

Statsvetenskaplig tidskrift

Årgång 122 · 2020 / 4

Ny följd, årg 99. Utgiven av Fahlbeckska stiftelsen.

REDAKTIONSSEKRETERARE *Magnus Jerneck* (ansvarig utgivare)

BITR. REDAKTIONSSEKRETERARE *Björn Badersten*

LITTERATURREDAKTÖR *Björn Östbring*

FÖRBUNDSREDAKTÖR *Katarina Roos*

REDAKTIONSÅD *Katarina Barrling*, Uppsala universitet; *Ann-Marie Ekengren*, Göteborgs universitet; *Edward Deverell*, Försvarshögskolan; *Rolf Hugoson*, Umeå universitet; *Mats Lindberg*, Örebro universitet; *Charlotta Söderberg*, Luleå tekniska universitet; *Peter Strandbrink*, Södertörns högskola; *Elin Wihlborg*, Linköpings universitet; *Susanne Wallman-Lundåsen*, Linköpings universitet; *Björn Badersten*, Lunds universitet, biträdande redaktionssekreterare; *Magnus Jerneck*, Lunds universitet, redaktionssekreterare; *Björn Östbring*, Lunds universitet, litteraturredaktör; *Katarina Roos*, Umeå universitet, förbundsredaktör

TEKNISK REDAKTÖR *Joshua Alvarez*, Sven Eighteen

Tidskriften utkommer med fyra nummer per år.

PRENUMERATIONSPRIS 2020 430 kr, enstaka nummer 100 kr. Medlemmar i Statsvetenskapliga förbundet och studenter erhåller tidskriften till rabatterat pris. Prenumeration sker via hemsidan, genom insättning på plusgiro 27 95 65-6 med angivande av namn och adress eller genom meddelande till tidskriftens expedition. Eftertryck av tidskriftens innehåll utan angivande av källan förbjudes.

ADRESS Statsvetenskaplig tidskrift, Box 52, SE-221 00 Lund, Sverige

TELEFON 046-222 97 77 (Jerneck) 046-222 01 59 (Badersten)

TELEFAX 046-222 40 06

E-POST statsvetenskaplig.tidskrift@svet.lu.se

HEMSIDA www.statsvetenskapligtidskrift.org

TRYCK Mediatryck, Lund 2020

ISSN 0039-0747

Statsvetenskaplig tidskrift

Statsvetenskaplig tidskrift är den svenska tidskriften för statsvetare. Tidskriften publicerar vetenskapliga uppsatser, översikter och litteraturgranskningar inom statsvetenskapens alla subdiscipliner och är därtill huvudorgan för Statsvetenskapliga förbundet (SWEPSA). Svenska är huvudspråket, men tidskriften publicerar också texter på danska och norska samt i undantagsfall på engelska av författare som inte har ett skandinaviskt språk som arbetspråk.

Statsvetenskaplig tidskrift utges av Fahlbeckska stiftelsen vid Lunds universitet, med fyra nummer per år. Tidskriften utkommer år 2020 med sin 122:a årgång. Stiftelsen leds av ett flervetenskapligt sammansatt kollegium (professorer i statsvetenskap, historia, skatterätt, förvaltningsrätt, antikens kultur och samhällsliv, nationalekonomi, ekonomisk historia, socialt arbete, sociologi och statistik). Kollegiet beslutar om tidskriftens budget och övergripande policyfrågor. Tidskriftens redaktion, som är helt fristående i publicistiska frågor, består av en redaktör, en biträdande redaktör, en litteraturredaktör och ett aktivt arbetande redaktionsråd på åtta personer som sammanträder minst två gånger per år. Redaktionsrådet utses i samråd med Statsvetenskapliga förbundet och representerar flertalet statsvetenskapliga universitetsinstitutioner i Sverige.

Samtliga artiklar kvalitetsgranskas internt av två personer i tidskriftsredaktionen. Den externa kvalitetsgranskningen, med två anonyma referees per artikel, omfattar alla publicerade artiklar utom litteraturgranskningar, kortare översikter/debattartiklar och sakkunnigutlåtanden. Författaren får i samband med redaktionens beslut om publicering/refusering ta del av utlåtanden från referees.

The Swedish Journal of Political Science

Statsvetenskaplig tidskrift (The Swedish Journal of Political Science) is the Swedish journal for political scientists. The journal publishes scientific essays, review articles and literature reviews in all the subsidiary disciplines of political science and is the principle organ of the Swedish Political Science Association (SWEPSA). While Swedish is its main language, the journal also publishes texts in Danish and Norwegian and, in exceptional cases where the author does not have a Scandinavian language as his/her working language, English.

Statsvetenskaplig tidskrift is published quarterly by Lund University's Fahlbeck Foundation, and in 2020 will be issuing its 122nd volume. The Foundation is led by a multidisciplinary board (professors of political science, history, fiscal law, administrative law, classical culture and social life, economics, economic history, social work, sociology and statistics), which decides on the journal's budget and overarching issues of policy. The editorial office, which is wholly independent as regards matters related to publishing, comprises an editor, a deputy editor, a literature editor, a technical editor and an active, eight-member editorial committee that meets at least twice a year. This committee is appointed in consultation with SWEPSA and represents the majority of political science faculties in Sweden.

All articles undergo an internal quality review by two members of the editorial office, and all published articles – excepting literature reviews, short review articles/opinion pieces and expert reports – an external quality review by two anonymous referees per article. Once the editorial office has made its decision to publish or reject a paper, a referees' statement will be sent to its author(s).

Innehåll 2020 / 4

Tema: Beslutsfattande

ANNIKA WALLIN Tema: Beslutsfattande.....	517
JESSICA ABBOTT Ett evolutionärt perspektiv på beslutsfattande.....	519
KALLE ÅSTRÖM Matematiska modeller för beslutsfattande hos människor, djur och maskiner	531
MARIA HEDLUND När maskiner fattar beslut – vem är ansvarig?.....	545
ANNIKA WALLIN Beslutsfattande ur tre perspektiv	567
LENA WAHLBERG Fri bevisprövning och evidensbaserad medicin. En jämförelse av två modeller för evidens och beslut.....	581
ARVID ERLANDSSON Hjälpdilemman: Beslutsfattande när man inte kan hjälpa alla i nöd	601
ERIK MOHLIN Spelteoretisk jämvikt som ett resultat av inläring eller evolution.....	625
TONY INGESSON Antagonistens roll. Vad Napoleon och KGB kan lära oss om beslutsfattandets konst inom militär ledning och underrättelseanalys	655

Översikter och meddelanden

JOHAN WENNSTRÖM & ÖZGE ÖNER Political Hedgehogs. The Geographical Sorting of Refugees in Sweden	671
KNUT MIDGAARD Demokrati med radvalg og fondvalg	687
KJELL Å MODÉER American Exceptionalism. Supreme Court och det historiska argumentet.....	691
LARS NIKLASSON FÖRVALTNINGSPOLITISKA BETRÄKTELSE: Skapar FN:s hållbarhetsmål ordning på den globala politiska nivån?	703

Essä

JON WITTRÖCK En kontroversiell klassiker731

Statsvetenskapliga förbundet

KATARINA ROOS Swepsa 2020737

Litteraturgranskningar

BENNY CARLSON De motade inte Adolf i grind

Rüdiger Barth & Hauke Friedrichs: *Dödgrävorna: Den sista vintern i Weimarrepubliken.*

Éric Vuillard: *Dagordningen.*

Tim Bouverie: *Appeasing Hitler: Chamberlain, Churchill and the Road to War.*

Pierre Gilly: *Europa mot katastrofen: Förspelet till andra världskriget.*743

CARSTEN ANCKAR

John Gerring & Wouter Veenendaal: *Population and Politics: The Impact of Scale.* 749

ULF LINDSTRÖM

Arild Stubhaug: *Stein Rokkan – fra periferi til sentrum.*..... 753

ANDERS SUNDELL

Stuart Ritchie: *Science Fictions: Exposing Fraud, Bias, Negligence and Hype in Science.* 759

**Statsvetenskaplig tidskrifts hemsida: www.statsvetenskapligtidskrift.org
Där finns bl. a.:**

- Utförliga anvisningar till författare om utformningen av manuskript för Statsvetenskaplig tidskrift (kan hämtas hem som pdf-dokument).
- Tidigare årgångar av Statsvetenskaplig tidskrift i fulltextformat – med sökfunktion till artikelarkivet.
- Information om innehållet i kommande nummer och en aktuell utgivningsplan för Statsvetenskaplig tidskrift.
- Information om prenumerationspriser och möjlighet att teckna prenumeration på Statsvetenskaplig tidskrift.
- Kontaktuppgifter till redaktionen och redaktionsrådet för Statsvetenskaplig tidskrift.

Tema: Beslutsfattande

Annika Wallin

Vad förtjänar att kallas ett beslut? Är det att en morgon plocka upp de randiga strumporna ur strumplådan, eller att avkunna dom i ett brottsmål? I båda fallen finns det mer än en sak att välja på, och något – utöver slumpen – som gör att det blir just detta utfall. Det kan anses vara grundstenarna i ett beslut.

Tanken bakom detta specialnummer är att beskriva beslutsfattande ur olika vetenskapliga perspektiv. Därför har vi bjudit in författare från så skilda ämnen som biologi, ekonomi, juridik, matematik, psykologi och statsvetenskap. En del har valt att brett presentera ämnets syn på beslutsfattande, andra har gett konkreta exempel på vad just deras forskningsinriktning kan säga om beslut. Alla har ändå ägnat sig åt samma grundläggande fenomen. Komponenterna är desamma: En agent eller flera, tillgänglig (faktiskt eller potentiellt) information, en mekanism eller process för att välja, ett kriterium för vad som är ett bra beslut, och i det sammanhanget, en metod för att utvärdera valet.

Agenten är i vissa fall en biologisk varelse – människa eller djur – och i andra fall artificiell. Den kan ses som någon som agerar på egen hand, eller i ett socialt sammanhang i samspel med en eller flera kontrahenter. Den fattar beslut antingen på individ- eller systemnivå. Agenten kan också vara lekman eller utövare av en profession.

Oavsett hur beslutsagenten definieras kommer den att behöva inhämta relevant information om situationen och på något sätt handla med utgångspunkt i detta. Det kan göras med hjälp av etablerade modeller eller utifrån erfarenhet och intuition. Information kan man hitta på egen hand, eller nå genom andras agerande. Ibland behöver man också värdera informationen – är den tillförlitlig eller kommer den med en viss osäkerhet? Är beslutsunderlaget bristfälligt? Kan man lita på motparten?

Vidare behöver agenten på något sätt förhålla sig till informationen för att avgöra hur den ska handla. Oftast krävs det någon form av inläring för att kunna bedöma informationens relevans och avgöra hur den bäst omsätts i beslut. För att kunna göra detta behöver vi (oftast) ha en aning om hur bra utfallet av beslutet är, antingen på kort, eller lång sikt (evolutionär tidsrymd). Men hos en enskild individ handlar beslutsfattandet också om hur den upplever

situationen; känslomässigt och strategiskt, särskilt ifall beslutsfattandet är kopplat till andra agenter.

Beslutsfattandet behöver också utvärderas. Utvärderingen behövs för individens egen skull – utan inlärning är det svårt att finslipa beslutsstrategier och träna sitt omdöme. Målet kan vara att fatta acceptabla eller optimala beslut. Utvärderingen behövs också av samhälleliga skäl. Målet kan vara moraliskt försvarbara beslut. När vi t ex använder AI-system för att fatta beslut, hjälper behövande eller överlåter åt en domare att fritt bevisvärdera innebär detta ett samhälleligt risktagande, särskilt om besluten har långtgående konsekvenser för enskilda individer. Detta risktagande måste komma med etiska och juridiska omdömen som berör frågor om autonomi, lagrum och ansvarsutkrävande.

Ett evolutionärt perspektiv på beslutsfattande

Jessica Abbott

Abstract

The ability to make well-reasoned and deliberate decisions is often seen as a uniquely human capacity. And it's true that most other animals can't think as far ahead, or understand the consequences of their decisions, as well as a human can. However most of the decisions that we make in our everyday life aren't actually very well-reasoned or deliberate, and are made more or less spontaneously. Which is a good thing! If we needed to do a major review of our internal mental and physical state every time we were faced with a decision of what type of pastry to have with our morning coffee (croissant or cinnamon bun?), we wouldn't have time for much else. It's likely that most animal decision-making is similar to this type of spontaneous decision, which is based on "gut feeling" rather than rational weighing of pros and cons. This means that we can potentially gain insight into daily human decision-making by studying animal decision-making from an evolutionary perspective. In this article, I will discuss three basic types of animal decisions: decisions based on self-acquired information (individual decisions), decisions based on information from other individuals (social information), and decisions that are made in a group (collective decisions).

Introduktion

Förmågan att fatta välgrundade och noga övervägda beslut ses oftast som typiskt mänskligt. Och det är sant att de flesta andra djur inte kan tänka lika långt fram i tiden och förstå konsekvenserna av sitt agerande på samma sätt som en människa. Men merparten av de beslut som vi fattar i vardagen är egentligen inte speciellt välgrundade och noga övervägda, utan kommer ganska spontant (Wallin, i detta nummer). Och det är tur! Hade man behövt grundligt rannsaka sig själv varje gång man ska bestämma sig för kanelbulle eller wienbröd till fiket, skulle man inte hinna med så mycket annat. Vi kan förmoda

Jessica Abbott är lektor vid Biologiska institutionen, Lunds universitet.
E-post: jessica.abbott@biol.lu.se

att de flesta beslut som djur fattar liknar dessa spontana beslut, som går mer på ”magkänsla” än på förnuftet. Det innebär att vi kanske kan få insikt om människors dagliga beslutsfattande genom att anlägga ett evolutionärt perspektiv på djurs. Här kommer jag att diskutera tre olika former av beslut: beslut som bygger på egeninhämtad information (individuella beslut), beslut som bygger på information från andra individer (social information), och de som fattas tillsammans i en grupp (kollektiva beslut).

Individuella beslut

Precis som vi, måste andra djur fatta många beslut varje dag. Ska jag leta efter mat eller försöka hitta en partner? Är det dags att gå en runda och vakta mitt territorium, eller ska jag vila mig ett tag? Ska jag hålla uppsikt efter rovdjur eller ska jag dricka vatten? I alla dessa fall har vi olika behov som måste ställas mot varandra. I allmänhet brukar djur ta hand om det mest brådskande eller viktigaste behovet först. Det innebär att när ett djur ska bestämma sig för att utföra ett visst beteende eller inte, måste både extern och intern information samlas in och bearbetas av hjärnan för att rangordna olika behov innan det kan välja det mest passande beteendet för stunden. Tvekan eller velande förekommer framförallt när flera olika konkurrerande behov väger ungefär lika tungt. Till exempel har försök med bananflugor visat att en hane som har utsatts för sömnbrist tar chansen att sova så fort han lämnas i fred, men inte om det finns en hona i närheten. Då håller han sig vaken för att försöka uppvakta henne istället, och prioriterar sex framför sömn (Beckwith et al. 2017). I detta fall måste flugan integrera extern information om en honas närvaro som kommer från ögonen och luktsinnet med intern information om trötthet. Om man stänger av nervcellerna som känner av honans doft kan man rubba beteendehierarkin, och få fram hanar som alltid väljer sömn först. Det bekräftar att insamling och integrering av information är viktigt för att styra beteendet hos djur.

Det finns all anledning att tro att olika individer skiljer sig åt när det gäller rangordningen av olika behov. Dels kan det handla om hur känslig man är för olika externa intryck; en hane som är blind eller som saknar luktsinne kommer att ha svårare att upptäcka när en hona är närvarande, och därmed kanske fatta ett annat beslut än en normal hane. Dels kan det handla om hur informationen integreras i hjärnan; till exempel kan vissa individer vara mer benägna att gå från att uppvakta honan till att sova vid en högre eller lägre trötthetsnivå än andra. Det innebär att djurs beslutsmekanismer kan anpassa sig evolutionärt till sin miljö. Om ett djur är välanpassat till sin miljö, och så länge informationen som kommer in är korrekt, kommer djuret oftast att bete sig på ett sätt som är till synes rationellt. Men vi vet att människor inte alltid beter sig på ett rationellt sätt, speciellt när vi tar spontana beslut (Kahneman 2011). Hur kan

det då komma sig? Finns det exempel på djur som fattar dåliga beslut, eller är det något som är speciellt mänskligt?

Tidspreferenser

Om den tillgängliga informationen är bristfällig eller om ett djur utsätts för en annorlunda miljö, kan detta såklart leda till missanpassat beslutsfattande. Det är en viktig anledning till varför det är svårt att studera djurs kognition i laboratoriet – den onaturliga miljön kan leda till att djur betar sig på ett sätt som inte speglar deras naturliga mönster. Ett exempel på detta är tidspreferenser. Både djur och människor visar en preferens för en liten belöning som kommer snabbt jämfört med en stor belöning som man måste vänta på. I en osäker värld är detta en bra strategi, för mycket kan hända medan man väntar på den stora belöningen. Man kanske blir uppäten av ett rovdjur, eller så kanske den stora belöningen försvinner under tiden. Man brukar betrakta oförmågan att vänta på en stor belöning som ett mått på impulsivitet, och allt tyder på att djur i så fall är mer impulsiva än människor. En duva kan till exempel vara beredd att vänta några sekunder på en stor belöning, och en apa kan vänta i flera minuter, men en människa är ofta beredd att vänta i flera månader (Hayden 2016).

Problemet är att ett typiskt försök kring tidspreferenser oftast utgör en ganska onaturlig situation för ett djur, där två olika belöningar (till exempel matbitar) presenteras samtidigt, men den mer attraktiva varianten blir tillgänglig först efter en tidsfördröjning. En mer naturlig utmaning är födosökning, där en individ måste bestämma sig för om den ska stanna kvar på samma område och leta mat, eller bege sig till ett nytt sökområde. Ju längre djuret stannar kvar på samma plats, desto mindre mat finns det kvar att hitta, så det är smart att röra på sig innan det blir för tidskrävande att hitta de sista smulorna som finns kvar. Det finns mycket data som visar att djur är duktiga på att ge sig av till nya sökområden vid nästan den optimala tiden (Hayden 2016). Dessa beslut om födosökning innebär också att en framtida belöning (mattillgång i det nya området) måste vägas mot en omedelbar belöning (mattillgång här och nu), men i detta fallet verkar djur kunna utgå ifrån ett längre perspektiv än bara ett par sekunder eller minuter. Det tyder på att det onaturliga upplägget med försök med tidsperspektiv kanske resulterar i mindre relevanta beslut, och att skillnaden i impulsivitet mellan djur och människor till viss del förmodligen har överskattats (Hayden 2016).

Sunk cost bias

Ett annat klassiskt exempel på irrationella beslut som har studerats hos både människor och djur är det så kallade "sunk cost bias", även ibland känt som "Concorde fallacy" efter överljudsflygplanet. Sunk cost bias (härefter kallad

SCB; kan översättas som felslutet om icke-återvinningsbara kostnader, men standardiserad svensk översättning saknas) innebär att man blir mer ovillig att ändra sina planer ju mer tid eller energi man redan har investerat, trots att det kanske hade varit bättre att avbryta processen. Så även om Concordeprojektets slutgiltiga kostnad var nästan tjugo gånger högre än den ursprungligen budgeterade (Seebass 1997), tillfördes projektet mer pengar gång på gång med motiveringen att man redan hade satsat så mycket att det vore dumt att avbryta projektet i förtid (Curio 1987). SCB är ett relativt vanligt förekommande fenomen hos människor, men har tidigare betraktats som osannolikt hos djur (Curio 1987, Arkes & Ayton 1999).

Det har mest undersökts i förhållande till omsorg om avkomman där storleken på kullen manipuleras och föräldrarnas benägenhet att fortsätta investera i omsorg jämförs med hur mycket som har investerats hittills. Man kan till exempel mäta hur lång tid det tar för en fågelhona att återvända till boet efter att hon har skrämts bort av ett rovdjur. En hona som återvänder snabbt äventyrar den egna säkerheten, eftersom rovdjuret kanske är kvar, men en hona som väntar längre på att återvända kan däremot vara mer säker på att rovdjuret har gått vidare och försvunnit. En hona som återvänder snabbt prioriterar därmed ungarnas väl och ve framför sig själv, och forskare kan se tiden det tar henne att återvända som ett mått på hennes benägenhet att fortsatt investera i ungarna. En hona som lider av SCB skulle basera sitt beslut på investeringen hon har gjort hittills, dvs hur många ägg hon har värpt, och inte kullens framtida chanser, dvs antal ungar i boet just då. Man kan testa vilken faktor som är viktigast genom att experimentellt öka eller minska storleken på kullen och sedan studera honans beteende. Flera olika studier har visat att det endast är det nuvarande antalet ungar i boet som är viktigt i detta sammanhang, inte hur många som har funnits tidigare, vilket är ett tecken på att djur inte drabbas av SCB när det gäller "barnomsorg" (Magalhães & White 2016).

Det är i dessa studier en utmaning att kontrollera för hur den förväntade belöningens storlek kan förändras över tid i naturen. Vissa försök har visat att djur, om de hotas av rovdjur, är mindre benägna att överge en kull när ungarna är äldre jämfört med när de är små. Detta skulle kunna tolkas som SCB, eftersom föräldrarna har investerat mer energi i äldre ungar. Men många djur har en begränsad tid tillgänglig för reproduktion varje år – när parningssäsongen eller sommaren är slut, får man vänta till nästa år. Eftersom ungarna är äldre senare under säsongen, är det svårt att särskilja effekten av den hittillsvarande investeringen, från effekten av tiden som finns kvar. Det är ingen idé att överge en hotad kull med äldre ungar om det ändå är för sent att försöka producera en ny kull. I de fallen där man har hittat bevis för SCB i förhållande till "barnomsorg" kan man oftast inte utesluta att det bara berodde på att det inte längre var lönt att byta strategi (Arkes & Ayton 1999).

Däremot finns det resultat som tyder på att djur också kan drabbas av SCB när det gäller födosökning. Försök med möss och råttor som fick välja mellan matstationer som erbjöd mer eller mindre smaskig mat med olika väntetider, visar att de beter sig på samma sätt som människor. Försöket var upplagt för att likna dilemmat som vi står inför när vi ska välja kö i mataffären. Den ena kön kanske är längre, men kan röra på sig snabbare ifall alla i kön bara har en liten korg och inte en stor kundvagn. Hos människor vet vi att ju längre tid man har stått kvar i en kö och väntat, desto mindre benägen är man att byta till en annan kö (vilket är ett tecken på SCB), även om den andra kön rör på sig snabbare. I försöket med råttorna och mössen, fick djuren till att börja med tränas till att koppla ihop en ljudsignal med en tidsfördröjning, där ljusare toner betydde en längre väntetid. På så sätt kunde forskarna förmedla till djuren hur länge de skulle behöva vänta i "kön". Sedan fick djuren en begränsad tid att röra sig mellan olika matstationer med olika väntetider, för att försöka samla in så mycket mat som möjligt. Väntetiden var i detta fall inte kopplad till hur attraktiv matbelöningen var, och djuren kunde avbryta och gå vidare till nästa matstation ifall de tyckte att väntan blev för lång. Man såg att både mössen och råttorna var mindre benägna att avbryta sin väntan ju längre tid som hade gått, precis som människor, även om de skulle kunna få mat snabbare genom att byta station (Sweis et al. 2018). Man har även sett liknande beteende hos duvor (Magalhães & White 2016).

Återigen, detta är en förhållandevis artificiell situation för djuren, så man kan undra om detta beteende faktiskt är gynnsamt i naturen. Det finns många exempel på dåliga lösningar inom naturen, som till exempel pandans "tumme", som egentligen är en förlängd bit av handleden och betydligt mindre rörlig än människans tumme. Man kan alltså inte utgå från att alla beteenden i naturen måste vara fördelaktiga på något sätt. Men att samma mönster finns hos så vitt spridda arter som möss, människor och duvor tyder på att det förmodligen finns en fördel med denna variant av SCB, alltså att inte vilja ge upp för snabbt. Det finns några olika möjliga fördelar. Om det överlag råder ett samband mellan hur svår en belöning är att få och hur attraktiv den är, så kan tiden som individen redan har investerat ge en fingervisning om hur bra utdelningen kommer att vara. I så fall kan det vara smart att vänta lite längre i förhoppningen om att få en bättre belöning. Inlärning kan också vara viktig. Har en strategi tidigare varit lyckad är det större chans att en individ provar samma sak igen. Det kan gynna individer som är uthålliga i olika sammanhang (Magalhães & White 2016). När det gäller människor kan det sociala samspelet även vara en viktig faktor. Vi vill till exempel inte framstå som slösaktiga, eller vill värna om vårt goda rykte och visa att vi kan genomföra det vi har lovat (Arkes & Ayton 1999).

Social information

Människor är sociala djur, och det innebär att vi som hjälp vid beslutsfattande kan samla in information från andra individer. Detsamma gäller andra grupplevande arter. En apa kan till exempel se var andra individer hittar mat, och använda den informationen för att själv bestämma sig för var den ska leta efter godbitar. Fåglar använder sig av speciella läten för att kommunicera när ett rovdjur är närvarande, och andra individer kan rätta sig efter det. Eftersom information som man inhämtar själv i de flesta fall är mest pålitlig, kan social information ses som ett tillägg till egeninhämtad information. Den ersätter egen information endast när sådan saknas (Mann 2018). Matematiska modeller visar att djur är mer benägna att nå någon form av konsensus i sitt beteende – till exempel samordnade rörelser inom flocken – om miljön är ”stökig”, dvs när den egna informationen inte är så pålitlig. I så fall är det bättre att lita på hela gruppens samlade information, eftersom den borde spegla verkligheten bättre än de egna bristfälliga intrycken. Djur kan också kopiera varandras beslut, speciellt i de fallen där individen själv saknar information om vad som är det bästa alternativet (Mann 2018). Konkreta exempel på detta är bananflugehonor, som kopierar varandras preferenser för att lägga ägg på en viss plats (Battesti et al. 2015), eller fågelhonor som väljer att parasig med samma hane bara för att andra gjort det (Höglund et al. 1995).

Social information används på ett speciellt sätt hos bebisar och unga schimpanser. Man har nämligen sett att bebisar är mer benägna att kopiera ett beteende om det utförs av flera olika personer. Experiment har visat att det är antalet personer som utför beteendet som är viktigt, inte hur många gånger bebisen får se det. Om en person utför ett beteende tre gånger är det inte alls lika sannolikt att bebisen kommer att kopiera beteendet som om tre olika personer utför beteendet en gång var. Detsamma gäller schimpansbebisar, som lever i grupp, men inte orangutanger, som lever ensamma (Haun et al. 2012). Det tyder på att förmågan att använda social information för att rätta sig efter gruppens normer är väldigt grundläggande hos människan, och med största sannolikhet har utvecklats långt innan vi ens blev mänskliga.

Många sociala djur har en dominanshierarki, med en ledare, eller ibland ett dominant par (en dominant hane och en dominant hona). Mer dominant individer kan till viss del bestämma över andra, till exempel genom att inte låta mindre dominant individer få fri tillgång till en attraktiv matkälla eller en partner. Men fram till relativt nyligen fanns det inga tecken på att information från en mer dominant individ skulle väga tyngre än information från en mindre dominant individ. En studie från 2015 har för första gången visat att schimpanser kopierar dominant individer mer än lågt rankade individer när de löser en ny uppgift (att öppna en låda för att få en vindruva) (Kendal et al. 2015). Eftersom många sociala djur kan känna igen mer dominant individer, är det troligt att andra arter kan ha en liknande tendens att följa auktoriteter.

Kollektiva beslut

När en grupp människor ska ta ett gemensamt beslut brukar det kräva en del diskussion och förhandlingar. Det kan därför framstå som svårt för djur att fatta gemensamma beslut, eftersom de saknar ett talat språk. Istället kanske man tänker att det är gruppleddaren som bestämmer över alla andra individer hos sociala djur, eller att varje individ sköter sig själv och att synkroniserat beteende bara är ett resultat av att alla individer använder sig av nästan samma information. Men det finns faktiskt exempel på grupper som fattar gemensamma beslut hos så vitt skilda arter som bakterier, honungsbin, och vilda hundar. Även om kommunikationssättet skiljer sig åt mellan arterna, måste i samtliga fall en kritisk massa inom gruppen uppnås för att ett beslut ska genomföras.

Bakterier kommunicerar sinsemellan med hjälp av olika signaleringsmolekyler. De kan ändra och samordna sin produktion av olika ämnen som ett resultat av denna kommunikation, som kallas för "quorum sensing" (svenskt ord saknas). Bakterien *Pseudomonas aeruginosa*, som bland annat kan orsaka lunginflammation, använder sig till exempel av quorum sensing (QS) vid övergången från att vara en ofarlig medpassagerare på huden till att orsaka en virulent infektion. QS anses framförallt vara ett sätt att känna av hur många andra bakterier som finns i närheten för att kunna anpassa sitt beteende därefter. Virulenta sjukdomsalstrande bakterier producerar oftast en mängd olika ämnen som skadar värdens vävnader och därmed främjar bakteriernas tillväxt. Den ökade tillväxten kan vara fördelaktig för bakterierna ifall den ökar chansen att de förs vidare till en ny värd, men produktion av de skadliga ämnena är kostsam. En ensam bakterie kan heller inte orsaka så mycket skada, så för att åstadkomma en betydande effekt måste många bakterier producera de skadliga ämnena samtidigt. Genom QS kan *Pseudomonas* känna av när det finns tillräckligt många artfränder i närheten så att det ska löna sig att slå om till en virulent livsstil. De kan också signalera till varandra att "nu har jag valt att bli virulent" för att uppmuntra andra till att följa efter. När en tillräckligt stor andel av populationen har blivit virulent, skiftar resten till virulens nästan omedelbart. Experiment har visat att *Pseudomonas*-bakterierna blir mer framgångsrika i sin överlevnad när de blir virulenta, och att mutanter som saknar QS-förmågan är mindre framgångsrika (Rumbaugh et al. 2009). Precis som hos mer avancerade djur har bakterierna en fördel av att kunna kommunicera och samordna sitt beteende!

Honungsbin kan kommunicera med varandra både genom doft och rörelse, och "dansar" för att visa vägen till nya matkällor, som till exempel en blomsteräng. Dansens riktning visar vilken väg man ska ta för att hitta blomsterängen, och dess intensitet (dvs hur länge och hur kraftigt arbetsbiet dansar) förmedlar information om hur matrik den nya ängen är. Ju högre intensitet, desto fler andra bin kommer att flyga till ängen. Efter ett tag med ständigt ökande antal besökare, kommer blommornas nektar att börja ta slut, och ängens värde som

matkälla sjunker. Då dansar bina mindre intensivt och återbesöksfrekvensen minskar. På så sätt är binas matsökningsystem självreglerande. I princip samma process pågår när bina flyttar till ett nytt bo. Arbetsbin börjar söka efter lämpliga boplatser, och förmedlar hur attraktiv en nyupptäckt plats är genom sin dans; fast i detta fall finns det ingen självreglering, utan processen är snarare självförstärkande. Ju fler bin som bedömer den nya platsen som lämplig, desto mer dansar de och ju fler följer efter. Även om det från början finns flera möjliga boplatser kommer, när tillräckligt många individer har följt med till den nya boplatzen och funnit den lämplig, en konsensus så småningom att uppnås (Seeley & Visscher 2004). Då flyttar drottningen och resten av bisamhället till den nya boplatzen genom ett kollektivt beslut.

Det är en vanlig missuppfattning att samhällets drottning är den som ”bestämmer” över ett bisamhälle; egentligen styrs gruppens beteende av en uppsättning enkla regler som alla individer rättar sig efter (som till exempel att man ska följa den individ som dansar mest intensivt). Bisamhällen och grupper av bakterier som *Pseudomonas* har därför det gemensamt att det inte finns någon specifik individ som är dominant och som bestämmer över hela gruppen, utan att kollektiva beslut fattas på ett nästan demokratiskt sätt. Fast även sociala djur med en dominanshierarki kan fatta mer eller mindre demokratiska beslut. Forskare har noterat att afrikanska vildhundar fattar gemensamma beslut med hjälp av nysningar (Walker et al. 2017). Alla som har ägt hund har väl märkt att hundar har en tendens till att nysa när de är upphetsade. När de afrikanska vildhundarna förbereder sig för jakt, träffas de i en grupp för att stämma av. De hälsar på varandra och börjar kommunicera med varandra genom nysningar för att visa hur sugna de är på att jaga; en nysning är en röst för att gå ut på jakt.

Forskarna märkte att sannolikheten att hela gruppen skulle bestämma sig för att jaga var korrelerad med antalet nysningar – grupper som bestämt sig för att jaga nös i genomsnitt sju gånger, jämfört med en eller ett par nysningar för en misslyckad träff (dvs en som upplöstes utan att det blev någon jakt). Hur många individer som deltog i träffen var mindre viktigt än det totala antalet nysningar; en eller två väldigt jaksugna individer kan ibland få igång hela gruppen genom att nysa mycket. Vem som nyser är också viktigt; en dominant individs nysning verkar väga lite tyngre. Här kan vi se paralleller till mänskliga grupper, där ett par eldsjälar kan lyckas övertala hela gruppen, och där en mycket respekterad persons åsikter kan betyda mycket för vad hela gruppen till sist kommer att göra. Genom att kommunicera kan fördelaktiga beslut uppnås för både individen och gruppen (se även spelteorins koordinationsspel Hjortjakten; Mohlin, i detta nummer).

Rent allmänt förväntar sig en evolutionsbiolog att organismer endast kommer att delta i kollektiva beslut om det ligger i deras eget intresse. Annars är det bättre att strunta i gruppen och ”köra sitt eget race”. De individuella fördelarna

med QS har visats hos *Pseudomonas*, men ett bisamhälle är lite speciellt eftersom det nästan kan betraktas som en individ som är uppbyggd av många små kroppar. Det är inte relevant att tänka på för- och nackdelar med kollektiva beslut för ett enskilt arbetsbi, eftersom det aldrig kommer att kunna överleva och föröka sig utan resten av bisamhället. Endast gruppens bästa räknas, och det är självklart att en samordnad grupp kommer att vara mer framgångsrik än en grupp utan samordning. Hos vildhundarna är det inte heller så svårt att förstå nackdelarna med att inte följa gruppens kollektiva beslut. Att ensam ge sig ut på jakt är förmodligen bortkastad energi, eftersom en ensam individ kommer ha svårt med att fälla ett byte på egen hand. Att omvänt stanna hemma när alla andra går ut på jakt betyder att man kommer att gå miste om maten om jakten är lyckad. Avstämningen före jakten är också ett bra sätt att uppskatta hur motiverade alla är – om jakten kommer att bli en halvhjärtad historia är det kanske bättre att låta bli då risken att misslyckas blir större.

Slutsatser

Även så enkla varelser som bananflugor och bakterier har utvecklat effektiva sätt att kommunicera sinsemellan och använda informationen för att fatta rationella beslut. Det är rätt imponerande, med tanke på att vi oftast ser människans rationalitet som ett av våra mer artspecifika kännetecken. Men evolutionen är förvånansvärt effektiv när det gäller att hitta bra lösningar på olika problem. Dåliga beslut resulterar i slöseri med energi, och de kommer på sikt att missgynnas. Därför förväntar vi oss att arter som är välanpassade till sin miljö kommer att bete sig på ett tillsynes rationellt sätt. En avancerad kognitiv förmåga behövs egentligen inte. Det räcker med ett fåtal beslutsregler, i kombination med interna system som kan mäta olika behov och uppskatta tillförlitligheten i den externa information som man tar in, för att klara av ändamålsenliga beslut.

Trots de många likheterna mellan oss och djuren, är det ändå uppenbart att människans beslutsprocesser måste vara mer komplexa, eftersom vi har en mycket mer omfattande förmåga att tänka framåt jämfört med andra djur. Det gör att vi kan förstå konsekvenserna av ett beslut på ett helt annat sätt och lägga an ett mycket längre tidsperspektiv än andra arter. Utan denna förmåga hade vi inte kunnat bygga stora samhällen, eftersom den omedelbara egna nyttan nästan alltid väger tyngre än gruppens långsiktiga väl och ve. En schimpans skulle förmodligen aldrig kunna förstå poängen med att betala skatt, till exempel.

Vi kan dock se många spår av hur människans beslutsförmåga har präglats av vår evolutionära historia. Att vi har en tendens att lyssna mest på dem som skriker högst i debatten eller på personer som har högst status framför dem som har bäst expertkunskaper, speglar förmodligen ganska allmänna beteenden hos sociala däggdjur. Kontexten spelar också stor roll. Folk är mindre

benägna till att begå sunk cost bias (SCB) när beslutet rör enbart ett finansiellt problem (till exempel att sälja av ett företag eller inte) jämfört med ett beslut som gäller en nära relation (till exempel att avbryta ekonomiskt stöd till en vän eller ett syskon) (Hrgović & Hromatko 2018). En rimlig tolkning här är att det är lättare att bedöma investeringens rationella för- och nackdelar när det gäller det mer abstrakta finansiella beslutet, medan beslutet om en nära relation kommer att påverkas väldigt mycket av hur det skulle ändra min status inom min sociala grupp och därmed betydligt mindre av ekonomiska faktorer. Vi kan därför fatta både väldigt bra och väldigt dåliga beslut på grund av gruppträck, eller det vi med en socialpsykologisk term kallar *groupthink* (Janis 1982).

Genom att veta hur andra djur fattar beslut kan vi förhoppningsvis bättre förstå vårt eget beteende, och kanske hitta sätt att uppmuntra till samhällsfrämjande individuella beslut. En metod som har visat sig vara effektiv är ”nudging” – att påverka folks spontana beslutsfattande på olika sätt (Kahneman 2011). Både djur och människor är mer benägna att välja objekt som lockar deras uppmärksamhet, och detta kan användas för att påverka individuella beslut. Folk köpte till exempel nyttiga snacks framför onyttiga, när de nyttiga varorna fick mer uppmärksamhet genom att ligga i anslutning till betalkön (Van Gestel et al. 2018). Forskning om hur enklare organismer som insekter och bakterier kommunicerar och tar kollektiva beslut är också värdefullt från en teknologisk synpunkt. Man kan till exempel koordinera beteendet hos små självständiga robotar genom att utgå ifrån hur sociala insekter koordinerar sina beslut. ”Svärmar” av dessa robotar skulle kunna användas bland annat till att hitta skadade människor vid jordskalv eller skörda grödor (Hamann 2018). Ett evolutionärt perspektiv på beslutsfattande kan alltså ge nya insikter inom vitt skilda fält.

Referenser

- Arkes, H. R. & P. Ayton, 1999. “The sunk cost and concorde effects: are humans less rational than lower animals?”, *Psychological Bulletin* 125, s. 591-600.
- Battesti, M., C. Moreno, D. Joly & F. Mery, 2015. “Biased social transmission in *Drosophila* oviposition choice”, *Behavioral Ecology & Sociobiology* 69, s. 83-87.
- Beckwith, E. J., Q. Geissmann, A. S. French & G. F. Gilestro, 2017. “Regulation of sleep homeostasis by sexual arousal”. *eLife* 6:e27445.
- Curio, E., 1987. “Animal decision-making and the ‘Concorde fallacy’”, *Trends in Ecology & Evolution* 2, s. 148-152.
- Hamann, H., 2018. *Swarm robotics: a formal approach*. Vienna: Springer-Verlag.
- Haun, D. B. M., Y. Rekers & M. Tomasello, 2012. “Majority-based transmission in chimpanzees and human children, but not orangutans”, *Current Biology* 22, s. 727-731.
- Hayden, B. Y., 2016. “Time discounting and time preference in animals: a critical review”, *Psychonomic Bulletin & Review* 23, s. 39-53.

