så att vem som helst skulle kunna göra om studien och därmed granska den. Eftersom den händelse som studeras är densamma hur många gånger den än studeras via det material och de metoder de primära forskarna har använt, skulle detta inte kunna bli detsamma som att göra replikationer i konventionell mening. Men det vore ett steg på vägen mot ökad metodologisk disciplin. Det vore också ett steg på vägen tillbaka mot det källkritiska arbetssätt som dagens historiker i alltför hög grad har övergett trots att de alltför framhåller källkritiken som sitt särskilda adelsmärke.

För det tredje skulle historikerna kunna iscensätta fysiska experiment som trots allt påminner om verkliga replikationsstudier. Till exempel skulle de kunna utföra upprepade odlingsexperiment under kontrollerat varierande förutsättningar för att på så sätt testa olika historiska syntesförsök. Tor Heyerdahls klassiskt djärva företag att korsa Atlanten med Kontiki var i begränsad mening ett sådant försök.

Slutligen vore det rent allmänt välgörande om humanvetare i allmänhet i högre grad än idag på ett uthålligt sätt vore villiga att koncentrera sig kring större och färre frågor. Humanvetenskapen kännetecknas idag både av stor tematisk och splittring och av tematisk ombytlighet. De samlas i för liten grad

kring gemensamma vetenskapliga angelägenheter och de tröttnar för fort. Detta är till nackdel för det kumulativa kunskapsbygge som vetenskapen är till för. En del framstående humanvetare gör redan vad som krävs, men ska det bli märkbar skillnad måste fler göra det och färre avvisa att det just är detta som är vårt överordnade ärende.

### *Noter*

- 1 <https://www.hjart-lungfonden.se/Sjukdomar/Hjartsjukdomar/Kvinnors-hjartfel/>.
- 2 Enquist et al. (2008).
- 3 Jarrick (2017), s. 51.
- 4 Se särskilt Kahnemann (2013), för åtskilliga exempel på detta.
- 5 Pehrsson (2018); <https://www.britannica.com/biography/Marcello-Malpighi>.
- 6 Jarrick & Wallenberg Bondesson (2018).
- 7 Milgram (1975).
- 8 Galtung & Inayatullah (1997).
- 9 Galtung & Inayatullah (1997).
- 10 Se t.ex. Durkheim (1983). Boken är full av mycket enkel statistik – och av ad hoc-resonemang som den enkla statistiken inte var en värn emot.
- 11 Scheidel (2017).
- 12 Piketty (2014).
- 13 Molander (2014); Turchin (2006).
- 14 Eisner (2014).
- 15 Milanovic (2016).
- 16 Jarrick & Wallenberg Bondesson (2018).

### *Referenser*

#### *Tryckta*

- Durkheim, Emile, *Själv mordet*, Lund: Argos 1983.
- Eisner, Manuel, "From Swords to Words: Does Macro-Level Change in Self-Control Predict Long-Term Variation in Levels of Homicide?", *Crime and Justice* 43:1 (2014).
- Enquist, Magnus, Stefano Ghirlanda, Arne Jarrick & Carl-Adam Wachtmeister, "Why does human culture increase exponentially?", *Theoretical Population Biology* 74 (2008).
- Galtung, Johan & Sohail Inayatullah, *Macrohistory and Macrohistorians*, Westport Connecticut & London 1997.

- Jarrick, Arne & Maria Wallenberg Bondesson, *The Dynamics of Law-Making: A World History*, Stockholm: Kungl. Vitterhets Historie och Antikvitets Akademien 2018.
- Jarrick, Arne, *Det finns inga häxor – en bok om kunskap*, Stockholm: Weyler 2017.
- Kahnemann, Daniel, *Tänka snabbt och långsamt*, Stockholm: Mån-pocket 2013.
- Milanovic, Branko, *Global Inequality: A New Approach for the Age of Globalization*, Cambridge (Mass) & London: Harvard University Press 2016.
- Milgram, Stanley, *Lydnad och auktoritet*, Stockholm: Wahlström & Widstrand 1975.
- Molander, Per, *Ojämlighetens anatomi*, Stockholm: Weyler 2014.
- Pehrsson, Kenneth, *William Harvey: En medicinsk revolutionär*, Stockholm: Fri tanke 2018.
- Piketty, Thomas, *Capital in the Twenty-First Century*, Cambridge (Mass.) & London 2014.
- Scheidel, Walter, *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century*, Princeton and Oxford: Princeton University Press 2017.

### **Digitala**

<https://www.britannica.com/biography/Marcello-Malpighi>

<https://www.hjart-lungfonden.se/Sjukdomar/Hjartsjukdomar/Kvinnors-hjartfel/>



*Hans Ruin*

## En delad blick – Husserl, Derrida och det teoretiska seendets kall

*Dick Haglund in memoriam*

Vetenskap vilar på igenkännande av mönster. Den kräver att vi kan se och fatta det enskilda som en instans av något allmänt. Människan uppfattar ett fenomen som förbryllar eller fascinerar. Hon anstränger sig att urskilja det som är generellt för att sedan kunna sätta det i relation till något annat, som dess orsak, effekt eller sammanhang. Denna drift att utvinna det generella i det enskilda leder hennes intellekt att hellre se gestalter i det gestaltlösa än tvärtom. Så byggs också vetandet genom att verklighetens flyktiga, partikulära och obestämda skeende samlas i fastare mönster och upprepbara regelbundenheter med hjälp av begrepp och numeriska relationer.

Det grekiska filosofiska genombrottet utmärks av hur en sedan tidigare utvecklad förmåga att omfatta och behärska naturen med språk och geometri blir till en strävan att formulera och formalisera de allmänna och underliggande mönstren för varat i dess helhet. ”Teori” härrör från grekiskans *theorein*: att se. Det är med ögat, först det yttre men i förlängningen med tankens inre öga, som människor från denna punkt i historien på allvar ger sig i kast med att i språket bemästra en rörlig, mångfaldig och ofta motsträvig natur. Aristoteles *Fysik* och *Metafysik* står kvar som tidiga monument över denna ansats, då de myntar de begrepp som sedan fortsatt att vägleda frågandet efter naturens grundläggande karaktär, som ”verklighet” (*entelecheia*), ”möjlighet” (*dynamis*), ”energi” (*energeia*), rörelse (*kinesis*) och ”substans” (*ousia*). Hos Aristoteles framhävs också denna strävan som idealet om ett visst slags *liv*: ett *bios theoretikos*, ett *vita contemplativa*.

*Finns* då dessa gestalter, väsen, lagar och samband, eller är de bara skapelser av vårt medvetande, som kastar sina nät över en ändlös mångfald singulara företeelser? Frågan har följt filosofin och vetenskapsteorin från första stund. Ja, den är på sätt och vis dess grundläggande dilemma. Spänningen mellan flöde och stillhet och mellan enskildhet och allmänhet samlas redan

från början i de heraklitiska fragmenten, som talar om en reversibel enhetens mångfald och mångfaldens enhet, ”ur allt ett och ur ett allt”.<sup>1</sup> Platon skall i sin ungdom ha varit heraklitiker, men han kom att bygga det mest berömda templet till allmänbegreppen genom sin idé- eller formlära. Alltifrån antikens skeptiker, till medeltidens nominalister, över den moderna tidens empirister och våra dagars begreppskritiker har dock den andra rösten alltid också funnits: den som Nietzsche gav ett förtäta uttryck när han i sin uppsats ”Om sanning och lögn i utomnormalisk mening” (1874) skrev om människan som det mest högmodiga djuret i universum, som uppfann begreppet som ett sätt att vinna kontroll över sin värld genom att till sist ”blunda för det individuella och verkliga”.<sup>2</sup>

Men även den mest radikala och vältaliga nominalist kan inte undgå att tala det allmännas språk för att förmedla sin skeptiska kritik. Det gäller också Nietzsche. I frammanandet av bilden av människan som det mest högmodiga djuret, och i den kritiska diskussionen av förhållandet mellan verklighetens enskildheter och det begreppsliga förnuftet, åberopar han en serie väsenstalter, däribland just allmänbegrepp som ”verklighet”, ”erfarenhet” och ”sanning”. Också han måste luta sig mot tanken att kunna bli förstådd och prövad, liksom mot förhoppningen att förnuftet genom denna kritiska blick ska väckas till självbesinning.

Vetandet – också det kritiska och självkritiska – är med andra ord aldrig en *enskilds* vetande, utan det är något som alltid redan från början räcks över som något *att dela med andra*. I tanken på en vetenskapskultur ligger tanken på en gemensam mänsklighet och ett gemensamt förnuft eller *logos*. I ett annat fragment skriver Herakleitos: ”Trots att förnuftet (*logos*) är gemensamt lever de många som om de hade ett eget förstånd”.<sup>3</sup> Vi är här på jorden tillsammans, och vårt vetande är något som vi – idealt sett – söker tillsammans, som ett sätt att orientera oss i varat, som seende, fria och ansvarstagande, bortom klass, nation, kön och religion.

De två begrepp som vi inbjudits till att reflektera över här – upprepbarhet och generaliserbarhet – pekar ytterst båda mot denna i grunden intersubjektiva dimension av vetandet. Det som någon sett och förstått skall – om det gör anspråk på vetande och vetenskap – i princip vara möjligt för en annan människa att också se och förstå. Äkta vetenskap kan inte förbli sluten inom en ensam individ eller en avskild grupp, även om den först kanske måste uppträda inom en sådan. Vetande är något som alltid redan från början åtminstone implicerar delbarhet.



Det som bekymrar vetenskapssamhället idag är en ökad medvetenhet om att mycket av det som i snabb takt publiceras som vetenskapliga resultat i den oöverblickbara flora av tidskrifter som konstituerar det globala vetenskapssamhället, inte klarar denna test. När vi återvänder till samma empiri och testsituationer visar sig inte samma mönster på nytt, eller beläggen för dem är mer tvetydiga än författarna gjort gällande. Vetenskap har blivit alltmer av en företagsverksamhet, som styrs av merkantila principer. När karriärer avgörs av publiceringsfrekvens ökar lockelsen att ta genvägar. Innehåll underordnas form. Produktivitet underordnas verklig nytta. Skenbara resultat ersätter välgrundade slutsatser. Det gamla intersubjektiva idealet om sanning, objektivitet och kvalitet genom ”granskningen av likar” – *peer review* – har alltmer kommit att instrumentaliseras. Det har förvandlats till en formaliserad auktoriserings-procedur som det gäller att ta sig igenom, för att kunna räkna sina resultat som en merit eller som en grund för lönebonus. Den filosofiskt sinnade vetenskapens ursprungliga *ethos*, det fria sökandet efter allmängiltiga mönster, förskjuts mot en affärsmässig jakt på publiceringspoäng, också bland filosofer. Ansvar för en delad sanning övergår i ansvar för karriär, ranking och vinst, där person tävlar mot person, lärosäte mot lärosäte och nation mot nation. Det leder till krav på att certifierarna utsättas för ännu en certifiering, i en ändlös kedja av legitimeringsprocedurer som kännetecknar ett samhälle där ansvar i växande grad läggs på entreprenad.