- Höglund, J., R. V. Alatalo, R. M. Gibson & A. Lundberg, 1995. "Mate-choice copying in black grouse", *Animal Behavior* 49, s. 1627-1633.
- Hrgović, J. & I. Hromatko, 2018. "The time and social context in sunk-cost effect", *Evolutionary Psychological Science* 4, s. 258-267.
- Janis, I. L., 1982. *Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes*. 2nd edition. Boston: Houghton Mifflin.
- Kahneman, D., 2011. *Thinking, fast and slow*. London: Penguin Random House.
- Kendal, R., L. M. Hopper, A. Whiten, S. F. Brosnan, S. P. Lambeth, S. J. Schapiro & W. Hoppitt, 2015. "Chimpanzees copy dominant and knowledgeable individuals: implications for cultural diversity", *Evolution and Human Behavior* 36, s. 65-72.
- Magalhães, P. & K. G. White, 2016. "The sunk cost effect across species: a review of persistence in a course of action due to prior investment", *Journal of the Experimental Analysis of Behavior* 105, s. 339-361.
- Mann, R. P., 2018. "Collective decision making by rational individuals", *Proceedings of the National Academy of Sciences, USA* 115, s. E10387-E10396.
- Rumbaugh, K. P., S. P. Diggie, C. M. Watters, A. Ross-Gillespie, A. S. Griffin, & S. A. West, 2009. "Quorum sensing and the social evolution of bacterial virulence", *Current Biology* 19, s. 341-345.
- Seebass, A. R., 1997. "The prospects for commercial supersonic transport", i H. Sobieczky (ed.), *New design concepts for high speed air transport*. Vienna: Springer-Verlag, s. 1-12.
- Seeley, T. D. & P. K. Visscher, 2004. "Quorum sensing during nest-site selection by honeybee swarms", *Behavioral Ecology & Sociobiology* 56, s. 594.
- Sweis, B. M., S. V. Abram, B. J. Schmidt, K. D. Seeland, A. W. MacDonald, III, M. J. Thomas & A. D. Redish, 2018. "Sensitivity to 'sunk costs' in mice, rats, and humans". *Science* 361, s. 178-181.
- Van Gestel, L. C., F. M. Kroese & D. T. D. De Ridder, 2018. "Nudging at the checkout counter – a longitudinal study of the effect of a food repositioning nudge on healthy food choice", *Psychology and Health* 33, s. 800-809.
- Walker, R. H., A. J. King, J. W. McNutt, & N. R. Jordan, 2017. "Sneeze to leave: African wild dogs (*Lycyaon pictus*) use variable quorum thresholds facilitated by sneezes in collective decisions", *Proceedings of the Royal Society London B Biological Sciences* 284: 20170347.

Matematiska modeller för beslutsfattande hos människor, djur och maskiner

Kalle Åström

Abstract

The science of decision making is studied in many areas of science and engineering, for example within, computer science, automatic control theory, mathematics and neurophysiology. The aim of this text is to provide a very brief overview of some of the models and tools used within mathematical decision making. These models can be studied in terms of an agent that observes the world and takes decisions. The techniques that have been developed, e.g. optimal control, dual control, machine learning, reinforcement learning, have found numerous uses in many applications, where machines take decisions, but they are also an inspiration for the question of understanding biological decision making, and could potentially be important for understanding decision making in political science.

Inledning

Typiskt för modellering och särskilt matematisk modellering är att man behöver göra förenklingar av de fenomen man vill betrakta. Det betyder ofta att en modell bara fångar vissa egenskaper hos det man vill efterlikna. En leksaksmodell av en bil kanske är användbar för att beskriva färg och form, men kan inte användas för att studera t.ex. bränsleförbrukning. Typiskt för modelleringsprocessen är att man, trots att man vet att modellen har brister, genom analys försöker förstå vilka slutsatser som har bäring på det väsentliga eller centrala problemet. Det är då viktigt att man försöker förstå vilka aspekter som är relevanta och vilka som är osäkra eller direkt irrelevanta. Inom maskininlärning pratar man t.ex. också om hur väl en modell generaliserar till nya data.

Kalle Åström är professor vid Matematikcentrum, Lunds universitet.
E-post: kalle@maths.lth.se

Syftet med den här artikeln är att diskutera matematiska modeller för beslutsfattande. Målet är att ge en översikt över modeller för agenter som fattar beslut i en värld av andra agenter. I den här kontexten vill jag kunna vara väldigt generös med vad som kan vara en agent. Jag vill kunna prata om människor och andra djur, och då även väldigt enkla organismer, men också konstruerade maskiner, robotar och datorprogram. Man skulle också kunna se en organisation som en agent. I vår modell tänker vi oss att agenten har möjlighet att få information om sin värld, sin omgivning, genom sinnesintryck. Vi tänker oss också att agenten har en möjlighet att fatta beslut om hur den vill påverka sin omgivning. Centralt för vår modellering är alltså följande fyra begrepp: agent, värld, sinnesintryck och beslut (se figur 1). Vår förenklade modell av människan är alltså att den mottar ett antal sinnesintryck som den får genom syn, hörsel, smak, lukt och känsel, att den har möjlighet att fatta beslut om hur muskler kan styras och på så sätt förflytta sig, att den kan manipulera saker i sin omgivning och att den kan styra ögonen och prata och äta. En intressant fråga är därför hur man konstruerar agenter som självständigt lär sig hur världen är beskaffad och sedan använder sig av sinnesintrycken för att fatta bra beslut. Hur fungerar människor och djur i det här avseendet? Och hur kan vi konstruera algoritmer och robotar som verkar på detta sättet? Inom maskininlärning och artificiell intelligens utvecklas ett antal metoder och verktyg som kan användas i sammanhanget.

Min ambition är här att introducera några centrala begrepp för att se hur dessa kan nyttjas i en analys av adekvat beslutsfattande. Vi går kort igenom övervakat lärande, självkodning, världsmodeller, dual reglering samt förstärkningsinlärning och diskuterar hur de kan användas för att utveckla *automatiserade* metoder för att skapa agenter som både lär sig *och* kan fatta beslut. Några av begreppen och idéerna om hur agenter kan utveckla sina beslutsprocesser automatiskt är gamla, välutvecklade och välkända och används mycket redan idag, medan andra är betydligt nyare. Jag tar därför särskilt upp några begrepp från nyligen publicerade arbeten.

Fixerade beslutsprocesser

För många agenter är processen att gå från sinnesintryck till beslut fixerat, vilket innebär att agenten inte lär sig av sina erfarenheter, även om det kanske sker ett lärande ur ett evolutionärt perspektiv, (Abbott i detta nummer). Ett exempel (Sterling & Laughlin 2015) är *e-coli* bakterien som kan känna av om det finns glucos och laktos i sin omgivning. Baserat på detta kan den ta beslut om hur mycket av tre olika proteiner, som ska tillverkas för att föra in glukos respektive laktos, samt hur mycket galactosidase som ska tillverkas för att bryta ner laktos. Om det finns glukos så tillverkas det protein som behövs för att bryta ner glukos. Om det inte finns glukos, men det finns laktos, så tar

bakterien beslut om att tillverka det proteinet som för in laktos samt det protein som behövs för att bryta ner laktos. För just e-coli bakterien vet man dessutom precis hur beslutsprocessen går till rent kemiskt.

Bakterien e-coli tar dessutom beslut om hur flageller ska användas för att navigera. Beslutsprocessen för rörelsen kan förenklat beskrivas som att varva beteendet att åka rakt fram med en beteendet om att slumpmässig rotera. Om det verkar finnas mer mat här än för en sekund sedan, och du fortfarande rör dig rakt fram, kör i så fall ytterligare lite längre.

Klassificeringsproblemet inom maskininlärning

Hur kan man ta fram en optimal beslutsprocess? Inom maskininlärning är s k övervakat lärande en central idé; man tänker sig att man tränar upp agentens kapacitet att gå från sinnesintryck till beslut genom att presentera ett stort antal exempel på sinnesintryck samt tillhörande korrekt beslut. Man använder sedan matematisk statistik och optimering för att koppla ihop de två för att på så sätt få till stånd en fungerande beslutsgång (se figur 2).

Detta har visat sig vara framgångsrikt för många tekniska tillämpningar. Erfarenheten är att det är svårt att för hand skriva datorprogram som tar rätt beslut givet en viss insignal. Däremot kan det vara möjligt att samla på sig exempel på indata med tillhörande korrekta beslut. Med övervakat lärande tar man sedan fram en beslutsprocess som fungerar. Här kommer några exempel på tekniska lösningar baserat på övervakat lärande.

- Företaget Hövding tillverkar en cykelhjälm med en uppblåsbar krockkudde. Baserat på sinnesintryck tar hjälmen hundratals gånger per sekund beslut om krockkudden ska lösas ut eller inte. Företaget har samlat in en stor mängd exempel på data (normal cykling), då den *inte* ska lösa ut, och ett stort antal iscensatta olyckor, som ger exempel på data när den väl *ska* göra det. Med maskininlärning tar man sedan fram en beslutsprocess som givet data bestämmer när krockkudden ska aktiveras och inte.
- Företaget Cellvision (Simonsson 2018), tillverkar ett automatiskt system för analys av blodprov. Systemet hittar vita och röda blodkroppar och känner t ex igen olika typer av vita blodkroppar; företaget har samlat in en stor mängd data om just vita blodkroppar. För varje bild har ett antal experter angett vad som är rätt klass. Med maskininlärning tar man sedan fram en beslutsprocess som på grundval av bilderna avgör vilken vit blodkropp det rör sig om.
- Vid patologen på sjukhus undersöker man biopsier genom att färga in cellproverna och sedan studera dem i mikroskop. För varje prov ska man ange

på en så kallad Gleasons skala hur irreguljärt cellerna växer. Med maskininlärning kan man konstruera en beslutsprocess som bedömer dessa prover automatiskt (Gummesson 2017).

- I fartkameror finns det system som, förutom att mäta hastigheten på bilarna, också hittar registreringskylten automatiskt. Ett system för att känna igen vilka bokstäver och siffror som finns på registreringskylten bygger på klassificering. Man har sedan tidigare ett stort antal bilder av t ex siffror, där man för varje bild har angett korrekt klass.

Inlärningsprocessen är enklast om man har få sinnesintryck och få beslutsmöjligheter. Låt oss ta ett klassiskt exempel hämtat från biologen Ronald Fisher i en artikel från 1936 (Fischer 1936). Där beskrivs ett dataset med 50 mätningar från tre sorters blommor. För varje blomma har man mätt kronbladslängd, kronbladsbredd, foderbladslängd och foderbladsbredd samt noterat vilken art det är. Som illustration studerar vi bara en av egenskaperna (kronbladslängd), som vi betraktar som sinnesintryck; vi tänker oss att vi ska försöka avgöra om det är arten *Iris Setosa* (klass 1) eller arten *Iris Versicolor* (klass 2) det rör sig om. Det finns många olika tillvägagångssätt för att klassificera. Vanligtvis har man någon sorts träningsdata, i det här fallet $T_a = \{(s_1, b_1), \dots, (s_n, b_n)\}$, där s_k är sinnesintryck (kronbladslängd) för exempel nr k och b_k är det korrekta beslutet, antingen klass 1 för *Iris Setosa* eller klass 2 för *Iris Versicolor*. Ett sätt att utföra analysen och komma fram till ett beslut är att studera träningsmaterialet enligt figur 3. Första steget är att skatta sannlikhetsfördelningen $f_x(x|Y = 1)$, dvs hur sannolikt det är att kronbladslängden är x om det är en mätning från en *Iris Setosa*. Överst i figur 3 illustreras just de mätningarna med stjärnor.

Genom att använda en metod som kallas kärnskattning, (Parzen 1962; Rosenblatt 1956), kan man skatta sannlikhetsfördelningen $f_x(x|Y = 1)$ för kronbladslängden för blad hos *Iris Setosa*. I grafen har vi multiplicerat den med sannolikheten $P(Y = 1)$ för att en slumpmässig växt är just *Iris Setosa*. I det här exemplet har vi antagit att $P(Y = 1)$ är 50 %. På motsvarande sätt har vi skattat sannlikhetsfördelningen $f_x(x|Y = 2)$ för kronbladslängden för *Iris Versicolor*. Analysen visar att kronbladslängderna typiskt är lite kortare för *Iris Setosa* jämfört med *Iris Versicolor*. Genom att summera $f_x(x|Y = 1)P(Y = 1)$ och $f_x(x|Y = 2)P(Y = 2)$ får man den totala sannlikhetsfördelningen $f(x)$. Genom att sedan dividera $f_x(x|Y = 1)P(Y = 1)$ med $f_x(x)$ får man, enligt Bayes sats (Bayes 1763), sannolikheten $P(Y = 1|x)$, dvs sannolikheten att det är *Iris Setosa* om kronbladslängden är x . Om denna sannolikhet överskrider 50 % tar man beslutet att det är *Iris Setosa*, annars att det är *Iris Versicolor*. I det här fallet ger analysen alltså att man ska gissa på *Iris Setosa* om kronbladslängden ligger under ett värde som är cirka 5.5.

Vissa klassificeringsmetoder försöker direkt göra en skattning av $P(Y = 1|x)$,

t ex linjär logistisk regression, genom att modellera den som en kombination av en linjär avbildning (som beror på några parametrar w) och en mjuk stegfunktion. Genom att formulera en felfunktion $L(w, T_A)$ som beror på parametrarna w och träningsdata T_A och minimera den, dvs försöka hitta de parametrar w som gör felet L så litet som möjligt, så får man en klassificerare som förhoppningsvis fungerar bra för framtida beslut. Kombinationen av en linjär avbildning och en olinjär klassificering ser vi i flera andra metoder, t ex perceptronen (Rosenblatt 1957) och stödvektormaskiner (Vapnik 1963). Genom att koppla flera perceptronmoduler i ett nätverk kan man konstruera en maskininlärningsalgoritm (ett artificiellt neuralt nätverk), som kan approximera olinjära beslutsytor i rumden. Man kan se det som att man ansätter $P(Y = 1|x)$ som ett nätverk av omväxlande linjära funktioner och mjuka stegfunktioner. En sådan funktion kan ha tusentals variabler och miljontals parametrar. Faltning (Hilbert 1906) är en matematisk operation som enklast beskrivs som ett viktat rullande medelvärde. Om man ersätter de linjära operationerna i ett artificiellt neuralt nätverk med faltningar så får man ett faltningsnätverk (LeCun et al 2015). Analogt med linjär regression ställer man in parametrarna w genom att minimera en felfunktion $L(w, T_A)$ som också beror på träningsdata T_A .

För att använda sig av klassificering med maskininlärning så behöver man tala om vad som är rätt beslut för en stor mängd sinnesintryck. Det är emellertid inte alltid lätt att veta vad som är rätt beslut och det kan vara dyrt och tidskrävande att samla in data, säkerställa att de har rätt kvalitet och att ange rätt beslut/svar. Trots dessa svårigheter används övervakat lärande framgångsrikt i ett stort antal tillämpningar för att få maskiner att nå fram till någorlunda korrekt beslut.

Självkodare

En svårighet med att använda övervakat lärande är att man måste ange rätt beslut för ett stort antal sinnesintryck. Självkodning är namnet på system som från indata (t ex sinnesintryck) försöker återskapa samma sinnesintryck som utsignal. Det här blir ointressant och oanvändbart om man får lov att skicka vidare insignalen direkt till utsignalen. Däremot blir det intressant om man konstruerar systemet med en analys/kodningsdel och en syntes/avkodningsdel. Analysdelen tar sinnesintrycken och returnerar några få tal, som fungerar som en komprimerad representation av sinnesintrycken. Syntesdelen tar den komprimerade representationen och återskapar sinnesintrycken. Både analys- och syntesdelen innehåller många parametrar som "tränas" med hjälp av övervakat lärande. Fördelen här är att varje exempel på sinnesintryck kan användas både som indata och som facit i det övervakade lärandet. Självkodande kan vara en viktig komponent för en agent som försöker förstå världen. Den kodade representationen kan ibland vara enklare att använda för inlärning och tolkning.

Både analys och syntesdelen kan vara användbara. T ex kan syntesdelen användas för att generera nya syntetiska sinnesintryck (eller bilder i det här fallet). Ett intressant exempel på framgångsrik syntetisering av realistiska ansiktsbilder kan man hitta på <https://thispersondoesnotexist.com>. Varje gång man klickar på sidan får man en ny slumpmässig ansiktsbild (Karras 2017).

Världsmodeller

I ovanstående exempel var agentens samspel med omvärlden begränsat. Mer intressant blir det om de beslut agenten tar påverkar vilka sinnesintryck den får i framtiden. Det här blir relevant när man ska styra saker, t ex inom reglerteori. Ett exempel på detta är om man vill ha en agent som håller temperaturen vid 22 grader i ett hus. Agenten får temperaturmätningar som "sinnesintryck" och kan "ta beslut" om hur mycket värme som elementen ska ge från sig. Andra exempel är självkörande bilar och självflygande drönare.

Ett vanligt tillvägagångssätt är att en människa analyserar och kommer fram till en världsmodell V' som kan användas för att förutsäga hur kommande sinnesintryck (temperaturen) påverkas av agentens beslut. Om man har en världsmodell V' , så kan den användas för att ta fram en bra beslutsprocess, som i sin tur leder till kloka beslut baserat på vissa sinnesintryck.

Optimal reglerteknik utvecklades under slutet av 1950-talet. Genom att formulera beslutsproblemet så att man tar de beslut som minimerar ett felmått för ett dynamiskt system över tiden, kan man för vissa typer av problem räkna ut vad som är ett optimalt beslut eller en optimal reglerteknik. En lösning utvecklades i mitten av 1950-talet av Richard Bellman och andra genom att utvidga teorier från 1800-talet framtagna av Hamilton och Jacobi. Metoden för att lösa optimala beslutsproblem genom att lösa den så kallade Bellman-ekvationen blev känd som dynamisk programmering (Bellman 1957). Bellman introducerade också den diskreta, stokastiska modell som kallas Markovska beslutsprocesser; och Ron Howard (1960) utformade i sin tur en iterationsmetod för sådana Markovska beslutsprocesser.

Dual reglering

Men hur ska man bära sig åt, om det inte finns någon bra världsmodell V' ? Inom reglerteori finns det ett forskningsområde som heter dual reglering, och som utvecklades av Alexander Aronovich Feldbaum 1960 (Feldbaum 1960). Idén är att man har ett reglersystem med insignaler (sinnesintryck) och utsignaler (beslut). Utsignalerna påverkar verkligheten och indirekt insignalerna, som illustreras i figur 1. Om man har en tillförlitlig modell av hur verkligheten fungerar så kan man konstruera regleralgoritmen (b beslutsprocessen) så att den gör ett så bra jobb som möjligt, det vill säga blir optimal. Men om man inte vet

hur den verkliga processen fungerar kan man behöva experimentera för att få fram mer information om världen. Med bättre information kan man i ett senare skede styra processen bättre. Om systemets egenskaper är okända från början och kanske också ändras över tid har man två olika mål. Man måste både styra och undersöka, dvs

- använda kunskapen om hur världen fungerar för att uppnå sina mål (*styra*).
- testa och studera världen för att lära sig mer om hur systemet fungerar (*undersöka*).

De två handlingarna kan stå i konflikt med varandra. Om man bara styr, lär man sig mindre om världen och gör kanske misstag senare. Om man bara undersöker uppnår man inte sina mål.

Ett praktiskt exempel är om man ska ut och köra en ny bil en kall vinterdag. Då behöver man kanske prova att gasa, bromsa och svänga lite försiktigt för att kunna köra mer effektivt och säkert senare. Ett annat exempel är när man ska ställa in styrparametrar i en fabrik. Då behöver man också göra lite experiment för att sedan se vilka beslutsprocesser som fungerar bäst. Ett tredje exempel är om man ska bära en bricka med glas. När man lyfter upp brickan gör man det försiktigt och känner efter hur stadigt glaset står. Baserat på denna undersökning så får man mer information. Känner man att glaset står stadigt, vågar man gå raskt, men om det känns ostadigt får man ta det försiktigt. Ett fjärde exempel är inom medicinen. För att förbättra vården och utveckla nya rutiner så gör man ibland studier på ett mindre patientmaterial. Baserat på resultatet på studien så väljer man det som man tror är bäst för framtida patienter.

För enklare system, med ett fåtal in- och utsignaler och enklare modeller, kan man analysera problemet matematiskt och komma fram till optimala strategier. Optimalt kan i det här sammanhanget t ex formuleras som att felet i den egenskap man vill styra blir så litet som möjligt över tiden. Feldbaum visade att den optimala lösningen kan hittas genom dynamisk programmering. Genom att analysera de optimala strategierna kan man se att det mycket riktigt finns tillfällen då man ska undersöka och andra då det är bäst att styra (se t ex Åström & Helmersson 1986).

Förstärkningsinlärning

En besläktad frågeställning finns inom det vi kallar *förstärkningsinlärning* (Sutton & Barto 2011). Även här tänker man sig en agent som tar emot insignaler (dvs sinnesintryck) och skickar ut utsignaler (dvs fattar beslut), så som i figur 1. I likhet med dual reglering har man ett tydligt mål. Inom förstärkningsinlärning är målet formulerat som en belöning som agenten får utifrån.

Förstärkningsinlärning bygger vidare på metoder inom optimal reglerteori, dual reglering, men också från forskning inom psykologin för djurs lärande (Abbott i detta nummer).

Om antalet möjliga beslut är få, antalet möjliga insignaler inte är för många och om systemet, miljön man befinner sig i, går att modellera som en partiellt observerbar Markov-beslutsmodell, POMDP (Åström 1965), så finns det bra algoritmer som löser problemet effektivt. Dessa algoritmer (precis som för dual reglering) löser i princip problemet att både utforska/undersöka systemet och hantera det så att agenten efter ett tag har en ganska bra modell av systemet/världen och därefter kan styra ganska bra. Tack vare utvecklingen inom djupa faltungsnetzverk har man kunnat konstruera approximativa lösningar, som även fungerar för lite större problem. Metoderna har varit väldigt framgångsrika, t ex för att konstruera en agent (ett datorprogram) som automatiskt spelar spel som exempelvis schack (Silver et al 2018) och go (Silver et al 2016).

Tillstånd eller minne

Många modeller använder sig av så kallade tillståndsmodeller. Man kan se tillståndet som ett sätt att lagra tidigare händelser i ett minne. För agenten består historien av alla tidigare beslut och sinnesintryck. En agents tillstånd skulle kunna bestå av all sparad historia, men ofta finns det mer effektiva sätt att lagra relevant information i ett tillstånd. Om den interna tillståndsmodellen är tillräckligt informativ räcker denna för att kunna modellera framtiden. För att automatiskt göra bra tillståndsmodeller kan man ibland använda sig av tekniker som självkodning. Man hushållar m a o med resurserna.

Världsmodeller, drömmar och beslutsträning

Det finns alltså flera exempel på att en agent kan skapa sig en modell av sinnesintrycken och dynamiken, t ex med hjälp av dual reglering eller förstärkningsinlärning. Men det finns fortfarande frågetecken om vad det är för målfunktion man ska ha. I dual reglering tänker man sig att det är ett reglerfel som ska minimeras. I förstärkande inlärning tänker man sig att det finns en väldefinierad belöningsfunktion.

Hur är det här kopplat till biologiska varelser och biologiskt beslutsfattande? För biologiska varelser är det inte helt tydligt vad målet är och vi är långt från att ha några som helst rimliga modeller för komplexa biologiska varelser, som människan.

Går det att få liknande resultat som för dual reglering och förstärkningsinlärning men utan externt definierade mål, och istället med internt formulerade mål eller åtminstone med andra principer?

I en färsk artikel (Ha & Schmidhuber 2018) föreslås en arkitektur där man

bygger upp en explicit modell av världen genom maskininlärning. Den består väsentligen av två delar. Den första är en självkodare, som tränas på ett stort antal sinnesintryck. Det gör att man får en mer komprimerad representation. Den andra delen är en dynamisk modell som "tränas upp" på tripplar av data (ett beslut samt sinnesintryck före och efter detta beslut). Den dynamiska modellen tar också fram en tillståndsmodell, som fungerar som ett minne för den dynamiska modellen. De två delarna tränas var för sig med maskininlärning. De är konstruerade så att de kan tränas på data som agenten själv kan ta fram genom att helt enkelt testa lite olika beslut. När de två delarna väl är inkörda, har agenten tillgång till en intern modell av världen som den kan använda för att prova olika scenarier, för att på så sätt utveckla en adekvat beslutsprocess.

En annan princip som tillämpats t ex i en artikel av Klyubin et al (2015) är att basera agentens inlärning på sk egenmakt. I artikeln utgår man från Gibsons princip om att kopplingen mellan sinnesintryck och beslut är centralt för en agents inlärning, men bygger vidare på tanken om att besluten är minst lika viktiga som sinnesintrycken och att det är viktigt att kvantifiera agentens förmåga att kontrollera sin situation. Agentens egenmakt mäts i bitar. Den är noll när agenten inte har någon kontroll över framtida sinnesintryck, men växer poängmässigt ju högre kännbart inflytande den har. Egenmakt kan också tolkas som den mängd information, som agenten potentiellt kan injicera i sin värld via sina beslut och senare fånga via sina sinnesintryck. Det har visat sig vara svårt att beräkna egenmakt, men i artikeln av Karl et al (2017) utvecklas metoder för att åtminstone approximativt göra just detta och för att sedan kunna använda dessa som principer för inlärning. I artikeln studerar författarna flera enkla dynamiska system. I exemplet "den svängande pendeln" har agenten störst egenmakt i det övre läget. Systemet både lär sig världsmodeller och att manövrera till detta övre läge. För ett annat exempel simuleras en tvåbent agent. Här visar det sig att det tillstånd med störst egenmakt råder, när agenten står upp. Egenmakten tillmäts ett högre värde i de tillstånd då agenten har stor handledningsfrihet, dvs när agenten inte är trängd i ett hörn eller befinner sig i en återvändsgränd.

Slutsatser

I det här kapitlet har vi studerat några olika matematiska/statistiska modeller för maskinellt – eller automatiserat – beslutsfattande. Även om kunskaperna, modellerna och metoderna för sådant autonomt lärande och beslutsfattande utvecklats enormt, används de för relativt enkla problem, ofta för spel där reglerna och systemen är relativt begränsade. De börjar dock användas i viss utsträckning även inom robotik, för självkörande fordon och för drönare. Problemet om automatiserat beslutsfattande beror på en mängd olika faktorer (Russel 2019). Även om dagens metoder är otillräckliga för att förstå mänskligt

beslutsfattande och beslutsfattande på en planetär nivå, så tror jag att det är värdefullt med kunskapsutbyte inom beslutsteori över traditionella ämnesgränser. Det skulle vara intressant att se hur metoderna skulle kunna användas inom statsvetenskapen och andra samhällsdiscipliner.¹

Referenser

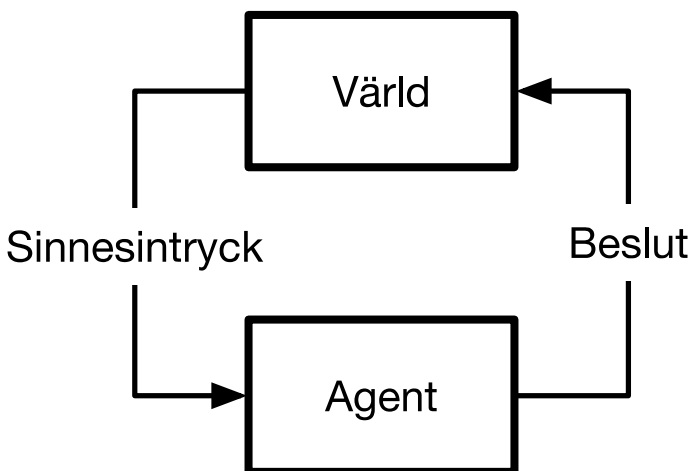
- Bayes, Thomas, 1763. "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS", *Philosophical transactions of the Royal Society of London* 53, s. 370–418.
- Bellman, Richard, 1957. "A markovian decision process", *Journal of mathematics and mechanics*, s. 679–684.
- Feldbaum, Alexander A., 1960. "Теория Дуального управления", *Автоматика и Телемеханика* 21(9), s. 1240–1249.
- Fisher, Ronald A, 1936. "The use of multiple measurements in taxonomic problems", *Annals of eugenics* 7(2), s. 179–188.
- Gummeson, Anna, Arvidsson, Ida, Ohlsson, Mattias, Overgaard, Niels Christian, Krzyzanowska, Agnieszka., Heyden, Anders, Bjartell, Anders & Åström, Kalle, 2017. "Automatic Gleason grading of H and E stained microscopic prostate images using deep convolutional neural networks", *Medical Imaging 2017: Digital Pathology* 101400S.
- Ha, David & Schmidhuber, Jürgen, 2018. "World models", *ArXiv preprint arXiv:1803.10122*.
- Howard, Ronald A., 1960. *Dynamic programming and Markov processes*. John Wiley.
- Hilbert, David, 1906. "Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. Fünfte Mitteilung", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, s. 439 – 480.
- Kaelbling, Leslie P. & Littman, Michael L. & Moore, Andrew W., 1996. "Reinforcement learning: A survey", *Journal of artificial intelligence research* 4, s. 237–285.
- Karl, Maximilian & Soelch, Maximilian & Becker-Ehmck, Philip & Benbouzid, Djalel & van der Smagt, Patrick & Bayer, Justin, 2017. "Unsupervised real-time control through variational empowerment", *ArXiv preprint arXiv:1710.05101*.
- Karras, Tero & Aila, Timo & Laine, Samuli & Lehtinen, Jaakko, 2017. "Progressive growing of GANs for improved quality, stability, and variation", *ArXiv preprint arXiv:1710.10196*.
- Klyubin, Alexander S., Polani, Daniel & Nehaniv, Chrystopher L., 2005. "Empowerment: A universal agent-centric measure of control", *2005 IEEE Congress on Evolutionary Computation* 1, s. 128–135.
- LeCun, Yann, Bengio, Yoshua & Hinton, Geoffrey, 2015. "Deep learning", *Nature* 521(7553), s. 436–444.
- Parzen, Emanuel, 1962. "On estimation of a probability density function and mode", *The annals of mathematical statistics* 33(3), s. 1065–1076.
- Rosenblatt, Frank, 1957. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory.

1 För en diskussion om de etiska implikationerna av detta, se Maria Hedlunds bidrag till detta nummer.

- Rosenblatt, Murray, 1956. "Remarks on some nonparametric estimates of a density function", *Annals of Mathematical Statistics* 27(3), s. 832–837.
- Russel, Stuart, 2019. "Human compatible: Artificial intelligence and the problem of control", *Viking, Penguin Random House*.
- Silver, David, Huang, Aja, Maddison, Chris J., Guez, Arthur, Sifre, Laurent, Van Den Driessche, George, Schrittwieser, Julian, Antonoglou, Ioannis, Panneershelvam, Veda & Lanctot, Marc et al., 2016. "Mastering the game of go with deep neural networks and tree search", *Nature* 529 (7587), s. 484.
- Silver, David, Hubert, Thomas, Schrittwieser, Julian, Antonoglou, Ioannis, Lai, Matthew, Guez, Arthur, Lanctot, Marc, Sifre, Laurent, Kumaran, Dharshan & Graepel, Thore, et al., 2018. "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play", *Science* 362(6419), s. 1140–1144.
- Simonsson, Per, 2018. *CellaVision Story: Innovation, människor & miljö*. Lund: Historiska Media.
- Sterling, Peter & Laughlin, Simon, 2015. *Principles of neural design*. Cambridge: MIT Press.
- Ström, Jakob, Åström, Kalle & Akenine-Möller, Tomas, 2017. *Immersive linear algebra*, <http://immersivemath.com/ila>.
- Sutton, Richard S. & Barto, Andrew G., 2011. *Reinforcement learning: An introduction*. Cambridge: The MIT Press.
- Vapnik, Vladimir, 1963. "Pattern recognition using generalized portrait method", *Automation and remote control* 24, s. 774–780.
- Åström, Karl Johan, 1965. "Optimal control of markov processes with incomplete state information", *Journal of mathematical analysis and applications* 10(1), s. 174–205.
- Åström, Karl Johan & Helmersson, Anders, 1986. "Dual control of an integrator with unknown gain", *Computers & Mathematics with Applications* 12(6), s. 653–662.

Figurer

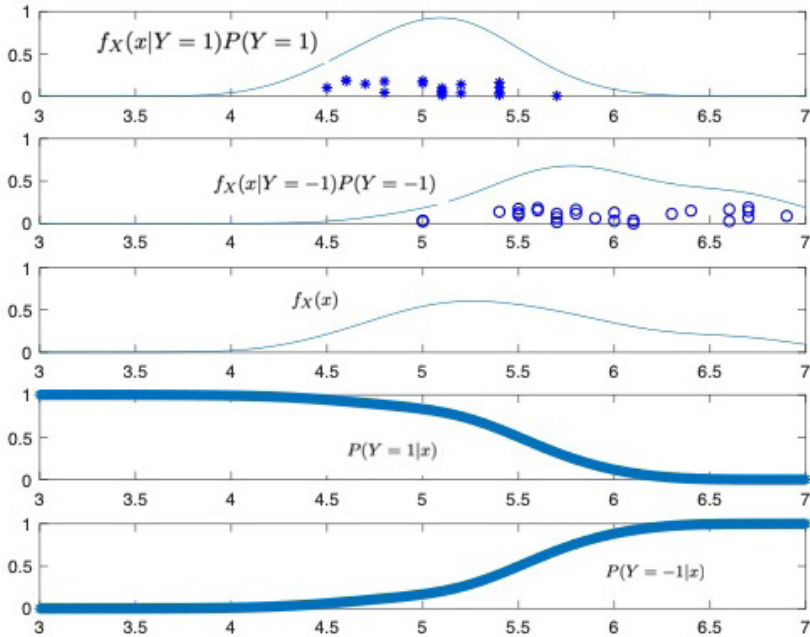
Figur 1. I den här texten presenterar vi några olika modeller för hur agenter fattar beslut i en värld baserat på sinnesintryck.



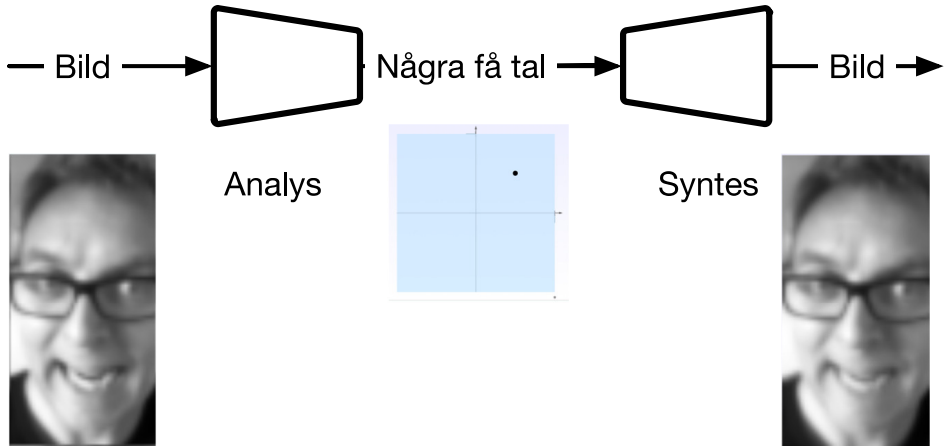
Figur 2. Inom övervakat lärande tränar man upp hur beslutsprocessen för en agent ska vara baserat på ett stort antal exempel på korrekta beslut för olika sinnesintryck. Man har då oftast inte en explicit modell hur världen fungerar och är inte alltid intresserad av återkoppling eller tidsaspekt.



Figur 3. Figuren illustrerar några av stegen i hur man tränar en klassificerare. De två översta graferna visar exempeldata och sannolikhetsfördelningar för två klasser. Mittenraden visar den totala sannolikhetsfördelningen. De två understa graferna visar a posteriorifördelningarna för de två klasserna.



Figur 4. En självkodare innehåller både en analys/kodningsdel och en syntes/avkodningsdel. Analysdelen tar sinnesintryck och returnerar några få tal. Syntesdelen tar den komprimerade representationen och återskapar sinnesintrycken. Exemplet är hämtat från Ström et al (2015).



När maskiner fattar beslut – vem är ansvarig?

Maria Hedlund

Abstract

Robots making decisions on social benefits, driverless cars causing traffic accidents, search engines presenting a selected narrow picture of the world – the rapid development of AI technology gives rise to machines that makes their own decisions, without direct influence from humans, but who is responsible for what a machine does? Can the machine itself be responsible? The aim of this article is to discuss and problematize responsibility relations when machines make decisions. The overarching question is whether machines can be responsible and if so, under which circumstances. Drawing on theories on responsibility, machine ethics, robot philosophy, and on recent AI development, the article demonstrates how functionalistic arguments can lead to the conclusion that machines are responsible for their actions, while approaches building on philosophical understandings of autonomy and agency rules out machine responsibility. Unless the machine is conscious, human actors always need to be responsible for decisions taking by machines. However, as self-improving systems increase machine autonomy and decrease human control, the question is raised whether we are witnessing an emerging responsibility gap, or if this development rather describes a situation of blurred responsibility, in which responsibility needs to be distributed between many different actors – AI developers, programmers, distributors, users, policy makers.

Artificiell intelligens (AI) har rötter i 1950-talets försök att utveckla datorprogram och robotar som kan efterlikna mänskliga förmågor som att använda språk, lösa problem och förbättra sig själva (Russell 2019; se även Kalle Åströms bidrag i denna volym) och har kommit att användas som ett samlingsnamn på teknik som med någon form av autonomi gör saker som människor gör, men ofta mycket snabbare. AI-teknik ligger bakom många tillämpningar som de flesta av oss möter mer eller mindre dagligen och snabbt har tagit för givna

Maria Hedlund är verksam vid Statsvetenskapliga institutionen, Lunds universitet.
E-post: maria.hedlund@svet.lu.se

såsom sökmotorer som vet var du befinner dig och minns vad du har sökt på tidigare, mobiltelefoner som låses upp med hjälp av ansiktsgenkänning, strömningstjänster som tipsar om musik eller teveprogram som du kan tänkas tycka om, eller sociala medier, som visar inlägg från användare med liknande åsikter. Andra exempel på områden där AI-teknik används är rekrytering, självkörande fordon, automatiserad börshandel och inom vården, där bildanalys används för att ställa diagnoser. AI har blivit något av ett *buzzword* i den offentliga debatten och man kan till och med stöta på uttrycket ”en AI” som i exempelvis ”En AI vi förstår” (Voister), ”Fejkade vd:ns röst med en AI – kom över 2,4 miljoner” (Techworld) och ”Kan en AI beräkna och förutsäga progressiv utveckling?” (Quora), vilket ger intrycket att AI i sig uppfattas som ett självständigt subjekt. Det är inte alldeles långsökt, eftersom AI, som antyds ovan, ofta förknippas med autonomi, dvs. självständigt agerande. Detta blir tydligt i den definition av AI som Europeiska kommissionen använder:

Artificiell intelligens (AI) syftar på system som uppvisar intelligent beteende genom att analysera sin omgivning och agera – med någon grad av autonomi – för att uppnå specifika mål (EU-kommissionen 2019, förf.:s översättning).

Exakt hur självständigt ett AI-system behöver vara för att kallas autonomt är emellertid omstritt. Räcker det att maskinen, så snart människan ”tryckt på knappen”, utför en förprogrammerad rutin? Eller krävs att maskinen inte bara gör något på egen hand, utan också självständigt kan bestämma vad det är den ska göra och hur den ska göra det (jfr Burri 2016)? Maskiner som kan agera (någorlunda) självständigt väcker frågor om ansvar. Vem är till exempel ansvarig för beslut som en självlärande algoritm fattar? När Googles sökmotor lär sig vad du brukar söka efter och ger dig träffar som kan tänkas ligga i linje med dina preferenser och inte låter dig se andra relevanta träffar, kan det begränsa din bild av världen och i förlängningen din förståelse av andra människors perspektiv. Kan vi då säga att Googles algoritm är ansvarig för dessa sökträffar? När en självkörande bil orsakar en olycka på grund av att smuts på stoppskylten får bilen att uppfatta trafikmärket som en 90-skyld och ökar farten – kan vi då säga att bilen är ansvarig? Och vad skulle det i så fall innebära för bilen? Den skulle knappast lida av ett straff, än mindre förstå att det var ett straff.