\*

I den första volymen av tidskriften *Logos* 1911 publicerade Edmund Husserl en berömd uppsats med den uppfordrande titeln *Filosofi som sträng vetenskap*. Tidskriften, som hade en bild av Herakleitos på omslaget, var ett försök av den dåvarande eliten bland tyskspråkiga human- och samhällsvetare att dra igång ett nationsövergripande samtal om vetenskapens mål och mening.<sup>4</sup> I redaktionen ingick, förutom Husserl, både Weber, Windelband, Meinecke och Simmel. För Husserl var den vetenskapliga filosofin trängd från två håll. Från ena hållet hotades den av en relativistisk historicism och världsåskådningsfilosofi, som inte längre menade sig behöva ta ansvar för vetandets övergripande rationalitet. Men från andra sidan, och detta var i hans ögon i grund och botten allvarligare, hotades den av ”naturalism”. Med detta begrepp åsyftade han hela den moderna vetenskapliga kulturens gängse världs bild som utgår från att naturen i dess helhet uteslutande består av rums-tidsliga företeelser underkastade beräkningsbara lagar.

Kritiken kan vid en första anblick te sig förbryllande, med tanke på att många tvärtom skulle vara benägna att framhålla denna naturvetenskapliga världsbild som ett fast fundament att stå på i kampen mot relativistiska strömningar. Men när detta naturalistiska perspektiv når fram till människan själv och hennes medvetande och rationalitet så blottas dess inre motsägelser. Då griper naturaliseringen efter även de ideala förnuftsnormerna i en själv-objektivering av medvetande och tänkande. Det förnuft som en gång öppnade världen genom sin fria, teoretiska blick blir i slutstadiet av detta förnuftsäventyr på samma gång infångat och osynliggjort av sig självt. Genom sin beräknande vetenskap har människan gjort sig själv till något hon också tycker sig kunna beräkna. Därmed har hon också skytt grunden för de ideal som från början vägledde henne mot vetandets oändliga mål.

För Husserl är vetenskap ytterst också ett etiskt projekt. I den meningen tänker han i Platons och Aristoteles anda, för vilka det handlade om ett liv präglad av oändliga förnuftsideal. Men det är just denna vetenskapens inre mening som riskerar att gå förlorad inom den moderna naturalismen, och som därför motiverar en fördjupad filosofisk kritik av vetandet självt. Det handlar inte om att bara lägga till en "subjektiv" dimension till den objektiva världen och det objektiva vetandet. Tvärtom: så länge subjektiviteten inte görs till ett tema i sin egen rätt, så förstår vi inte heller objektivitet och verklighet. För de vilar ytterst på ett allmänmänskligt icke-naturaliserbart intersubjektivt medvetandeliv.

Om vi följer naturalismen konsekvent blir till sist även kunskapen en biologisk process bland andra biologiska processer, som återfinns i skiftande grad hos olika livsformer. Men då blir det omöjligt att förstå hur samma natur också kan bli föremål för kunskap med ett anspråk på sanning om hur det är. Hur kan naturen dubblera sig själv i ett vetande med anspråk på giltighet, om den från början själv är en lagbunden entitet bland andra? Naturvetenskaperna, däribland biologi och psykologi, har svårt att omfatta eller ens se denna problematik, eftersom den ligger utanför deras disciplinärt-begreppsliga räckvidd.

Vissa skulle säga att det går att komma runt det genom att sociologisera och historisera vetandet, och på så vis förklara dess framväxande föränderliga anspråk. Men Husserls poäng i uppsatsen är att detta inte löser problemet. Förvisso har alla andliga formationer, inklusive vetenskapen, en historia. Det är en historia som det är viktigt att bekanta sig med för att också kunna sätta dess anspråk, misstag och missgrepp i ett kritiskt ljus. Förvisso kan vetande

därtill utgå från många olika perspektiv och kunskapsintressen. Men historisering och situering rymmer också sina fällor, då de riskerar att lösa upp vetandet i ett flöde av skiftande och kanske motstridiga narrativ, som befriar såväl historikern som fysikern från inte behöva ställa den egna förnuftspraktiken under belysning.

I formuleringar som bitvis föregriper vad Weber några år senare skulle uttrycka i sitt andliga testamente "Vetenskap som kall/yrke (*Beruf*)", tar Husserl istället sikte på ett filosofiskt vetande-ideal som inte ska vara slutet inom sig självt, utan som ska rymma sin egen oändliga förvandling. Det handlar inte om "visdom" i en äldre mening av livskloket, utan om något mer svårfångat: om att ta ansvar för ett liv väglett av en oändlig idé om sanning och förnuft. Liksom Weber, ser Husserl nämligen vetenskapen som en kraft som inte bara bidrar till framsteg utan som också riskerar att urholka världen, något som Weber skulle komma att tala om som "avförtrollning". Han talar rentav om en "andlig nöd", som bottnar i en "fakta-vidskepelse". Men till skillnad från dem som bara lägger skulden på vetenskapen och förnuftet, insisterar han på att vetenskapen till sist ändå rymmer en räddning. Det kräver emellertid att den förmår återuppta ansvaret för sina ursprungliga filosofiska motiv: ett skådande av "sakerna själva" och deras väsenssammanhang, som ytterst bottnar i ett intersubjektivt fritt och delat medvetandeliv. Däri, skriver han, ligger till sist den enda "sanna positivismen".