Vanligtvis brukar vi säga att maskiner inte kan vara ansvariga för sina handlingar. Det har dock från politiskt håll rests krav på att robotar ska hållas ansvariga för skador som de orsakar. I januari 2017 föreslog EU-parlamentets juridiska utskott att robotar skulle få status som ”elektroniska personer” för att kunna hållas ansvariga för sina handlingar (EP 2017: 59F). Detta förslag var dock inget som togs vidare av EU-kommissionen, som sedan hösten 2017 arbetar med att komma fram till ett europeiskt förhållningssätt till AI (EC

Communication 2018).¹ Det faktum att politiska aktörer lyfter frågan, kan dock ses som ett erkännande av att utvecklingen inom artificiell intelligens och robotik utmanar synen på ansvar (Gunkel 2017: 5).

För att kunna ta ansvar, krävs ett handlande subjekt och någon form av agentskap, eller agens, det vill säga någon form av kontroll över sina handlingar (Kane 2002: 698). När vi talar om juridiskt ansvar, är också subjektets intentionalitet avgörande (Lanteri 2009). Den (eller det) som saknar båda dessa egenskaper (agens och intentionalitet) kan därför inte hållas juridiskt ansvarig (HRW 2015: 18). Men även om en maskin inte kan hållas juridiskt ansvarig, kanske den kan ha moraliskt ansvar för vad den gör?

Denna artikel tar visserligen utgångspunkt i observerbara tekniska landvinningar (inom AI och robotik), men syftet är främst att på ett principiellt plan diskutera och problematisera ansvarsförhållanden när maskiner fattar beslut. Den övergripande frågan som ställs är om maskiner kan vara ansvariga för beslut som de fattar, och i så fall, under vilka omständigheter. Med maskin avses här såväl mjukvara (algoritmer) som hårdvara (robotar) som utför uppgifter med en viss grad av artificiell intelligens.

Framställningen tar sin början i ansvarsbegreppet – vad vi vanligtvis menar med ansvar och vilken sorts ansvar som är relevant här – för att därefter diskutera i vilken utsträckning beslutsfattande maskiner kan ses som ansvarstagande eller ansvarsbärande agenter. Den teknologiska utvecklingen har inneburit allt självständigare maskiner och allt mindre direkt mänsklig kontroll över vad maskinerna faktiskt gör och hur de kommer fram till vad de ska göra. Innebär detta att människors ansvar successivt övergår till maskiner? Vad skulle det i så fall betyda, rent konkret, att maskiner har ansvar? Har det överhuvudtaget någon betydelse för ansvarsfrågan vad maskiner kan göra? Aspekter som autonomi och agens blir här centralt, men också relationen människa – maskin.

Vilken sorts ansvar?

När vi talar om ansvar tänker vi ofta på kausalt ansvar, det vill säga att den som har orsakat ett visst utfall också är ansvarig – svarar för – detta utfall. För att vara ansvarig, räcker det dock inte att ha orsakat utfallet. Om en aktör ska kunna hållas ansvarig krävs också att hon har eller kan förväntas ha adekvat kunskap om de kausala förhållandena och konsekvenserna av att handla eller att inte handla (Thompson 1987: 47). En förutsättning för ansvar är också att aktören har haft möjlighet att handla annorlunda (Thompson 1987: 40). För att tillskrivas ansvar måste aktören alltså kunna bindas till en handling vars

1 Ett annat exempel är Saudiarabien, som i oktober 2017 gav medborgarskap till roboten Sofia (Hart 2018). Det väckte en hel del kritik att ett av världens mest förtryckande länder gav en robot rättigheter som vida överstiger de rättigheter som landets kvinnor har, men det hela var sannolikt ett sätt att få uppmärksamhet och att symbolisera att Saudiarabiens ekonomiska framtid inte bara handlar om olja.

kausala förhållanden hon förstår (Adam & Groves 2011: 18) och som hon kan kontrollera (Fischer & Ravizza 1998: 14). Det går med andra ord inte att hålla någon ansvarig för en handling som vederbörande utför under tvång (Talbert 2008: 517).

En distinktion kan göras mellan kausalt ansvar och moraliskt ansvar, där skillnaden ligger i att moraliskt ansvar också involverar skuld och klander (Miller 2001: 458). Moraliskt ansvar förutsätter en moralisk agent, alltså en aktör som är medveten om – har kunskap om – en handlingens moraliska natur (Held 1970: 97). Med moralisk avses här för enkelhetens skull normer om vad som är rätt och gott i en viss kontext, normer som kan, men inte behöver vara juridiskt bindande. En aktör som oavsiktligt orsakar en dålig situation kan således tillskrivas kausalt men inte moraliskt ansvar, eftersom hon inte har gjort något klandervärd (Talbert 2008: 516). Juridiskt ansvar skulle kunna ses som ett specialfall av moraliskt ansvar, eftersom det är en viss typ av normer som avgör om en handling är klandervärd eller inte, nämligen legala normer. Dessutom är legala normer något vi är skyldiga att känna till; vi befrias inte från juridiskt ansvar med hänvisning till att vi inte kände till lagarna.

Resonemanget så här långt utgår från att vi talar om huruvida ansvar kan läggas på en individ. Flera individer som agerar tillsammans komplicerar ansvarsfördelningen. Det kan ifrågasättas varför vi överhuvudtaget behöver bry oss om ansvarsfördelning mellan flera individer när vi funderar över om maskiner kan vara ansvariga. Frågan om maskiners ansvar kommer dock till viss del att behöva handla om deras ställföreträdarens ansvar, vilket aktualiserar en diskussion om kollektivt ansvar. Ett kollektiv är något mer än ett aggregat av flera individer, då medlemmarna i ett kollektiv har någon form av band till varandra som aggregatet saknar (Gilbert 2006: 95, 102). Frågan är om kollektivet kan hållas ansvarigt som ett kollektiv, eller om ansvaret ska fördelas mellan de olika medlemmarna, och i så fall hur.

Ett omstritt sätt att hålla ett kollektiv ansvarigt är att tillskriva ansvar till kollektivet som en enhet (Thompson 1987: 44). Huvudfrågan här är huruvida ett kollektiv kan betraktas som en moralisk agent, vilket är en förutsättning för moraliskt (och juridiskt) ansvar. En organiserad grupp skulle möjligen kunna betraktas som en moralisk agent och därmed kunna tillskrivas ansvar på motsvarande sätt som en individ (Held 1970: 97), det vill säga på grundval av kausalitet, medvetenhet, och kontroll. En annan komplikation med kollektivt ansvar är huruvida och under vilka förutsättningar individer kan hållas ansvariga för andra individers handlingar (McGarry 1986: 79). En sådan förutsättning är gruppssolidaritet, vilket, även om inte alla medlemmar är klandervärda, skulle utgöra en grund för ställföreträdande kollektivt ansvar (Feinberg 1970: 61–62, 73). I hierarkiska organisationer kan formella positioner avgöra vem som är (juridiskt) ansvarig för vad kollektivet gör (May and Hoffman 1991: 11). En arbetsgivare kan exempelvis vara ansvarig för vad de anställda gör som

om arbetsgivaren själv hade utfört handlingen. Ett sådant ställföreträdande ansvar motiveras av att arbetsgivaren kontrollerar vem som anställs (kontrollargument) och har större möjlighet att ta (ekonomiskt) ansvar (positionsargument) (Feinberg 1968: 675–676).

Det som hittills har sagts om ansvar avser bakåtblickande ansvar, som handlar om att identifiera den som är ansvarig för något dåligt som redan har hänt i syfte att på något sätt straffa den felande. I politikens värld syftar detta ofta på att den ansvariga ska avgå. Allmänna val sägs till exempel vara medborgarnas sätt att ”utkräva ansvar” av politikerna. Inom juridiken hittar vi bakåtblickandet i olika skäl till varför den som brutit mot lagen ska straffas: dels handlar straffet om ett sätt för brottslingen att sona sitt brott genom någon form av återbetalning (retribution) till samhället, dels syftar straffet till att ”kommunicera samhällets fördömande av det klandervärda beteendet” (Wahlberg 2019: 125). Rättfärdigande av straff genom att framhålla dess preventiva funktion är mer framåtblickande, både när det handlar om att förhindra den individuella brottslingen från att begå brott, i alla fall temporärt, och om att avskräcka andra från att begå brott (Wahlberg 2019: 125).

En framåtblickande syn på ansvar, såväl juridiskt som moraliskt, tar fasta på just vad som ska hända i framtiden. Distinktionen mellan å ena sidan att *bära* ansvar, med betydelsen att klandras för något som redan har hänt, och å andra sidan att *ta* ansvar, med betydelsen att göra saker nu med en blick på deras framtida konsekvenser (Rovane 2014: 12), är belysande. På engelska görs denna åtskillnad i språket, där *accountability* eller *liability* står för bakåtblickande ansvar, medan *responsibility* kan, men inte behöver, syfta på framåtblickande eller prospektivt ansvar. På svenska talar vi ibland explicit om ansvarsutkrävande när vi syftar på bakåtblickande ansvar, men ofta är den bakåtblickande betydelsen underförstådd när vi pratar om ansvar. Det blir därför viktigt att särskilt framhålla när framåtblickande ansvar åsyftas.

En poäng med framåtblickande ansvar är att det, istället för att invänta att dåliga saker händer, identifiera den som har gjort fel och straffa den felande, riktar fokus mot hur vi kan *undvika* att dåliga saker händer. Framåtblickande ansvar handlar alltså om att ta ansvar för att göra gott i framtiden, vilket innebär att kausalitet inte får samma framträdande roll i bedömningen av vem det är som är ansvarig. Kausalitet åberopas ibland som skäl för att peka ut vilka aktörer som bör ta ansvar för framtiden, exempelvis i debatten om kompensation för klimatförändringarnas effekter. Det som då åsyftas är att de som har orsakat dessa effekter också bör vara ansvariga för att kompensera dem som drabbas av dem. Kausala förhållanden är komplexa, men även om de skulle vara svåra eller i praktiken omöjliga att spåra (jfr Albertsen 2015), finns de. Svårigheter att identifiera de ansvariga aktörerna, som när det gäller klimatförändringar både är många – länder, regioner, företag, beslutsfattare – och inte alltid längre existerande, gör emellertid att kausalitet inte förefaller vara någon

framkomlig väg för att fördela detta ansvar. Dessutom kan man fundera över om detta överhuvudtaget handlar om framåtblickande ansvar, utan snarare är ett fall av retribution (jfr ovan och Wahlberg 2019).

Andra grunder än kausalitet förefaller rimligare för fördelning av framåtblickande ansvar. Makt och resurser kan utgöra en sådan grund. En aktör som befinner sig i en position som gör att hon har möjlighet att agera eller som besitter kapacitet att agera kan ha en moralisk skyldighet att förbättra en dålig situation, oavsett just den aktörens roll i uppkomsten av den dåliga situationen (Miller 2001: 460; Young 2006: 126; jfr Nihlén Fahlquist 2009: 111). Aktören kan tillhöra en grupp av aktörer som, nu eller i förfluten tid, har bidragit till exempelvis klimatförändringar. Principen om gruppsolidaritet (jfr ovan) kan då rättfärdiga tillskrivandet av ansvar.

Maskinen som expert

Den AI som finns idag är specialiserad på en enda uppgift. Ansiktsgigenkänningsprogram, stöd för cancerdiagnoser inom sjukvården och rekryteringsprogram är exempel på olika tillämpningar som har det gemensamt att de är tränade på att känna igen en viss sorts mönster. Detta gör de med större precision ju fler ansikten, concertumörer eller rekryteringar de har exponerats för, och man skulle kunna säga att de blir experter på sin enda uppgift. Kanske kan vi jämföra maskinen med den mänskliga experten när det gäller frågan om ansvar? En expert brukar definieras som någon med överlägsen specialistkunskap som kan tala med auktoritet inom ett väl avgränsat område, men inte på andra områden (Schudson 2006: 499; Turner 2003: 6), så i det avseendet finns en parallell. Skulle vi kunna överföra resonemang om mänskliga experters ansvar till ett resonemang om expert-maskinens ansvar?

När är experten ansvarig? Från ett idealtypiskt rationalistiskt perspektiv är experter i politiska beslutsprocesser neutrala leverantörer av kunskap, medan politiker gör värderingar och fattar beslut, som de sedan kan ställas till ansvar för. Detta synsätt, som också ligger till grund för den formella uppdelningen av roller i politiska beslutsprocesser, befriar experten från ansvar för de beslut som den levererade kunskapen bidrar till. Överfört till expert-maskinen skulle alltså inte heller den vara ansvarig för beslut som bygger på den kunskap som maskinen levererat. Det kan tyckas som att saken därmed borde vara avgjord, men åtminstone två aspekter komplicerar ansvarsfrågan.

Den ena komplikationen är att det rationalistiska synsättet är idealtypiskt (och formellt) och inte nödvändigtvis beskriver vad som faktiskt sker vid (politiskt) beslutsfattande. Ett konstruktivistiskt synsätt ifrågasätter förekomsten av neutral kunskap och pekar på vikten av att uppmärksamma den politiska roll som experten i praktiken kan spela. Dels kan experten konsulteras för att ge en rekommendation, och en rekommendation bygger alltid på någon form av

värdering. Dels kan experten i kraft av sin kognitiva auktoritet (Turner 2003: 23) avsiktligt eller oavsiktligt smyga in värderingar i det som kan förefalla vara neutral information. Forskares kognitiva, eller epistemiska, auktoritet i samhället ger dem stora möjligheter att påverka politiska beslut (Gundersen 2018: 54–55), och om experter de facto spelar en politisk roll, är det inte självklart att de ska vara befriade från ansvar. Om en expert-maskin på motsvarande sätt som en mänsklig expert ger rekommendationer, som exempelvis sker inom missbruksvården i Helsingborgs kommun (AI Lund 2020-01-29; MVTe 2020-01-21) men också i flera andra kommuner i Sverige (Svensson 2019: 27–32) och i Norden (Nordiska ministerrådet 2019: 18–20), eller till och med fattar beslut, som den så kallade handlägggarroboten inom socialtjänsten i Trelleborg (Trelleborg 2015), skulle man alltså kunna hävda att även den borde ha ett ansvar.² Eller kan man det? Är det inte skillnad på maskiner och människor? Detta leder in på en andra aspekt som komplicerar ansvarsfrågan.

Den andra komplikationen är just att maskinen är en maskin. Även om maskinen bearbetar information och drar slutsatser som resulterar i en rekommendation, eller till och med ett beslut, gör den det på grundval av någon form av förprogrammerad logik. Det kan se ut som att maskinen fattar ett självständigt beslut, men den gör det på premisser som en människa (eller flera människor) har bestämt. Och till skillnad från den mänskliga experten, kan inte maskinen gå utöver sin uppgift att leverera svaret på den fråga som ställts, till exempel huruvida en viss person är berättigad till ett visst bidrag. Där den mänskliga handläggaren har möjlighet och förmåga att ta individuella hänsyn, genom att beakta faktorer som inte finns med i 'formeln', kan maskinen bara göra det den är programmerad att göra. Som vi såg i ansvarsdiskussionen ovan, kan möjligheten att göra annorlunda vara en förutsättning för ansvar. Att jämföra maskinen med en mänsklig expert är alltså inte helt adekvat för frågan om ansvar.

2 Huvudregeln är att automatiserade individuella beslut inte är tillåtna (GDPR art. 22.1). Med detta avses beslut med rättsliga följder (t.ex. diskriminering eller att en anställning sägs upp) eller som på liknande sätt i betydande grad påverkar en människa och som enbart grundas på automatiserad behandling. En människa måste fatta själva beslutet (på riktigt, inte bara symboliskt), men beslutsunderlaget kan tas fram av AI, algoritmer etc. Detta gäller både privata företag och offentlig sektor. Det finns dock ett antal undantag till denna huvudregel. Ett sådant undantag är om automatiserat beslutsfattande är tillåtet enligt nationell lag (GDPR art. 22.2). Den nya förvaltningslagen (2017:900) som började gälla 1 juli 2018 öppnar upp för automatiserat beslutsfattande (28 §). Rättsläget är dock inte helt klart, och olika aktörer gör olika tolkningar. Enligt SKR är automatiserat beslutsfattande tillåtet i staten, men inte inom kommuner och regioner, och SKR bedriver ett påverkansarbete för att automatiserat beslutsfattande ska vara lagligt även i kommuner och regioner (skr.se). I skrivande stund pågår en utredning (dir. 2020:10) som bland ser över hur detta kan införas i kommuner och regioner. Utredningens förslag väntas i december 2020.

Om maskinen är autonom?

Maskiner kan alltså inte vara ansvariga, även om de i teknisk mening kan fatta beslut eller ge rekommendationer. Till skillnad från människor, saknar de självbestämmande och förmåga att reflektera över sina handlingar. Maskiner har heller ingen uppfattning om rätt och fel, och saknar därför förmåga att göra moraliska bedömningar. Utvecklingen av självlärande maskiner och autonoma system som agerar utan direkt mänsklig inblandning väcker dock frågor om huruvida maskiner kan och bör vara ansvariga. Autonomi, eller självbestämmande, brukar anföras som en förutsättning för ansvar. Det man inte själv har bestämt, kan man inte heller svara för, men om en maskin är autonom, kan den då också vara ansvarig för de beslut den fattar? Obemannade fordon, finansrobotar och robotar som kan ställa medicinsk diagnos är exempel på maskiner som ofta omnämns som autonoma system (EGE 2018), men kan vi säga att dessa maskiner är autonoma i den mening som förutsätts för ansvar?

Ordet autonomi kommer av grekiskans *autos* (själv) och *nomos* (lag) och kan uttydas som att den som är autonom, bestämmer sina egna lagar (Hakli & Mäkelä 2019: 264). Autonomi betyder alltså självstyre. En autonom stat eller region har friheten att styra sig själv, och en individ som är autonom har förmåga att agera och fatta beslut utan att kontrolleras av någon annan – men kan en maskin vara autonom? En fabriksrobot som dag efter dag, dygnet runt, utför sina uppgifter utan inblandning av någon människa, är den autonom? Och hur är det med självlärande system, som självkörande fordon eller spelet AlphaGo, som blir bättre och bättre på att utföra sina uppgifter ju mer de tränar på dem? Även om maskininlärning har effekten att fordonets eller spelets autonomi ökar, bygger denna autonomi på förprogrammerad kod (Burri 2016: 349), och kan vi då tala om autonomi i betydelsen självständig och oberoende? Det kan här vara på sin plats att påpeka att termen 'autonom' används på olika sätt i olika sammanhang.

Den tekniska förståelse av autonomi som forskare inom AI och robotik använder kan uttryckas som maskinens förmåga att uppfatta omgivningen, att lära från sina erfarenheter och att agera oberoende av en extern operatör (Hakli & Mäkelä 2019: 264). Autonomi tolkas här relativt till mängden mänsklig inblandning medan roboten utför sin uppgift (Haselager 2005: 518). En robot är alltså inte autonom om en människa styr den medan den agerar. Däremot anses maskinen ha autonomi när den kan utföra en förprogrammerad uppgift utan inblandning från en människa (Burri 2016: 346; Jain & Pratihari 2010: 1). Även när en människa har bestämt vad maskinen ska göra, anses den i teknisk mening vara autonom så länge själva genomförandet av uppdraget sker utan mänsklig vägledning (Haselager 2005: 518).

Ur ett filosofiskt perspektiv däremot, är en robot som följer en rutin som bestämmer robotens handlingar inte autonom, utan följer bara de direktiv som en människa (eller flera människor) på förhand har satt upp (Burri 2016: 354).

En filosofisk, konceptuell förståelse av autonomi betonar kapaciteten att agera självständigt och att göra sina egna val (Haselager 2005: 519). Att vara autonom innebär att vara självbestämmande, oberoende och inte styrd av någon extern kraft. I Kants mening syftar autonomi på grundläggande villkor för den fria viljan, nämligen viljans kapacitet att följa moraliska lagar som den har givit sig själv (Gunkel 2017: 4). Det är således inte hur, utan varför, en uppgift utförs som är av intresse. För att vara autonom i filosofisk mening måste en maskin alltså ha förmåga att välja sina mål och att agera fritt utan mänsklig kontroll, varken under själva genomförandet av målet eller dessförinnan (Hakli & Mäkelä 2019: 265). Sådana förmågor hänger ihop med attityder och värderingar, principer och önskningar, och den självkontroll och det omdöme som är en förutsättning för att verkligen kunna göra det som krävs för att uppnå sitt mål (Hakli & Mäkelä 2019: 265), egenskaper som förknippas med agentskap (mer om det nedan).

Skillnaden mellan den ingenjörsmässiga och den filosofiska förståelsen av autonomi har delvis sin grund i olika temporala perspektiv. Det tekniska synsättet är inriktat på vad som är genomförbart just nu, medan det filosofiska synsättet är mer inriktat på framtiden (Burri 2016: 355) och på vad som är principiellt möjligt. Detta aktualiserar frågan om vilka maskiner som diskussionen om ansvar bör handla om: maskiner som finns just nu, eller alla maskiner, även sådana som kan komma att finnas i framtiden. Vi bevittnar nu en oerhört snabb utveckling av maskiner som lär sig själva utan att få nya instruktioner från dem som designar och programmerar dem. Några exempel är det ovan nämnda spelet AlphaGo, som är designat för att formulera sina egna instruktioner (Gunkel 2017); rekryteringsprogram som lär sig att hitta lämpliga kandidater genom att exponeras för tidigare rekryteringar (O'Neill 2016); och sociala medier som Facebook, som hela tiden anpassar vad som visas i varje användares flöde utifrån vars och ens interaktionsmönster (Bucher 2018). Sjävlärande maskiner av detta slag lär sig inte genom att få nya instruktioner från programmeraren, utan genom att interagera med omgivningen, med användarna och med olika situationer, och blir därmed allt mer autonoma i filosofisk mening (Laukyte 2017: 12).

När den omgivande miljöns påverkan ökar och programutvecklaren successivt förlorar kontroll, kan vi verkligen fråga oss vem som är ansvarig för vad maskinen gör, och inte minst vem som är ansvarig för konsekvenserna av maskinens beslut. När det gäller spelet AlphaGo kan konsekvenserna betraktas som relativt harmlösa (om man inte heter Lee Sedol och inte längre är en oslagbar Go-spelare), men konsekvenserna av diskriminerande rekrytering eller att var och en presenteras olika världsbilder i sociala medier är betydligt allvarigare. Vem, om någon, kan rimligen hållas ansvarig för dessa effekter? Bevittnar vi, som Matthias (2004) hävdar, en ansvarslucka – *a responsibility gap* – där det inte är möjligt att förutse maskinens beteende och därmed inte att hålla någon

ansvarig för vad maskinen gör? Eller kan en maskin, som utför sina uppgifter med en viss grad av autonomi vara ansvarig för vad den gör?

I en strikt kausal och momentan mening skulle man möjligen kunna säga att en autonom maskin är ansvarig (jfr Fischer 2007: 65; Hakli & Mäkelä 2019: 269). I det specifika ögonblick som den fattar ett beslut – till exempel att visa dig reklam för en viss produkt i ditt Facebook-flöde – orsakar maskinen, i det här fallet algoritmen, att du får se just det du får se. Men det är svårt att hävda att maskinen också skulle ha ett moraliskt ansvar för att du får se just denna produkt. Moraliskt (eller juridiskt) ansvar förutsätter en moralisk agent, och autonomi i teknisk mening eller en viss grad av autonomi i filosofisk mening är otillräckligt för moraliskt agentskap. Men om vi släpper tanken på de maskiner som existerar här och nu, och istället funderar mer principiellt över möjliga framtida maskiner som kan agera med ännu större autonomi: kan en maskin någonsin vara en moralisk agent?

Om maskinen har moralisk agens?

Enligt ett instrumentalistiskt synsätt är det alltid människan som är ansvarig för vad ett teknologiskt system gör, hur oberoende, automatiskt eller interaktivt det än är, eftersom det alltid kommer att vara en produkt av mänskligt beteende, mänskliga institutioner och mänskliga beslut (Gunkel 2017: 3). Men självlärande system, med tillämpningar som exempelvis prognoser, datorspel eller mönsterigenkänning, och maskiner som har förmåga att förbättra sig själva – gör de inte något på egen hand och som är oberoende av mänsklig inblandning? Det finns de som hävdar att maskiner under vissa omständigheter kan ha moralisk agens, medan andra menar att det varken är möjligt eller önskvärt. Men vad innebär det att vara en moralisk agent?

Typiskt sett brukar en moralisk agent förstås som någon (eller möjligen något) som kan skilja på rätt och fel eller som kan hållas ansvarig för sina handlingar (Malle 2016: 244). För att vara en moralisk agent måste en aktör alltså vara medveten om en handlings moraliska natur (Held 1970: 97). Ofta anförda villkor på moralisk agens är olika kapaciteter såsom rationalitet, intentionaltet, förmåga till självreflexion, känslighet för moraliska skäl, förmåga att erfara känslor, social förmåga och att vara en person (*personhood*) (Hakli & Mäkelä 2019: 262). En moralisk agent måste också ha möjlighet att handla annorlunda (Fischer 2006: 65). Moraliskt agentskap förutsätter både handlingsfrihet och handlingsförmåga, annars kan individen inte på ett övertygande sätt prisas eller klandras för sitt beteende (Sondermann m.fl. 2018: 3; jfr Lundquist 1992: 41). Agentskap är emellertid i stor utsträckning villkorad av sociala praktiker och vanor, och inte ett resultat av fri vilja (Sondermann m.fl. 2018: 2).³ Den

3 För en översikt över den fria viljans problem, se Fischer 1999.

dominerande uppfattningen är ändå att vuxna människor är moraliska agenter, medan icke-mänskliga agenter som djur, robotar och grupper inte är det (Hakli & Mäkelä 2019: 260, 265). Med den ökande betydelsen av artificiell intelligens och robotik, är det emerllertid viktigt att uppmärksamma också icke-mänsklig agens, inte bara därför att dessa intelligenta system utvecklas till att agera och att lära sig mer och mer självständigt (Sondermann m.fl. 2018: 4), utan också därför att maskiner som agenter utmanar hur vi vanligtvis ser på ansvar. Frågan är alltså om maskiner kan ha moralisk agens, och om de i så fall kan eller bör vara ansvariga för de beslut de fattar.

Vissa forskare menar att moralisk agens inte bör ses som något antingen – eller, utan som ett kontinuum på motsvarande sätt som vi inte har några problem med att tillskriva barn särskilda kompetenser i grader eller stadier (Sharkey 2017: 211). Sullins (2006) menar till och med att robotar kan vara fullständiga moraliska agenter.⁴

Utgångspunkten för Sullins resonemang är att medvetande inte är en förutsättning för moral, och givet denna utgångspunkt, skisserar han tre förutsättningar: 1) maskinen måste vara signifikant autonom, i teknisk bemärkelse, i förhållande till programmerare eller operatörer, 2) den måste kunna uppvisa ett beteende som kan uppfattas som en avsikt att göra gott eller ont, och 3) den måste bete sig på ett sådant sätt att beteendet kan uppfattas som att den har ansvar för någon annan moralisk agent (2006: 28–29). Även om det just nu inte finns några maskiner som kan uppvisa dessa beteenden, blir AI-utvecklare allt mer kapabla att genom programmering imitera mänskligt beteende. De förutsättningar som Sullins anför är rent funktionalistiska och innebär inte att maskinen är autonom, har intentioner eller är ansvars-kännande i filosofisk mening, bara att den betar sig på ett sådant sätt att beteendet kan tolkas som att den agerar autonomt med intentioner och ansvarsfullt gentemot andra. Om exempelvis en vådrobot betar sig på detta sätt, skulle beteendet kunna tolkas som att den har en övertygelse om att den ska bry sig om sina patienter, och om beteendet kan tolkas så, då kan vi enligt detta sätt att resonera tillskriva den moralisk agens (Sullins 2006: 29). Det skulle alltså vara tillräckligt att maskinen är moralisk i funktionell mening för att kunna tillskriva den ansvar, ungefär på samma sätt som företag ses om ”personer” i juridisk mening. Samma moraliska och legala överväganden skulle enligt detta resonemang vara möjliga att överföra till en AI eller robot som Google’s DeepMind eller IBM:s Watson (Gunkel 2017: 10).

Till detta kan läggas vår tendens att antropomorifiera saker, och att vi

4 En extrem position är att robotar, men inte människor, är moraliska agenter, en uppfattning som faktiskt en person har sagt sig ha. I en artikel från 2006 hävdar Joseph Emile Nadeau att den frihet som krävs för moraliskt agentskap förutsätter förmåga att tänka strikt logiskt och fullständigt tänka igenom alla skäl, och eftersom den mänskliga hjärnan saknar den förmågan, kan vi inte vara moraliska agenter (Sullins 2006: 27).

tenderar att reagera emotionellt på sociala robotar. En social robot är social på ett människolikt sätt och att interagera med den kan i många avseenden vara som att interagera med en annan människa. Flera empiriska studier har visat att mänskliga användare, som ett resultat av social interaktion, ger social ställning till datorer som liknar den som vi ger till andra människor. Till exempel har människor visat förmåga att empatisera med simulerad "robotsmärta" (Gunkel 2017: 7–8). Men detta säger rimligtvis ingenting om maskinens moraliska agens. Det pekar snarare på hur sociala robotar suddar ut gränsen mellan moraliska subjekt och maskinen som ett rent instrument, vilket är kärnan i det problem som denna text tar sin utgångspunkt i.

Ett annat sätt att funktionalistiskt argumentera för att en maskin skulle kunna vara en agent är att jämföra med hur vi ser på grupper agent-skap. Liksom grupper, kan maskinen beskrivas som en icke-mänsklig agent, så om den uppfyller de villkor som vi ställer på grupper agent-skap, skulle även maskinen kunna erkännas som en agent, går resonemanget. Laukyte (2017) pekar på ett antal sådana krav, av vilka förmågan att agera rationellt och förmågan att interagera med sin miljö i sammanhanget är relativt oproblematiske. Dessa förmågor är utmärkande för de maskiner vi talar om här, nämligen robotar och algoritmer som bygger på artificiell intelligens. Mer utmanande är kravet på att vara en person (*personhood*). Vad det innebär att vara en person skiljer sig mellan olika kulturer och kontexter och är en kontroversiell fråga inom filosofi och juridik.⁵ Relevant i detta sammanhang är hur en funktionalistisk syn på vad det innebär att vara en person kan göra det möjligt att argumentera för att en maskin kan erkännas som en person, om än en artificiell sådan, och därmed tillerkännas agent-skap.

Laukyte utgår från ett performativt personbegrepp, utifrån vilket vi kan bedöma maskinen utifrån vad den gör eller kan göra. Vi har redan konstaterat att en maskin kan interagera med sin miljö, och detta kan nu kvalificeras till att handla såväl om den fysiska miljön som om den sociala miljön. Enligt det performativa personbegrepp som Laukyte använder, kan maskinen betraktas som en person om den kan agera i båda dessa miljöer: den fysiska världen och den sociala världens interaktioner. För att kunna agera i den sociala världen krävs kapacitet att operera inom ett system av krav och förväntningar och förmåga att interagera med människor och grupper inom etablerade konventioner. Om maskinen klarar detta, kan den enligt Laukyte (2017: 8–9) betraktas som en person, och uppfyller därmed även detta krav på agent-skap. Detta performativt-relationella synsätt påminner om Turing-testet, som handlar om möjligheten att skilja på en dator och en människa.⁶ Att vara en person

5 För en översikt, se Carrithers et al. 1985.

6 Det ursprungliga Turing-testet lanserades 1950 av matematikern och kryptoanalytikern Alan Turing, som satte upp ett imitationsspel där en utfrågare skulle avgöra om svaren kom från en man eller en kvinna. Turing ersatte sedan mannen (som det var) med en dator. Ansatsen att använda Turing-test för

handlar alltså om vad en agent kan göra i en social miljö och om den kan bete sig på ett sätt som överensstämmer med sociala förväntningar, och en funktionalistisk ansats gör det enligt Laukyte möjligt att låta komplexa socio-tekniska system – som teknologier baserade på artificiell intelligens – ansvara för skador som de orsakar eller bidrar till (2017: 9).

Andra menar att ett teknologiskt system aldrig, hur oberoende, autonomt eller interaktivt det än är, kan vara en moralisk agent. Invändningar mot det funktionalistiska synsätt på moralisk agens som Sullins och Laukyte ger uttryck för betonar social förmåga och betydelsen av att vara en biologisk varelse och förmågan att erfara känslor. Maskinens brist på känslor skulle göra den till en moraliskt blind, artificiell psykopat (Gunkel 2017: 10), medan biologin sägs vara grunden för anknytning och empati, för utveckling av komplexa sociala relationer, för förmåga att förstå och förutse andras handlingar, och för att erfara social smärta. Detta sägs leda till att människor – och i stor utsträckning även andra däggdjur – har en inneboende känsla för rättvisa (Sharkey 2017: 211). Att bry sig om sig själv och andra är grunden för utvecklingen av moral, men robotar bryr sig inte om sitt självbevarande eller att undvika smärta, och ännu mindre andras smärta (Sharkey 2017: 211). Enligt det biologiska argumentet beror det på att maskinen inte har en kropp på samma sätt som levande varelser. Man skulle kunna ta bort delar av roboten utan att den känner någon smärta eller oro, och ännu mindre skulle den bry sig om skada eller smärta som en människa drabbas av. Och även om roboten kan programmeras att bete sig som om den brydde sig om att skydda sig själva eller andra, skulle det bara vara möjligt genom mänsklig intervention (Sharkey 2017: 212).

Den mänskliga inblandningen återkommer i olika argument mot att tillskriva agentskap till maskiner. En sådan invändning är att det är logiskt inkorrekt att tillskriva agentskap, och därmed ansvar, till något som enbart är ett objekt under mänsklig kontroll (Gunkel 2017: 3). Det skulle också vara moraliskt problematiskt, då det skulle göra det möjligt att skylla våra misstag på datorer (Gunkel 2017: 3).

Som vi har sett, tas ofta den tekniska utvecklingen som intäkt för att maskiner bör kunna betraktas som agenter. När maskiner i allt större utsträckning klarar av sådant som brukar förknippas med mänskligt agentskap, dras slutsatsen att också maskiner kan tillskrivas agentskap med allt vad det innebär av rättigheter och ansvar. En invändning mot att på detta sätt utgå från kapaciteter tar fasta på hur kapaciteterna har uppstått. Även om robotar skulle ha alla kapaciteter som krävs för moraliskt agentskap, skulle enligt Hakli & Mäkelä (2019) deras historia beröva dem autonomi på ett sätt som underminerar ansvar. Även om maskiner kan ha kapaciteter som intentionalitet och rationalitet, är det tveksamt om de skulle kunna ha förmåga till självreflexion, känslor,

att tillskriva *personhood* till robotar istället för att helt enkelt säga att de gör något, har kritiserats (se till exempel Bringsjord 1992). Turing beskrivs ofta som den artificiella intelligensens fader.

och normativ förståelse. Och även om alla dessa kapaciteter kommer att kunna implementeras i robotar, skulle detta inte göra robotarna till moraliska agenter, eftersom det spelar roll hur kapaciteterna har uppnåtts (Hakli & Mäkelä 2019: 264). Liksom Laukyte ovan, liknar Hakli & Mäkelä roboten vid grupper, men de kommer till en annan slutsats. De pekar på att kollektiva agenter, som företag och andra grupper, inte har den kontroll över sina handlingar som krävs för moraliskt ansvar, utan kontrollen ligger i de individuella medlemmarnas händer (2019: 266–267). På samma sätt är robotens agentskap beroende av sin skapare. Utan en historia av att själv erövra kapaciteter, värderingar och attityder, är robotens agentskap inte autentiskt, och det går därför inte att tillskriva roboten agentskap. Manipulationen undergräver autonomi, och därmed möjligheterna att hålla roboten ansvarig (2019: 268). Medan Laukyte (2017: 3) hävdar att vi inte behöver bry oss om maskinens ”inner workings”, det vill säga hur den har kommit fram till vad den ska göra, utan bara vad den faktiskt gör, menar Hakli & Mäkelä att det är just ”the inner workings” som är relevant för om vi kan tillskriva en maskin moralisk agens.

Maskinen som husdjur

I diskussioner om huruvida framtida robotar ska ha rättigheter är det vanligt att jämföra maskinen med ett djur (Sullins 2006; Gunkel 2017; Köhler m.fl. 2018). Jämförelsen mellan maskin och djur vilar på antagandet att det inte är någon signifikant moralisk skillnad mellan en maskin och ett djur (Hogan 2017: 29). En ofta anförd analogi är den mellan sociala robotar och familjehunden. Familjehunden agerar utifrån sina egna önskningar, men är också tränad att utföra vissa uppgifter i vissa situationer. På motsvarande sätt försätts den sociala roboten eller drönaren i särskilda omständigheter i vilka den sedan agerar autonomt (Köhler m.fl. 2018: 57–58, 63). Programmeringen av roboten är här att jämföra med dresserandet av hunden (Johnson & Verdicchio 2018: 296).

Den huvudsakliga invändningen mot husdjursanalogin är att den är missledande. Djur har medvetande, vilket roboten har, och djur har förmåga att uppleva känslor och kan lida, vilket roboten inte heller kan (Johnson & Verdicchio 2018: 294). En invändning mot denna invändning är att framtida robotar skulle kunna ha medvetande och lida. Djur får moralisk status från sin kapacitet att lida, och om robotar har kapacitet att lida, bör de också ha moralisk status (Johnson & Verdicchio 2018: 295). Men så länge maskiner inte har denna förmåga, leder jämförelsen fel. Djurens fysiska existens är ett skäl till deras moraliska status (jfr diskussionen om biologi ovan), medan maskinens eventuella moraliska status vilar på deras mentala kvaliteter: autonomi, intelligens och rationalitet (Hogan 2017: 31).

Andra argument för att robotar och djur inte är jämförbara utgår från hur deras relation till människor skiljer sig åt. För maskiner är frågan om huruvida

de ska behandlas etiskt, men när det handlar om djur, är frågan på vilket sätt de ska behandlas etiskt (Hogan 2017: 29). Vidare, i diskussioner om maskinens moraliska status handlar frågan om huruvida de ska betraktas som moraliska agenter, det vill säga om de ska kunna hållas ansvariga för skador som de orsakar, medan det för djur handlar om huruvida de ska betraktas som moraliska patienter (eng. *patient*), det vill säga, huruvida de förtjänar moralisk hänsyn (Hogan 2017: 30). En standarduppfattning inom etiken är att alla agenter också är patienter, medan alla patienter inte nödvändigtvis är agenter. Familjehunden förtjänar etisk hänsyn trots att den inte tar ansvar för sina handlingar. Robot-etiken besvarar frågan om robotens patientskap med hänvisning till robotens agentskap, medan djuretiken erkänner moraliskt patientskap till varelser som vanligtvis förnekas moraliskt agentskap (Hogan 2017: 30–31). Djur och maskiner är således inte jämförbara när det gäller frågan om moralisk agens.⁷

Om maskinen är medveten?

Men om maskinen är medveten, då borde den väl kunna vara ansvarig för beslut som den fattar?⁸ Ja, givet att den också har adekvat kunskap om orsaksförhållanden och moral, och att den har möjlighet att agera annorlunda – det vill säga de krav på ansvar som vi ställer på mänskliga beslutsfattare – då skulle medvetande så att säga putta den över gränsen och göra maskinen till ett ansvarstagande subjekt. Medvetna maskiner är, om något, förknippat med science fiction, men det finns ändå tänkare, exempelvis inom transhumanismen, som tror att den snabba tekniska utvecklingen inom AI kan leda till att maskiner så småningom blir medvetna.