När Husserl skriver sin programskrift står världen på randen till vad som skulle bli det mest fruktansvärda och förödande krig som någonsin skådats och som sargade även den internationella vetenskapskultur som byggts upp under det föregående seklet. Direkt efter kriget görs ansatser att återskapa denna värld, men de ekonomiska kriserna mot slutet av tjugotalet sliter sönder den sköra väven på nytt. Det leder till en ny våg av nationalistisk politik med förnufts- och vetenskapsfientliga strömningar. När antisemitismen blir lagstadgad politik i Tyskland 1933 fråntas Husserl gradvis sina medborgerliga rättigheter och möjligheter att verka som emeriterad professor. I ett sista försök att ta till orda, bara något år före sin död, håller han ett föredrag i Wien 1935, om "Den europeiska mänsklighetens kris och filosofin". Hans budskap då var dock återigen att denna kris i första hand inte var politisk eller ekonomisk, utan snarare andlig eller intellektuell (*geistig*).

I dess centrum såg han nämligen förnuftet och vetenskapskulturen själv. Trots dess enorma landvinningar och framgångar hade den förlorat kontakten med sina egna rötter i en teoretisk livshållning med oändliga mål. Istället

hade den alltmer kommit att förvandlas till en instrumentell-teknisk verksamhet med det övergripande syftet att behärska naturen. Den uppgift han formulerade var inget mindre än att upprepa hela vetenskapskulturens grundläggande ursprung och syfte genom en ”tillbakafråga” eller ”återfråga” (*Rückfrage*). Att fråga ”tillbaka” är att besinna sig över vetenskapskulturens egen yttre och inre utveckling, och vad som skett med dess egna mål. En sådan historisk blick behöver inte stå i motsatsställning till ett principiellt filosofiskt mål. Tvärtom hänger de samman. Endast genom att reflektera kritiskt över vetenskapskulturens betingelser och genom att återvända till dess ursprungliga motiv kan vi hoppas på att kunna iscensätta den på nytt för en framtid. Premissen för hans kritiska resonemang är att många vetenskapsutövare inte längre bryr sig om dessa övergripande motiv. De fortlever i ett passivt övertagande av vetenskap som gällande tradition, inom vilken de levererar sin ”resultat” inom en given form. Mitt i sina framgångar har vetenskapen börjat förlora vad han ser som dess ursprungliga etiskt-intellektuella kompass. Husserl knyter också detta till en viss idé om Europa, som en kultur idealt sett präglad av en öppen universalism, som samtidigt inbegriper sin egen förvandling i en rörelse mot oändliga mål.

Från denna sista tid i Husserls liv finns en också en annan berömd uppsats på dryga tjugo sidor med titeln ”Geometrins ursprung som intentionalhistoriskt problem”. Där tar han de ideala vetenskaperna som exempel för att ställa frågan hur en kultur baserad på traderbara universella sanningar ens är möjlig. Texten publicerades av hans assistent Eugen Fink 1939, året efter Husserls död. År 1962 översattes den till franska av en då drygt trettioårig Jacques Derrida, tillsammans med en 200 sidor lång inledning, vilket var Derridas första egna publicerade bok. Ända sedan dess har den haft en nyckelställning som en av de mest omdiskuterade texterna i modern filosofi och som den kanske enskilt viktigaste bryggan mellan fenomenologi och dekonstruktion och det som ibland kallats ”poststrukturalism”.

Jag har två skäl att vilja lyfta fram denna text här. Det första skälet – som jag strax ska utveckla vidare – är att den på ett exemplariskt sätt fördjupar den tematik som skisserats ovan, och som knyter an till det övergripande temat uppreparhet och generaliserbarhet. Men det finns också en mer personlig bakgrund. I mitten av åttiotalet, när jag själv ännu var en ung student i matematik och teoretisk filosofi vid Stockholms universitet, hade jag ett samtal med Dick Haglund om fenomenologi och om förhållandet mellan Husserl och Derrida. Han föreslog då att jag till svenska skulle översätta just

Husserls uppsats om ”Geometrins ursprung” från tyska tillsammans med Derridas långa ”Introduktion” från franska. Han lovade att verka för att den i så fall skulle kunna publiceras på det då relativt nygrundade förlaget Thales med stöd av Hultengrens fond för filosofi. Jag var nog egentligen inte helt mogen denna uppgift, vare sig språkligt eller filosofiskt, men jag antog utmaningen. Under de följande åren levde jag med dessa båda texter och andades in deras tankevärld. Fem år senare kom boken ut och fick ett positivt mottagande.<sup>5</sup> I min avhandling skulle jag ta upp ett besläktat tema, om filosofins egen historicitet.<sup>6</sup> Avhandlingen skrevs inom ett forskningsprojekt som leddes av den polske filosofen Aleksander Orłowski där Dick också ingick. Bland svenskfödda filosofer i den äldre generation fanns ingen annan som hade en lika djup förståelse för fenomenologins bidrag till den moderna filosofin. Också senare under min bana skulle Dick komma att spela en viktig roll. Det är därför jag valt att dedicera denna text till hans minne.