En avgörande skillnad verkar finnas mellan å ena sidan biologiska naturalister, som hävdar att förmågan att vara medveten är unik för biologiska organismer, och å andra sidan teknologiska optimister, som menar att medvetande är *substrate independent*, alltså oberoende av vilket material någonting är

7 Detta förhindrar inte att det ändå kan vara relevant att jämföra robotar med djur. Det kantianska argumentet att hur vi behandlar djur har betydelse för vad det gör med oss som människor kan utvidgas till att också gälla hur vi behandlar robotar: Om vi behandlar robotar (och djur) illa, blir vi inhumana människor. Vår behandling av robotar (och djur) kan förstärka mänskligt beteende som är moraliskt korrekt eller inkorrekt. Att behandla robotar (och djur) på ett grymt sätt gör oss känslomässigt avtrubbade och kan underminera den empati vi har för andra människor. Men vi vet inte om det finns en sådan överföringseffekt när det gäller djur, och om den finns, hur kraftfull den i så fall är. Och överfört till robotar blir argumentet ännu svagare, eftersom robotar inte lider. Eftersom roboten inte lider när någon slår den, är det inte adekvat att säga att detta beteende är grymt. Inte heller denna analogi mellan maskinen och djuret håller, eftersom maskinen inte lider (Johnson & Verdicchio 2018: 298). Men att ge robotar rättigheter kan ändå ha betydelse för vilka vi blir som människor, då roboten kan tjäna som en etisk förebild för människan (Levy i Hogan 2017: 31).

8 Det är inte helt lätt att definiera medvetande, men känslor, inre mentalt liv, inre erfarenheter och subjektiv erfarenhet ringar in vad det handlar om (jfr Schneider 2019: 16; Tegmark 2017: 315) och fångar på ett ungefär in vad medvetande är. För en mer ingående diskussion, se t.ex. Blackmore & Troscianko 2018.

uppbyggt av (jfr Schneider 2019: 17–26; Tegmark 2017: 315). Om och när människor utvecklar sofistikerade generella AI:ar som flexibelt skiftar mellan olika intellektuella uppgifter och som överträffar människans förmåga att resonera, så kommer dessa att ha medvetande, enligt denna position. Då – men inte förr – skulle vi kunna tillskriva maskiner ansvar.

Vi kan förstås inte veta om medvetna maskiner någonsin kommer att utvecklas och hur långt in i framtiden en sådan utveckling i så fall ligger (Bostrom 2014: 18–21). Än mindre kan vi veta hur vi kan veta att en maskin är medveten. Om vi accepterar slutsatsen att medvetande är en förutsättning för att vara ett ansvarstagande subjekt, är det en avgörande fråga om vi kan veta att maskinen verkligen är medveten. Även om maskinen betar sig som om den är medveten, behöver det inte innebära att den också är det. Den – eller de som utvecklar och säljer dem – kanske bara vill få oss att tro det. Medvetande är inte en oundviklig konsekvens av teknisk utveckling, och även om framtida maskiner skulle kunna vara medvetna, skulle självförbättrande system kanske utveckla bort medvetandet – kanske för att slippa ta ansvar? Svårigheten att verkligen avgöra om en maskin är medveten föranleder filosofen Susan Schneider att förespråka en försiktighetsprincip: från ett etiskt perspektiv är det bäst att anta en sofistikerad AI kan vara medveten (2019: 149).

Är maskinen ett hinder för ansvar?

Om nu slutsatsen blir att maskiner – såvida de inte är medvetna – *inte* kan vara ansvariga, bör vi fundera över vilka uppgifter som vi ska överlämna till maskiner, hur effektivt och kostnadsbesparande det än kan vara. Det kan till exempel, som Sharkey (2017) hävdar, vara rimligt att begränsa maskiners roll i situationer där maskinen kan fatta beslut som har konsekvenser för människors liv och välbefinnande: stridsrobotar eller robotpoliser kan orsaka död eller allvarlig skada, robotlärare kan behöva fatta beslut om att straffa eller begränsa barn, robotar inom äldreården kan vara tvungna att fatta beslut om att dela information om de äldre eller att hindra dem från att göra något farligt. Sådana beslut innebär moraliska bedömningar av sociala situationer och bör enligt Sharkey (2017: 214) göras av människor, som både har förmåga att fatta ansvarsfulla beslut (framåtblickande ansvar) och kan ställas till svars för dåliga beslut (bakåtblickande ansvar).

Även om försök görs att, utifrån vad allt komplexare robotar och algoritmer kan göra, argumentera för att maskiner kan tillskrivas ansvar, kan vi konstatera att det inte är helt enkelt att göra det på ett övertygande sätt. Även om maskinerna betar sig som mänskliga aktörer och kan få oss att tro att de verkligen är autonoma agenter, är det svårt att komma ifrån att de är skapade av människor och att allt som de gör, gör de därför att en människa, eller flera människor tillsammans, har bestämt att de ska göra just det de gör. Medvetande verkar vara

förutsättningen för att kunna ta ansvar på ett meningsfullt sätt. Ändå är frågan om maskiners ansvar högst aktuell. Den snabba tekniska utvecklingen av artificiell intelligens och självlärande maskiner väcker onekligen många frågor om vem som är ansvarig för beslut som vi konfronteras med dagligen. Vi ska nu, avslutningsvis, närma oss frågan om maskinen kan utgöra ett hinder för ansvar genom att återkomma till debatten om en ansvarslucka, som nämndes i förbigående tidigare.

Diskussionen om en ansvarslucka tar avstamp i utvecklingen av autonoma maskiner. Det handlar här alltså om autonomi i den tekniska betydelsen att utföra en förprogrammerad uppgift utan inblandning från en människa. Självlärande system gör också vad de är programmerade att göra, men exakt hur de gör det och exakt vart det leder kan inte designern veta. För att återvända till spelet AlphaGo, så är det utformat för att formulera sina egna instruktioner. Ingenjörerna som utformade och byggde det har liten eller ingen idé om vad det kommer att göra när det körs. Självlärande system är avsiktligt utformade för att göra saker som deras programmerare inte kan förutse eller helt kontrollera (Gunkel 2017: 6). Ansvarsluckan uppkommer alltså därför att ingen har tillräcklig kontroll över vad maskinen gör. Matthias (2004: 177) menar att autonoma, självlärande maskiner skapar en situation där tillverkaren eller operatören i *princip* inte är kapabel att förutse den framtida maskinens beteende och därför inte kan hållas moraliskt ansvarig för den.

Ett exempel som illustrerar svårigheten med ansvar är den mobila leksaksroboten Aibo. Den har förmåga att lära sig, till exempel en ny gångstil, när den inbyggda algoritmen för att gå visar sig inte vara optimal för tjocka persiska mattor. Efter litet experimenterande upptäcker den att batteritiden kan förlängas om den galopperar, vilket minskar friktionen till underlaget. Antag att roboten när den galopperar omkring råkar kollidera med ett litet barn, som ramlar omkull och skadas. Vem är ansvarig? Tillverkaren? Varför, i så fall? Barnets föräldrar för att ha placerat leksaksroboten i lägenheten? Eller är detta en oförutsebar utveckling, som skedde på grund av robotens adaptiva förmågor, så att ingen kan sägas vara ansvarig (Matthias 2004: 177)? Om detta är en ansvarslucka, är det något som behöver hanteras i lagstiftning och moralisk praktik (Matthias 2004: 183), men det finns invändningar mot resonemanget.

Köhler och kolleger (2018) menar att det inte finns någon ansvarslucka, och lyfter fram två skäl. För det första pekar de på att det visserligen inte är någon som kontrollerar autonoma maskiner medan de opererar. Däremot är kontroll inblandad när maskinen försätts i olika omständigheter i vilka man vet att den kommer att agera autonomt (2018: 57–59). Jämförelsen med dresserade djur kommer här tillbaka, men här för att illustrera att djuret *inte* kan vara ansvarigt. För det andra, om det bara handlar om att det är svårare att *veta* vem som är ansvarig eller i vilken grad, finns inget problem (2018: 55; jfr Young 2006). Att det är svårt att reda ut vem som är ansvarig för vad skapar inte en lucka.

Det implicerar bara att många personer behöver hållas ansvariga, vilket snarare handlar om oklart ansvar (Köhler m.fl. 2018: 63–65).

Men vad skulle det innebära om maskiner *kan* vara ansvariga? Även om nu existerande maskiner inte har sådana egenskaper som krävs för ansvar, kan vi inte veta om framtida maskiner skulle kunna ha det. Vad skulle en sådan maskin göra med ansvaret? Såvida inte kännande maskiner utvecklas, förefaller bakåtblickande och straff föga meningsfullt. Å andra sidan har vi precis konstaterat att medvetande är en förutsättning för ansvar, och då torde det vara rimligt att maskinen också uppfattar straffet som ett straff. Men *om* medvetna maskiner skulle bli verklighet, är det inte helt otänkbart att det är maskinerna, och inte vi människor, som avgör ansvarsfrågor. När det gäller de icke-kännande maskiner vi har omkring oss idag, och sannolikt under lång tid framöver, är det inte särskilt fruktbart att fundera över hur vi lämpligen straffar maskinen för klandervärda handlingar den fattar beslut om. Det är människor som måste bära ansvaret för vad maskinerna gör. *Vilka* människor som ska bära detta ansvar och hur ansvaret ska fördelas mellan dem är en fråga som får behandlas i ett annat sammanhang.

Maskiner som fattar beslut bidrar definitivt till att försvåra fördelningen av ansvar, men för att kunna hävda att maskinen är ansvarig, krävs att den är medveten, med allt vad det innebär av att kunna känna smärta och uppleva andra känslor. Och dit är det sannolikt en bit på väg.

Referenser

- Adam, Barbara & Groves, Chris, 2011. "Futures tended: Care and future-oriented responsibility", *Bulletin of Science, Technology & Society* 31(1), s. 17–27.
- AI Lund, 2020-01-29. Presentation på workshop. <http://ai.lu.se/events/?event=ai-lund-fika-till-fika-workshop-om-ai-i-offentlig-sektor>, <http://ai.lu.se/tv/ai-i-offentlig-sektor/>.
- Albertsen, Andreas, 2015. "Feiring's concept of forward-looking responsibility: a dead end for responsibility in health care", *Journal of Medical Ethics* 41(2), s. 161–164.
- Blackmore, Susan & Troscianko, Emily T., 2018. *Consciousness*. 3:e upplagan. Abingdon, Oxon and New York, NY: Routledge.
- Bostrom, Nick, 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Bringsjord, S., 1992. *What robots can and can't be*. Dordrecht: Kluwer Academic.
- Bucher, Taina, 2018. *If...Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.
- Burri, Thomas, 2016. "The Politics of Robot Autonomy", *European Journal of Risk Regulation* 7(2), s. 341–360.
- Carrithers, Michael (red.), 1986. *The category of the person: Anthropology, philosophy, history*. Cambridge: Cambridge University Press.

- EC Communication, 2018. "Artificial Intelligence for Europe", Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, COM (2018) 237.
- EGE, 2018. *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*. European Commission: European Group on Ethics in Science and new Technologies.
- EP, 2017. European Parliament resolution "Civil Law Rules on Robotics", P8_TA(2017)0051.
- EU-kommissionen, 2019. *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence.
- Feinberg, Joel, 1968. "Collective responsibility", *The Journal of Philosophy* 65(21), s. 674-688.
- Feinberg, Joel, 1970. "Collective responsibility", i May, Larry & Hoffman, Stacey (red.), *Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics*. Lanham, MD: Rowman & Littlefield Publishers Inc.
- Fischer, John Martin, 1999. "Recent work on moral responsibility", *Ethics* 110(1), s. 93-139.
- Fischer, John Martin, 2007. "Free will and moral responsibility", i Copp, David (red.), *The Oxford Handbook of Ethical Theory*, s. 321-354.
- Fischer, John Martin & Ravizza, Mark, 1998. *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Gilbert, Margaret, 2006. "Who is to blame? Collective moral responsibility and its implications for group members", *Midwest Studies in Philosophy* 30(1), s. 94-114.
- Gundersen, Torbjørn, 2018. "Scientists as experts: A distinct role?", *Studies in History and Philosophy of Science* 69, s. 52-59.
- Gunkel, David J., 2017. "Mind the gap: responsible robots and the problem of responsibility", *Ethics and Information Technology*, <https://doi.org/10.1007/s10676-017-9428-2>.
- Hakli, Raul & Mäkelä, Pekka, 2019, "Moral responsibility of robots and hybrid agents", *The Monist* 102, s. 259-275.
- Haselager, Willem F. G., 2005, "Robotics, philosophy and the problems of autonomy", *Pragmatics & Cognition* 13(3), s. 515-532.
- Hart, David Robert, 2018. "Saudi Arabia's robot citizen is eroding human rights", *Quartz*, 2018-02-18: <https://qz.com/1205017/saudi-arabias-robot-citizen-is-eroding-human-rights/>.
- Held, Virginia, 1970. "Can a random collection of individuals be morally responsible?", i May, Larry & Hoffman, Stacey (red.), *Collective responsibility: Five decades of debate in theoretical and applied ethics*. Lanham, MD: Rowman & Littlefield Publishers Inc.
- Hogan, Katharyn, 2017. "Is the machine question the same question as the animal question?", *Ethics of Information Technology* 19, s. 29-38.
- HRW 2015: Human Rights Watch, 2015. *Mind the Gap: The Lack of Accountability for Killer Robots*. Harvard Law School: International Human Rights Clinic.
- Jain, Lakhmi C. & Pratihar, Dilip Kumar, 2010. *Intelligent Autonomous Systems*. Berlin/Heidelberg: Springer.
- Johnson, Deborah G. & Verdicchio, Mario, 2018. "Why robots should not be treated like animals", *Ethics and Information Technology* 20(4), s. 291-301.

- Kane, Robert, 2002. "Responsibility, reactive attitudes, and free will: Reflections on Wallace's theory", *Philosophy and Phenomenological Research* 64(3), s. 693–698.
- Köhler, Sebastian, Roughley, Neil & Sauer, Hanno, 2017. "Technologically blurred responsibility? Technology, responsibility gaps and the robustness of our everyday conceptual scheme", i Ulbert, Cornelia, Finkenbusch, Peter, Sondermann, Elena & Debiel, Tobias (red.), *Moral Agency and the Politics of Responsibility*. London: Imprint Routledge.
- Lanteri, Alessandro, 2009. "Judgements of intentionality and moral worth: Experimental challenges to hindriks", *The Philosophical Quarterly* 59(237), s. 713–720.
- Laukyte, Migle, 2017. "Artificial agents among us: Should we recognize them as agents proper?", *Ethics of Information Technology* 19, s. 1–17.
- Lundquist, Lennart, 1992. *Förvaltning, stat och samhälle*. Lund: Studentlitteratur.
- Malle, Bertram F., 2016. "Integrating robot ethics and machine morality: the study and design of moral competence in robots", *Ethics of Information Technology* 18, s. 243–256.
- Matthias, Andreas, 2004. "The responsibility gap: Ascribing responsibility for the actions of learning automata", *Ethics and Information Technology* 6, s. 175–183.
- May, Larry & Hoffman, Stacey (red.), 1991. *Collective responsibility: Five decades of debate in theoretical and applied ethics*. Lanham, MD: Rowman & Littlefield Publishers Inc.
- McGary, Howard, 1986. "Morality and collective liability", i May, Larry & Hoffman, Stacey (red.), *Five Decades of Debate in Theoretical and Applied Ethics*. Lanham, MD: Rowman & Littlefield Publishers Inc.
- Miller, David, 2001. "Distributing responsibilities", *The Journal of Political Philosophy* 9(4), s. 453–471.
- MVTe: Mötesplats välfärdsteknologi och e-hälsa, konferenspresentation, 2020-01-21. <https://www.mvte.se/sv/activity/kan-missbruksvarden-forbattras-med-hjalp-av-ai/>.
- Nihlén Fahlquist, Jessica, 2009. "Moral responsibility for environmental problems—Individual or institutional?", *Journal of Agricultural and Environmental Ethics* 22(2), s. 109–124.
- Nordiska ministerrådet, 2019. *Nordiske kommuners arbeid med kunstig intelligens*. Ulf Andreasson & Truls Stende.
- O'Neill, Cathy, 2016. *Weapons of Math Destruction*. New York: Crown.
- Quora, <https://sv.quora.com/Kan-en-AI-ber%C3%A4kna-och-f%C3%B6ruts%C3%A4ga-progressiv-utveckling>.
- Rovane, Carol, 2014. "Forward-Looking Collective Responsibility: A Metaphysical Reframing of the Issue", *Midwest Studies In Philosophy* 38(1), s. 12–25.
- Russell, Stuart, 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking Press.
- Schneider, Susan, 2019. *Artificial You: AI and the Future of Your Mind*. Princeton: Princeton University Press.
- Schudson, Michael, 2006. "The trouble with experts – and why democracies need them", *Theor Soc* 35(5/6), s. 491–506.
- Sharkey, Amanda, 2017. "Can robots be responsible moral agents? And why should we care?", *Connection Science* 29(3), s. 210–216.

- Sellers, Mortimer, 2007. "An Introduction to the Value of Autonomy in Law", i Sellers, Mortimer (red.), *Autonomy in the Law*. Springer.
- SKR, 2018. "Automatiserat beslutsfattande i den kommunala förvaltningen". Stockholm: Sveriges kommuner och Landsting.
- Sneddon, Andrew, 2005. "Moral responsibility: The difference of strawson, and the difference it should make", *Ethical Theory and Moral Practice* 8(3), s. 239–264
- Sondermann, Elena, Ulbert, Cornelia & Finkenbusch, Peter, 2017. "Introduction: Moral agency and the politics of responsibility", i Ulbert, Cornelia, Finkenbusch, Peter, Sondermann, Elena & Debiel, Tobias (red.), *Moral agency and the politics of responsibility*. London: Imprint Routledge.
- Sullins, John P., 2006. "When is a robot a moral agent?", *International Review of Information Ethics* 6(12), s. 23–36.
- Svensson, Lupita, 2019. "Tekniken är den enkla biten': Om att implementera digital automatisering i handläggningen av försörjningsstöd", *Research report in social work 2019:12*. Lund: Socialhögskolan, Lunds universitet.
- Talbert, Matthew, 2008. "Blame and responsiveness to moral reasons: Are psychopaths blameworthy?", *Pacific Philosophical Quarterly* 89(4), s. 516–535.
- Techworld, <https://techworld.idg.se/2.2524/1.722934/vd-rost-fejk-ai>.
- Tegmark, Max, 2017. *Life 3.0: Being human in the age of Artificial Intelligence*. Penguin Random House.
- Thompson, Dennis F., 1987. *Political ethics and public office*. Cambridge, MA, och London: Harvard University Press.
- Trelleborg, 2015. <https://www.trelleborg.se/sv/omsorg-hjalp/hjalp-i-hemmet/for-en-smartare-valfard/>
- Turing, Alan M., 1950. "Computing machinery and intelligence", *Mind* 59, s. 433–460.
- Turner, Stephen P., 2003. *Liberal Democracy 3.0*. London, Thousand Oaks & New Delhi: Sage Publications.
- Voister, <https://www.voister.se/artikel/2019/04/en-ai-vi-forstar/>.
- Wahlberg, Lena, 2019. "Skäligt straffansvar för misstag inom medicinsk vård och behandling", *Nordisk Socialrättslig Tidskrift*, särtryck ur NST, s. 19–20.
- Young, Iris Marion, 2006. "Responsibility and global justice: A social connection model", *Social Philosophy and Policy* 23(1), s. 102–130.

Beslutsfattande ur tre perspektiv

Annika Wallin

Abstract

In this paper I present two normative standpoints that have been used to define and evaluate human decision making: 'economic man' and 'ecological rationality'. I argue that both norms are unsuitable to understand decision making in a natural environment. I do so by trying to apply them to a real-life decision situation; that of buying pasta in a Swedish supermarket. Neither account is sufficiently flexible to handle how decisions unfold in this setting. Consumers' behaviours are flexible and consist of an interplay between the shoppers' preferences, the structure of the shelf and the visually salient attributes of various packages.

I den här artikeln tänker jag beskriva hur vår förståelse av beslutsfattande – kanske i onödan – påverkas av vad vi anser vara det rätta sättet att fatta beslut på. Jag kommer att introducera de två dominerande normativa perspektiven (på beslutsfattande): det som representeras av den så kallade "heuristics and biases"-traditionen (Kahneman, Slovic & Tversky 1982), och konkurrenten "ekologisk rationalitet" (Gigerenzer, Todd & ABC Group 1999). Jag kontrasterar sedan dessa perspektiv med ett fall av verkligt – och vardagligt – beslutsfattande: att köpa pasta i en matvaruaffär. Min slutsats blir att inget av de normativa perspektiven är särskilt väl lämpat för att vare sig förstå eller förklara vad som händer i matvaruaffären.

Nationalencyklopedin säger följande om rationalitet: "det som bygger på förnuftet. Att en handling (eller åsikt) är rationell betyder att man medvetet och metodiskt väljer de bästa medlen för att uppnå sina mål (jämför irrationalitet). I anslutning till beslutsteorin utgår en riktning inom beteendevetenskaplig metodologi från att allt mänskligt beteende kan förklaras som styrt av rationella val (engelska rational choice theory)."

När beslutsfattande studeras gör man oftast detta i förhållande till vad man anser är normativt korrekt. En uppfattning om det normativt korrekta behövs

Annika Wallin är verksam vid Kognitionsvetenskap, Lunds universitet.
E-post: annika.wallin@lucs.lu.se

eftersom det är betydligt lättare att ta reda på hur beslut fattas om de på något sätt avviker från det vi förväntar oss. På samma sätt tar vi reda på hur vår syn fungerar med hjälp av visuella illusioner, och hur vårt minne fungerar genom att studera glömska (Kahneman, Slovic & Tversky 1982). Ett felfritt beteende kan uppstå på många olika sätt, men en systematisk avvikelse från det förväntade är en ledtråd till vilken mekanism som ligger bakom det. I det som följer tänker jag jämföra tre olika beskrivningar av mänskligt beslutsfattande. Två är redan väl etablerade: Dessa är det psykologiska forskningsprogrammet som fokuserar på "heuristics and biases" (Tversky & Kahneman 1974) och det som istället diskuterar s.k. ekologisk rationalitet eller "simple heuristics that make us smart" (Gigerenzer, Todd & ABC 1999). Den tredje beskrivningen utgår från en verklig situation. Hur ser ett faktiskt beslut i en vanlig svensk matvaruaffär ut? Jag kommer att undersöka i vilken mån de två etablerade beskrivningarna av mänskligt beslutsfattandet går att applicera i en sådan situation. Målet är att illustrera vilka krav som ställs på en beslutsmodell när den ska användas i verkliga livet.

I det inledande citatet är det som är normativt riktigt samma sak som det rationella valet, det vill säga det medvetna och metodiska väljandet av rätt medel för att uppnå ett visst mål. Citatet pekar på en beteendevetenskaplig metodologi som antar att denna norm är rimlig och befogad. Exempelvis föreslog Ward Edwards (1954) att vi använder idealet "economic man" när mänskligt beslutsfattande studeras. Måhända är antagandena *bakom* idealetorealistiska (fullständig information, ordnade preferenser och beräkningar som inte tar tid eller energi), men, påpekar Edwards, detsamma gäller för många psykologiska teorier: "the most useful thing to do with a theory is not to criticize its assumptions but rather to test its theorems. If the theorems fit the data, then the theory has at least heuristic value" (sid. 382). Med andra ord ger ekonomisk teoribildning upphov till "a new and rich field for psychologists, in which a theoretical structure has already been elaborately worked out and in which many experiments need to be performed" (sid. 411). Med det rationella valet som utgångspunkt och norm har vi en metod för att utforska det mänskliga beslutsfattandet. Det föreslagna förhållningssättet anammades av Daniel Kahneman och Amos Tversky (1974), och gav upphov till det mycket inflytelserika "heuristics and biases"-forskningsprogrammet. Här valde man ut en version av "economic man" och testade den systematiskt mot mänskliga beslutsfattare.

Perspektiv ett: "Heuristics and biases"

Att anta en version av "economic man" ger oss ett första normativt perspektiv att utgå ifrån, ett som innehåller antaganden om vilka principer ett gott eller rationellt beslutsfattande ska följa. Detta är det första perspektiv på

beslutsfattande jag tänker beskriva. Principerna kan, exempelvis, vara att ett beslut (eller ett beslutssystem) följer kraven på invarians, dominans och transitivitet (Tversky & Kahneman 1986). Kravet på invarians innebär att olika beskrivningar av ett och samma alternativ inte bör påverka vilket beslut som fattas, så länge beskrivningarna innehåller samma information. Det faktum att ett utfall beskrivs som $\frac{1}{2}$ chans till krona får alltså inte ge upphov till andra val än om det beskrivs som $\frac{1}{2}$ chans till klave. Det är sedan länge etablerat att hur utfall beskrivs påverkar beslutsfattandet och detta anses ibland utgöra ett brott mot invariansprincipen. Fenomenet går oftast under termen "framing". Så här beskrev Tversky och Kahneman framing när de först presenterade det (1981): "Veridical perception requires that the perceived relative height of two neighboring mountains, say, should not reverse with changes of vantage point. Similarly, rational choice requires that the preference between options should not reverse with changes of frame. Because of imperfections of human perception and decision, however, changes of perspective often reverse the relative apparent size of objects and the relative desirability of options".

Beskrivningen av "framing" visar hur beroende angreppssättet är av vad man anser vara normativt korrekt. Det är avvikelserna – att perspektivskiftet förändrar vilket alternativ man föredrar – som definierar fenomenet. När väl fenomenet är etablerat kan man studera det i naturliga kontexter i den bemärkelse att man kan notera att formuleringar väljs med omsorg på ett sätt som verkar bygga på framing-effekten (en maträtt beskrivs kanske som 90% fettfri snarare än som 10% fet). För att etablera att det verkligen handlar om en framing-effekt måste man dock ställa olika formuleringar *mot* varandra och konstatera att de ger upphov till olika beteenden. Det är ganska svårt, om än inte omöjligt, att visa på framing inom en enda individ. Ofta kontrasterar man istället gruppens beteenden. Den bakomliggande logiken är då att två tillräckligt stora och slumpmässigt valda grupper rent teoretiskt borde välja likadant på en aggregerad nivå. Grupperna får välja mellan två alternativ som så när som på framingen är identiska. När grupperna (på en aggregerad nivå) väljer olika säger man att man har observerat en framing-effekt. Men även om framing är svårt att observera utanför den experimentella situationen kan det knytas till en teoribildning som är relativt oberoende av rationalitetsantagandet, exempelvis den s.k. "prospect theory" som beskriver hur beslutsfattande påverkas beroende på om det handlar om vinster eller förluster.

Figur 1. Ett beslut inom 'heuristics and biases'-programmet.

The Asian Disease Problem (Kahneman & Tversky 1981) originalstimuli	
"Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people".	
½ av deltagarna väljer mellan nedanstående alternativ	½ deltagarna väljer mellan nedanstående alternativ
"If Program A is adopted, 200 people will be saved. If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved"	"If Program A is adopted, 400 people will die. If Program B is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600"
72% väljer A	22% väljer A
Ursprunglig tolkning: Beroende på vilken frame ("saved"/"die") program A och B har kommer försökspersonerna att vara mer eller mindre benägna att chansa och välja det osäkra alternativet. "Frame" är här ifall vi ser utfallet som en vinst ("save") eller en förlust ("die").	

Under sjuttio- och åttioalet studerades framing, och andra systematiska avvikelser från rationalitetsantaganden (de principer som beskrevs ovan). Resultaten gav upphov till en lång och hetsig debatt som främst pågick under åttio- och nittioalet. Debatten, som ibland getts namnet "the rationality wars", handlar om ifall vi på basis av forskningsresultaten bör dra slutsatsen att människan är en dålig, eller rent av irrationell, beslutsfattare.

Perspektiv två: Ekologisk rationalitet

I samband med debatten utkristalliserade sig en alternativ syn på vad "rätt sorts" norm borde vara när man bedömer beslutsfattande. Denna alternativa syn på det normativt korrekta formulerades på ett sätt som blev tydligt (och begripligt?) för de kognitiva vetenskaperna i och med att begreppet "ekologisk rationalitet" myntades. Det här är det andra perspektivet på beslutsfattande som jag vill beskriva. Med denna syn på vad rationellt beslutsfattande är flyttas fokus. Företrädarna är inte intresserade av att undersöka i vilken utsträckning det mänskliga beslutsfattandet följer de principer på vilka (en version av) rationalitetsbegreppet vilar. Istället koncentrerar man sig på hur *framgångsrikt* beslutsfattandet är och hur *mycket* information och kognitiva resurser det kräver. Dessa krav motiveras ofta genom att de är "evolutionärt" eller "biologiskt" rimliga, men kopplingen till naturligt urval uttrycks i relativt vaga termer (Giglerenzer, Todd & ABC 1999). Tanken är att våra beslutsprocesser har slipats över långa tidsrymder och att detta gett upphov till enkla och snabba genvägar till vettiga beslut. Den kända igenkänningsheuristiken ("the recognition heuristic", se nedan) presenteras exempelvis ofta genom att man hänvisar till att unga brunråttor bara äter mat de har ätit förut eller har känt lukten av i vuxna råttors andedräkt.

En välkänd illustration av igenkänningsheuristiken är en studie av amerikanska och tyska universitetsstudenters förmåga att avgöra vilken stad, i ett par av (slumpmässigt ihopsatta) städer av en viss storleksordning, som är den största. När studenterna får bedöma ett flertal ihop-parade amerikanska eller tyska städer visar det sig att de tyska studenterna generellt sett var bättre än de amerikanska studenterna på att bedöma den relativa storleken på *amerikanska* städer. Trots att amerikaner är välkända för att vara mindre hemma på geografi medan tyskar ofta antas ha goda kunskaper på området är detta inte den mest troliga förklaringen av resultatet. Om inte annat har man hittat det omvända resultatet för tyska städer, i den meningen att de amerikanska studenterna gör bättre bedömningar än tyska studenter (Goldstein & Gigerenzer 2002, se Figur 2). Förklaringen är snarare att de tyska studenterna hade hört talas om färre amerikanska städer än de amerikanska studenterna hade. De kunde alltså – i de fall där de bara hade hört talas om en i ett par av städer – anta att den stad de hört talas om också var den större staden. De amerikanska studenterna hade hört talas om fler städer och kunde därmed inte använda sig av igenkänning för att bedöma storlek. Studiens mest välciterade exempel rör stadsparet San Diego och San Antonio. För dessa två städer gjorde de tyska studenterna 100% korrekta bedömningar av vilken stad som var störst. De amerikanska studenterna å andra sidan låg strax över sextio procent (Gigerenzer & Goldstein 1996). Om vi går till förstasidan på Die Zeit detta år ser vi att San Diego nämnts 47 gånger på första sidan medan San Antonio enbart nämnts 4 gånger, ett mönster som motsvarar deras (vid den tidpunkten) faktiska storlek.

Detta antyder att det i en miljö med slumpmässigt ihopsatta städer av en viss storleksordning kan vara en fördel att känna igen färre städer, eftersom igenkänning är kopplat till storlek. Slutsatsen som Gigerenzer och medförfattare drog var att det kan vara *fördelaktigt* att ha mindre kunskap. Detta rimmar rätt illa med nationalencyklopedins medvetna och metodiska väljande eller "economic man"s fullständiga tillgång till information. Forskargruppen arbetade vidare med resultatet och visade att beslutsstrategier som ignorerar information ibland klarar sig lika bra eller till och med bättre än mer sofistikerade strategier som utnyttjar hela den tillgängliga informationsmängden. Dessa alternativa beslutsstrategier är inte bara framgångsrika utan även billiga eftersom de använder sig av minimalt med information och inte kräver några avancerade beräkningar. Å andra sidan måste vi notera att strategiernas framgång beror på i vilken miljö de verkar. Igenkänningsheuristiken fungerar inte särskilt väl om den används på en välkänd men väldigt liten stad, som Tjernoby (Oppenheimer 2003). Men, resonerar heuristikens förespråkare, används den med omsorg – eller till och med utan större omsorg – kommer den att ge upphov till fler fördelaktiga än ofördelaktiga beslut i det långa loppet. Mängden tid, information och kognitiv kraft en sådan beslutsstrategi sparar gör det värt det.

Figur 2. Ett beslut i den ekologiska rationalitetens anda.

Recognition heuristic (Goldstein & Gigerenzer 2002)	
"Which city has the largest population?"	
Hamburg	Köln
För studenter från University of Chicago gäller att de oftast känner igen den större staden i 300 slumpmässigt dragna par av de 25 största städerna i Tyskland (recognition validity \approx 80%). När de är tvungna att välja, brukar deras svar vara i linje med heuristiken (medelantal svar som stämmer med denna förklaring \approx 90%)	
Igenkänningsheuristiken: "If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value" (s. 41).	

I och med den ekologiska rationalitetens inträde på scenen har vi nu tillgång till två normer, eller perspektiv, enligt vilka vi kan bedöma det mänskliga beslutsfattandet. Ett garanterar ett bästa möjliga beslut i varje situation. Det andra fokuserar på generellt sett framgångsrika beslut som är relativt billiga för en kognitiv organism. Ett fokuserar på struktur: att vi följer de principer på vilka (en version av) rationalitetsbegreppet vilar. Det andra på utfall. Båda har sina för- och nackdelar.

Fungerar perspektiven?

Den främsta kritik som riktats mot det rationalitetsbegrepp som "heuristics and biases" baseras på är att det ställer omänskliga krav på den mänskliga beslutsfattaren. Detta har på senare år ofta bemötts genom att man inom teoribildningen ser beslutsfattandet som något som består av två eller flera processer där minst en är snabb och intuitiv och minst en följer det anammade rationalitetsbegreppets principer i görligaste mån. Den främsta kritik som riktats mot den ekologiska rationaliteten är istället att den inte kan garantera framgång. Bara för att några fingerfärdiga forskare kan konstruera en situation där okunskap är en fördel, betyder det inte att vi generellt sett borde låta bli att skaffa oss en så fullständig bild vi kan av det beslut som ska fattas. Personligen tycker jag dock att det största problemet med den ekologiska rationaliteten är att den behöver kunna definiera vad ett framgångsrikt beslut är. I vissa fall, som de tyska och amerikanska städerna, vet vi vad det rätta svaret är, men det finns en stor mängd situationer där vi inte har sådan kunskap. Det går att undersöka hur en individ väljer exempelvis arbetsplats, och vi kan jämföra strategierna med de normer som finns inom exempelvis "economic man". Men hur ska man kunna avgöra ifall beslutet var framgångsrikt? Genom att fokusera på framgång så behöver den ekologiska rationaliteten också kunna definiera vad detta är, och definitionen måste vara något annat än att säga att det individen väljer är det som kommer att ge den störst framgång (eftersom ett sådant påstående inte går att motbevisa).

Den ekologiska rationaliteten kan möjligtvis stödja sitt fokus på framgång genom att hänvisa till evolutionära urvalsprocesser, men det blir ganska långsökt för den typ av beslut som studeras i beslutsforskningen idag. Få av de beslut som studeras (exempelvis valet mellan städer) har konsekvenser av en sådan natur att det är rimligt att tänka sig att den evolutionära pressen på dem är särskilt stor. Beslutsfattande är dessutom en så komplex aktivitet att det bör finnas många olika (tillräckligt) framgångsrika sätt att hantera det på. Detta innebär att de som använder sig av ekologisk rationalitet oftast blir tvungna att hålla sig till uppgifter som snarare handlar om omdömen än beslut: uppgifter som frågar försökspersonen vilken stad som är störst men inte vilken stad den väljer att bosätta sig i.

Debatten går vidare, och som så ofta i akademiska sammanhang har båda ståndpunkterna modifierats för att bättre stå emot den kritik man fått utstå utan att man egentligen når någon enighet om hur beslut *bör* studeras. Jag tänker försöka mig på något lite annorlunda här. Jag tänkte titta på ett faktiskt, vardagligt beslut och se hur väl de olika idéerna om vårt beslutsfattande passar in på den struktur vi faktiskt ser i det här sammanhanget. Fungerar något av våra två alternativa sätt att definiera det normativt riktiga beslutsfattandet, och i vilken utsträckning kan de hjälpa oss att förstå vad som händer?

Om vi tittar på de två studier som beskrivs i figurerna ovan ser vi att de har vissa likheter. Den mest slående är att de behandlar beslutsfattande med två alternativ. Även om val mellan fler än ett alternativ ibland diskuteras i besluts-litteraturen (exempelvis 'elimination by aspects' av Amos Tversky, 1972), är de i klar minoritet. En annan är att båda studierna starkt begränsar den mängd information som försökspersonen behöver ta ställning till. Ytterligare en likhet är att studierna är designade så att de producerar beteenden som lätt utvärderas av den norm som förespråkas av forskningstraditionen. I fallet med ADP (Asian Disease Problem) anses tillförlitliga skillnader mellan de två gruppernas svar bryta mot invariansprincipen. I fallet med igenkänningsheuristiken kan man med hjälp av ett uppdaterat uppslagsverk enkelt avgöra vilka relativa storleksbedömningar som är korrekta och vilka som är felaktiga.

Ett tredje perspektiv: Beslutsfattande i det fria

Vad händer egentligen när beslutssituationen inte konstruerats av en forskare med en viss syn på vad rationellt beslutsfattande är, utan finns i ett naturligt förekommande sammanhang? Tillsammans med kollegor (främst Kerstin Gidlöf) har jag ägnat en hel del tid åt att försöka förstå hur man går tillväga när man exempelvis köper ett paket pasta i en vanlig livsmedelsbutik. Vi har naturligtvis våra egna åsikter om beslutsfattande och rationalitet, men vi har inte medvetet manipulerat den information som försökspersonerna har att ta ställning till när de fattar sitt beslut. Jag tänker nu ta ett av dessa beslut ur sitt sammanhang och se hur väl det passar med de två perspektiven på beslutsfattande som beskrevs ovan.

Figur 3. Ett vanligt beslut, att köpa pasta i en svensk livsmedelsaffär (Gidlöf, Anikin, Lingonblad & Wallin 2018).

Att köpa pasta i en vanlig svensk affär (2017)	
Antal olika sorters pastaförpackningar (exklusive spaghetti och lasagne)	82
Totalt antal paket på hyllan (exklusive spaghetti och lasagne)	213
Antal egenskaper listade på paketet (pris, vikt, innehåll, ursprungsland etc.)	112

Vad som omedelbart blir uppenbart är hur komplex situationen är. I våra två tidigare exempel valde man mellan två alternativ. I pastafallet är siffran mer än fyrtio gånger större. Likaså finns det betydligt mer information tillgänglig om alternativen och den information som finns tillgänglig är också mer komplex. Vi kan redan vid första anblicken se att det är orimligt att förvänta sig att någon konsument skulle gå igenom all tillgänglig information på ett medvetet och metodiskt sätt för att sedan välja ut den bästa pastan. Detta skulle vara alltför tidskrävande.

Figur 4. Hur konsumenten betar sig (i snitt) när de köper pasta i en svensk livsmedelsaffär (Gidlöf, Anikin, Lingonblad & Wallin 2018).

Att köpa pasta i en vanlig svensk affär (2017)	
Antal pastaköp per år	30
Antal besök i affären	Varannan dag
Antal (olika) pastapaketer de fäster blicken på	32
Tid med blicken fäst på ett paket	30 sekunder

I studien som beskrivs ovan lät vi konsumenterna bära portabla ögonrörelsemätare. De visar var konsumenten fäster blicken, och därmed vilken information de har tillgång till visuellt (rimligtvis har de också förkunskaper om pasta). Precis som man kunde förvänta sig ignorerar konsumenterna en stor del av den information som finns tillgänglig. Exempelvis fäster de inte ens blicken på hälften av de tillgängliga alternativen. Beslutet tar inte lång tid, och i snitt får varje pastapaketer som konsumenten fäster sin blick på inte ens en sekunds (visuell) uppmärksamhet. Men snittvärdet är missvisande eftersom ett fåtal paket får den största mängden visuell uppmärksamhet. Det är alltså mycket troligt att konsumenterna förenklar beslutssituationen när de väljer pastapaketer på ett vis som passar med de modeller som föreslagits av "heuristics and biases" och ekologisk rationalitet. Men vilken sorts metod de använder är i princip omöjligt att avgöra. Ett förslag jag fått i olika sammanhang är att konsumenterna redan vet vad de vill ha. I butiken letar de bara reda på pastapaketer och vad vi ser är hur ett redan taget beslut implementeras. Teoretiskt sett kan alltså valet av

pasta vara fullständigt genomtänkt. Detta kan inte vara hela sanningen. Rent praktiskt ändrar sig priser och utbud hela tiden, och ett tidigare optimalt alternativ kan snabbt sluta vara det. Konsumenter har också en tendens att i rätt stor utsträckning köpa livsmedel de inte egentligen hade planerat att handla (Inman, Winer & Ferraro 2009) och de brukar dessutom veta rätt lite om de produkter de köper (Dickson & Sawyer 1990).