\*

För att förstå hur Husserls uppsats knyter an till vårt aktuella tema är det viktigt att säga något om hans vetenskapliga bakgrund. Husserl skolades som matematiker, bland annat som elev till Weierstrauss. Hans tänkande vägledades av de exakta vetenskapernas idealiserande variationsmetod. Inom matematik och logik handlar ju vetande om att utveckla principiella modeller genom att formulera *eidetiska* samband eller väsenssammanhang. Matematik bedrivs med hjälp av ett tänkande som har sitt säte i en ändlig kropp och ett ändligt psyke, men dess mål är upptäckten och formulerandet av ideala strukturer med en universell räckvidd. Trots sitt tankemässiga och psykologiska ursprung är matematik inte något psykologiskt. Matematiska sanningar är inte sanningar *om* ett psyke eller hur det fungerar. De äger en *ideal* natur. Av detta skäl ägnar Husserl också en omfattande del av sina *Logiska undersökningar* åt en polemik mot den så kallade ”psykologismen”, tanken att matematiska principer ytterst sett skulle vara empiriska generaliseringar om det mänskliga psykets funktion.

Hela det fenomenologiska programmet, som det sedan skulle utvecklas, från *Logiska undersökningar* (1900) över *Ideer* (1913) och *Cartesianska meditationer* (1929) fram till de senare kulturfilosofiska texterna i *De europeiska vetenskapernas kris* (postumt publicerade 1954) handlar i grund och botten om att försöka utvidga detta eidetiska seende till en allmän kunskaps-

och medvetandeteori. Det finns en filosofisk blick, ett kategoriskt seende, menar Husserl, varigenom vi kan öva oss i att se mönster och väsen, som använder sig av detta slags väsens-variation. Det kan låta oss kartlägga de generella strukturerna eller principerna för ett visst fenomen-område, det kan gälla varseblivning, affektliv, etik, vetenskap, eller det som Dick Haglund studerade i sin avhandling: tid och tidsmedvetande.<sup>7</sup> Vad vi då söker avtäckta är ett fält av ideala sammanhang och gestalter. Efter hand skulle Husserl, med ett begrepp lånat från Kant och den tyska idealismen, komma att kalla detta fält för det *transcendental*. Filosofin blev för honom just detta: ett transcendental-fenomenologiskt väsenkådande utforskande av de intentionala strukturer inom vilka vårt vetande och erfarenhetsliv tar gestalt och vidmakthålls.

Husserl var inte ensam om att kritisera dem som menade att ideala sanningar kunde reduceras till psykologiska fakta. Striden om den så kallade ”psykologismen” var i själva verket en definierande fråga för hela hans generation. Om vi på ena sidan hittar en radikal empirist som John Stuart Mill så har vi på den andra sidan inte bara Husserl utan i än högre grad Gottlob Frege, den moderna logikens fader. För Frege bestod ideala sanningar just i deras tidlösa objektiva sanning, oavsett om de inses och förverkligas av ett mänskligt intellekt eller ej. Men till skillnad från Frege, insisterade Husserl på vikten av att ändå beakta hur denna objektivitet och giltighet är förbunden med ett tänkande medvetande som åskådar eller ”intenderar” den och som därmed ytterst *förverkligar* den i kunskapsakter. En sanning som inte är en sanning *för* någon, varken för ett subjekt eller för en gemenskap av subjekt, är till sist ingenting. Dess giltighet lever nämligen i det sammanhang inom vilket det realiserar och åsyftas, utan att för den sakens skull vara något blott subjektivt eller relativt.<sup>8</sup>

Det är just detta försök till balansgång mellan en relativistisk psykologism och en objektivistisk idealism som gör Husserls ansats så originell och så löfresrik. Hur kan vi på ett systematiskt tänka oss detta ontologiska *mellanrum* där objektivitet uppstår, lever och manifesteras? I den korta texten ”Geometris ursprung” ställer han just denna fråga som ett speciellt slags *historiskt* problem. Han utgår från det självklara, att de geometriska och matematiska sanningarna inte alltid har varit kända. Men på vilket sätt ska vi tänka oss att de blev till de sanningar de är? I de flesta fall vet vi inte vem den person – denne imaginära Thales – var som en gång formulerade dem. Men vi måste utgå ifrån att de någon gång tagit form i en enskild människas medvetande.

Men om så är fallet, hur ska vi förstå att de kom att få status av en allmänt omfattad ideal sanning?

I sin text reflekterar Husserl över detta skeende i principiella termer, sökandes efter vad vi skulle kunna kalla den vetenskapliga traditionsbildningens eidetik eller väsenslära.

Han tänker sig att det till att börja med kräver en *livsvärld* i vilken människor var infogade, som ett kommunikativt sammanhang av åtaganden och föremålsligheter. Det måste därtill föreligga ett gemensamt språk, varigenom upptäckten och insikten redan från början har karaktär av något kommunicerbart. Med andra ord: en eidetisk sanning som manifesterar sig inom ett medvetande eller psyke är till sin mening redan från början kommunicerbar och därmed upprepbar. Den pekar på så vis redan från början mot det *intersubjektiva* i det subjektiva, som en underliggande premis för att mönster och allmängiltigheter alls ska kunna uppträda och omfattas. Men för att en sådan sanning verkligen ska få karaktären av något bestående över tid som en traderbar och fortlevande sanning krävs något mer. Det räcker inte med ett delat språk och en kommunikationsförmåga. Det behövs också ett bestående materiellt fundament i vilket den kan avsättas och som kan hålla över tid, med andra ord något slags *skrift*.