Problemet med den typ av data som vår studie av pastaköp ger oss är att det inte finns något normativt korrekt beslut att jämföra med. Det är orimligt att begära en fullständig genomgång av all tillgänglig information ihop med perfekt formulerade preferenser. Detta gör att "heuristics and biases"-traditionen saknar medel för att identifiera beslutsstrategier – det finns inget (rimligt) att jämföra med. Tyvärr verkar det lika svårt att avgöra vad ett framgångsrikt beslut är i matvaruaffären. Alla tillgängliga alternativ är sådana att vi kan förvänta oss framgång ur ett evolutionärt perspektiv: ingen pasta är så dyr att den ruinerar ens en fattig student, och även om snabba kolhydrater ofta beskrivs som rent gift, är det få konsumenter som ramlar ner döda för att de har valt fel paket. Även om det kan finnas starka åsikter om snabbmakaroner så lär deras inflytande på möjligheten att skaffa avkomma inte vara så stark att den går att upptäcka. Eftersom pastainköpet handlar om ett äkta beslut snarare än en rangordning kan vi inte heller använda oss av ett uppslagsverk för att avgöra vilka val som borde göras. Det finns säkerligen en pasta som är mest näringsriktig på hyllan, och kanske också en som är godast, eller ger mest prestige. Men vilket av dessa val är det rätta? Det blir svårt att avgöra vilka snabba och enkla beslutsregler som är framgångsrika i miljön.

Går det att identifiera beslutsstrategier för pastaköpet?

Hur ska man då kunna avgöra hur konsumenterna egentligen går till väga när de köper pasta? Det verkar inte som att vi kan utgå ifrån någon norm, utan istället får vi titta på (delar av) processen och vilka val som görs.

Det finns strategier inom bland annat heuristics and biases traditionen som passar bättre med den typ av uppgift våra konsumenter står inför. Det kanske bästa alternativet är en strategi som föreslagits för sekventiellt beslutsfattande. Beslutsstrategin Elimination By Aspects (EBA), skär bort alternativ egenskap för egenskap (Tversky 1972). Hypotetiskt kanske en konsument först väljer ut de billigaste alternativen och sorterar bort allt som är dyrare för att sedan eliminera alla som är oekologiska eller som tillhör fel märke, tills bara ett alternativ kvarstår.

Våra försökspersoners ögonrörelser passar dock inte med den här beslutsmodellen. En stor andel alternativ fick ingen uppmärksamhet alls och

ögonrörelserna indikerade inte heller ett systematiskt eliminerande av alternativ egenskap för egenskap. Det är inte så förvånande eftersom pastahyllan är en miljö där det är svårt att göra jämförelser. Den enda egenskap som är lättillgänglig för alla pastapaketer är priset (och kilopriset) som står angivet på hyllkanten. Pastan är emellertid inte sorterad efter pris, och det är ansträngande att ens hitta pastapaketen med lägst kilopris. Bland annat kommer en sådan uppgift att ställa stora krav på arbetsminnet. Det går att hitta paketet, kanske främst eftersom stora förpackningar med lågt kilopris ofta står längre ner på hyllan, och konsumenterna är relativt medvetna om de principer man brukar använda sig av när man organiserar butikshyllor (Valenzuela, Raghuram & Mitakakis 2013). Men även en systematisk jämförelse av produkterna i ett visst segment av hyllan är kognitivt och tidsmässigt krävande och vi ser inte mycket som talar för att detta är vad som händer.

Låt oss uppehålla oss ett litet ögonblick vid vad vi just såg: Beroende på hur informationen är strukturerad - hur beslutsalternativen presenteras - blir olika typer av beslutsstrategier mer eller mindre lätta att implementera. När det finns många alternativ och en relativt rörig presentation av information verkar man behöva gå mer sekventiellt till väga. En beslutsstrategi som skulle kunna fungera relativt väl i ett sådant sammanhang är Herbert Simons välkända förslag på "satisficing" (Simon 1956). Här förväntas beslutsfattaren leta efter, eller undersöka, alternativ bara fram tills dess att hon stöter på ett som motsvarar hennes krav eller behov. Ibland lägger man till att beslutsfattaren först bekantar sig med ett antal alternativ innan hon bestämmer sig för vilka kraven är. Detta ger beslutsfattaren viss information om ifall de krav hon ställt upp är rimliga eller om de är för högt eller för lågt ställda. Att konsumenterna i vår studie ägnar ett fåtal alternativ stor visuell uppmärksamhet skulle kunna ses som kompatibelt med en liknande strategi. Att strategin är tydligt sekventiell passar också bra med hur miljön är strukturerad. I en matvaruaffär är man nästan tvungen att huvudsakligen värdera alternativen ett efter ett, eftersom hyllan innehåller så många alternativ som i sin tur presenterar information på vitt skilda sätt. Det är dock svårt att avgöra ifall det verkligen är strategin satisficing som konsumenterna använder sig av.

En annan typ av beslutsmetod som på sätt och vis fungerar väl med beteendet framför hyllan är en som ser beslutsfattande som en form av "sampling", där vi överväger alternativ och känner efter hur mycket vi tycker om dem. Genom att flytta uppmärksamheten mellan alternativen skaffar vi mer information om vad vi egentligen tycker. Så fort det överväger för det ena alternativet så fattas ett beslut (Shimojo m.fl. 2003; Glaholt & Reingold 2009). Det är tilltalande att den här modellen av mänskligt beslutsfattande tar den aktiva aspekten av informationsinhämtning på allvar. Modellen stöds också av studier som antyder att om man manipulerar den visuella uppmärksamheten så att någon "råkar" titta mer på ett alternativ så väljs detta oftare än när uppmärksamheten

inte manipuleras på detta sätt (Pärnamets, Johansson, Hall, Balkenius, Spivey & Richardson 2015). Det som är mindre tilltalande är att de modeller man använder sig av i dagsläget är komplicerade och att de är svåra att applicera på fler än två eller tre alternativ. Så som de ser ut idag är det svårt att se hur de skulle kunna användas när man köper pasta.

Hur gör då konsumenten som står framför pastahyllan? Förmodligen är det så att den typ av strategier som beskrivs ovan inte kan fånga detta beteende. Det handlar om ett mer komplext samspel mellan omgivning och individ.

Beslut: ett komplext samspel mellan omgivning och individ

När vi undersökte konsumenters beteende framför pastahyllan använde vi oss av ögonrörelsemätare eftersom vi var intresserade av de visuella aspekterna av vårt beslutsfattande. Från början handlade intresset mest om *vilken* information som deltagarna tittade på och *när* de gjorde det, men efter hand blev spelet mellan det visuellt tillgängliga och beslutsfattandet mer och mer intressant. Det är ju inte bara så att vi ”plockar på oss” den information vi behöver för att fatta beslut. Det som finns på och runtomkring butikshyllan förändrar och påverkar vår visuella uppmärksamhet. En aspekt av detta är det som redan nämnts, att hur varor placeras är en ledtråd till vilka egenskaper de har. En annan aspekt är att marknadsförare, förpackningsdesigners och andra arbetar aktivt med att fånga vår uppmärksamhet. ”Syns du inte finns du inte” gäller inte bara fotgängare i trafiken. Framför pastahyllan finns det information som vi aktivt letar efter och information som vi får vare sig vi vill det eller inte. Vi styr vår egen uppmärksamhet men den blir också styrd. Det här betyder till exempel att vi i affären tittar både på de varor som genom sitt utseende fångar vår uppmärksamhet genom att utnyttja hur vårt synsystem fungerar (genom att till exempel sticka ut från omgivningen vad det gäller färg eller form) och på de varor som motsvarar våra önskemål. När det som sticker ut visuellt också motsvarar våra önskemål (exempelvis genom pris, storlek, märke) så ökar det drastiskt vår benägenhet att fästa blicken vid varan. Det verkar som att de som går och handlar utnyttjar de automatiska delarna av sitt synsystem för att snabbt sortera bort det som inte är relevant och därmed lättare hitta det som passar deras önskemål. Detta kan man till exempel göra genom att (beroende på önskemål) helt bortse från den blåa delen av pastahyllan, eller snabbt rikta blicken och uppmärksamheten just dit (Gidlöf m.fl. 2017).

Slutord

De två perspektiv på beslutsfattande som jag presenterat här är båda beroende av en normativ ståndpunkt. Antingen definieras det goda beslutet av vilka

principer det följer ("heuristics and biases") eller så definieras det av hur framgångsrikt det är (ekologisk rationalitet). Behovet av en norm är dock inte det enda som dessa perspektiv har gemensamt. De beskriver båda oftast beslutsfattande som ett val mellan ett väldigt begränsat antal alternativ. De beslutsproblem som beskrivs och förklaras av perspektiven är också statistiska: en viss information finns tillgänglig och beslutsfattaren har att ta ställning till den eller ignorera den. Mot detta har jag kontrasterat det vardagliga beslutsfattandet. Vi kan konstatera att ett vardagligt beslut som det du fattar när du handlar i en affär inte alltid är lätt att beskriva. Ett problem är att många av de beslutsmodeller som utvecklats inom de kognitiva vetenskaperna kräver en normativ definition av det goda beslutsfattandet att utgå ifrån, och det finns inga tillräckligt precisa normer för pastahyllan för att kunna använda dem här. Ett annat problem är att flertalet modeller utgår ifrån val mellan två alternativ med all information lättillgänglig. Så ser det inte ut i affären. Vidare har vi en situation som är mer komplex än man brukar erkänna i och med att vi har ett aktivt samspel mellan det som finns på hyllan och de kompetenser vi tar med oss in i beslutssituationen. Vi människor verkar kunna utnyttja det här samspelet på ett intelligent sätt, men inte på ett sätt som passar vidare bra med hur beslutsfattande beskrivs i litteraturen.

Referenser

- Dickson, P. & Sawyer, A., 1990. "The price knowledge and search of supermarket shoppers", *Journal of Marketing*, 54, s. 42-53.
- Edwards, W., 1954. "The theory of decision making", *Psychological Bulletin*, 51(4), s. 380-417.
- Gidlöf, K., Anikin, A., Lingonblad, M. & Wallin, A., 2017. "Looking is buying. How visual attention and choice are affected by consumer preferences and properties of the supermarket shelf", *Appetite*, 116, s. 29-38.
- Gigerenzer, G. & Goldstein, D., 1996. "Reasoning the fast and frugal way: Models of bounded rationality", *Psychological Review*, 103(4), s. 650-669.
- Gigerenzer, G., & Todd, P. & the ABC Research Group, 1999. *Simple Heuristics that make us smart*. New York: Oxford University Press.
- Glaholt, M. & Reingold, E., 2009. The time course of gaze bias in visual decision tasks. *Visual Cognition*, 17(8), s. 1228-1243.
- Goldstein, D. & Gigerenzer, G., 2002. "Models of ecological rationality: the recognition heuristic", *Psychological Review*, 109(1), s. 75-90.
- Inman, J. J., Russell, S. W. & Ferraro, R., 2009. "The interplay among category characteristics, customer characteristics, and customer activities on in-store decision making", *Journal of Marketing*, 73, s. 19-29.
- Kahneman, D., Slovic, P. & Tversky, A. (eds), 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Oppenheimer, D. M., 2003. "Not so fast! (and not so frugal!): rethinking the recognition heuristic", *Cognition*, 90, s. B1-B9.

- Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J. & Richardson, D. C., 2015. "Biasing moral decisions by exploiting the dynamics of eye gaze", *Proceedings of the National Academy of Sciences*, 112(13), s. 4170-4175.
- Shimojo, S., Simion, C., Shimojo, E. & Scheier, C., 2003. "Gaze bias both reflects and influences preference", *Nature neuroscience*, 6(12), s. 1317-1322.
- Simon, H., 1956. "Rational choice and the structure of the environment", *Psychological Review*, 63(2), s. 129-138.
- Tversky, A., 1972. "Elimination by aspects: A theory of choice", *Psychological Review*, 79, s. 281-299.
- Tversky, A. & Kahneman, D., 1974. "Judgment under uncertainty: Heuristics and biases", *Science*, 27, s. 1124-1131.
- Tversky, A. & Kahneman, D., 1986. "Rational choice and the framing of decisions", *The Journal of Business*, 59(4), s. 251-78.
- Valenzuela, A., Raghurir, P. & Mitakakis, C., 2013. "Shelf space schemas: Myth or reality?", *Journal of Business Research*, 66(7), s. 881-888.

Fri bevisprövning och evidensbaserad medicin

En jämförelse av två modeller för evidens och beslut¹

Lena Wahlberg

Abstract

This paper compares two seemingly very different models for evidence management and decision making: the model of evidence based medicine, EBM, and the model of free evaluation of evidence (as applied in the Swedish judiciary system). Whereas EBM identifies and ranks the relevant kinds of evidence that a decision-maker should consider, the model of free evaluation of evidence relies on the idea that truth is best pursued freely, and that the decision maker's evaluation of evidence should not be governed by rules. The paper compares the levels of freedom in the two models with respect to what evidence to consider, how to assess the value of different pieces of evidence, and how to estimate the combined value of all the evidence. It concludes that the two models diverge in other ways than first assumed, and that the juxtaposition of them can be beneficial to both models, by casting light on underlying but seemingly unwarranted assumptions relating to the role of intuition and life experience in the evaluation of evidence.

1. Inledning

I svensk rätt, liksom i flera andra rättsordningar, sätter principen om fri bevisprövning (nedan förkortad FBP) ramarna för vilken evidens som får läggas till

¹ Artikeln har skrivits inom ramen för forskningsprogrammet *Vetenskap och beprövad erfarenhet – VBE* (www.vbe.lu.se), som finansieras av Riksbankens Jubileumsfond M14-0138:1. Jag vill tacka Sten Anttila, Christian Dahlman, Magnus Jerneck, Peter Westberg och medlemmarna i forskningsprogrammet för värdefulla synpunkter.

grund för en dom och hur denna evidens ska värderas. Enligt förarbetena till rättegångsbalken (1942:740) innebär principen att ”någon begränsning ej uppställs i fråga om de kunskapskällor som får användas vid sanningens utledande, som ock att domaren vid värderingen av den föreliggande bevisningen icke är bunden av legala regler” (SOU 1938:44, s. 377). FBP innebär alltså att det som huvudregel inte finns några begränsningar i fråga om vilken typ av bevisning som tillåts eller hur denna ska värderas. I litteraturen framhålls ofta beslutsfattarens (domarens) intuition och erfarenhet som viktiga element i bevisvärderingen.²

FBP kontrasteras ofta mot ”legala” bevisprövningsmodeller, där bevisprövningen i högre grad styrs av rättsregler. En legal modell låg till grund för 1734 års rättegångsbalk, och kom till tydligt uttryck i dess 29 §, som stadgade att ”Tu witne äro fullt bewis, ther i the sammanstemma. Ett witne om sielfwa målet gälle ej mer, än som halft bewis”. För de flesta i Sverige skolade jurister framstår FBP nog som ett klart bättre alternativ än en legal bevisvärderingsmodell för en modern svensk rättsordning.

FBP kan emellertid kontrasteras inte bara med legala bevisvärderingsmodeller utan också med modeller för att hantera evidens inom andra discipliner. Ett exempel på en mycket inflytelserik icke-rättslig modell är principen om evidensbaserad medicin (nedan förkortad EBM) som tar sikte på medicinskt beslutsfattande. En grundtanke inom EBM är att läkarens beslutsfattande ska vila på bästa möjliga evidens. För att åstadkomma detta finns tydliga riktlinjer både för vilken evidens som får användas och hur denna ska värderas. Inom EBM höjs också ofta ett varnande finger för att tillmäta beslutsfattarens (läkarens) intuition och erfarenhet betydelse vid bevisvärderingen.³

Vid en yttlig betraktelse framstår EBM och FBP närmast som varandras motsatser: I FBP finns inte någon begränsning för vilken typ av bevis som ska tillåtas, och inte heller några på förhand givna riktlinjer för styrkan hos olika typer av bevis. I EBM finns i stället tydliga riktlinjer både för vad som ska räknas som evidens, och vilken styrka olika typer av evidens ska tillmätas. I litteraturen om FBP lyfts betydelsen av beslutsfattarens intuition och erfarenhet ofta fram; i litteraturen om EBM avfärdas dessa tvärtom som opålitliga. Även om läkarens och domarens beslutssituationer skiljer sig åt är det inte uppenbart, i vart fall inte för mig, på vilket sätt dessa skillnader motiverar användningen av två så till synes olika modeller för att hantera evidens och fatta beslut.

Syftet med denna uppsats är att bättre förstå skillnaderna mellan FBP och EBM genom att undersöka i vilka avseenden och på vilka sätt de reglerar respektive inte reglerar hanteringen av evidens, och om de skillnader som kan konstateras låter sig förklaras av olikheter mellan den rättsliga och den

² Se vidare avsnitt 3 nedan.

³ Se vidare avsnitt 3 nedan.

medicinska beslutscontexten. I vissa delar kommer jag även att relatera de två modellerna till det svenska begreppet *Vetenskap och beprövad erfarenhet* (VBE), som sedan slutet av 1800-talet fungerat som ett kvalitetskriterium för svensk hälso- och sjukvård,⁴ och som intressant nog uppvisar drag av både FBP och EBM. Även om olikheterna mellan den rättsliga och den medicinska beslutsprocessen gör det svårt att göra en rättvisande jämförelse av FBP och EBM utan att också gå djupare in i deras respektive historiska och disciplinära kontexter än det funnits utrymme för i denna uppsats, är min förhoppning att jämförelsen ska bidra till att belysa väsentliga skillnader och likheter mellan modellerna, och till att de i förlängningen kanske rentav kan lära sig något av varandra.

2. Läkaren och domaren

Läkarens och domarens beslutsfattande skiljer sig åt med avseende på såväl föremål som ändamål. Medan läkaren avgör vilka medicinska åtgärder som kan anses befogade och ska erbjudas en patient, tar domaren i stället ställning till vilka rättsliga rekvisit som kan anses uppfyllda och vilka rättsliga påföljder dessa ska föranleda. Medan, enkelt uttryckt, läkarens beslut främst syftar till att främja hälsa, är domarens beslut i högre grad inriktat på att åstadkomma rättvisa.

Vid sidan om skillnaderna mellan domarens och läkarens beslutsfattande finns emellertid också stora likheter: Både läkarens och domarens beslut handlar om konkreta fall och får som regel mycket stor betydelse för de enskilda människor som fallet gäller. I båda beslutssituationerna är kunskapsunderlaget ofullständigt – oavsett om osäkerheten handlar om hur den konkreta patienten kommer att reagera på behandlingen, eller om den handlar om huruvida den tilltalade faktiskt varit på platsen vid den tidpunkt som åklagaren hävdar. Vidare beror frågan om hur osäkerheten ska hanteras i beslutet i båda fallen på vad som står på spel i den konkreta situationen. Inom straffrätten måste åklagaren bevisa den tilltalades skuld *utom rimligt tvivel*, vilket hänger samman med att det anses betydligt värre att fälla en oskyldig än att fria en skyldig.⁵ Inom skadeståndsrätten, där det inte finns samma värdemässiga asymmetri mellan att oriktigt utdöma och att oriktigt underlåta att utdöma skadestånd, räcker det att bevisningen ”styrker” eller rentav att den visar med ”övervägande sannolikhet” att svaranden vållat skadan.⁶ Beviskraven har inte lika etablerade verbala

4 Kravet på att läkaren ska utföra sitt arbete i överensstämmelse med vetenskap och beprövad erfarenhet infördes i svensk författning genom Oscar den II:s läkarinstruktion. Se Pontin 1891.

5 William Blackstone's berömda uttalande (1769: 352): "It is better that ten guilty persons escape than that one innocent suffer". I svensk rätt användes uttrycket "utom rimligt tvivel" första gången i rättsfallet NJA 1980 s. 725.

6 Se t.ex. 32 kap. 3 § 3 st miljöbalken (1998:808).

uttryck inom medicinen som inom juridiken, men också medicinska beslut måste med nödvändighet vila på en avvägning mellan risk och förväntad nytta med olika behandlingsalternativ, och denna avvägning förutsätter i sin tur en värdering av möjliga utfall.⁷ Detta framgår tydligt i framtagandet av generella riktlinjer där risk och nytta avvägs på gruppnivå,⁸ men gäller förstås även läkarens behandlingsbeslut i den konkreta situationen där också den individuella patientens förutsättningar (och värderingar) vägs in.

Läkarens och domarens beslutssituationer liknar alltså varandra på det sättet att båda beslutsfattarna måste *ta ställning till om evidensen är tillräcklig för att rättfärdiga ett visst handlingsalternativ i en konkret situation*. Inför denna bedömning verkar emellertid de två modellerna, EBM och FBP, ge läkaren respektive domaren helt olika svar på frågan om vilken evidens som ska beaktas och hur denna ska värderas.

3. Två till synes olika modeller

Enligt principen om *evidensbaserad medicin* (EBM) ska läkarens beslutsfattande vila på bästa möjliga evidens. I litteraturen om evidens-baserad medicin illustreras detta ofta med en pyramid, där systematiska forskningsöversikter placeras överst, följda av randomiserade kontrollerade studier, och observationsstudier (se tex Wilkinson m. fl., 2017: 22). Fallstudier och expertutlåtanden väger lätt och placeras långt ner i pyramiden i den mån de alls tillåts vara med. I Sverige tillhandahålls systematiska forskningsöversikter av Statens beredning för medicinsk och social utvärdering (SBU). SBU:s evidensbaserade använder det så kallade GRADE-systemet för att klassificera det vetenskapliga underlagets styrka. Av SBU:s handbok framgår att randomiserade kontrollerade studier anses ha högst preliminär evidensstyrka, medan observationsstudier har lägre styrka och fallstudier allra lägst (SBU:s handbok, 2017: 127 ff.). Den slutliga evidensstyrkan kan justeras upp eller ned beroende på studiens kvalitet och resultat. Principen om evidensbaserad medicin innebär därmed att det på förhand gjorts ställningstaganden både ifråga om *vad* som ska räknas som evidens och i fråga om *vilken styrka* olika slags evidens ska tillmätas.

EBM verkar för att medicinska beslut ska fattas på aktuell och välbelagd forskningsevidens och är en reaktion mot en tradition av medicinskt beslutsfattande som baserades på läkarens egen intuition och erfarenhet och på äldre auktoriteters uppfattningar. Enligt förespråkare för EBM visar det faktum att

7 Jag gjorde en mer omfattande jämförelse mellan rättsliga och icke-rättsliga beviskrav i min avhandling: *Legal Questions and Scientific Answers: Ontological Differences and Epistemic Gaps in the Assessment of Causal Relations* (2010).

8 Socialstyrelsen, Nationella riktlinjer: Metodbeskrivning, s. 6 ff. <https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/dokument-webb/nationella-riktlinjer/nationella-riktlinjer-metodbeskrivning.pdf>. Citerad 2020-05-28.

flera tidigare utbredda metoder (t.ex. åderlåtning) som *uppfattats* som mycket effektiva i själva verket var ineffektiva eller rentav skadliga, att läkares upplevda erfarenheter inte är en tillförlitlig källa. Drivkraften i EBM-rörelsens framväxt kommer till uttryck i följande idag välkända ord, som inledde en banbrytande artikel av Gordon Guyatt och andra i the EBM working group, publicerad i JAMA (Journal of the American Medical Association) (1992: 2420):

A NEW paradigm is emerging. Evidence-based medicine de-emphasizes intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research.

EBM framstår som en tydlig motpol till principen om fri bevisprövning (FBP), som anger ramarna för den svenska domarens evidenshantering. FBP kommer till uttryck i 35 kap. 1 § rättegångsbalken (1942:740), enligt vilken rätten "efter samvetsgrann prövning av allt, som förekommit, [ska] avgöra, vad i målet är bevisat". FBP kan i sin tur delas upp i principen om fri bevisföring och principen om fri bevisvärdering. Enligt principen om fri bevisföring finns som regel ingen begränsning för vilken bevisning som får läggas fram i en rättegång. Parterna är alltså normalt fria att åberopa all bevisning som de kan få fram. En princip om fri bevisföring förespråkades redan av Jeremy Bentham (1827:743) som det bästa sättet att nå sanningen men gäller inte i amerikansk rätt, där det istället finns ett omfattande regelverk som anger vilken bevisning som är "admissible" och får föras fram av parterna.⁹ Principen om fri bevisvärdering innebär att värdet av den bevisning som förts fram provas fritt av domstolen. Fri bevisvärdering kontrasteras ofta mot "legal" bevisvärdering, som ovan exemplifierades med bestämmelsen om vittnesmåls bevisverkan i 29 § i 1734 års rättegångsbalk, och inom vilken rättsregler på förhand anger vilka bevis som får föras fram och hur dessa ska värderas. Till skillnad från en legal bevisprövning innebär den fria bevisprövningen att lagstiftaren som huvudregel inte tagit ställning till vare sig *vad* som ska räknas som bevis eller *vilken styrka* olika former av bevis, till exempel ett vittnesmål, ska tillmätas.

Möjligheten att låta domaren fritt värdera bevisningen i stället för att vara bunden av legala regler diskuterades i Sverige redan i början av 1800-talet (Modéer 1987: 248, 257). Här, liksom på kontinenten, kritiserades den legala bevisvärderingens oförmåga att ta hänsyn till sammanhangets betydelse vid bedömningen av värdet av ett visst bevis (Damaska 1995). Kritik av detta slag framfördes bland annat av den svenska kriminallagskommittén, som menade

9 Det amerikanska rättsväsendet skiljer sig dock från det svenska bland annat på det sättet att bevisningens styrka bedöms av en lekmanjury, och det har ansetts angeläget att inte släppa in bevisning som riskerar att vilseleda denna.

att det var svårt att formulera regler som skulle vara tillämpliga för alla fall, och att det var orimligt att låta vissa bevis vara ovillkorligt bindande (Dereborg 1990: 35–36). Principen om den fria bevisprövningen fick genomslag i HD:s praxis från 1850-talet och framåt – en utveckling som kan ses i ljuset av en generell förstärkning av domarens makt (Modéer 1999: 401; 1987: 256–257) – och kom till fullt uttryck i 1942 års rättegångsbalk som citerades ovan.

I motsats till vad som är fallet i EBM finns alltså inom FBP som huvudregel inte någon begränsning för vilken typ av bevis som ska tillåtas, och inte heller några på förhand givna riktlinjer för styrkan hos olika typer av bevis. På ett sätt som står i tydlig kontrast till den roll EBM working group tillskrev läkarens intuition och kliniska erfarenhet i citatet ovan framhåller många praktiskt verksamma jurister och rättsvetare dessutom betydelsen av domarens intuition och erfarenhet inom FBP. I en bok om bevisvärdering skriver till exempel en tidigare domare, tillika rektor för Domstolsakademin¹⁰ och forskare i processrätt (Nordh 2013: 97):

[Vid] den slutliga sammanvägningen och avstämningen mot beviskravet är det mest betydelsefulla verktyget domarens intuition. [...] Med intuition i detta sammanhang åsyftas domarens kunskap, erfarenhet, klokskap och goda omdöme.

En annan skribent pekar på intuitionens betydelse för att identifiera det rimliga tvivel som räcker för en friande dom (Mellqvist 2013: 759, not 12):

Om jag som domare *känner* att bevisningen för en fällande dom inte är tillräckligt övertygande är det min skyldighet att frikänna! Motiveringen må bli den det blir; varje läsare av domen måste inte instämma i slutsatsen.

Å andra sidan framgår av förarbetena till rättegångsbalken att den fria bevisvärderingen inte medger att evidens får värderas på helt subjektiva grunder och att domaren måste klargöra vilka grunder som stöder övertygelsen (SOU 1938:44, s. 377–378) och många rättsvetare har betonat vikten av att bevisvärderingen respekterar såväl empiriska forskningsrön (Schelin 2006) som logiska och matematiska principer för sammanvägning av evidens (Gärdenfors 1983, Ekelöf m.fl. 2009, Dahlman 2018). Sammanfattningsvis kan sägas att det inte är klart vilket utrymme som finns för domarens intuition och erfarenhet i FBP, en fråga som kommer att diskuteras ytterligare nedan.

10 Domstolsakademin ansvarar för all juridisk utbildning av domare i Sverige.

4. Ramar för en fördjupad jämförelse

På ytan finns alltså tydliga skillnader mellan EBM och FBP. I FBP finns som vi sett som huvudregel inte någon begränsning för vilken typ av bevis (hädanefter "evidens") som ska tillåtas, och inte heller några på förhand givna riktlinjer för styrkan hos olika typer av evidens. Inte sällan framhålls beslutsfattarens (domarens) intuition och erfarenhet som viktiga verktyg för bevisvärderingen. I EBM finns i stället tydliga riktlinjer både för vad som ska räknas som evidens, och vilken styrka olika typer av evidens ska tillmätas. Där höjs också ett varnande finger för att tillmäta beslutsfattarens (läkarens) intuition och erfarenhet betydelse vid bevisvärderingen.

I det följande kommer jag att fördjupa jämförelsen av de två modellerna genom att närmare undersöka i vilka avseenden och på vilka sätt hanteringen av evidens är reglerad inom de två paradigmen. Jag kommer att skilja mellan å ena sidan frågan om vilken evidens som får beaktas (avsnitt 5 nedan) och å andra sidan frågan om hur evidensen ska värderas (avsnitt 6 nedan). Den senare frågan kommer i sin tur att delas upp i frågan om bedömningen av relevansen och värdet av *enstaka evidens* tagen för sig (avsnitt 6,1), och frågan om det *sammanvägda värdet* av all evidens (avsnitt 6.2). Jämförelsen kommer att relateras till en diskussion om de intressen som gör sig gällande i en rättslig respektive medicinsk beslutscontext.

5. Vilken evidens får beaktas?

5.1 FRIHET/BUNDENHET FÖR VEM OCH VARFÖR?

Vad gäller frågan om vilken evidens som får beaktas kan först konstateras att motsättningen mellan EBM:s riktlinjer för vilken evidens som får beaktas och den fria bevisföringens avsaknad av riktlinjer för vilken evidens som får föras fram är delvis skenbar. Den fria bevisföringen är förvisso en central del av FBP men avser *parternas* frihet att föra fram den bevisning de önskar och innebär inte att domaren är fri att på eget initiativ beakta vilken bevisning som helst. Tvärtom uttalas i processlagsutredningen, som låg till grund för införandet av FBP i den nuvarande rättegångsbalken, att den svenske domaren "är begränsad till det material, som framkommit i rättegången, och [får] sålunda icke på annat sätt skaffa sig kännedom om relevanta omständigheter" (SOU 1938:44, s. 378). Av rättegångsbalken framgår vidare att domen i tvistemål och brottmål ska "grundas på vad som förekommit i målet", den så kallade omedelbarhetsprincipen.¹¹ I motsats till vad den initiala jämförelsen i avsnitt 3 antydde finns här alltså snarare en likhet mellan den svenska domarens roll i FBP och

11 17 kap. 2 § och 30 kap. 2 § rättegångsbalken.

läkarens roll i EBM, i den meningen att båda är föremål för riktlinjer om vilken evidens som ska beaktas.

Den svenska domarens bundenhet behöver dock ses i ljuset av ramarna för den svenska rättsliga *processen*. Till att börja med är den svenska processordningen kontradiktorisk, på så sätt att parterna ska ha möjlighet att själva föra fram, ta del av och bemöta argument och bevisning som kan komma att ligga till grund för en dom.¹² Den svenska processordningen är också i hög grad akusatorisk, vilket innebär att parterna bär det huvudsakliga ansvaret för utredningen och för att relevant evidens förs fram.¹³ Till skillnad från vad som är fallet i mer inkvisitoriska system, där domaren själv spelar en aktiv roll i utredningen, är den svenska domarens roll inriktad mot att bedöma den evidens som presenteras av parterna (även om det i vissa måltyper, t.ex. brottmål också finns inkvisitoriska inslag, bl.a. på så sätt att domstolen på eget initiativ kan inhämta kompletterande expertutlåtanden). Den svenska domarens bundenhet till den evidens som förts fram av parterna tycks därmed gå att härleda till intresset av processuell rättvisa, opartiskhet etc., och kanske inte i första hand till det intresse för materiellt riktiga domar som anfördes som skäl för en fri bevisprövning.

Även om domarens bundenhet i fråga om vilken evidens som får beaktas därmed inte utan vidare kan sägas vara en nödvändig del av FBP, visar det svenska exemplet emellertid att FBP ansetts *förenlig* med en sådan begränsning. I förarbeten och doktrin har dessutom framhållits att regeln att domaren bara får beakta vad som framkommit under rättegången även främjar sanningsökandet. Detta genom att regeln antas underlätta bedömningen av vittnens tillförlitlighet och tillförandet av olika perspektiv på bevisningen samt skapa ”goda möjligheter för värdering av hela materialet i ett sammanhang”, och eftersom parterna förmodas ha bäst kännedom om omständigheterna i det konkreta fallet (SOU 1938:44, s. 22). För både den svenske domaren inom FBP och för läkaren inom EBM tycks alltså gälla att bundenheten till viss evidens ansetts främja en i sak riktig (med verkligheten överensstämmande) bedömning. Det finns därför anledning att göra en närmare jämförelse av vilken evidens dessa beslutsfattare får beakta och på vilket sätt den kan antas främja en riktig bedömning.

5.2 FRIHET/BUNDENHET TILL VILKEN EVIDENS?

Det finns, som redan antytts, väsentliga skillnader mellan *vilken* evidens som domaren respektive läkaren får beakta. Läkaren får förstås beakta de observationer som gjorts om den enskilda patienten (inklusive hänsyn till patientens

12 Detta anses framgå av rätten till en rättvis rättegång i art. 6 i den Europeiska konventionen om skydd för de mänskliga rättigheterna och de grundläggande friheterna.

13 35 kap. 6 § rättegångsbalken.

preferenser). Utöver detta får läkaren emellertid bara beakta den typ av evidens som på förhand erkänts som tillförlitlig. Här ingår i första hand systematiska forskningsöversikter, randomiserade kontrollerade studier och, i viss mån, kohortstudier och fall-kontrollstudier. I praktiken sker läkarens användning av denna evidens ofta indirekt genom behandlingsriktlinjer som vilar på den tillgängliga evidensen. Domaren, i sin tur, är som vi sett huvudsakligen begränsad till vad parterna fört fram som bevis eller till vad som annars framkommit i rättegången (till exempel hur ett vittne uppträtt under ett förhör). Även om bundenheten i fråga om vilken evidens som får beaktas tjänar flera intressen i den svenska rättsprocessen, anses den både där och inom EBM också främja en materiellt riktig bedömning. Det är emellertid intressant att konstatera att bundenheten verkar tänkt att främja sanningssökandet på olika sätt: inom EBM genom att läkaren är bunden att beakta endast *tillförlitlig evidens*, för den svenske domaren genom att bundenheten skapar förutsättningar för en sammanhållen bedömning och belysning, och därmed för en *tillförlitlig process*, där evidens och argument för respektive mot det som ska bevisas kan värderas i ett sammanhang.

Som jag återkommer till i nästa avsnitt behöver både läkaren och domaren göra en bedömning av evidensens relevans i det individuella fallet. FBP innebär emellertid – till skillnad från EBM – att ingen evidens på förhand exkluderas som otillförlitlig eller otillåten: i princip vad som helst får alltså beaktas som evidens under förutsättning att det framkommit under rättegången. Ett resultat av detta är att FBP – trots att den fria bevisföringen är en frihet för parterna snarare än för domaren – i högre grad än EBM överläter åt beslutsfattaren, domaren, att själv värdera relevansen och värdet hos olika typer av evidens, vilket kan uppfattas som ett FBP i högre grad än EBM har tilltro till beslutsfattarens förmåga att själv göra en sådan värdering.

Domaren är inte heller helt bunden till vad som framkommit under rättegången, utan får i viss utsträckning använda sin kunskap om världens beskaffenhet i sin bevisvärdering. Av 35 kap 2 § rättegångsbalken framgår att det inte krävs bevis för omständighet ”som är allmänt veterlig”. Till omständigheter som är allmänt veterliga hör så kallade ”notoriska fakta” och ”allmänna erfarenhetssatser” (SOU 1938:44, s. 379). Med ”notoriska fakta” avses allmänt kända konkreta omständigheter, som att jorden är rund, och att Stockholm är Sveriges huvudstad. Med ”allmänna erfarenhetssatser” avses allmänt kända och på ”livserfarenheten” baserade generella påståenden om sambandet mellan olika företeelser, till exempel att ”mörker gör det svårt att se” och ”om det regnar blir det blött” (Ekelöf m.fl. 2009: 281 ff.; Lindell 2007). En distinktion görs mellan *allmänna* och *särskilda* erfarenhetssatser, där de senare till skillnad från de förra förutsätter någon form av specialkunskap och måste tillföras målet explicit, till exempel genom att en av parterna åberopar ett utlåtande av en medicinsk expert. Det anses dock finnas ett visst utrymme för domstolen

att i bevisvärderingen använda sin egen kunskap också om förhållanden som domaren lärt känna genom yrkeserfarenheten eller har någon slags specialkunskap om, under förutsättning att denna kunskap kommuniceras till parterna (Ekelöf m.fl. 2009: 60-61; jfr. SOU 1938:44: 378). Domarens möjlighet att på dessa sätt använda en del av sin egen kunskap i bevisvärderingen motiveras av processekonomiska skäl (Ekelöf m.fl. 2009: 58).

Den svenska domarens möjlighet att använda kunskap baserad på egen livs- och yrkeserfarenhet kan jämföras med synen på läkarens erfarenhet inom EBM. Som framgått anses klinisk erfarenhet som inte processats i vetenskapliga studier inte utgöra tillförlitlig evidens inom EBM, och ska därmed inte ingå i beslutsunderlaget. Detta väcker frågan om den svenska domarens möjlighet att använda erfarenhetssatser i dömandet ska tolkas som att rättsordningen här har offrat sanningsintresset till förmån för intresset av processekonomi. Eller innebär det att FBP ser annorlunda på erfarenhetens tillförlitlighet än vad EBM gör? Noga taget tycks ju redan uttrycket ”allmänt veterlig erfarenhetssats” implicera inte bara att erfarenhetssatsen är allmänt utbredd, utan också att den är sann.

5.3 ERFARENHETENS ROLL

I detta sammanhang är det intressant att relatera EBM och FBP till det inom svensk hälso- och sjukvård centrala kravet att den vård som ges ska stå i överensstämmelse med vetenskap och beprövad erfarenhet (VBE).¹⁴ I Sverige fungerar VBE som ett grundläggande kvalitetskriterium inte bara för medicinskt beslutsfattande utan också inom utbildning, socialt arbete, miljövard m.m. Till skillnad från det första ledet (vetenskap) saknar det andra ledet (beprövad erfarenhet) motsvarighet i EBM:s evidensbegrepp. Att erfarenheten ska vara ”beprövad” innebär att inte vilken personlig erfarenhet som helst duger, och väcker förstas frågan om vad beprövad erfarenhet egentligen är. Studier av hur begreppet *beprövad erfarenhet* används inom den medicinska professionen, liksom i rättsliga bedömningar av medicinsk evidens, tyder på att det har flera dimensioner (Persson & Wahlberg 2017; Wahlberg 2018; Sahlin m.fl. 2020). Ett par av dessa är av särskilt intresse här: Ofta verkar beprövad erfarenhet innebära att något är *noga prövat* – erfarenheten har *visat* att något fungerar eller är fallet. Med denna innebörd, där VBE lyfter fram erfarenhetens roll som evidens blir skillnaden mot EBM tydlig. Andra gånger verkar tyngdpunkten hos den beprövade erfarenheten snarare ligga i att något är *utbrett*: att en uppfattning eller praktik delas av många. Att något är utbrett kan förstas vara ett tecken på att det är *noga prövat* och fungerar, men det finns inget nödvändigt samband mellan utbredning och prövning (Persson & Wahlberg 2015).