För Husserl är utforskandet av dessa villkor ett led i sökandet efter de ideala meningarnas ursprung. Men syftet är inte bara vetenskapshistoriskt. Han vill med sin analys bereda vägen för vad han talar om som ett "självbevisande över vår samtidssituation". Att ställa detta slags frågor om vetenskapens principiella betingelser och historiska framväxt är nämligen också att bedriva en form av förnuftskritik. Vetenskapen samlar på sanningar som den samlar på resultat och bedrifter. När vi sitter mitt uppe i denna pågående sannings- och resultatjakt, inställda på leverans av prövbara och hållbara värden, vill han väcka frågan om vetandets inre historicitet, som också är en påminnelse om vad vetenskapen är just som ett levande traditionssammanhang.<sup>9</sup>

Genom att peka på hur trading vilar på språk och teckensystem, vill han samtidigt inskräpa att det som möjliggör det idealas fortbestånd också är det som på sikt kan undergräva det. Själva nedtecknandet och det tekniska överförandet av vetande rymmer både möjlighet och hot. Genom att bli till förvärv frikopplas vetandet från sin upphovsperson och sitt upphovssammanhang. Traditionen förs vidare, men utan sammanhang och grund. Arvtagarna möter inte längre frågan om meningarnas ursprungsbetingelser, istället tas de för givna som erövrade bestånd. Att tänka *historiskt* i denna mening



handlar inte om att historisera och relativisera eller om att reducera matematik och fysik eller för den delen filosofi till kulturhistoria. Tvärtom handlar det om att försöka avtäcka det principiella i just meningens historiska tillblivelse, som ett förlopp inom vilket vårt vetande upprätthålls och där vi bär ett individuellt ansvar för att dess mening och meningssammanhang inte förloras på vägen.

\*

Vad gör då Derrida med denna text och dess ansats? Han sympatiserar med Husserls försök att återanknyta till en levande, historisk besinnande rationalism. Den universellt strävande vetenskapskulturen behöver beakta sitt eget ändliga livssammanhang och sin upprinnelse i historia, kropp och språk. Men medan Husserl ser reaktivering av ursprungsbetingelserna som ett sätt att till sist återvinna den utopiska kraften i en rationalitet som han menar hotas av en inneboende meningsförlust och nihilism, så driver Derrida argumentet i en annan riktning. Medan Husserl konstaterar att all idealitet vilar på en reaktiverbar språk- och teckenstruktur, så trycker Derrida på den roll som i synnerhet skriften här tilldelas av Husserl själv, som inte bara en hjälpfunktion för idealisering utan som en irreducerbar betingelse för det rum inom vilket generaliserbar mening alls kan uppstå.

Det är en struktur som bygger på en principiell upprepbarhet. Ett meningsbärande tecken är redan i sig en inrättad upprepningsstruktur. En mening som förstås och som sedan kommuniceras har redan präglats av upprepningens möjlighet. En begriplig mening är inte något som en individ först kommer på och sedan med hjälp av språket kommunicerar till en annan. I den mån den framträder som en insikt om något och hur det är, så pekar den redan utöver sig själv, i riktning mot någon – kanske ännu okänd – för vilken den också – i princip – ska kunna te sig just så som den först uppträder, nämligen med en inre giltighet. Upprepbarheten har på så vis alltid redan föregripits, innan den faktiskt överförs. Varje insikt är alltid redan ett spår av sig själv, som pekar tillbaka mot något som inte kan gripas med mindre än att det redan pekar vidare mot sin upprepbarhet. Det är i denna mening som dekonstruktionen kan hävda att ett meningsursprung aldrig kan sammanfalla med sig självt. Som förnuftsvarelse bebor vi en *skillnad* mellan tecken och mening som alltid redan ägt rum. Intersubjektiviteten – den gemensamma meningen – framstår på så vis som både sanningens grund och dess hotande förfall.



Liksom den öppnar ett levande delande, så kan den också stelna till konvention, ett dött innehåll utan personligt ansvar. Men Derridas poäng är att denna meningens inre ändlighet och sårbarhet inte är något som vi en gång för alla kan bemästra med hjälp av en historiskt besinningsfull rationalism, utan snarare ett villkor som vi måste öva oss i att bebo som en oavvislig aspekt av all intellektuell kultur. Traditionens död och liv är oavvisligt förbundna.

Dessa vid det här laget bekanta, om än kanske inte alltid så väl förstådda, dekonstruktiva grundidéer kan locka till både nihilistiska och relativistiska slutsatser. Om vi drar slutsatsen att vi aldrig kan komma ner till en kärna av mening och verifierbar sanning, utan för alltid är förvisade till tecknens ändlösa skillnadsspel, kan ju själva idén om en mer grundad vetenskapskultur förefalla hopplös. Men poängen med att lyfta upp Husserl och Derrida tillsammans här är snarare påpeka hur nära besläktade de egentligen är. Steget från det fenomenologiska grundläggningsprojektet till den dekonstruktiva kritiken handlar till sist snarare om en förskjutning i betoning. Liksom Husserl, är Derrida angelägen om att förnuftet besinnar sin ändlighet och sin situation. Liksom Husserl ser han hur vetenskapen bara vidmakthålls inom ramen för en intersubjektiv livsvärld, inom ett språk och ett tecken-system, som också rymmer ett etiskt åtagande om ett delande med andra.

Genom att betona den oåtkomliga grunden hos en alltid redan inträffad upprepning och materiell rest, riktar Derrida till sist bara ett ännu strängare *memento mori* till hela vetenskapskulturen. Så snart den tror sig rätt och slätt äga det universella, och så snart den tror sig vila på en uppövad och säkrad begreppsram, så är det viktigt att påminna sig om att bakom det skenbart uppenbara har alltid redan ett beslut fattats. En mening har tagit gestalt, traditionen har fått en form, en begreppsram har fixerats.

I uppmärksammandet av detta behöver vi inte stanna som inför en absolut gräns. Snarare kan vi se det som en påminnelse om att vetandet och erfarenheten tvärs igenom sin universella strävan fortsätter att vila på något ändligt. Det rör sig dock inte om en slutet, en gång för alla fastslagen ändlighet, utan för något som öppnar sig för en omprövning och för ett experimenterande med själva grunden. Det allmänna och universella finns och fortlever genom en upprepbarhet som ytterst har sin ort i människans språkliga eidetiska medvetande och i hennes teoretiska seende. Genom språket sträcker hon sig ut bortom sig själv i riktning mot ett seende av varat, som öppnar för ett delande med andra. Därmed ställer det oss också inför tänkandet som en alltid utestående möjlighet att tänka annorlunda.

Vi befinner oss i ett språk som alltid redan har format oss. Men vi är också fria att inom detta system tänka över det och pröva det med hjälp av dess egna medel. Det handlar till sist, i både Husserls och Derridas fall, om att bli varse hur varje medvetande- och kunskapsliv befinner sig inom ett fortlöpande intersubjektivt traditionssammanhang. För den yngre Husserl bestod hotet mot vetenskapens idealitet i en relativistisk historicism och en naturalism som inte ser kunskapens och medvetandets unika plats i naturen. För den äldre Husserl hade hoten mot vetenskapens självurholkning blivit ännu mer påtagliga. Vetenskapens vilja att ta kontroll över naturen och över allt levande får till sist grepp också om människan, vars förnuftsverksamhet framstår som ett instrument för makt.

Till Husserls analys av vetandets inneboende frestelser och risker kan vi idag lägga nya. Inte bara mellan vetenskaperna, utan mellan olika grenar inom en och samma vetenskap, tenderar utövare att alltmer sluta sig kring sig själva och några normerande tidskrifter. Ansvar för att sträcka sig ut mot andra och mot en större gemenskap överlämnas till journalister. Vetenskap blir till små intressegemenskapers angelägenhet, som inte behöver försvara sig inför en bredare kritik. Dess resultat är varor som räknas hem i ett alltmer reglerat poängsystem knutet till resurstilldelning. Ett högkvalitativt tvärdisciplinärt tidskriftsprojekt som *Logos* vore knappast längre möjligt.

Dagens system bidrar till att den vetenskapliga värdegemenskapen med dess outtalade normer för excellens riskerar att urholkas ytterligare. Det gynnar inte övergripande filosofisk reflektion och inom-disciplinär kritik. Forskare reser förvisso mer och träffar fler forskare än någonsin tidigare, eftersom systemet premierar ”internationalisering” (så länge den är anglofon). Men måttet på vetenskapens prestanda – dess antal konferenser, papers och artikelpubliceringar – frikopplas från dess grund i en fri förnuftsgemenskap med oändliga mål. Inom ramen för detta globala merkantila poängsystems fungerar den dominerande universitetsfilosofin ofta bara som dess beskedliga försvarare, eftersom den själv också gynnas av dess tilldelningsmatriser. När den tar till orda för att försvara vikten av sitt och vetenskapens uppdrag når den sällan längre än att avvisa imaginära filosofiska fiender till något som brukar kallas ”upplysningens värden”. I denna diskussion behövs andra filosofiska röster, som utifrån en historisk kritisk reflektion kan lyfta frågan om den vetenskapliga kulturens egen inneboende strävan, riktning, och hot. Husserl har fortfarande något viktigt att förmedla, kanske i ännu högre grad nu än när han först tog till orda i början av det förra seklet.

Det är angeläget att fortsätta att blottlägga empirisk forskning som inte håller måttet genom att pröva den experimentella grunden för dess generaliseringar. Men det är också viktigt att fortsätta det mer principiella samtalet om forskningens djupare fäste i ett teoretiskt seende av allmängiltiga sanningar. En kultur som fortfarande implicit vägleds av detta ideal om ett *vita contemplativa* och som sätter av stora resurser för att upprätthålla dess institutioner, bär nämligen också inom sig risken att korrumpas. Den enskilda vetenskapsutövaren frestas att låta sig reduceras till leverantör av mätbara varor inom en upprättad ekonomi, där incitamenten flyttas från det vetenskapliga seendets praktik till en fiktiv marknadsekonomi med prissatta varor.

Kritiken av medvetandets naturalisering i Husserls tidigare arbeten vittnar inte om en förlegad idealism som modern psykologi och kognitionsvetenskap lagt bakom sig, vilket ibland hävdas. Tvärtom uttrycker det en bestående utmaning till den enskilda forskaren att axla den teoretiska vetenskapskulturens *ethos*. Att besinna sin plats inom ett menings- och traditionssammanhang är att inse att vetandet har sin inre historicitet och genealogi samtidigt som det utgör en ändlig öppning mot det oändliga. Inom denna horisont fortlever det bara som ett ansvarstagande för det egna seendets grund och räckvidd, som ett delat kall till förmån för en öppen gemenskap i vardande.