Överfört på erfarenhetssatserna i den rättsliga processen kan frågan ställas om det avgörande är att dessa erfarenhetssatser är utbredda (allmänt kända

14 6 kap. 1 § patientsäkerhetslagen (2010:659), 1 kap. 7 § patientlagen (2014:821).

eller i vart fall kommunicerade till parterna) eller om de också måste vara noga prövade. I den mån tonvikten ligger på utbredningen kan erfarenhetssatserna förmodas tjäna intressen av processuell rättvisa och processekonomi. I så fall finns egentligen ingen konflikt mellan synen på erfarenheten i EBM och synen på erfarenheten i FBP – det är snarare så att erfarenheten har en särskild funktion i den rättsliga beslutsprocessen som inte är relevant i den medicinska. I den mån FBP också förutsätter att en erfarenhetssats är noga prövad och kan utgöra tillförlitlig evidens blir skillnaden mot EBM:s syn på den osystematiserade erfarenheten som otillförlitlig tydligare.

Även om det inte tycks finnas något uttryckligt krav på att de erfarenhetssatser domaren får använda i bevisvärderingen verkligen är sanna, måste det vara rimligt att anta att de åtminstone förutsätts vara någorlunda tillförlitliga. Men i så fall ställs FBP, liksom VBE, inför frågan om vilka omständigheter som närmare bestämt måste föreligga för att erfarenheten ska kunna fungera som en prövningsmekanism som ger oss tillförlitlig kunskap. VBE ger självt inget svar på den frågan, men inom psykologin brukar tillgången till *feedback* och en tillräckligt *regelbunden miljö* pekas ut som avgörande för möjligheten att lära sig av erfarenheten (Kahneman & Klein 2009). En viktig fråga är förstås i vad mån domarens erfarenhetsbaserade kunskap har förvärvats under sådana omständigheter (Lindell 2007: 103). Jag återkommer till denna fråga i uppsatsens sista avsnitt.

6. Hur ska evidensen värderas?

6.1 VÄRDET AV ENSTAKA EVIDENS

Som framgick av avsnitt 5 är den fria bevisföringen en frihet för parterna att föra fram den bevisning de vill, men den ger ingen motsvarande frihet för domaren att på eget initiativ beakta annan bevisning. Att parterna är fria att föra fram den bevisning de önskar är ändå relevant i detta sammanhang, eftersom det visar att rättsreglerna inom FBP inte på förhand avfärdar viss evidens som värdelös, utan ger domaren förtroendet att bedöma värdet och relevansen av denna. Den svenska domarens frihet i detta avseende är visserligen inte helt oinskränkt: det finns alltså enstaka lagstadgade begränsningar för vilken bevisning som får åberopas och därmed ligga till grund för en dom, och vilket värde som ska tillmätas denna. Till exempel framgår av 35 kap 14 § rättegångsbalken att en skriftlig berättelse som avgetts inför en rättegång som huvudregel inte får åberopas, och 10 kap. 2 § 2 stycket ärvdabalken (1958:637) stadgar att "Hava vittnena å handlingen intygat att vid testaments upprättande så tillgått som i 1 § sägs, skall intyget, där klander väckes, äga tilltro, såframt omständigheter ej förekomma som förringa intygets trovärdighet". Högsta domstolen har dessutom i vissa avgöranden formulerat

kriterier för bedömningen av värdet av bland annat vittnesmål. HD:s avgöranden betraktas som rättskällor och är vägledande för tolkningen av rättsregler. Det råder dock oenighet om huruvida HD även kan begränsa domstolarnas frihet att värdera evidensen (Zetterquist 2019:180; jfr. Andersson 2014) och en del av de kriterier som föreslagits har senare ändrats efter att ha fått kritik för att stå i strid med forskning på området (Jfr. NJA 2010 s. 671 och NJA 2017 s. 316. Se också Schelin 2006).

På det hela taget är det fortfarande ovanligt att lagtext och prejudikat stadgar undantag från den fria bevisvärderingen. Innebär detta att domaren är fri att göra en skönsmässig bedömning av värdet och relevansen av enskild evidens? Är det kanske här domarens "intuition" kommer in? Såvitt jag kan se är svaret på den frågan inte självklart. För att bedöma värdet av viss evidens, behöver domaren bedöma sambandet mellan evidensen och det som ska bevisas, till exempel sannolikheten för att den tilltalade befunnit sig i närheten av brottsplatsen om hon är skyldig jämfört med om hon är oskyldig. Detta samband verkar vara en erfarenhetssats, och i så fall ska den ju som huvudregel vara allmänt veterlig. Oavsett hur vi närmare bestämt tolkar kravet på "allmänt veterlig" – som noga prövad, utbredd och/eller något annat – kommer detta krav att kraftigt begränsa domarens frihet att göra en egen intuitiv bedömning av evidensens värde. Däremot verkar innebörden av det rättsliga kravet på "allmän veterlighet" inte särskilt klart utan i behov av tolkning (se t.ex. Wahlberg 2019). Här finns sannolikt visst utrymme för en skönsmässig bedömning (Hart 1994). Det innebär dock bara att domaren, som så ofta, måste tolka innebörden av de rättsregler hon tillämpar – inte att hon har utrymme att göra en intuitiv värdering av evidensen *givet* den tolkning hon väl gör.

Som vi sett tonar EBM kraftigt ner erfarenhetens roll som evidens, till förmån för forskningsstudier. Beslut om vården av en enskild patient ska fattas med utgångspunkt från information om patienten och information från relevanta studier. Studiernas värde som evidens bestäms i på förhand lagda riktlinjer. En studies värde som evidens i ett konkret fall är emellertid inte bara en fråga om studiens vetenskapliga värde, utan också en fråga om i vad mån dess resultat går att överföra till omständigheterna i just det fallet. En väl genomförd studie som med säkerhet visar att ett läkemedel ger en viss förhöjd risk för hjärtsvikt på gruppnivå för personer över 70 innebär till exempel inte utan vidare att en 70-åring med utmärkt kondition eller en fyrtioåring löper samma risk. Det är tydligt att det behövs något mer än forskningsstudier och information om patienten för att tillämpa och bedöma forskningens relevans i det konkreta fallet. Enligt EBM spelar läkarens *expertis* en stor roll i denna tillämpning, och läkarens erfarenhet verkar tillerkännas rollen av medel för att uppnå den *expertis* som behövs för bedömningen. I den välkända artikeln "Evidence based medicine: what it is and what it isn't" förklarar Sackett med flera (1996: 71) att

The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. By individual clinical expertise we mean the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice.

It is [clinical expertise] that decides whether the external evidence applies to the individual patient at all and, if so, how it should be integrated into a clinical decision.

Som påpekats av Johannes Persson m.fl. tycks läkarens erfarenhet i EBM alltså tillerkännas en roll för att implementera evidens, men *inte* som evidens *i sig* (Persson m.fl. 2019). Erfarenhetens roll inom EBM skiljer sig alltså väsentligt från dess roll i VBE, och sannolikt även från dess roll i FBP.¹⁵

Man kan emellertid fråga sig om det är möjligt att, som EBM, utmönstra erfarenheten som evidens. En spännande fråga är vad det ”proficiency and judgment” som Sackett m.fl. hänvisar till närmare bestämt består i, om inte just erfarenhet av vilka faktorer som påverkar olika behandlingars utfall i konkreta fall. På senare tid har också ett mer inkluderande evidensbegrepp börjat framskynta i litteraturen om EBM. I tredje upplagan av Guyatt’s ”Users’s guide to the medical literature: a guide for evidence-based clinical practice” finns faktiskt en passage, där den osystematiserade erfarenheten erkänns som – om än svag – evidens, och det står att läsa: ”The available evidence may warrant very low confidence – it may be the unsystematic observation of a single clinician [...] – but there is always evidence” (Guyatt m.fl. 2015).

Frågan om vad som fyller ut kunskapsglappet mellan forskningsevidens och det konkreta fallet inom EBM (expertis eller kunskap förvärvat genom erfarenhet) väcker frågan om det behövs något för att fylla ut glappet mellan allmänna erfarenhetssatser och det konkreta fallet i FBP. Är det kanske här – i tillämpningen av den allmänna erfarenhetssatsen på den konkreta evidensen – som domarens intuition kommer in? En mer grundläggande fråga är förstås i vilken utsträckning glappet alls ska få lov att fyllas ut. Inom EBM menar vissa att i den mån det konkreta fallet inte avviker på vissa i förhand angivna sätt från den grupp som forskningsevidensen direkt avser, så ska denna evidens anses tillämplig också i det konkreta fallet (Guyatt m.fl. 2008). En möjlig invändning är att ett sådant tillvägagångssätt inte tar tillvara den lokala och/eller specifika kunskap som kan finnas i olika fall. Inom FBP framstår det som naturligt att vara av uppfattningen att beviskravet inte kan vara uppfyllt om de allmänna erfarenhetssatserna och det som framkommit i målet inte förslår för att uppfylla det. Mot den uppfattningen kan möjligen invändas att beviskravet, i vart fall det straffrättsliga, i så fall nog mycket sällan skulle vara uppfyllt eftersom

15 Se avsnitt 5.1 ovan.

det ofta får antas råda brist på allmänt veterliga erfarenhetssatser om evidensvärde i det konkreta fallet.

6.3 SAMMANVÄGNINGEN AV DEN TOTALA EVIDENSEN

Inom såväl FBP som EBM finns vissa riktlinjer för hur värdet av den totala evidensen ska bestämmas. I förarbetena till rättegångsbalken finns några mycket allmänt hållna anvisningar av relevans för sammanvägningen och vid sidan om detta har även Högsta domstolen uttalat sig om hur bevisvärderingen bör struktureras i vissa typer av fall (se t.ex. NJA 2015: 702). Av förarbetena framgår att en dom inte får grundas på totalintrycket av det föreliggande målet. Bevisvärderingen måste alltså bygga på en sammanvägning av värdet av all enskild evidens. Att så måste ske verkar rimligt givet vad som ovan sagts om användningen av erfarenhetssatser, då det framstår som otänkbart att det skulle finnas en allmänt veterlig erfarenhetssats om sambandet mellan det som ska bevisas i ett konkret fall och all evidens som förekommer i just det fallet. I den rättsvetenskapliga litteraturen har frånvaron av sådana mycket specifika erfarenhetssatser ibland ansetts innebära att sammanvägningen av evidensen måste ske intuitivt (Ekelöf m.fl. 2009: 193).¹⁶ Är det alltså här i sammanvägningen, slutligen, som vi hittar utrymmet för domarens intuition i bevisvärderingen?

Av förarbetena framgår emellertid också att domen ska vara objektivt grundad och stöddas av skäl som kan godtas av andra förståndiga personer (SOU 1938:44, s. 378). Vad detta närmare bestämt innebär – till exempel vem som ska räknas som en förständig person (Lindell 1987: 22) kan förstås diskuteras men det kan hur som helst ifrågasättas om uppfattningen att sammanvägningen ska ske intuitivt är förenlig med förarbetenas krav på objektivitet och intersubjektivitet. Det finns heller ingen motsättning i att mena att bevisprövningen är fri i den meningen att det inte är på förhand reglerat vilken evidens som får föras fram och beaktas eller vilket värde som ska tillmätas denna, men samtidigt ofri i fråga om hur värdet av all evidens ska sammanvägas. Inom FBP-litteraturen finns många som menar att bevisvärderingen inte får stå i strid med logiska principer, och dessutom flera som förespråkat användningen av matematiska regler för att bedöma hur evidensen samverkar (Damaska 1985: 435). Visserligen finns olika metoder för att värdera evidens. Redan bland svenska förespråkare för en matematiskt inriktad bevisvärdering finns såväl anhängare av bayesianska metoder¹⁷ (Dahlman 2018) som anhängare av icke-bayesi-

16 Se även Nordh 2013, citerad i avsnitt 3 ovan.

17 Med bayesianska metoder används Bayes' teorem för att uppdatera sannolikheten för bevismat (hypotesen) i ljuset av ny evidens. Sannolikheten för bevismat givet evidensen är en funktion av bevismats ursprungssannolikhet, sannolikheten för att evidensen ska föreligga när bevismat är sant samt sannolikheten för att evidensen ska föreligga när bevismat är falskt. En vanlig kritik mot Bayes' teorem är att en hypotes inte har någon given ursprungssannolikhet. I ett rättsligt sammanhang kan detta problem eventuellt undvikas genom att ursprungssannolikheten sätts normativt. Till exempel kan ursprungssannolikheten i ett straffrättsligt sammanhang sättas mycket lågt med hänvisning till

anska sammanvägningsregler (Gärdenfors 1983).¹⁸ Kanske kan man hävda att intuitionen har en roll i valet mellan dessa metoder. Den praktiskt verksamma domaren, däremot, är sällan tillräckligt insatt i någon av dessa metoder för att kunna välja mellan dem eller tillämpa dem i praktiken. Sedan några år tillbaka kan juriststudenter i Lund läsa en fördjupningskurs om hur bayesianska nätverk kan användas i rättslig bevisvärdering, men de flesta nu verksamma jurister har inte haft någon sådan möjlighet. Att sammanvägningen av evidensen i den fria bevisprövningen i hög grad sker intuitivt är därför förmodligen än så länge en riktig lägesbeskrivning. Men det är förstås inte detsamma som att en sådan ordning är önskvärd.

Också inom EBM tycks förutsättas att det sker en sammanvägning av de olika delarna av evidensen och att värderingen inte baseras på totalintrycket av beslutsunderlaget. I litteraturen betonas betydelsen av den beslutande läkarens förmåga att förstå "certain rules of evidence [...] the properties of diagnostic tests and [...] use a quantitative approach to these tests" (Guyatt 1992: 2421, 2424) vilket tyder på att en sammanvägning av evidensen är tänkt att ske enligt någon på förhand etablerad metod (se också Guyatt m.fl. 2013). Inte heller inom EBM är det såvitt jag förstår hugget i sten vilken metod som bör användas, men många förespråkare för EBM verkar utgå från att bayesianska resonemang ska användas vid sammanvägningen.

7. Slutsatser av en reviderad jämförelse

Jag började skriva denna uppsats för att jag upplevde det som förbryllande att två discipliner utgick från till synes olika modeller för att hantera evidens och fatta beslut. EBM:s avfärdande av beslutsfattarens intuition och erfarenhet, och riktlinjer för vad som räknas som evidens och med vilken styrka föreföll mig stå i uppenbar kontrast till FBP:s avsaknad av begränsningar i fråga om vilken evidens som får föras fram och hur denna ska värderas, och till uttalanden i litteraturen om betydelsen av domarens intuition och erfarenhet i bevisvärderingen. När jag nu närmar mig slutet på uppsatsen kan jag konstatera att de två modellerna liknar varandra i högre grad än jag förväntade mig att se. I stället för att fokusera på modellernas skillnader, skulle en kort jämförande beskrivning lika gärna kunna ta fasta på deras likheter och konstatera att det inom såväl FBP som EBM finns riktlinjer som på förhand anger vilken slags evidens som får beaktas, och som samtidigt begränsar det underlag som får ligga till grund för värderingen. En sådan beskrivning skulle också kunna påpeka att

oskyldighetspresumtionen. För att underlätta beräkningarna kan program för att konstruera bayesianska nätverk användas (se vidare Dahlman 2018).

18 "Bevisvärdeemetoden" som utvecklas i Gärdenfors 1983 är liksom den bayesianska metoden matematisk men ställer krav på en fungerande mekanism (t.ex. en orsaksmekanism) mellan bevisetemat (hypotesen) och evidensen. Se även Sahlin 2004.

båda modellerna utgår från att sammanvägningen av evidensen ska ske på ett strukturerat sätt men lämnar visst utrymme för beslutsfattarens skön (om än oklart vari detta består). Vid sidan om en lite mer nyanserad syn på förhållandet mellan de två modellerna, har jämförelsen också gett en bättre förståelse för deras särdrag. Jag avslutar därför uppsatsen med att kort summera några iakttagelser som jag tar med mig från denna övning.

En första iakttagelse är att FBP:s tilltro till domarens förmåga att värdera evidensen verkar större än EBM:s tilltro till motsvarande förmåga hos läkaren. Till skillnad från EBM avfärdar FBP inte på förhand någon typ av evidens som otillförlitlig utan överlåter helt åt domaren att bedöma evidensens relevans och värde. Domarens roll motsvarar i medicinen därmed både SBU:s (genom att avgöra vilka erfarenhetssatser som kan användas) och läkarens (genom att tillämpa dessa erfarenhetssatser i det konkreta fallet). I avsnitt 5.2 konstaterade jag att medan EBM fokuserar på att *evidensen* i sig är tillförlitlig, verkar FBP i vart fall delvis fokusera på att *processen* för att värdera evidensen är tillförlitlig. Den skillnaden finns dock bara om vi jämför domarens bevisvärdering med läkarens. Om vi i stället vidgar perspektivet till att vid sidan av läkaren omfatta även SBU och andra aktörer för att främja ett gott medicinskt beslutsfattande blir det tydligt att processens tillförlitlighet är viktig också i en medicinsk kontext. Den observerade skillnaden mellan EBM:s fokus på tillförlitlig evidens och FBP:s fokus på tillförlitlig process kan alltså i vart fall delvis förklaras av att domarens roll i bevisvärderingen omfattar fler led än läkarens.

En andra iakttagelse är att FBP, liksom VBE men i högre grad än EBM verkar utgå från att erfarenhet kan utgöra evidens av värde för bevisvärderingen även om den inte blivit föremål för en systematisk vetenskaplig prövning. Å ena sidan verkar EBM:s närmast totala utmönstring av erfarenheten som evidens väl drastisk. Å andra sidan är det en utmaning för såväl FBP som VBE att klargöra när erfarenheten är tillräckligt tillförlitlig för att kunna tjäna som evidens, men också när den inte är det. Eftersom de feedback-processer som forskningen visat är så viktiga för att vi ska lära oss av erfarenheten typiskt sett saknas vid juridisk bevisvärdering (en läkare kan se sin patient tillfriskna men en domare får sällan någon återkoppling som indikerar om den tilltalade verkligen begått det brott han dömts för) bör vi nog vara försiktiga med att utgå från att domarens erfarenhet av bevisvärdering per automatik ger kunskap om vad som utmärker ett trovärdigt vittnesmål eller hur ett våldtäktsoffer normalt beter sig.¹⁹ För FBP:s del kan en del av förklaringen till att erfarenheten får betydelse som evidens kanske vara att det sällan finns vetenskapliga studier om hur sannolikt det är att någon stannar kvar i en lägenhet efter att ha blivit

19 Jfr tex. Bolding 1953:312 "Det är [...] rimligt att anta att domaren har tillägnat sig en viss förmåga att - dock icke utan felmarginal- uppskatta bevisvärdet av sådana vaga omständigheter som en persons "sätt att vara" och "sätt att uttrycka sig".

hotad eller om andra liknande sociala samband som är relevanta för att döma i en rättegång. Att det saknas vetenskapliga studier om flera av de samband som är relevanta vid bevisvärderingen kan kanske också delvis förklara den högre tilltro som FBP tycks uppvisa till beslutsfattarens förmåga att värdera evidensen och tolka riktlinjerna för bevisvärderingen – det är sannolikt betydligt svårare att på förhand ange kriterier för en erfarenhetssats tillförlitlighet än att ranka olika typer av vetenskapliga studier. Det finns dock skäl att påpeka att om svårigheten att fånga sociala samband vetenskapligt beror på att mänskligt agerande är oregelbundet, kan det nog vara minst lika svårt att förvärva tillförlitlig kunskap om sådana samband genom erfarenheten.

En tredje iakttagelse, som nog hänger samman med den första iakttagelsen, är att domaren i större utsträckning än läkaren tolkar och därmed formar sina egna riktlinjer. Riktlinjerna för domarens bevisvärdering är förhållandevis öppna för tolkning. Begreppet *allmän erfarenhetssats* verkar till exempel betydligt mer öppet för tolkning än begreppet *randomiserad studie*, och det råder långt ifrån enighet om vilken metod domaren bör använda för att sammanväga bevisningen. Mot bakgrund av att domarens uppdrag är just att tillämpa och tolka rättsregler, är det förstås inte särskilt konstigt att hon har i uppgift att tolka också de rättsliga riktlinjerna för bevisvärderingen. Domarens utrymme att tolka den rättsliga regleringen av bevisvärderingen bör dock inte misstas för ett utrymme att göra en intuitiv värdering av evidensen.

En fjärde och sista iakttagelse rör just intuitionens roll i bevisvärderingen. Sammantaget kan jag inte se att FBP lämnar större utrymme för intuitionen än vad EBM gör, i vart fall inte i teorin. Till att börja med verkar det inte finnas utrymme att låta intuitionen påverka värderingen av enskilda bevis, som ju i stället ska ske med hjälp av allmänt veterliga erfarenhetssatser. Eftersom det finns regler att använda för sammanvägningen av evidens – låt vara att det återstår en del arbete för att implementera dem i den rättsliga processen – verkar det inte heller finnas behov av att åberopa intuitionen i sammanvägningsledet. En annan sak är förstås att intuitionen kan få stor betydelse i praktiken, om beslutsfattaren exempelvis bortser från de sammanvägningsregler som finns. I teorin liknar intuitionens roll i FBP nog snarare dess roll i EBM, och kan träda in vid bedömningen av om en generell erfarenhetssats är tillämplig på evidens i det konkreta fallet. Något som emellertid hade kunnat nödvändiggöra ett större utrymme för intuitionen i den rättsliga bevisvärderingen är om det skulle visa sig att den tillgängliga evidensen (inklusive de erfarenhetssatser domaren kan använda) med de regler som finns för sammanvägning inte räcker för att uppfylla det rättsliga beviskravet i någon meningsfull utsträckning. För egen del håller jag det visserligen inte för osannolikt att det skulle kunna vara på det viset. Om det är så, och om det i så fall är acceptabelt, är öppna och sannolikt svarbesvarade frågor, men inte desto mindre viktiga.

Referenser

- Andersson, S., 2014. "Rättsdogmatisk metod och bevisvärdering", i Andersson, S. & Lainpelto, K. (red.), *Festskrift till Christian Diesen*. Stockholm: Norstedts Juridik.
- Bentham, J., 1978 (1827). *Rationale of Judicial Evidence*, Vol. V. New York: Garland.
- Blackstone, W., 1769. *Commentaries on the Laws of England*. Oxford: Clarendon Press.
- Bolding, P.-O., 1953. "Sannolikhet och bevisvärdering i brottmål", *SvJT*, s. 305.
- Dahlman, C., 2018. *Beviskraft: Metod för bevisvärdering i brottmål*. Stockholm: Norstedts Juridik.
- Damaska, M., 1995. "Free Proof and Its Detractors", *The American Journal of Comparative Law* 43, s. 343-357.
- Dereborg, A., 1990. *Från legal bevisteori till bevisprövning i svensk straffprocess*. Stockholm: Juristförlaget.
- Ekelöf, P.-O., Edelstam, H. & Heuman, L., 2009. *Rättegång Häfte 4*. Stockholm: Norstedts Juridik.
- Europeiska konventionen om skydd för de mänskliga rättigheterna och de grundläggande friheterna.
- Gärdenfors, P., 1983. "Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory", i Gärdenfors, P. (red.), *Essays dedicated to Sören Halldén on his sixtieth birthday*. Theoria.
- Guyatt, G., m.fl., 1992. "Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine", *JAMA*, 268, s. 2420-2425.
- Guyatt, G., Rennie, D., Meade, M. & Cook, D., 2008. *Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 2nd ed., McGraw Hill Education.
- Guyatt m.fl., 2013. "GRADE Guidelines: 11. Making an Overall Rating of Confidence in Effect Estimates for a Single Outcome and for All Outcomes", *Journal of Clinical Epidemiology* 66, s. 151-157.
- Guyatt, G., Rennie, D., Meade, M. & Cook, D., 2015. *Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed., McGraw Hill Education.
- Hart, H. L. A., 1994. *The Concept of Law*. Oxford: Clarendon.
- Kahneman, D. & Klein, G., 2009. "Conditions for Intuitive Expertise: A Failure to Disagree", *American Psychologist* 64, s. 515-526.
- Lindell, B., 1987. *Sakfrågor och rättsfrågor: En studie av gränser, skillnader och förhållanden mellan faktum och rätt*. Uppsala: Iustus.
- Lindell, B., 2007. *Notorietet och kontradiktion*. Uppsala: Iustus.
- Mellqvist, M., 2013. "Om tro, tyckande och vetande: Högsta domstolens värdering av utsagor", *SvJT*, s. 753.
- Miljöbalken (1998:808).
- Modéer, K. Å., 1987. "En god domare är bättre än en god lag: Om juridiska yrkesroller i ett historiskt perspektiv", *SvJT*, s. 248.
- Modéer, K. Å., 1999. "Den stora reformen: Rättegångsbalkens förebilder och förverkligande", *SvJT*, s. 400.
- NJA 1980, s. 725.
- NJA 2010, s. 671.

- NJA 2015, s. 702.
- NJA 2017, s. 316.
- Nordh, R., 2013. *Bevisrätt C: Bevisvärdering*. Uppsala: Iustus.
- Patientlagen (2014:821).
- Patientsäkerhetslagen (2010:659).
- Persson, J. & Wahlberg, L., 2015. "Vår erfarenhet av beprövad erfarenhet: Några begreppsprofiler och ett verktyg för precisering", *Läkartidningen* 112, s. 2230 -2232.
- Persson, J., Väreman, N., Wallin, A., Wahlberg, L. & Sahlin, N. E., 2019. "Science and Proven Experience: a Swedish Variety of Evidence Based Medicine and a Way to Better Risk Analysis?", *Journal of Risk Research* 22, s. 833-843.
- Pontin, D. M., 1891. *Författningar m.m. angående medicinalväsendet i Sverige, omfattande år 1890*. Stockholm: P.A. Norstedt & Söner.
- Rättegångsbalken i 1734 års lag.
- Rättegångsbalken (1942:740).
- Sackett, D., Rosenberg, W., Muir Gray, J., Haynes, R. & Scott Richardson, W., 1996. "Evidence Based Medicine: What It Is and What It Isn't", *BMJ* 312, s. 71-72.
- Sahlin, N.-E., 2004. *The Evidentiary Value Model: Too Short an Introduction (book)*. <http://www.nilsericsahlin.net/book/evidentiary-value/>. Citerad 2020-06-01.
- Sahlin, N.-E., Dewitt, B., Persson, J., Wahlberg, L. & Wallin, A., 2020. "Så uppfattar tandläkare beprövad erfarenhet: några resultat från en enkätstudie", *Tandläkartidningen* 7, s. 50-54.
- Socialstyrelsen, *Nationella riktlinjer: Metodbeskrivning*. <https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/dokument-webb/nationella-riktlinjer/nationella-riktlinjer-metodbeskrivning.pdf>. Citerad 2020-05-28.
- Schelin, L., 2006. *Bevisvärdering av utsagor i brottmål*. Stockholms universitet.
- SOU 1938:44. Processlagsberedningens förslag till rättegångsbalk, II.
- SBU:s handbok, 2017. *Utvärdering av metoder i hälso- och sjukvården och insatser i socialtjänsten*.
- Wahlberg, L., 2010. *Legal Questions and Scientific Answers: Ontological Differences and Epistemic Gaps in the Assessment of Causal Relations*. Lund.
- Wahlberg, L., 2018. "Rätten till ersättning för gränsöverskridande vård och kravet på vetenskap och beprövad erfarenhet", *Förvaltningsrättslig tidskrift* 4, s. 789-817.
- Wahlberg, L., 2019. "Begreppet lagtolkning", i C. Dahlman & L. Wahlberg (red.), *Juridiska grundbegrepp: en vänbok till David Reidhav*. Studentlitteratur, s. 141-162.
- Wilkinson, I., Raine, T., Wiles, K., Goodhart, A., Hall, C., O'Neill, H., 2017. *Oxford Handbook of Clinical Medicine*. Oxford: Oxford University Press.
- Zetterquist, O., 2019. "Begreppet prejudikat", i C. Dahlman & L. Wahlberg (red.), *Juridiska grundbegrepp: en vänbok till David Reidhav*. Studentlitteratur, s. 163-190.
- Ärvdabalken (1958:637).

Hjälpdilemman: Beslutsfattande när man inte kan hjälpa alla i nöd

Arvid Erlandsson

Abstract

Helping dilemmas occur when it is impossible to help everyone in need, and when one must decide how to allocate resources across multiple beneficiaries. Deciding which patient that should be connected to the only available respirator, or deciding which charitable organization to donate to, are both examples of real-life helping dilemmas. This paper examines the meaning of moral dilemmas and especially helping dilemmas, discusses different normative perspectives of helping dilemmas as well as the influential effective altruism movement, and summarizes findings from my own and other's empirical research related to how people behave when faced with helping dilemmas.

Introduktion

Denna artikel handlar om hjälpdilemman, som uppkommer när de tillgängliga resurserna inte räcker till för att hjälpa alla som behöver det, och det är upp till en beslutsfattare att bestämma hur resurserna ska fördelas mellan de behövande, eller vilka som ska få hjälp och inte. *Beslutsfattande i hjälpdilemman* är ett av flera perspektiv som man kan utgå från när man, som jag, bedriver forskning om "hjälpandets psykologi".

Först definierar jag begreppet "hjälpande" och argumenterar varför det, från ett samhällsperspektiv, är minst lika viktigt att få människor att fatta kloka beslut i hjälpdilemman som att få dem att hjälpa mer. Därefter förklarar jag närmare vad som menas med ett dilemma och mer specifikt ett hjälpdilemma. Efter det så följer en diskussion om vad som räknas som moraliskt rätt agerande

Arvid Erlandsson är docent vid Institutionen för beteendevetenskap och lärande och JEDI-Lab, Linköpings universitet.

E-post: arvid.erlandsson@liu.se

i hjälpdilemman. Sen följer en lista med attribut, eller villkor som kan få människor att föredra att rädda färre snarare än fler liv i hjälpdilemman (så kallad antalsökänslighet), och en längre beskrivning av en av våra studier om just detta. Jag avslutar med att beskriva min egen och andras forskning om hjälpdilemman utifrån några andra relevanta perspektiv (individ- och situationskillnader, psykologiska mekanismer samt hur olika beslut i hjälpdilemman påverkar hur man uppfattar personer), och rundar av med en kort normativ kommentar.

Det bör tydliggöras att denna artikel skrevs innan Coronaviruset lamslog världen. På många sätt är texten än mer aktuell nu eftersom fenomenet hjälpdilemma under pandemin exemplifierats tydligt genom prioriteringar inom vård (t.ex. vilka sjuka patienter ska få förtur till de begränsade intensivvårdsplatserna), och inom samhällsekonomi (t.ex. vilka aktörer ska kompenseras mer med ekonomiska stödpaket). Jag kommer dock inte att diskutera hjälpdilemman specifikt i en COVID-19 kontext.

Vad är hjälpande?

Hjälpande är ett brett begrepp som inbegriper flera olika typer av handlingar men som här definieras som *”En medveten handling som förbättrar välfärden för den som hjälps, och som innebär någon form av kostnad för den som hjälper”*. Hjälpande kan ske direkt (Mariella ger tiggaren 100kr) eller indirekt (Mariella ger 100kr till Stadsmissionen som i sin tur hjälper tiggaren). Hjälpande kan dessutom ske genom olika medel i form av pengar (donationer), tid (volontärbete), uppoffringar - ansträngning och minskad bekvämlighet - (leva mer miljövänligt), risk för kroppsskada (livräddande), eller med delar av sin egen kropp (blod och organdonationer).

Tre frågor för att förbättra människors hjälpande

Förutom att vara ett intressant grundforskningsämne så kan ökad kunskap om hjälpankets psykologi åtminstone indirekt bidra till att göra världen till *”en bättre plats”*. Genom att få bättre insikt i människors beslutsfattande kommer vi även ha möjlighet att med hjälp av så kallad nudging (Thaler & Sunstein 2009) påverka deras framtida beslut i en riktning som kan göra att det samlade lidandet minskar och den samlade lyckan ökar. När det kommer till människors beteende i hjälpsituationer så finns det – som jag ser det – tre frågor som vi kan fokusera på om vi vill förbättra människors beslutsfattande i hjälpsituationer.

1. Hur kan vi få människor som inte hjälper att hjälpa överhuvudtaget?

Konkret handlar det t.ex. om vad man kan göra för att öka sannolikheten att människor som tidigare inte hjälpt börjar göra det (genom att t.ex. donera pengar till välgörenhet eller att utföra volontärarbete).

2. Hur kan vi få människor som redan hjälper att hjälpa mer?

Denna fråga fokuserar på skillnaden mellan att hjälpa lite och att hjälpa mycket. Vissa hjälpbeteenden är vad man kan kalla för symboliska; de innefattar en väldigt liten personlig kostnad och har ingen större betydelse för andra. Andra hjälpbeteenden är omfattande, uppoffrande och kan ha stor betydelse. Välgörenhetsorganisationer är medvetna om att det i de flesta fall är enklare att få en befintlig givare att öka sina insatser än vad det är att rekrytera nya givare och att de tekniker som fungerar bäst för att rekrytera nya hjälpare sällan är de mest effektiva när det gäller att få existerande hjälpare att hjälpa mer. Det har även föreslagits att beslutet att hjälpa och beslutet om *hur mycket* man vill hjälpa drivs av olika typer av emotioner (Dickert, Sagara & Slovic 2011).

De två första frågorna handlar i princip om att öka mängden hjälpande i världen, antingen genom att öka antalet som hjälper eller genom att öka mängden hjälp som varje person bidrar med. Båda är dessutom aktuella för alla som jobbar inom frivillighets- och välgörenhetssektorn då dessa har ett gemensamt intresse av att öka storleken på den kaka alla ska dela på. Det finns dock en tredje fråga där enigheten inte är lika stor och som därför inte diskuteras lika ofta bland välgörenhetsorganisationer och som inte heller undersöks lika ofta i forskningssammanhang (Breeze 2013; Ord 2013). Den lyder:

3. Hur kan vi få människor som redan hjälper att hjälpa mer effektivt?

Denna tredje fråga fokuserar på hur människor beslutar när de ställs inför situationer där de inte kan hjälpa alla och när resurser behöver fördelas mellan olika hjälpprojekt, dvs hur man delar kakan (Gordon-Hecker, Chosen-Hillel et al. 2017; Soyer & Hogarth 2011). Dessa situationer kallas för hjälpdilemman, vilket också är temat för denna artikel.

Personliga och moraliska dilemman vs. hjälpdilemman

För att bättre förstå vad ett hjälpdilemma är så är det klokt att jämföra det med andra typer av dilemman. Definitionen av ett dilemma är enligt Wiktionary: ”en besvärlig situation där man tvingas göra ett val mellan två eller flera alternativ som vart och ett frambringar olika oangenäma följder” och enligt NE: ”svår valsituation”; och vidare: ”Att stå inför ett dilemma innebär förenklat att hur

man än gör så blir det fel. I en valsituation verkar båda alternativen lika dåliga eller lika bra". När jag använder begreppet så är dock inte "besvärligheten" eller "svårigheten" en central del av definitionen. Beslutande i ett dilemma kan upplevas som svårt, men det kan även upplevas som enkelt och göras snabbt och utan tvekan. Dessutom bör det betonas att dilemman uppkommer inte bara när man ställs inför två oangenäma alternativ utan även när man ställs inför två eller fler angenäma alternativ (även om själva beslutssituationen upplevs som obehaglig).

Vidare kan dilemman handla om olika saker. Ett personligt dilemma råder när du ställs inför ett valbeslut som primärt rör dig själv. Dessa personliga dilemman kan vara vardagliga och triviala (ska jag välja bubbelvatten eller kranvatten till min lunch) eller storskaliga och livsavgörande (ska jag stanna och jobba i Tokyo eller ska jag åka tillbaka och börja forskarprogrammet i Lund).

Ett moraliskt dilemma inträffar när du ställs inför ett val som primärt påverkar andra, men där framförallt insatserna är höga (Christensen & Gomila 2012). Ett exempel på ett klassiskt moraliskt dilemma är det så kallade trolley-dilemman (Thomson 1985).

En förarlös tågsvagn rullar okontrollerat fram på ett spår. Om inget görs så kommer den köra över fem oskyldiga arbetare som befinner sig längre fram på spåret. Du står vid en spårväxlare och kan enkelt växla över tåget på ett alternativt spår. Problemet är att på detta spår befinner det sig en oskyldig arbetare. Frågan är om du bör växla över tåget på det alternativa spåret och på så sätt aktivt orsaka en persons död för att undvika att fem personer dör?

Detta dilemma och dess varianter har använts flitigt inom både filosofi och moralpsykologi (Greene 2007; 2008). En viktig aspekt av dilemman är dock att det ställer två olika typer av moraliska handlingar mot varandra, nämligen hjälpsamt (av de fem på huvudspåret) mot aktivt skadande (av den ensamma personen på det alternativa spåret). Denna så kallade act-omission (handlingsunderlåtelse) distinktion (Baron 2008; Baron & Ritov 2009) är mycket intressant, men inte på något sätt en nödvändig del av ett moraliskt dilemma. Man kan konstruera minst lika intrikata moraliska dilemman, där aktivt skadande inte är en del, utan där man bara fokuserar på hjälpsamt. Denna typ av dilemman kallar jag här för hjälpdilemman. Ett artificiellt exempel på ett sådant är:

Du är ute på en bergsbestigningsexpedition med din nära vän. Ni är båda vana bergsbestigare. På samma berg finns dessutom en grupp med fem bergsbestigare som är okända för er. Plötsligt sker ett ras. Du befinner dig på stabil mark och klarar dig bra men alla andra ramlar ner och hänger nu i sina respektive linor. Din vän hänger i en lina och de fem okända bergsbestigare hänger i en gemensam lina. Du kan säkra linan som din vän hänger i eller linan som de fem okända hänger i, men du kan inte säkra båda. Vad bör du göra?

Detta "bergsbestigardilemma" är ett hjälpdilemma, eftersom det inte innehåller någon form av aktivt skadande. Istället handlar frågan i princip om vilka av de drabbade, som du moraliskt prioriterar att hjälpa – i det här fallet din vän eller fem okända människor.

Verkliga hjälpdilemman

Bergsbestigardilemmat är såklart ett konstruerat och närmast extremt scenario som de flesta människor lyckligtvis aldrig kommer ställas inför. Det är emellertid viktigt att inse att de flesta av oss ställs inför hjälpdilemman mer eller mindre dagligen. Läkare prioriterar vilka patienter som ska få hjälp i vilken ordning och hur mycket pengar som ska läggas på olika undersökningar och behandlingar. Politiker beslutar vilka av de många utsatta samhällsgrupperna som det ska satsas mer pengar på i nästa budget. Människor som jobbar med bistånd beslutar om vilka föreslagna biståndsprojekt som ska finansieras och vilka som inte ska få pengar. Dessutom ställs vi alla inför hjälpdilemman när vi medvetandegörs om att det finns oräkneliga behov runt om i världen, som olika hjälporganisationer verkar för att mildra. Hur mycket pengar vi än har så kan vi inte hjälpa alla, utan måste fördela vårt hjälpande mellan olika behövande och varje hundralapp som du skänker till en välgörenhetsorganisation skulle du kunna ha skänkt till en annan organisation. Även om det oftast inte upplevs som ett moraliskt dilemma, så är det just det vi hamnar i när vi beslutar oss för att donera till en organisation, men inte till en annan. Detta är något som de så kallade "effektiva altruisterna" försöker göra människor medvetna om.

Effektiv altruism

Effektiv altruism är en social rörelse som framhäver viktigheten av att fatta rätt beslut i hjälpsammanhang (www.effectivealtruism.org). Förutom att uppmåna alla som kan (vilket innefattar de flesta svenskar med en inkomst) att donera minst 5–10 % till välgörenhet (Singer 2009), så betonar rörelsen hur viktigt det är att donera till rätt ändamål och till rätt organisationer (Singer 2015; MacAskill 2015). Ett extremt exempel som Scott Alexander presenterar på rörelsens hemsida handlar om hur man genom en insamling lyckades få in 550 000 pund för att kunna behålla Turners berömda målning *Blue Rigi* på Tate Gallery i England. Vi kan anta att åtminstone de flesta av de 11 000 personer som donerade pengar för detta ändamål bryr sig genuint om konst, men att de ändå tycker att människoliv är viktigare än en tavla. Enligt noggranna uträkningar skulle pengarna som samlades in istället nämligen ha kunnat användas för att förbättra saniteten i många fattiga afrikanska byar – och på det sättet rädda över 1000 liv om man räknar konservativt. Ifall man tvingar folk att välja mellan att rädda en tavla och att rädda 1000 människoliv så skulle nästan

alla välja människoliven – ett ganska enkelt moraliskt dilemma med andra ord. Problemet är att i människors medvetande så upplevs fallet med konsten inte ens som ett hjälpdilemma; i det aktuella fallet sågs hjälpen som en generös och osjälvisk handling för att rädda en vacker tavla, inget mer.