### Noter

- 1 Herakleitos, *Fragment*, övers. H. Rehnberg & H. Ruin (Lund: Propexus, 1997), fragment DK 10, s. 33.
- 2 Nietzsche, "Om sanning och lögn i utomoralisk mening", övers. M. Holmqvist, i Band 2 av *Samlade skrifter* (Stockholm: Symposion, 2005), ss. 501-516, s. 501 och s. 507.
- 3 Herakleitos *ibid.*, DK2, s. 19.
- 4 På svenska finns den i övers. av K. Weigelt, i volymen Edmund Husserl, *Fenomenologin och filosofins kris* (Stockholm: Thales, 2002).
- 5 Jacques Derrida, *Geometrins ursprung*, övers. H. Ruin (Stockholm: Thales, 1990).
- 6 *Enigmatic Origins: Tracing the Theme of Historicity through Heidegger's Works* (Stockholm: Almqvist & Wiksell, 1994)
- 7 Dick Haglund, *Perception, time, and the unity of mind: problems in Edmund Husserl's philosophy* (Göteborg: Institutionen för filosofi, 1977).

- 8 För en utmärkt genomgång av det husserlska intentionalitets-begreppet och dess historiska sammanhang, se Dick Haglund "Intentionalitet som medvetandets väsen", i Orlowski & Ruin (red.), *Fenomenologiska perspektiv. Studier i Husserls och Heideggers filosofi* (Stockholm: Thales, 1997), ss. 35-64.
- 9 För en aktuell analys av förhållandet mellan Husserl, hermeneutik, Heidegger och Derrida på detta tema, se även min "Historicity and the Hermeneutic Predicament: from Yorck to Derrida", in Dan Zahavi (ed.), *The Oxford Handbook of the History of Phenomenology* (Oxford: Oxford UP, 2018), ss. 717-733.

### Referenser

- Derrida, Jacques, *Husserl och geometrins ursprung*, övers. H. Ruin (Stockholm: Thales, 1990).
- Haglund, Dick *Perception, time, and the unity of mind: problems in Edmund Husserl's philosophy* (Göteborg: Institutionen för filosofi, 1977).
- Haglund, Dick, "Intentionalitet som medvetandets väsen. Det fenomenologiska synsättet", i *Fenomenologiska perspektiv. Studier i Husserls och Heideggers filosofi*, red. A. Orlowski & H. Ruin (Stockholm: Thales, 1997), ss. 35-64.
- Herakleitos, *Fragment*, övers. H. Rehnberg & H. Ruin (Lund: Propexus, 1997).
- Husserl, Edmund, "Geometrins ursprung", övers. H. Ruin, i Jacques Derrida, *Geometrins ursprung* (Stockholm: Thales, 1990) ss. 175-202.
- Husserl, Edmund, *Fenomenologin och filosofins kris*, övers. K. Weigelt & H. Ruin (Stockholm: Thales, 2002).
- Nietzsche, "Om sanning och lögn i utomoralisk mening", övers. M. Holmqvist, i Band 2 av *Samlade skrifter*, utg. H. Ruin m.fl. (Stockholm: Symposion, 2005).
- Ruin, Hans, *Enigmatic Origins: Tracing the Theme of Historicity through Heidegger's Works* (Stockholm: Almqvist & Wiksell, 1994).
- Ruin, Hans, "Historicity and the Hermeneutic Predicament: from Yorck to Derrida", in Dan Zahavi (ed.), *The Oxford Handbook of the History of Phenomenology* (Oxford: Oxford UP, 2018) ss. 717-733.

## De medverkande:

**Anna Dreber Almenberg** är professor i nationalekonomi vid Handelshögskolan i Stockholm. Hon forskar främst om beteendekonomi och replicerbarhet. Ledamot av Sveriges unga akademi.

**Elisabet Engdahl** är professor i svenska språket vid Göteborgs universitet och ledamot av Vetenskaps- och Vitterhetssamhället i Göteborg samt av Vitterhetsakademien.

**Jan-Eric Gustafsson** är professor em. i pedagogik vid Göteborgs universitet. Hans forskning har huvudsakligen rört frågor inom pedagogisk psykologi, med särskilt fokus på individuella differenser i kognitiva förmågor och andra förutsättningar för lärande.

**Peter Jagers** är professor em. i matematisk statistik vid Chalmers tekniska högskola och Göteborgs universitet samt ledamot av Vetenskapsakademien och Vetenskaps- och Vitterhets-Samhället i Göteborg. Han har författat skrifter om stokastiska processer, särskilt förgreningsprocesser och populationsmodeller, samt strödda bidrag till sannolikhetsteori och statistisk slutledning.

**Arne Jarrick** är sedan 1997 professor i historia vid Stockholms universitet. Han har varit generalsekreterare för humaniora och samhällsvetenskaperna vid Vetenskapsrådet och är ledamot av Vetenskapsakademien och Vitterhetsakademien. Hans forskningsintressen är mentalitetshistoria, kulturell evolution samt världshistoria.

**Magnus Johannesson** är professor i nationalekonomi vid Handelshögskolan i Stockholm. Han forskar främst om beteendekonomi och replicerbarhet. Han är ledamot av Vetenskapsakademien.

**Sven-Eric Liedman** är professor em. i idé- och lärdoms historia vid Göteborgs universitet, med en omfattande produktion om den moderna erans filosofi och vetenskap bakom sig. Han är ledamot av Vetenskaps- och Vitterhets-Samhället i Göteborg.

**Lars Nyberg** är professor i neurovetenskap vid Umeå universitet och föreståndare för Umeå centrum för funktionell hjärnabbildning (UFBI). Hans forskning handlar om sambandet mellan hjärnans funktioner och minnesfunktioner och hur detta samband förändras över livsspannet.

**Hans Ruin** är professor i filosofi vid Södertörns högskola, med inriktning mot fenomenologi, fransk filosofi samt försokratiker. Han är medlem av Moderna museets forskningsråd