Ett annat exempel som tas upp av moralfilosofen Toby Ord (2013) handlar om olika sätt att minska spridningen av HIV och AIDS. Ords syfte var att jämföra effektiviteten av olika interventioner och detta mättes med det förväntade antal friska livsår, som kunde fås om 1000 dollar spenderades på endast en av de olika interventionerna. Den mest effektiva interventionen visade sig vara att satsa pengarna på utbildning för högriskgrupper (cirka 27 extra livsår för 1000 dollar), medan operationer för Kaposis sarkom gav mindre än 0.05 extra livsår för samma belopp. Även kondomdistribution var mindre än hälften så effektivt som utbildning (cirka 12 extra livsår).

Man bör definitivt vara tydlig med att dessa uträkningar är väldigt komplexa och långt ifrån exakta, men en konsekvenskalkyl som predicerar med 20-30 % felmarginal ger oss fortfarande långt mycket mer information än ingen konsekvenskalkyl alls. Dessutom behöver inte varje enskild person sätta sig in i exakt hur effektiva olika hjälpprojekt är. Hemsidan www.givewell.org ger tips åt människor som vill donera så effektivt som möjligt. Enligt effektiv altruismrörelsen så är många hjälpprojekt mer eller mindre verkningslösa (globalt sett) medan andra är väldigt effektiva, och vi är då moraliskt förpliktade att allokera pengar till något av de mer effektiva hjälpprojekten.

En annan slutsats är, att åtminstone människor som har en medelstor eller hög inkomst, av effektivitetsskäl bör undvika att hjälpa direkt, genom t ex volontärarbete, och istället använda tiden för att tjäna mer pengar, som sen kan doneras till välgörenhet. Volontärarbete kanske känns bättre för den som utför det, men det är sällan det mest effektiva sättet att hjälpa. Ytterligare ett tips från rörelsen är att begåvade människor, som vill hjälpa effektivt, inte i första hand bör utbilda sig till ett omvårdnadsyrke (t ex sjuksköterska) utan snarare skaffa sig ett yrke som är så välbetalt som möjligt, och därefter donera en stor del av sin inkomst till effektiva välgörenhetsorganisationer.

Den effektiva altruismrörelsen har fått stor genomslagskraft och inspirerat många mycket rika människor (t ex Bill och Melinda Gates och Warren Buffet) att donera enorma summor till de mest effektiva välgörande ändamålen. Samtidigt har rörelsen, och idéerna som den för fram, kritiserats från många håll (Gabriel 2017; McMahan 2016). En typ av kritik är 1) att rörelsen är elitistisk och skuldbelägger och nedvärderar de som hjälper ineffektivt, småskaligt eller "följer sitt hjärta", 2) att den fokuserar oproportionerligt mycket på mätbara utfallsvariabler och 3) att den bekämpar t ex fattigdom snarare än de politiska system som orsakar fattigdomen. En fjärde typ av kritik är att effektiva altruister är beredda att offra jämlikhet och rättvisa i sin kamp för att maximera effektiviteten.

Maximera effektivitet eller maximera rättvisa

Tänk dig en läkare med ansvar att bestämma vilka kritiskt sjuka patienter som ska få förtur till existerande njurar. Du vet att njurtransplantationer har 95 % chans att lyckas om det råder en perfekt blodgruppsmatchning mellan donatorer och mottagare, och 75 % chans om det är en partiell blodgruppsmatchning. Idag har du fått tillgång till tio nya, donerade njurar och du läser att samtliga dessa representerar en perfekt matchning med dina tio patienter som har blodgrupp A (dvs. 95 % chans att lyckas) och en partiell matchning med dina tio patienter med blodgrupp AB (dvs. 75 % chans att lyckas). Du vet också att nya njurar kommer ganska sällan, så sjuka patienter som inte får en transplantation riskerar att dö. Hur många patienter från respektive grupp väljer du ut till njurtransplantation (Baron & Szymanska 2011)?

Om ditt mål, liksom en effektiv altruist, är att maximera effektiviteten så borde du välja ut samtliga tio patienter med blodgrupp A, eftersom chansen att överleva är högre för dem. Om man kalkylerar den beräknade nyttan så är det 0.95 (sannolikheten att patienten överlever) $\times 10$ (antalet patienter) = 9.5 räddade liv. Samtidigt så innebär ju detta beslut att du bestämmer vilka av patienterna som ska leva och vilka som skall löpa risk att dö.

Om ditt mål istället är att maximera rättvisa så är det istället troligt att du kommer att behandla fem slumpmässigt utvalda patienter med blodgrupp A och fem slumpmässigt utvalda patienter med blodgrupp AB. På så sätt så har alla patienter en exakt lika stor chans att få en ny njure. Samtidigt så är den beräknade nyttan $(0.95 \times 5) + (0.75 \times 5) = 8.5$, alltså ett liv mindre än om du låtit alla med blodgrupp A få en njure.

I en studie som testade ett liknande dilemma valde 56 % av vanliga människor och 41 % bland de medicinska experterna hjälpprojektet som var mer rättvist men ineffektivt (alla i riskgruppen kunde testas med normaltestet och 1000 dödsfall beräknades därmed kunna förhindras) framför hjälpprojektet, som var mer effektivt men inte kunde erbjudas till alla (hälften av de i riskgruppen kunde testas med det uppdaterade testet och 1100 dödsfall ansågs kunna förhindras; Ubel et al. 1996). Studier har även visat att människor ogillar att välja hur en extra, odelbar resurs ska fördelas mellan två eller fler exakt lika förtjänta personer, eftersom detta skulle innebära en orättvisa och eventuellt signalera att man är partisk. I dessa lägen är många villiga att kasta bort den extra resursen, vilket innebär att man offrar effektiviteten för att slå vakt om rättvisan (Gordon-Hecker, Rosensaft-Eshel et al. 2017; Chosen-Hillel Shaw & Caruso 2015).

Olika moralfilosofiska skolor kommer fram till tämligen olika slutsatser i denna fråga. För de klassiska utilitaristerna är de beräknade konsekvenserna och effektiviteten det enda som spelar roll i ett hjälpdilemma och det är därför inte konstigt att de flesta effektiva altruister även kallar sig utilitarister (Singer 2009; 2015). För andra, mer regel- och rättvisefokuserade,

moralfilosofier så spelar inte de beräknade konsekvenserna någon roll alls, utan om valet står mellan ett hjälpprojekt som kan behandla en patient eller ett som kan behandla tio patienter, så bör man singla en slant om vilket projekt som ska genomföras, eftersom detta ger alla patienter en lika stor chans att få behandling (Taurek 1977).

För de flesta människor så spelar dock både de beräknade konsekvenserna (effektivitet) och andra värden (t ex rättvisa, jämlikhet och ens magkänsla) en viktig roll. Dessutom bör det betonas att människor ofta ogillar att bli tipsade eller rekommenderade om hur de ska hjälpa på det mest effektiva sättet, då detta upplevs inkräkta på ens fria vilja (Berman et al. 2018). Alla dessa fenomen leder till att man tämligen ofta ser allokeringsbeslut som hjälper färre personer snarare än fler. Detta kallas ibland för antalsokänslighet (Dickert & Slovic 2011; Dickert et al 2015; Slovic 2007; Baron 1997; 2008).

Vad gör människor antalsokänsliga i hjälpdilemman?

Allt annat lika så brukar människor föredra att göra mer nytta än mindre och på samma linje så föredrar de flesta att rädda fler liv än färre. Om du får välja mellan Projekt A som kan rädda 100 storkar och det lika kostsamma Projekt B som kan rädda 80 storkar så kommer de flesta att välja Projekt A, förutsatt att de ser något som helst värde i att rädda storkar.

Det är dock sällan som två hjälpprojekt är exakt lika, förutom antalet. Oftast skiljer sig de tillgängliga hjälpprojekten på flera sätt, och vissa aspekter av en hjälpsituation tenderar att göra oss mer antalsokänsliga än andra. Nedan följer några attribut, eller omständigheter som vi vet gör att människor föredrar hjälpprojekt som kan hjälpa färre snarare än fler individer.

BARN SOM OFFER

De flesta människor värdesätter ett barns liv högre än en vuxens liv och detta innebär att många kommer att föredra ett projekt som kan rädda färre barn framför ett projekt som kan rädda fler vuxna (Goodwin & Landy 2014; Tsuchiya, Dolan & Shaw 2003). Detta kan förklaras på flera sätt: (1) Barn påminner oss om vår egen avkomma som vi genom evolutionen utvecklats till att beskydda (Popp et al. 2006). (2) Barn upplevs som oskyldigare och mer förtjänta av hjälp än vuxna (Back & Lips 1998). (3) Barn har längre kvar att leva, vilket innebär att man kan maximera antalet friska livsår genom att hjälpa dem (Goodwin & Landy 2014; Bravo Vergel & Sculpher 2008).

KVINNOR SOM OFFER

Ett vanligt, traditionellt utrop i akuta evakueringsituationer är ”Kvinnor och barn först” vilket kan tolkas som att kvinnor (liksom barn) bör värderas högre

än vuxna män, när man allokera hjälp. I linje med detta tyder forskning på att kvinnliga behövande i många situationer får mer hjälp än manliga. En förklaring till detta är att män (i högre grad än kvinnor) använder hjälpanvändning som ett sätt att kontakta och imponera på potentiella partners, och för att signalera att de har hjärtat på rätta stället (van Vugt & Iredale 2013; Raihani & Smith 2015), och dessutom resurserna som krävs för att hjälpa (t.ex. pengar eller mod). En annan förklaring till att människor ofta föredrar projekt som räddar kvinnor framför män är relaterad till upplevd oskyldighet. Enligt stereotypen upplevs kvinnor vara mindre kapabla att skydda sig själva och ha ett större behov av hjälp. Dessutom upplevs de som mindre hotfulla och kan därför förtjäna att bli hjälpta före män (Curry, Lee & Rodriguez 2004).

OFFER SOM INTE HAR NÅGON SKULD I PROBLEMET

Ett tredje sätt som oskyldighet kan göra folk antalsokänsliga är när vissa drabbade lider på grund av omständigheter helt utanför deras kontroll, medan andra lider på grund av tidigare dåliga beslut (Edlin, Tsuchiya & Dolan 2012). I en serie studier fann Hanna Zagefka och kollegor (2011) att människor som lider på grund av en naturkatastrof hjälps i högre grad än människor som lider lika mycket på grund av ett inbördeskrig. Deras förklaring till detta var att en naturkatastrof upplevs vara en orsak som ingen kan beskyllas för, medan ett inbördeskrig är resultatet av mänskliga val och att det därför är någons (och eventuellt även den behövandes) fel.

Ett annat, mer direkt sätt att manipulera oskyldighet är att specificera varför en person lider. En person som behöver hjälp delvis på grund av sitt eget leverne (t.ex. cancer orsakad av missbruk eller osunda matvanor), uppfattas som lite mindre förtjänt av hjälp än en person som lider på grund av ren och skär otur (t.ex. en person som varken röker eller dricker och som håller sig i form, men som ändå drabbats av cancer). Antalsokänslighet uppkommer när vi föredrar att hjälpa färre ”oskyldiga” patienter än fler patienter som delvis orsakat sitt eget lidande.

INGRUPPSOFFER

Ingruppen består av de individer som du upplever att du har ett speciellt band till och som du upplever dig bilda en gemensam enhet med. Människor är generellt sett inte bara del av *en* ingrupp utan flera, och det är möjligt att betona olika ingrupper vid olika tillfällen. Den starkaste och mest universella ingruppen är oftast ens familj. De flesta människor skulle inte tveka att välja ett hjälpprojekt som räddar ens eget barn, även om detta skulle innebära att ett mycket större antal andra barn blev utan hjälp. Exempel på andra starka ingrupper är ens vänskapskrets, ens religiösa tillhörighet, vilket idrottslag man hejar på och vilken nationalitet man har (Fiedler et al. 2018). Människor som inte är i din ingrupp är per definition i din utgrupp (även om man kan ha olika attityder mot olika utgrupper).

Antalsokänslighet sker när vi föredrar att hjälpa färre ingruppsoffer framför fler utgruppsoffer (Fiedler et al. 2018). Denna ingruppsseffekt kan förklaras av att man har en positiv attityd till sin ingrupp, att man har en negativ attityd till sin utgrupp (Brewer 1999), eller på grund av att man upplever ett större ansvar att hjälpa sin ingrupp än sin utgrupp (Erlandsson, Björklund & Bäckström 2015; 2017).

EXISTERANDE OFFER

Nästan alla människor värdesätter ett liv som kan räddas här och nu högre, än ett liv som kan räddas i framtiden. Detta kan förklaras med en mer allmänmänsklig tendens att positiva händelser tappar i värde i takt med att tidpunkten för deras inträffande skjuts framåt (Samuelson 1937). Denna tendens kan förklara varför många föredrar att få 10 000 kr idag hellre än 20 000 kr om ett år, och varför inte människor pensionssparar mer. Denna tendens leder till dåliga beslut när det rör beslut om konsekvenser för mitt nuvarande eller mitt framtida jag, men till ännu sämre beslut när det rör beslut om konsekvenser för existerande eller framtida behövande människor.

De flesta människor skulle föredra att en beslutsfattare räddade ett oskyldigt barn som behövde hjälp nu, även om detta beslut skulle innebära att 2, 10 eller till och med 10 000 lika oskyldiga barn inte fick hjälp om 20 år. Man kan med lätthet påstå att denna tendens förklarar varför vi bryr oss mycket mer om problem som finns på jorden nu, jämfört med troligtvis mycket värre klimatrelaterade problem som kommer finnas inom några få generationer (Wade-Benzoni & Tost 2009).

EN HÖG RÄDDNINGSPROCENT

Ett räddningsprojekt som kan rädda 100 % av alla drabbade (t.ex. en räddningsaktion för att rädda barn instängda i en grotta) upplevs av de flesta som väldigt mycket bättre och viktigare än ett projekt som kan rädda 2 % av de drabbade. Detta kan vara rationellt men också leda till att man föredrar projekt som räddar färre snarare än fler personer om hjälpprojektet med låg räddningsprocent är ett väldigt storskaligt projekt (du kan rädda 2 % av de 300,000 sjuka barn som har cancertyp A), samtidigt som hjälpprojektet med hög räddningsprocent är ett väldigt småskaligt projekt (du kan rädda 100 % av de 12 barn med cancertyp B).

Denna effekt kallas proportionsdominans-effekten (Bartels & Burnett 2011; Fetherstonhaugh et al. 1997; Mata 2016), och i forskning med kolleger har jag funnit att den förklaras bäst av att människor ofta felaktigt upplever att småskaliga projekt med hög räddningsprocent är mer effektiva och gör mer nytta än storskaliga projekt med låg räddningsprocent (Erlandsson, Björklund & Bäckström 2014).

SPEKTAKULÄRA, NYA OCH FÖRVÄRRANDE PROBLEM

Människors uppmärksamhetsförmåga och minne är begränsat och detta gör att vi ofta är dåliga på att utvärdera allvarligheten av olika problem. Detta gör också att problem som är lättare att minnas på grund av att de är spektakulära (Epstein 2006), eller på grund av att de inträffat nyligen, kommer tilldelas oproportionerligt mycket resurser i hjälpdilemman. I en studie så lät man alla deltagare fördela en summa pengar mellan fyra olika hjälpprojekt (Huber et al. 2011). Hälften av deltagarna läste ett projekt i taget och efter varje projekt fick de uppge hur mycket av den kvarvarande summan de ville allokera till just det projektet. Den andra hälften tog del av alla fyra projekten samtidigt, och fick först därefter allokera summan mellan dem. Resultaten visade att deltagare som blev informerade om ett projekt i taget allokerade störst del av summan till projektet de läste om först och minst till det projekt de läste om sist (eftersom de hade minst pengar kvar då). Deltagare som läste om alla projekten innan de allokerade pengar gav mest till projektet som presenterades sist, eftersom det var det som var färskast i minnet när de ombads fördela medlen.

I en annan artikel (Erlandsson et al. 2018) undersökte jag tillsammans med norska kollegor i ett stort antal delstudier hur upplevda trender påverkade både upplevd allvarlighetsgrad av olika cancertyper och hur människor fördelade donationer till forskningsprojekt som fokuserade på att bekämpa dessa cancertyper. Vi fann att människor upplever att en cancertyp som har en något lägre dödsrisk men där dödsrisken tycks stiga (ex. dödsrisken för de som drabbades av Hodgkins lymfom var 8 % 2014 och 15 % 2016) uppfattas som mer allvarlig och allokeras en större del av resurserna, än en cancertyp som har en högre dödsrisk, men där risken tycks sjunka med tiden (ex. dödsrisken för de som drabbades av sarkom var 24 % 2014 och 17 % 2016).

INGEN OVERHEAD

Potentiella givares negativa attityder till overhead (dvs kostnader för att kunna samla in pengar till välgörenhet) är ett väldokumenterat problem inom den ideella sektorn. Trots att de flesta människor vet att overheadkostnader behövs för att kunna hjälpa effektivt och långsiktigt, så är tanken på att en del av de donerade pengarna går till reklam och chefslöner något som på ett känslomässigt plan gör många privata donatorer negativt inställda och omotiverade (Caviola et al. 2014; Newman et al. 2019; Portillo & Stinn 2018). Denna ”overhead-aversion” kan göra att givare hellre engagerar sig och ger pengar till småskaliga privata aktioner som t ex ”Vi gör vad vi kan”, ett initiativ som uppstod i samband med flyktingkrisen 2015 (Engvall 2015).

Gneezy, Keenan och Gneezy (2014) menar dock att det inte är overheadkostnaderna i sig som är problemet utan snarare tanken på att det är just mina pengar som försvinner på vägen som uppfattas negativt. I ett fältexperiment

skickade de ut fyra olika versioner av ett brev till potentiella givare. I ”bit på väg-versionen” läste mottagaren en text i stil med: *”2 miljoner kronor behövs för att nå vårt mål. Det positiva är att vi redan kommit en bra bit på väg. En miljon har redan donerats av en rik privatperson så det är bara hälften kvar tills vi når målet.”*

I ”dubblings-versionen” läste de istället att *”2 miljoner kronor behövs för att nå vårt mål. Det positiva är att din donation kommer vara dubbelt så mycket värd. För varje hundralapp som du donerar så kommer nämligen en rik privatperson att matcha din donation och också ge 100kr tills att vi har nått målet.”*

Till sist, i ”ingen overhead-versionen”, läste mottagarna: *”2 miljoner kronor behövs för att nå vårt mål. Det positiva är att alla overheadkostnader redan är täckta av en rik privatperson. Detta innebär att 100 % av det som du donerar kommer att gå till de behövande.”*

Resultaten visade att alla tre versionerna drog in mer pengar än kontrollversionen (som inte innehöll någon information om en rik privatperson) men att ”ingen overhead-versionen” drog in klart mest pengar (23 100 dollar jämfört med 13 200 för ”bit på väg-versionen”, 12 210 för ”dubblingsversionen” och 8 040 för kontrollversionen).

Prominenseffekten i hjälpdilemman

Flera av de attribut som listades ovan, ingick i en serie studier som jag och mina kollegor nyligen utförde för att undersöka hur olika sätt att uttrycka sig påverkar människors preferenser i hjälpdilemman (Erlandsson et al. 2020). I dessa studier undersökte vi även hur folk beslutar när de väljer mellan två hjälpprojekt som är exakt lika attraktiva enligt dem själva.

I en första fas så lät vi deltagare matcha två hjälpprojekt genom att skriva t ex hur många vuxna som Projekt A måste kunna rädda för att vara lika värdefullt som Projekt B, som kan rädda 100 barn och ungdomar. Som väntat värderade en majoritet av deltagarna barns liv högre än vuxnas liv och skrev därför att Projekt A måste kunna rädda t ex 150 vuxna för att vara lika värdefullt som det som kunde rädda 100 barn.

I ett senare läge (antingen en månad senare, eller direkt efteråt) så lät vi samma deltagare se de två projekt som vederbörande tidigare bedömt som lika attraktiva och uppgiften denna gång var att välja vilket av hjälpprojekten som skulle finansieras. Om deltagare verkligen upplevde att de valde mellan två lika värdefulla projekt så skulle de välja slumpmässigt, till exempel med hjälp av en tärning eller slumpgenerator, och de två projekten skulle bli valda ungefär lika ofta på gruppnivå. Resultaten visade att så inte var fallet. En klar majoritet (70–80 %) valde projektet som kunde rädda barn framför projektet som kunde rädda vuxna, trots att de två projekten tidigare hade ansetts vara lika attraktiva och att projektet som behandlade vuxna kunde rädda ett större antal.

Dilemman ovan med barn och vuxna var som sagt bara ett av flera dilemman som testades på samma sätt. Vi fann liknande resultat när vi jämförde ett projekt som kunde behandla patienter som rökte och drack alkohol mot ett projekt som kunde behandla patienter som åt nyttigt och motionerade (deltagare valde projektet som räddade "oskyldiga patienter" klart oftare trots att de två projekten skattats som lika attraktiva och att fler rökare kunde räddas). Även i ett dilemma, där existerande offer ställdes mot framtida offer, såg vi samma mönster. I matchningsuppgiften resonerade en majoritet av deltagarna så att existerande patienter värderades högre än framtida patienter. När samma personer sen skulle välja mellan två lika attraktiva hjälpprojekt, så valde en klar majoritet projektet som hjälpte färre existerande patienter än projektet som hjälpte fler framtida patienter.

Vår tolkning av dessa resultat är att barn, oskyldiga patienter samt "existerande" patienter är så kallade *prominenta attribut*, vilket innebär att de väger tyngre när man tvingas göra ett val än när man ombeds jämföra två alternativ. När man jämför (dvs matchar projekten) så påverkar både de *prominenta attributen* och antalet patienter (ett *icke-prominent attribut*) ens bedömning, men när man tvingas välja så ökar "viktigheten" hos de *prominenta attributen* och därför väljs det projekt som kan hjälpa barn, oskyldiga och existerande patienter oftare än vad det skulle göra om ens preferenser i matchningsuppgiften var stabila.

I två andra dilemman (kön och nationalitet) fann vi ett annat intressant mönster. Deltagare var här tydligt neutrala i matchningsuppgiften (över 90 % tyckte att 100 män är lika mycket värda som 100 kvinnor och över 75 % tyckte att 100 patienter från det egna landet är lika mycket värda som 100 patienter från ett annat land). Dock, när vi bad deltagare att välja mellan ett projekt som räddade 100 män eller 100 kvinnor, så valde inte människor slumpmässigt, utan föredrog i 85 % av fallen projektet som räddade kvinnor (effekten fanns hos både kvinnor och män). En nästan lika stor procent av deltagarna valde att rädda 100 landsmän framför att rädda 100 från ett annat land när de tvingades välja. Dessa resultat illustrerar tydligt att preferenser i hjälpdilemman påverkas av sättet som de uttrycks på. Preferenser som är dolda i matchningsuppgiften (där man kan uttrycka en neutral preferens) kan komma upp till ytan när människor tvingas välja.

Vår förklaring till detta är att människor i första hand vill uttrycka politiskt korrekta och försvarbara preferenser när de ställs inför hjälpdilemman. Majoriteten av deltagare uttrycker att ingrupp och utgrupp och män och kvinnor är lika värda när de kan uttrycka *indifferens* i matchningsuppgiften, eftersom detta är det mest försvarbara och politiskt korrekta. När de sen tvingas välja så gör de inte det slumpmässigt utan väljer istället det hjälpprojekt som är näst mest politiskt korrekt och näst lättast att försvara (liknande förklaringar ges av Chosen-Hillel, Shaw & Caruso 2015; Tetlock & Boettger 1994 och Capraro & Rand 2018).

Fyra andra forskningsfrågor om hjälpankets psykologi

Jag brukar ofta presentera mitt forskningsområde med hjälp av fem breda frågor som kan undersökas separat eller i kombination med varandra. Denna artikel har fokuserat på endast en av dessa fem, nämligen ”Hur beslutar människor när de måste fördela hjälp?” De andra fyra frågorna som avhandlas i min forskning om hjälpankets psykologi är:

Vem hjälper? Denna fråga fokuserar på vilket sätt individuella skillnader (till exempel kön, ålder, utbildning, inkomst, intelligens, personlighet, politisk åskådning eller världs bild) predicerar graden av hjälpande. Vilka slags människor hjälper mycket och vilka hjälper lite?

När hjälper vi? Denna fråga fokuserar på vilka situationer som leder till att människor hjälper mer eller mindre. Situationsfaktorer kan vara till exempel vilket typ av problem som man har framför sig (jordbävning eller svältkatastrof), karakteristika hos dem som behöver hjälp (barn eller vuxna), eller på vilket sätt som man ber om hjälp (en sorglig eller en hoppingivande välgörenhetsannons).

Varför hjälper vi? Denna fråga fokuserar på de psykologiska mekanismer som kan göra människor mer eller mindre motiverade att hjälpa andra. Psykologiska mekanismer definieras här som känslor, tankar och trosföreställningar, som är medvetna för den som ställs inför ett hjälpbeslut och som kan motivera (öka) eller demotivera (minska) hjälpande. Det finns många psykologiska mekanismer som påverkar graden av hjälpande, och nästan alltid är flera mekanismer inblandade samtidigt. Vissa psykologiska mekanismer som motiverar hjälpande räknas generellt som mer altruistiska (t.ex. empati och ett upplevt personligt ansvar), medan andra har tydligt egoistiska inslag (t.ex. att man vill slippa känna skuld känslor eller att man vill förbättra sitt sociala rykte).

Vad är konsekvenserna av att hjälpa? Denna fråga fokuserar på vad som händer efter ett hjälpbeslut och kan handla antingen om personen som utför hjälpbeteendet eller om personer som observerar hjälpbeteendet. Det har till exempel forskats på om människor blir lyckligare av att hjälpa (Dunn, Aknin & Norton 2008; 2014; Wang & Tong 2015), och om hjälpande vid tidpunkt 1 ökar eller minskar sannolikheten för att samma person hjälper vid tidpunkt 2 (Conway & Peetz 2012; Mazar & Zhong 2010). Det finns även forskning om vad människor tänker och känner när de ser andra som hjälper, och det är inte tydligt om de mest generösa donatorerna och de mest självupppoffrande moraliska hjältarna i första hand inspirerar eller irriterar andra med sitt beteende (Reinstein & Riener 2012; Tosi & Warmke 2016).

Förutom att dessa fyra frågor kan undersökas separat eller i kombination med varandra, så kan man även kombinera dem med vår fokusfråga om allokeringsbeslut. Nedan går jag igenom egen och andras forskning som undersökt dessa kombinationer.

1. INDIVIDUELLA SKILLNADER OCH ALLOKERINGSBESLUT

Olika grupper av människor tenderar att fördela resurser och hjälpa på olika sätt och ofta så speglar detta kulturella skillnader (t ex olika välfärdssystem; Vamstad & von Essen 2013) eller individuella skillnader (t ex personlighet och världsbild; Bennet 2003; van Lange et al. 2007; Winterich, Mittal & William 2009). I ett par artiklar undersökte vi om olika grupper av människor i Sverige tenderar att donera pengar till *olika* välgörenhetsorganisationer (Erlandsson et al. 2019; Nilsson, Erlandsson & Västfjäll, 2016; 2020; se också Micklewright & Schnepf 2009; Rajan, Pink & Dow 2009 och Wiepking 2010 för liknande studier i andra länder).

Det tydligaste resultatet vi fann var att människor som står till vänster politiskt, och människor som hjälper tiggare, är klart överrepresenterade när det handlar om att donera till organisationer som primärt jobbar för att hjälpa människor i andra länder (ex. Röda Korset, UNICEF, Rädda Barnen, Läkare Utan Gränser). Människor som står till höger politiskt och som är negativt inställda till tiggare är däremot mer motiverade att donera till organisationer som primärt jobbar för att hjälpa människor i Sverige (ex. Barncancerfonden, Cancerfonden, Hjärt-Lungfonden, Sjöräddningssällskapet).

Studierna illustrerade tydligt att vi lär oss mer om en person, om vi vet hur vederbörande fördelar donationer, än om vi vet hur mycket han eller hon donerar, och kan kopplas till forskning om att vänsterliberala och högerkonservativa baserar sina respektive moraluppfattningar på olika moraliska intuitioner eller "fundament" (Graham, Haidt & Nosek 2009). Moral för vänsterliberala baseras, enligt denna teori, framför allt på en aversion mot skadande och på rättvisa, medan moral för högerkonservativa dessutom baseras på lojalitet, respekt gentemot auktoriteter och en slags renhet/dygdighet. Vänster och högermänniskor skiljer sig lite åt när det handlar om mängden hjälpande, men de skiljer sig mycket åt när det handlar om hur de *fördelar* hjälp.

2. SITUATIONSSKILLNADER OCH ALLOKERINGSBESLUT

I en serie opublicerade studier utgick jag och två uppsatsstudenter från bergsbestigardilemman som presenterades ovan (Erlandsson, Berg & Dahlén 2012). I en studie varierade vi ordalydelsen av den sista meningen så att den antingen formulerades som: "Vad skulle du göra?", "Vad borde du göra?" eller "Vad är det moraliska valet". Vi fann tydliga indikationer på att majoriteten av människor hävdar att de *borde* rädda 5 okända och att detta även ses som det moraliska

valet, men att de som får frågan om vad de själva skulle göra, ganska unisont menar att de *skulle* rädda sin vän eller partner.

I två andra experiment undersökte vi hur svar i bergsbestigardilemmat (och varianter av det) ändras om deltagare tror att deras svar kommer att göras publika, jämfört med om de tror att svaren är anonyma. Vi fann tydliga tecken på att deltagare som skattade dilemmat publikt sa att de var än mer övertygade om att de skulle rädda sin romantiska partner, jämfört med deltagare som skattade dilemmat privat. När vi ändrade dilemmat så att det istället presenterades som att du måste välja mellan att rädda din partner eller en elev som du har ansvaret för, så var deltagare däremot mer övertygade om att de skulle rädda sin partner när svaren var anonyma, än när svaren var publika. Vår preliminära tolkning av resultaten är att det ses som mer socialt acceptabelt att lita på magkänslan (dvs att rädda sin romantiska partner) när alternativet är att maximera antalet räddade liv, än när alternativet är att uppfylla sin plikt.

I Erlandsson & Ivan (2015) testade vi om ett s k inducerat känslomindset eller kalkylativt mindset påverkade deltagares beslut både i donations-situationer och i allokeringssituationer. Mindsetet inducerades genom att deltagare ombads att skriva ett stycke text för hand i tron att deras handstil skulle analyseras. Hälften av deltagarna ombads att skriva en text om något som gjorde dem känslomässigt berörda (känslomindset), medan den andra hälften ombads att skriva en text om hur de räknar ut en restaurangnota (kalkyleringsmindset).

Vi fann att ett emotionellt mindset gjorde att folk i större utsträckning donerade pengar i donationssituationen, men också att folk allokerade mer till organisationer som hjälper få personer väldigt konkret och direkt (t.ex. Min Stora Dag), snarare än organisationer som hjälper många personer indirekt via ökad forskning (t.ex. Cancerfonden). Slutsatsen är att emotioner är bra för att få folk att donera mer, men även kan vara problematiskt, då de kan få oss att hjälpa färre snarare än fler i hjälpdilemman. I linje med dessa resultat har flera studier visat att människor ofta är beredda att allokera om resurser på sätt som gynnar enstaka offer som de känner empati för eller skuld känslor över, även när detta sker på bekostnad av hjälpanDET av en större grupp (Batson et al. 1995a; 1995b; 1999; Oceja 2008; Oceja & Jimenez 2007; de Hooge et al. 2011).

3. PSYKOLOGISKA MEKANISMER OCH ALLOKERINGSBESLUT

I en av mina avhandlingsartiklar (Erlandsson, Björklund & Bäckström 2017) undersökte jag tillsammans med mina handledare dels om tre olika hjälpeffekter existerar även när man testar dem som hjälpdilemman, dels hur folk rättfärdigar sina beslut när de fördelar pengar till olika projekt. Vi undersökte effekterna i en djurkontext och deltagare valde mellan projektet Rätts Uttrarna och projektet Rätts Igelkottarna, och dessa projekt formulerades på olika sätt för de olika hjälpeffekterna.

Vi fann stöd för proportionsdominanseffekten när den testades som ett hjälpdilemma, då en tydlig majoritet hellre valde att donera till projektet som kunde rädda 100 % av 30 djur än till det som kunde rädda 4 % av 800 djur, trots att det senare projektet innebar fler räddade djur. Detta motiverades framför allt med nyttoargument (ex. "Mina pengar kan göra en större skillnad där"). Vi fann också stöd för ingruppsseffekten, då en majoritet valde projektet som kunde rädda djur i Sverige (egna landet) framför projektet som kunde rädda lika många djur i Polen, och detta motiverades framförallt med ansvarsargument (ex. "Jag har en större skyldighet att hjälpa"). Vi fann dock inget stöd för den så kallade identifierbarhetseffekten när den testades med ett hjälpdilemma, då ungefär lika många valde projektet som räddade djur illustrerat med gulliga bilder, som projektet där det inte förekom bilder. Dock fann vi, att de som valde att hjälpa identifierade djur i högre grad motiverade detta med emotionsargument (ex. "Jag kände starkare medkänsla").

4. ATTITYDER GENTEMOT HJÄLPARE SOM UTFÖR OLIKA ALLOKERINGSBESLUT

Som vi tidigare sett så verkar det ofta som att människor som ställs inför hjälpdilemman tenderar att välja det alternativ som är lättast att försvara och motivera. Detta kan i sin tur innebära att människor först och främst vill att andra ska uppfatta dem som moraliska, och är beredda att anpassa sina beslut i hjälpdilemman för att uppnå detta (Capraro & Rand 2018). Inom den psykologiska forskningen kallas fenomenet att människor ändrar sitt beteende och sina beslut för att bli uppfattade som mer sympatiska för "impression management". Forskning visar att folk hjälper mer när de vet att de observeras av andra och att hjälpare i många situationer uppfattas som mer sympatiska än icke-hjälpare (Milinski, Semmann, & Krambeck 2002).

Samtidigt vill människor ge skenet av att de inte hjälper av "fel" skäl. Deltagare i en studie gav pengar till en välgörenhetsorganisation (genom att trycka på en knapp) i mindre grad när de blev observerade, om de visste att observatörerna visste att deltagarna själva också skulle tjäna pengar på knapptryckandet. Om deltagarna däremot visste att observatörerna trodde att endast hjälporganisationen tjänade på knapptryckandet, så hjälpte de mer om de blev observerade jämfört med när de inte blev det (Ariely, Bracha & Meyer 2009).

Man kan dock tänka sig att människor misstar sig angående hur deras hjälpande uppfattas av andra, och därför är det av intresse att undersöka hur människor faktiskt uppfattar andra som fattar beslut i hjälpdilemman. Detta brukar kallas för "impression formation". Det är till exempel en vanlig missuppfattning att tro att hjälpande och gillande är linjärt relaterade. Folk gillar varken de som hjälper ovanligt lite eller de som hjälper ovanligt mycket, eftersom det senare kan få andra att verka snåla och ohjälpsamma i jämförelse, vilket påverkar deras självbild och rykte (Klien & Epley 2014; Parks & Stone 2010).

När det kommer till attityder gentemot resursfördelare så finns det t.ex. mycket som tyder på att en person som fördelar på så sätt att det finns en tydlig förlorare ($A = 49$, $B = 49$, $C = 2$) uppfattas som klart mer osympatisk än en som fördelar så att det finns en tydlig vinnare ($A = 98$, $B = 2$, $C = 2$), trots att skillnaden mellan den som får mest och den som får minst är mindre i det senare fallet (Eriksson & Simpson 2011).

I en studie undersökte jag och kollegor hur olika hjälpare uppfattades när vi manipulerade både hur mycket de hjälpte (t ex donerar 100kr/månaden eller 600kr/månaden) och deras olika sätt att hjälpa (Erlandsson, Wingren & Andersson 2020). Ett av dessa sätt var om man fokuserar (ger till endast en av organisationerna som ber om hjälp) eller diversifierar sitt hjälpande (ger till alla organisationer som ber om hjälp). I två studier fann vi att en hjälpare som diversifierar en mindre summa mellan alla organisationer uppfattas som lika sympatisk som en person som fokuserar en sex gånger större donation till en enda hjälporganisation, när de båda hjälparna bedömdes var för sig.

I en uppföljningsstudie som vi gjorde nyligen fann vi dock att om man ställde de två hjälparna bredvid varandra (så att de enkelt kunde jämföras) så uppfattades hjälparen som gav en större summa till endast en organisation som mer sympatisk. I samma studie fann vi, att när hjälpare bedöms separat så uppfattas den som minskar sitt givande, från 100 till 90 dollar, som mindre sympatisk än den som ökar hjälpanDET från 10 till 15 dollar. Men när man bedömer båda hjälparna samtidigt anses den som ger 90 dollar som mer sympatisk.

Sammanfattande normativ kommentar: Är vi, och bör vi vara, konsekvensetiker, regeletiker eller dygdetiker?

Tidigare i denna artikel diskuterades huruvida effektivitet *eller* rättvisa är det viktigaste vid hjälpdilemman. Med andra ord, är människor överlag konsekvensetiker (utilitarister) eller rättvise/regeletiker när de fattar beslut i hjälpsituationer? Jag (och andra) menar att det närmaste svaret på frågan är "varken eller".

Konsekvensetiken och regeletiken har länge dominerat normativ moralfilosofi men en tredje moralfilosofisk ansats är dygdetiken. Denna skiljer sig från både konsekvens- och rättvise-/regeletiken genom att det moraliska fokuset inte är på människors handlingar utan snarare på människors karaktärsdrag. Det viktiga enligt en normativ dygdetiker är inte om en specifik handling är "rätt" eller "fel", utan snarare om en specifik människa är "god" eller "ond".

Om vi rör oss från den normativa frågan (hur bör människor tänka/agera i moraliska situationer?) till den deskriptiva (hur tänker/agerar människor i moraliska situationer?), så finns det mycket stöd för att människor är intuitiva dygdetiker. Enligt teorin "The Person-Centered Morality" (Uhlmann, Pizarro

& Diermeier 2015; Landy & Uhlmann 2018), är målet med moralisk kognition oftast inte att bedöma handlingar utan snarare att bedöma människor, eftersom det senare är viktigare för vår överlevnad. Detta innebär också att utförandet av den mest moraliska handlingen kan innebära att personen som utför den verkar osympatisk.

I en studie bedömde deltagare att den *moraliskt korrekta handlingen* var att offra en skadad person för att de många övriga personerna i livbåten skulle överleva, men samtidigt så upplevdes de som faktiskt slängde den skadade personen överbord ha sämre *moralisk karaktär* (Uhlmann, Zhu & Tannenbaum 2013). Flera av de studier som jag har tagit upp här ger stöd för att den personcentrerade moralpsykologiska teorin förklarar beteende, inte bara i traditionella dilemman utan även i hjälpdilemman. De flesta människor tänker, känner och betar sig inte i linje med vare sig rättvis/regeletiken eller konsekvensetiken när de ställs inför hjälpbeslut, utan snarare i linje med dygdetiken.

Det är dock viktigt att komma ihåg att denna slutsats inte utgör ett argument för att den normativa dygdetiken är att föredra. Det är inget inkonsekvent med att å ena sidan acceptera att människor de facto oftast följer sin magkänsla och fördelar hjälp på sätt som får dem att må bra eller verka sympatiska, men å andra sidan arbeta för att människor ska basera sina hjälpbeslut mer på konsekvenser och antal, och mindre på känslor och identitetsskapande. Just detta är också min personliga inställning.

Referenser

- Ariely, D., Bracha, A. & Meier, S., 2009. "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially", *American Economic Review*, 99, s. 544-55.
- Back, S. & Lips, H. M., 1998. "Child sexual abuse: Victim age, victim gender, and observer gender as factors contributing to attributions of responsibility", *Child Abuse & Neglect*, 22, s. 1239-1252.
- Baron, J., 1997. "Biases in the quantitative measurement of values for public decisions." *Psychological Bulletin*, 122, s. 72-88.
- Baron, J. 2008. *Thinking and deciding (4th ed.)*. New York: Cambridge University Press.
- Baron, J. & Ritov, I., 2009. *Protected values and omission bias as deontological judgments. Moral judgment and decision making*. San Diego: Elsevier Academic Press.
- Baron, J. & Szymanska, E., 2011. "Heuristics and biases in charity", i Oppenheimer, D. M. & Olivola, C. Y. (red.), *The Science of Giving: Experimental Approaches to the Study of Charity*. New York: Psychology Press.
- Bartels, D. M. & Burnett, R. C., 2011. "A group construal account of drop-in-the-bucket thinking in policy preference and moral judgment", *Journal of Experimental Social Psychology*, 47, s. 50-57.

- Batson, C. D., Ahmad, N., Yin, J., Bedell, S. J., Johnson, J. W. & Templin, C. M., 1999. "Two threats to the common good: Self-interested egoism and empathy-induced altruism", *Personality and Social Psychology Bulletin*, 25, s. 3-16.
- Batson, C. D., Batson, J. G., Todd, R. M., Brummett, B. H., Shaw, L. L. & Aldeguer, C. M. R., 1995. "Empathy and the collective good: Caring for one of the others in a social dilemma", *Journal of Personality and Social Psychology*, 68, s. 619-631.
- Batson, C. D., Klein, T. R., Highberger, L. & Shaw, L. L., 1995. "Immorality from empathy-induced altruism: When compassion and justice conflict", *Journal of Personality and Social Psychology*, 68, s. 1042-1054.
- Bennett, R., 2003. "Factors underlying the inclination to donate to particular types of charity", *International Journal of Nonprofit and Voluntary Sector Marketing*, 8, s. 12-29.
- Berman, J. Z., Barasch, A., Levine, E. E. & Small, D. A., 2018. "Impediments to effective altruism: The role of subjective preferences in charitable giving", *Psychological science*, 29, s. 834-844.
- Bravo Vergel, Y. & Sculpher, M., 2008. "Quality-adjusted life years", *Practical Neurology*, 8, s. 175-182.
- Breeze, B., 2013. "How donors choose charities: The role of personal taste and experiences in giving decisions", *Voluntary Sector Review*, 4, s. 165-183.
- Brewer, M. B., 1999. "The psychology of prejudice: Ingroup love and outgroup hate?", *Journal of Social Issues*, 55, s. 429-444.
- Capraro, V. & Rand, D. G., 2018. "Do the right thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality", *Judgment & Decision Making*, 13, s. 99-111.
- Caviola, L., Faulmüller, N., Everett, J., Savulescu, J. & Kahane, G., 2014. "The evaluability bias in charitable giving: Saving administration costs or saving lives?", *Judgment and Decision Making*, 9, s. 303-315.
- Choshen-Hillel, S., Shaw, A. & Caruso, E. M., 2015. "Waste management: How reducing partiality can promote efficient resource allocation", *Journal of Personality and Social Psychology*, 109, s. 210-231.
- Christensen, J. F. & Gomila, A., 2012. "Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review", *Neuroscience and Biobehavioral Reviews*, 36, s. 1249-1264.
- Conway, P. & Peetz, J., 2012. "When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior", *Personality and Social Psychology Bulletin*, 38, s. 907-919.
- Curry, T. R., Lee, G. & Rodriguez, S. F., 2004. "Does victim gender increase sentence severity? Further explorations of gender dynamics and sentencing outcomes", *Crime & Delinquency*, 50, s. 319-343.
- De Hooge, I. E., Nelissen, R. M. A., Breugelmans, S. M. & Zeelenberg, M., 2011. "What is moral about guilt? Acting 'prosocially' at the disadvantage of others", *Journal of Personality and Social Psychology*, 100, s. 462-473.
- Dickert, S., Sagara, N. & Slovic, P., 2011. "Affective motivations to help others: A two-stage model of donation decisions", *Journal of Behavioral Decision Making*, 24, s. 361-376.
- Dickert, S. & Slovic, P., 2011. "Unstable values in lifesaving decisions", *Frontiers in Psychology*, 2, <https://doi.org/10.3389/fpsyg.2011.00294>.

- Dickert, S., Västfjäll, D., Kleber, J. & Slovic, P., 2015. "Scope insensitivity: The limits of intuitive valuation of human lives in public policy", *Journal of Applied Research in Memory and Cognition*, 4, s. 248-255.
- Dunn, E. W., Aknin, L. B. & Norton, M. I., 2008. "Spending money on others promotes happiness", *Science*, 319, s. 1687-1688.
- Dunn, E. W., Aknin, L. B. & Norton, M. I., 2014. "Prosocial spending and happiness: Using money to benefit others pays off", *Current Directions in Psychological Science*, 23, s. 41-47.
- Edlin, R., Tsuchuya, A. & Dolan, P., 2012. "Public preferences for responsibility versus public preferences for reducing inequalities", *Health Economics*, 21, s. 1416-1426.
- Engvall, C., 2015. "Vi gör vad vi kan: En del av en folkrörelse", *Patos*. <https://www.givasverige.se/patos/vi-gor-vad-vi-kan-en-del-av-en-folkrorelse/>.
- Epstein, K., 2006. "Crisis mentality: why sudden emergencies attract more funds than do chronic conditions, and how nonprofits can change that", *Stanford Social Innovation Review*, 4, s. 48-57.
- Eriksson, K. & Simpson, B., 2011. "Perceptions of unfairness in allocations between multiple recipients", *Cognitive Psychology*, 62, s. 225-244.
- Erlandsson, A., Berg, E. & Dahlén, L., 2012. "Impression formation of intuitive and utilitarian helping", *International Society of Utilitarian Studies*. New York.
- Erlandsson, A., Björklund, F. & Bäckström, M., 2014. "Perceived utility (not sympathy) mediates the proportion dominance effect in helping decisions", *Journal of Behavioral Decision Making*, 27, s. 37-47.
- Erlandsson, A., Björklund, F. & Bäckström, M., 2015. "Emotional reactions, perceived impact and perceived responsibility mediate the identifiable victim effect, proportion dominance effect and in-group effect respectively", *Organizational Behavior and Human Decision Processes*, 127, s. 1-14.
- Erlandsson, A., Björklund, F. & Bäckström, M., 2017. "Choice-justifications after allocating resources in helping dilemmas", *Judgment and Decision Making*, 12, s. 60-80.
- Erlandsson, A., Hohle, S. M., Löhre, E. & Västfjäll, D., 2018. "The rise and fall of scary numbers: The effect of perceived trends on future estimates, severity ratings, and help-allocations in a cancer context", *Journal of Applied Social Psychology*, 48, s. 618-633.
- Erlandsson, A. & Ivan, C.-E., 2015. "Feeling or calculating in helping situations: Induced mindset influence both helping intentions and money allocations", *Society of Judgment and Decision Making*. Chicago.
- Erlandsson, A., Lindkvist, A., Lundqvist, K., Andersson, P. A., Dickert, S., Slovic, P. & Västfjäll, D., 2020. "Moral preferences in helping dilemmas expressed by matching and forced choice". Manuscript accepted by *Judgment and Decision Making*.
- Erlandsson, A., Nilsson, A., Tinghög, G., Andersson, D. & Västfjäll, D., 2019. "Donations to outgroup charities, but not ingroup charities, predict helping intentions toward street-beggars in Sweden", *Nonprofit and Voluntary Sector Quarterly*, 48(4), s. 814-838.
- Erlandsson, A., Wingren, M. & Andersson, P. A., 2020. "Type and amount of help as predictors for impressions of helpers". Manuscript accepted by *Plos One*.
- Fetherstonhaugh, D., Slovic, P., Johnson, S. & Friedrich, J., 1997. "Insensitivity to the value of human life: A study of psychophysical numbing", *Journal of Risk and Uncertainty*, 14, s. 283-300.

- Fiedler, S., Hellmann, D. M., Dorrrough, A. R. & Glöckner, A., 2018. "Cross-national in-group favoritism in prosocial behavior: Evidence from Latin and North America", *Judgment & Decision Making*, 13, s. 42-60.
- Gabriel, I., 2017. "Effective altruism and its critics", *Journal of Applied Philosophy*, 34, s. 457-473.
- Gneezy, U., Keenan, E. A. & Gneezy, A., 2014. "Avoiding overhead aversion in charity", *Science*, 346, s. 632-635.
- Goodwin, G. P. & Landy, J. F., 2014. "Valuing different human lives", *Journal of Experimental Psychology: General*, 143, s. 778-803.
- Gordon-Hecker T., Choshen-Hillel S., Shalvi S., Bereby-Meyer Y., 2017. "Resource allocation decisions: When do we sacrifice efficiency in the name of equity?" i Li M., Tracer D. (eds.), *Interdisciplinary perspectives on fairness, equity, and justice*. Springer.
- Gordon-Hecker, T., Rosensaft-Eshel, D., Pittarello, A., Shalvi, S. & Bereby-Meyer, Y., 2017. "Not taking responsibility: Equity trumps efficiency in allocation decisions", *Journal of Experimental Psychology: General*, 146, s. 771-775.
- Greene, J. D., 2007. "Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains", *Trends in Cognitive Sciences*, 11, s. 322-323.
- Greene, J. D., 2008. "The secret joke of Kant's soul" i Sinnott-Armstrong, W. (ed.), *Moral psychology. Vol 3: The neuroscience of morality: Emotion, brain disorders, and development*. Cambridge: MIT Press.
- Huber, M., Van Boven, L., McGraw, A. P. & Johnson-Graham, L., 2011. "Whom to help? Immediacy bias in judgments and decisions about humanitarian aid", *Organizational Behavior and Human Decision Processes*, 115, s. 283-293.
- Landy, J. F. & Uhlmann, E. L., 2018. "Morality is personal" i J. Graham & K. Gray (eds), *The Atlas of moral psychology: Mapping good and evil in the mind*. Guilford Press.
- Macaskill, W., 2015. *Doing good better: Effective altruism and a radical new way to make a difference*. Guardian Faber Publishing.
- Mata, A., 2016. "Proportion dominance in valuing lives: The role of deliberative thinking", *Judgment and Decision Making*, 11, s. 441-448.
- Mazar, N. & Zhong, C.-B., 2010. "Do green products make us better people?", *Psychological Science*, 21, s. 494-498.
- McMahon, J., 2016. "Philosophical critiques of effective altruism", *The Philosophers' Magazine*, s. 92-99.
- Micklewright, J. & Schnepf, S. V., 2009. "Who gives charitable donations for overseas development?", *Journal of Social Policy*, 38, s. 317-341.
- Milinski, M., Semmann, D. & Krambeck, H., 2002. "Donors to charity gain in both indirect reciprocity and political reputation", *Proceedings of the Royal Society of London B: Biological Sciences*, 269, s. 881-883.
- Newman, G. E., Shniderman, A., Cain, D. M. & Sevel, K., 2019. "Do the ends justify the means? The relative focus on overhead versus outcomes in charitable fundraising", *Nonprofit and Voluntary Sector Quarterly*, 48(1), s. 71-90.
- Nilsson, A., Erlandsson, A. & Västfjäll, D., 2016. "The congruency between moral foundations and intentions to donate, self-reported donations, and actual donations to charity", *Journal of Research in Personality*, 65, s. 22-29.
- Nilsson, A., Erlandsson, A., & Västfjäll, D., 2020. "Moral foundations theory and the psychology of charitable giving", *European Journal of Personality*, 34, s. 431-447.
- Oceja, L., 2008. "Overcoming empathy-induced partiality: Two rules of thumb", *Basic and Applied Social Psychology*, 30, s. 176-182.

- Oceja, L. & Jiménez, I., 2007. "Beyond egoism and group identity: Empathy toward the other and awareness of others in a social dilemma", *The Spanish Journal of Psychology*, 10, s. 369-379.
- Ord, T., 2013. "The moral imperative toward cost-effectiveness in global health", *Center for Global Development* [Online]. Available: www.cgdev.org/content/publications/detail/1427016.
- Popp, F., Eggert, N., Hoy, L., Lang, S., Obed, A., Piso, P., Schlitt, H. & Dahlke, M., 2006. "Who is willing to take the risk? Assessing the readiness for living liver donation in the general German population", *Journal of Medical Ethics*, 32, s. 389-894.
- Portillo, J. E. & Stinn, J., 2018. "Overhead aversion: Do some types of overhead matter more than others?", *Journal of Behavioral and Experimental Economics*, 72, s. 40-50.
- Raihani, N. J. & Smith, S., 2015. "Competitive helping in online giving", *Current Biology*, 25, s. 1183-1186.
- Rajan, S. S., Pink, G. H. & Dow, W. H., 2009. "Sociodemographic and personality characteristics of canadian donors contributing to international charity", *Nonprofit and Voluntary Sector Quarterly*, 38, s. 413-440.
- Reinstein, D. & Riener, G., 2012. "Reputation and influence in charitable giving: An experiment", *Theory and Decision*, 72, s. 221.
- Samuelson, P. A., 1937. "A note on measurement of utility", *The Review of Economic Studies*, 4, s. 155-161.
- Singer, P., 2009. *The life you can save: Acting now to end world poverty*, New York, Random House.
- Singer, P., 2015. "The logic of effective altruism", *Boston Review*. <http://bostonreview.net/forum/peter-singer-logic-effective-altruism>.
- Slovic, P., 2007. "If I look at the mass I will never act: Psychic numbing and genocide", *Judgment and Decision Making*, 2, s. 79-95.
- Soyer, E. & Hogarth, R. M., 2011. "The size and distribution of donations: Effects of number of recipients", *Judgment and Decision Making*, 6, s. 616-628.
- Taurek, J. M., 1977. "Should the numbers count?", *Philosophy & Public Affairs*, s. 293-316.
- Tetlock, P. E. & Boettger, R., 1994. "Accountability amplifies the status quo effect when change creates victims", *Journal of Behavioral Decision Making*, 7, s. 1-23.
- Thaler, R. H. & Sunstein, C., 2009. *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin Books.
- Tosi, J. & Warmke, B., 2016. "Moral Grandstanding", *Philosophy & Public Affairs*, 44, s. 197-217.
- Tsuchiya, A., Dolan, P. & Shaw, R., 2003. "Measuring people's preferences regarding ageism in health: some methodological issues and some fresh evidence", *Social Science & Medicine*, 57, s. 687-696.
- Ubel, P. A., Dekay, M. L., Baron, J. & Asch, D. A., 1996. "Cost-effectiveness analysis in a setting of budget constraints—is it equitable?", *New England Journal of Medicine*, 334, s. 1174-1177.
- Uhlmann, E. L., Pizarro, D. A. & Diermeier, D., 2015. "A person-centered approach to moral judgment", *Perspectives on Psychological Science*, 10, s. 72-81.
- Uhlmann, E. L., Zhu, L. L. & Tannenbaum, D., 2013. "When it takes a bad person to do the right thing", *Cognition*, 126, s. 326-334.

- Wade-Benzoni, K. A. & Tost, L. P., 2009. "The egoism and altruism of intergenerational behavior", *Personality and Social Psychology Review*, 13, s. 165-193.
- Vamstad, J. & Von Essen, J., 2013. "Charitable giving in a universal welfare state – Charity and social rights in Sweden", *Nonprofit and Voluntary Sector Quarterly*, 42, s. 285-301.
- Van Lange, P. A. M., Bekkers, R., Schuyt, T. N. M. & Van Vugt, M., 2007. "From games to giving: Social value orientation predicts donations to noble causes", *Basic and Applied Social Psychology*, 29, s. 375-384.
- Van Vugt, M. & Iredale, W., 2013. "Men behaving nicely: Public goods as peacock tails", *British Journal of Psychology*, 104, s. 3-13.
- Wang, X. & Tong, L., 2015. "Hide the light or let it shine? Examining the factors influencing the effect of publicizing donations on donors' happiness", *International Journal of Research in Marketing*, 32, s. 418-424.
- Wiepking, P., 2010. "Democrats support international relief and the upper class donates to art? How opportunity, incentives and confidence affect donations to different types of charitable organizations", *Social Science Research*, 39, s. 1073-1087.
- Winterich, K. P., Mittal, V. & Ross Jr, W. T., 2009. "Donation behavior toward in-groups and out-groups: The role of gender and moral identity", *Journal of Consumer Research*, 36, s. 199-214.
- Zagefka, H., Noor, M., Brown, R., De Moura, G. R. & Hothrow, T., 2011. "Donating to disaster victims: Responses to natural and humanly caused events", *European Journal of Social Psychology*, 41, s. 353-363.

Spelteoretisk jämvikt som ett resultat av inläring eller evolution

Erik Mohlin

Abstract

One way of justifying the use of equilibrium concepts in game theory is by appeal to learning or evolution. This paper provides a non-technical introduction to evolutionary game theory. A short summary of basic standard (non-evolutionary) game theory sets the stage and introduces a number of simple games. Evolutionary game theory is then introduced via a simple learning rule, according to which individuals occasionally revise their strategy choices by imitating currently successful peers. The resulting dynamic is studied in the previously introduced games, investigating whether behaviour converges to an equilibrium, and if so, which one. Finally, some empirical evidence is discussed.

Spelteori beskrivs ofta som en teori om perfekt rationella aktörer vars beteende är i jämvikt. Var och en fattar ett beslut som är *rationellt* i den mening att det är optimalt utifrån respektive beslutsfattarens målsättning och vad de *tror* att andra kommer göra. Teoretiskt fångas beslutsfattarens målsättning i en så kallad nyttofunktion; rationalitet består i att välja det alternativ som maximerar den förväntade nyttan givet beslutsfattarens subjektiva probabilistiska förväntningar om hur andra betar sig. I en *jämviktsanalys* måste vi dessutom förut-sätta att vars och ens förväntningar om vad motparten kommer att göra visar sig vara fullständigt korrekta. Därmed fattar var och en ett optimalt (nyttomaximerande) beslut givet sin målsättning och givet vad motparterna *faktiskt* gör.

Det finns definitivt situationer där aktörerna fattar övertänkta beslut som kännetecknas av en mycket hög grad av rationalitet. Det finns till och med sällsynta fall där man kan motivera de korrekta jämviktsförväntningarna med

Erik Mohlin är docent i nationalekonomi vid Lunds universitet och Institutet för Framtidsstudier.
E-post: erik.mohlin@nek.lu.se

hänvisning till rent logiska resonemang om vad gemensam kunskap om rationalitet implicerar för vilka beteenden man kan förvänta sig av sin motpart.¹ I allmänhet verkar det dock rimligare att tänka sig att rationellt beslutsfattande är erfarenhetsbaserat, dvs att beslutsfattaren haft tid och möjlighet att observera, pröva och utvärdera olika handlingsalternativ. Framförallt verkar det svårt att se hur aktörerna skulle kunna skaffa sig korrekta förväntningar om vad andra kommer att göra ifall de inte har erfarenhet av interaktionen - och andras beteende i den.

Lyckligtvis är det inte så att spelteorin behöver börja och sluta med perfekt rationalitet och jämvikt. Det finns en omfattande teoretisk och experimentell forskning om hur inläring (exempelvis i form av upprepade ”försök och misslag”) och gradvis anpassning av beteende (i ljuset av vad man lär sig om andras beteende och handlingars konsekvenser) kan leda till jämviktstillstånd där alla inblandade beter sig på ett sätt som är rationellt (nyttomaximerande) för var och en, givet de andras beteende.² I föreliggande uppsats ger jag en kort introduktion till detta forskningsfält. Jag kommer mestadels att hålla mig på ett teoretiskt plan, men även ge hänvisningar till experimentella studier som testat teorin.³ Det matematiska inslaget hålls på ett absolut minimum, även om en del resonemang utvecklas i ett matematiskt appendix. Mitt mål är att texten skall kunna läsas även av dem som på sin höjd såg några spelmatriser i en kurs för många år sedan. Därför följer en kort repetition av klassisk spelteori.

Lite klassisk spelteori

SPELARE, STRATEGIER OCH NYTTOR

Spelteorin är ett verktyg för att analysera interaktiva beslutssituationer, det vill säga situationer där konsekvenserna av en persons beslut även beror på andra personers beslut. Namnet till trots är teorin inte begränsad till att handla om spel i vardaglig mening, som poker (även om också sådana spel kan analyseras med hjälp av spelteori). Tvärtom är spelteorins räckvidd mycket stor. Teorin kan i princip användas till att studera alla sociala interaktioner – från

1 En strategi i ett spel sägs vara dominerad om den ger en lägre nytta än en annan strategi, oavsett vad de andra spelarna gör. En rationell spelare kommer naturligtvis aldrig att välja en dominerad strategi. En spelare som vet att motparten är rationell kan sluta sig till att motparten aldrig väljer en dominerad strategi. En spelare som dessutom vet att de andra spelarna vet att han eller hon är rationell kan eventuellt dra ytterligare slutsatser om vilka strategier de andra spelarna kommer undvika. Vi kan fortsätta resonemanget för godtyckligt komplexa hierarkier om kunskap om andras kunskap. För mig framstår detta angreppssätt som en fullständigt orealistisk väg att rättfärdiga jämvikt.

2 Redan i sin doktorsavhandling (Nash 1950) föreslog Nash en evolutionär tolkning av det jämviktsbegrepp som senare skulle komma att bära hans namn.

3 Den experimentella litteraturen om inläring i spel sammanfattas väl av Camerer (2003), även om en del har hänt sedan den kom ut.

marknadsutbyten och förhandlingar till krig och samarbete, liksom interaktioner mellan djur.⁴

I spelteorin kallas en interaktiv situation för ett spel och deltagarna i interaktionen kallas spelare. Spelarna väljer var sin strategi. Varje spelares utfall beror på vad alla andra spelare gör. Spelarnas drivkrafter representeras av nyttofunktioner som anger hur en spelare värderar vart och ett av de möjliga utfallen. Närmare bestämt antas i allmänhet von Neumann-Morgenstern nyttor, vilken innebär att tilldelningen av nyttor till de olika utfallen är sådan att spelarna strävar efter att maximera förväntad nytta.⁵

Om du exempelvis är ute och kör bil kan strategierna vara dina olika färdvägar och spelarna kan utgöras av alla andra trafikanter som är ute och rör sig samtidigt som du. Din nytta kan tänkas bero på om du kan undvika köer och slippa olyckor. Dina möjligheter att komma fram fort och säkert beror på vilka färdvägar andra trafikanter väljer. På samma sätt beror dina medtrafikanters utfall på hur du betar dig.

Ett vanligt missförstånd är att tolka nytta som det används i vardagsspråket, kanske i termer av materiella nyttigheter. Nytt skall istället ses som en teknisk term som representerar vad helst spelaren värderar och motiveras av. Spelteorin tillåter såväl altruistiska och idealistiska som själviska och materialistiska drivkrafter. Med detta sagt måste det tilläggas att när man försöker testa spelteori i laboratorieexperiment är det ofta bekvämt att motivera deltagarna med monetära belöningar. När man tolkar resultaten från sådana experiment måste man naturligtvis vara medveten om att de subjektiva nyttorna, eller inre belöningarna, kan skilja sig systematiskt från de objektiva materiella belöningarna.

För att förstå eller förutsäga vad som händer i sådana här interaktioner använder spelteorin vanligen olika jämviktsbegrepp. Det vanligaste och mest grundläggande jämviktsbegreppet inom spelteorin är Nashjämvikt. En Nashjämvikt är en uppsättning strategier (en strategi per spelare), sådana att ingen spelare kan uppnå en högre nytta genom att ensidigt byta strategi, det vill säga genom att själv byta strategi samtidigt som de andra spelarna inte byter strategier.

Låt oss nu konkretisera detta abstrakta ramverk genom att studera två enkla (men viktiga) exempel på spel. Det så kallade "Hök och Duva"-spelet beskriver en konfliktsituation med delvis motstridiga intressen. Spelet "Hjortjakten" beskriver istället en situation med sammanfallande intressen.

4 Huruvida analysen är fruktbar är en annan fråga som bör bedömas från fall till fall.

5 Man kan visa matematiskt att om en individs preferenser rörande de olika utfallen i spelet uppfyller vissa axiomer, så finns det någon nyttofunktion som är sådan att individerna betar sig som om de maximerade förväntad nytta.

KONFLIKTSPEL: HÖK OCH DUVA

Tänk dig två länder som gör anspråk på en obebodd ö. På ön finns naturresurser till ett värde av 6. Varje land kan välja om det ska följa en aggressiv strategi för att lägga beslag på hela ön, något som vi kan kalla en hökstrategi, eller följa en mera generös strategi som syftar till fredlig samexistens på ön, något som vi kallar för en duvstrategi. Om båda länder väljer att bete sig som hökar så utbryter en konflikt som minskar öns värde med 8, och i slutändan får båda länder dela på ön, så att vardera landet får nytta $-1 = (6 - 8) / 2$. Om ett land väljer hökstrategin och det andra landet väljer duvstrategin, får höklandet hela ön medan duvlandet får ingenting. Om båda länder följer duvstrategin så delar de på ön utan att behöva genomlida konflikten som minskar värdet på ön. Vi kan sammanfatta detta med matrisen i figur 1.

Den första siffran i varje ruta i figur 1 anger nyttan för radspelaren, den spelare som väljer mellan strategierna som representeras av raderna. Den andra siffran i varje ruta anger nyttan för kolumnspelaren, den spelare som väljer mellan strategierna som representeras av kolumnerna.

Figur 1. Konfliktspel: hök och duva.

	Hök	Duva
Hök	-1, -1	6, 0
Duva	0, 6	3, 3

En alternativ tolkning av matrisen i figur 1 är att det handlar om två spelare som är inbegripna i en förhandling. Det skulle kanske kunna handla om en arbetsgivare och en arbetstagare som förhandlar om hur mycket av vinsten som ska tillfalla respektive part. Om konflikt uppstår (i form av strejk eller lock-out) så förlorar båda sidor på det. Man kan också tolka spelet som en förhandling mellan två företag.⁶

Oavsett vilken tolkning vi gör visar matrisen att om motspelaren spelar hök så är det rationellt att spela duva. Att spela duva när motparten spelar hök ger visserligen bara nyttan 0, men att spela hök när motparten spelar hök ger en ännu lägre nytta, nämligen -1. På motsvarande sätt är det, om motspelaren spelar duva, rationellt att spela hök eftersom man då får 6 medan man bara skulle fått 3 ifall man spelade duva. Det betyder att både kombinationen duva-hök och kombinationen hök-duva är jämvikter.

Man kan fråga sig hur de två kontrahenterna bestämmer sig för vem som ska spela duva och vem som ska spela hök. Båda spelare föredrar ju den jämvikt

6 Rent allmänt är det naturligtvis så att en och samma spelmatris kan tolkas som en beskrivning av många olika situationer.

där de själva spelar hög och motparten spelar duva. Går det att finna en annan jämvikt?

Hittills har vår diskussion utgått från att varje spelare måste bestämma sig för en strategi. Det verkar dock rimligt att tillåta att spelare kan välja mellan sina handlingsalternativ på ett mera slumpmässigt sätt och därmed välja olika alternativ med olika sannolikhet. En spelare som utför flera handlingar med positiv sannolikhet sägs följa en blandad strategi. Motsatsen, att spela en enda handling med sannolikheten ett, kallas för att följa en ren strategi. För att förstå hur blandade strategier fungerar kan det vara bra att tänka på straffsparkar i fotboll. En framgångsrik målvakt slänger sig inte alltid åt samma håll och en framgångsrik straffläggare skjuter inte alltid åt samma håll. Just möjligheten att blanda sin strategi är avgörande i sammanhanget.

Låt oss anta att kolumnspelaren följer en blandad strategi som slumpmässigt alternerar mellan två rena strategier, så att vederbörande väljer högalternativet med sannolikheten q och duvalternativet med sannolikheten $1 - q$. Radspelarens förväntade nytta av att bete sig som hög blir då

$$U_{\text{Hök}} = (-1) \cdot q + 6 \cdot (1 - q) = 6 - 7q$$

och radspelarens förväntade nytta av att bete sig som duva blir

$$U_{\text{Duva}} = 0 \cdot q + 3 \cdot (1 - q) = 3 - 3q.$$

Notera att de båda handlingsalternativen ger samma förväntade nytta om och endast om $6 - 7q = 3 - 3q$, vilket är fallet om och endast om $q = 0,75$. Det betyder att ifall kolumnspelaren slumpmässigt väljer mellan rena strategier på ett sådant sätt att man väljer hög med sannolikheten $q = 0,75$, då kan radspelaren inte förbättra sin förväntade nytta genom att ändra sitt beteende; skälet är att båda strategier ger exakt samma förväntade nytta.

Situationen är likadan för kolumnspelaren. Anta att kolumnspelaren slumpmässigt väljer mellan de två alternativen hög och duva, så att hon agerar som hög med sannolikheten p och som duva med sannolikheten $1 - p$. Om radspelaren väljer hög med sannolikheten $q = 0,75$, så kan kolumnspelaren inte förbättra sin förväntade nytta genom att ändra sitt beteende; återigen är skälet att båda strategier ger exakt samma förväntade nytta.

Det betyder att om radspelaren väljer hög med sannolikheten $p = 0,75$, samtidigt som kolumnspelaren väljer hög med sannolikheten $q = 0,75$, så står båda spelarna inför situationen att båda deras strategier ger samma förväntade nytta. Således kan ingendera spelaren förbättra sin förväntade nytta genom att ensidigt ändra sitt beteende. Detta är ju precis vad definitionen av Nashjämvikt kräver. Alltså har vi identifierat en tredje jämvikt. Vi säger att denna tredje jämvikt är blandad eftersom den innebär att spelarna använder blandade strategier. De två övriga jämviktterna sägs vara rena eftersom spelarna där bara utför en handling var.

Man kan fråga sig hur det går till när spelare väljer ett alternativ med en viss sannolikhet. Kanske kan spelarna låta magkänslan avgöra vilken alternativ de väljer: ett mera precist sätt skulle kunna vara att använda en fyrsidig tärning och sedan spela duva om tärningen landar på siffran 1, 2, eller 3 (eftersom sannolikheten för att en fyrsidig tärning landar på siffran 1, 2 eller 3 är 0,75). I avsnittet om hjortjakten nedan kommer vi se att inlärning ger oss en annan (och ofta rimligare) tolkning av vad en blandad strategi är.

Man kan även fråga sig vilken, eller vilka, av de tre jämvikterna i konfliktspelet som troligast kommer att blir utfallet i en verklig situation. Innan vi undersöker dessa frågor ska vi stifta bekantskap med ytterligare ett spel som innehåller tre jämvikter, men som skiljer sig från konfliktspelet på avgörande sätt.

KOORDINATIONSSPEL: HJORTJAKTEN

Två jägare kan välja mellan att jaga hjort eller hare. För att jaga hjort framgångsrikt krävs det att båda jägare hjälps åt; om en jägare försöker jaga hjort medan den andra jagar hare så får den förre inte något byte. Att jaga hare ger däremot ett visst byte även om den andre spelaren jagar hjort, men bytet blir mindre ifall båda jagar hare eftersom de då konkurrerar om samma djur. Jägarna bryr sig endast om att komma hem med så mycket kött som möjligt. Vi kan sammanfatta dessa preferenser i matrisen i figur 3.⁷

Figur 2. Koordinationsspel: hjortjakten.

	Hjort	Hare
Hjort	6, 6	0, 3
Hare	3, 0	2, 2

Båda spelare föredrar att jaga hjort ifall den andre spelaren jagar hjort. På samma vis föredrar båda att jaga hare ifall den andre spelaren jagar hare. Därför finns det en jämvikt där båda jagar hjort och en annan jämvikt där båda jagar hare.

Spel med flera symmetriska jämvikter kan illustrera hur sociala normer eller konventioner, det vill säga outtalade beteenderegler, fungerar. Många olika konventioner är möjliga men om man förväntar sig att motparten följer en viss konvention så ligger det i ens intresse att också följa just den konventionen. Huruvida man kör på höger eller vänster sida om vägen är ett enkelt men tydligt exempel. I Sverige kör vi på höger sida om vägen för att lagen kräver det men framför allt därför att vi annars riskerar att råka ut för en olycka. Före 1967 körde svenskar istället på vänster sida. Det krävdes ett regeringsbeslut och en

7 Situationen, om än inte spelmatrisen, beskrevs redan av Jean-Jacques Rousseau.

noga planerad övergång för att genomföra bytet av konvention.⁸ För det mesta växer dock konventioner fram över lång tid och är resultatet av gradvis förändrat beteende hos en grupp av människor. Evolutionär spelteori är ett sätt att modellera dessa långsamma processer genom vilka konventioner uppstår.⁹

Liksom i konfliktspelet finns det ytterligare en (blandad) jämvikt. Låt oss anta att kolumnspelaren följer en blandad strategi och väljer slumpmässigt mellan sina två rena strategier: hjort med sannolikheten q och hare med sannolikheten $1 - q$. Radspelarens förväntade nytta av att jaga hjort blir då

$$U_{\text{Hjort}} = 6 \cdot q + 0 \cdot (1 - q) = 6q$$

och radspelarens förväntade nytta av att jaga hare blir

$$U_{\text{Hare}} = 3 \cdot q + 2 \cdot (1 - q) = 2 + q$$

Vi ser att de båda strategierna ger samma förväntade nytta om och endast om $6q = 2 + q$, det vill säga om och endast om $q = 0,4$. Situationen är precis likadan för kolumnspelaren. Den tredje, blandade, jämvikten är alltså sådan att båda spelare följer den blandade strategin som spelar hjortalternativet med sannolikhet 0,4.

Evolutionär spelteori

Det är nu dags att bli tydligare med hur vi tänker oss att inläring går till och hur det sker en gradvis anpassning av beteende som eventuellt kan hamna i jämvikt. Vi behöver inte anta att spelarna är fullt rationella. Det räcker att de följer tumregler som i någon mån anpassas efter om en strategi är framgångsrik eller inte. Det kan exempelvis handla om att man härmar de medspelare som verkar lyckas bättre med sina val av strategier, eller att man provar nya strategier på måfå när man tycker att ens nuvarande strategi ger ett dåligt utfall.

Den formella teorin som ligger till grund för diskussionen här kallas för *evolutionär spelteori*. Termen evolutionär ska dock inte förstås bokstavligt. Den matematiska teorin är tillräckligt generell för att kunna beskriva både biologiska evolutionära processer och sociala eller psykologiska eller sociala inlärningsprocesser. I både det biologiska och det kulturella fallet handlar det om populationer vars sammansättning och beteenden förändras gradvis på ett sätt som åtminstone delvis beror på hur, i någon mening, framgångsrika olika beteende är. I det följande kommer vi huvudsakligen att använda termen inläring och inte evolution.

8 Man kan eventuellt invända att anledningen till att vi kör på höger sida är att det är en trafikregel och att man bötfälls om man bryter mot regeln. Jag skulle dock gissa att det inte främst är rädslan för böter utan rädslan för krock som får oss att respektera trafikregeln och den konvention den kodifierar.

9 Young (1998) analyserar många konventioner på detta sätt, inklusive konventionen om vilken sida av vägen man kör på.

EN ENKEL TUMREGEL FÖR INLÄRNING

Vi tänker oss en stor population av individer som varje dag delas in i olika par för att spela ett spel med varandra. För det mesta följer var och en av individerna samma strategi som de följde igår. Då och då får de dock för sig att undersöka om de kanske ska pröva en annan strategi.¹⁰ En individ som funderar på att byta strategi jämför sig då med en slumpvist vald annan individ i populationen. Om den individen följer en strategi som ger en högre förväntad nytta så är den jämförande individen benägen att byta till den mer framgångsrika strategin. Åtminstone är sannolikheten för ett byte till en strategi högre ju större skillnaden i förväntad nytta är.

Man kan visa att om alla individer betar sig på det beskrivna sättet kommer andelen som spelar en viss strategi att öka om och endast om den strategi de spelar ger en förväntad nytta som ligger över den genomsnittliga, förväntade nyttan i populationen.¹¹ När det bara finns två strategier betyder det att den strategi som för närvarande ger mest nytta ökar i popularitet och att den som ger minst nytta minskar i popularitet. Det finns många andra tumregler för inläring än den som beskrivits här, som också leder till att strategier som ger mer än andra växer i antal medan de som ger mindre minskar i antal.¹² I det följande utgår vi helt enkelt från att den strategi som för ögonblicket ger högst förväntad nytta ökar i popularitet i populationen och att den strategi som ger lägst förväntad nytta minskar i popularitet.

Vi kan även definiera en biologisk evolutionär process med i stort sett identiska egenskaper. Återigen tänker vi oss en stor population av individer som varje dag delas in i olika par för att spela ett spel med varandra. Varje individ är programmerad att välja en av handlingarna i spelet. Då och då dör en individ. Vederbörande ersätts då av en nyfödd individ. I och med den darwinistiska principen om att de bäst anpassade överlever (*survival of the fittest*) antar vi att sannolikheten för att den nyfödda individen ska vara programmerad att följa en viss strategi är proportionell mot hur hög förväntad nytta strategin ger. Mer exakt kan man specificera reproduktionen på ett sådant sätt att denna evolutionära process beskrivs med precis samma ekvationer som den tidigare beskrivna

10 Notera likheten med det evolutionsbiologiska perspektivet på beslutsfattande som Jessica Abbott beskriver i sin uppsats i detta nummer.

11 Björnerstedt & Weibull (1996) samt Weibull (1995) visar att om den beskrivna inlärningsprocessen följs av individerna i populationen så kommer populationsandelarna att utvecklas i enlighet med den så kallade replikatorodynamiken. Olika former av imitation, med åtföljande dynamiska konsekvenser, undersöks även av Schlag (1998).

12 Till exempel visar Börgers & Sarin (1998) att om individer lär sig genom positiv och negativ förstärkning när de provar olika strategier så kommer populationen att utvecklas enligt replikatorodynamiken. Det finns självfallet många inlärningsregler som inte fångas av denna dynamik; se Fudenberg & Levine (1998) och Sandholm (2010).

inlärningsprocessen.¹³ I ett appendix går jag igenom mer exakt på vilket sätt matematiken för biologisk evolution och social inläring kan sammanfalla.

HJORTJAKT

Låt oss nu tillämpa vår enkla inlärningsmodell på hjortjaktsspelet som beskrevs ovan. För att kunna beskriva hur beteenden ändras över tid använder vi oss av en funktion $x(t)$ som för varje tidpunkt t anger hur stor andel av populationen som följer hjortstrategin.

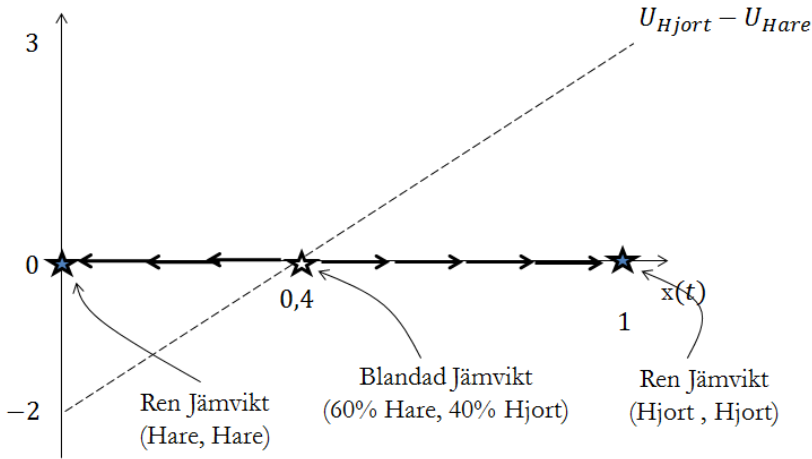
Eftersom alla individer paras ihop slumpvis för att spela mot varandra så är sannolikheten att möta en individ som spelar hjort lika med $x(t)$. Om man själv spelar hjort så kommer man därför att uppnå den förväntade nyttan $U_{Hjort} = 6x(t)$. Om man istället spelar hare är den förväntade nyttan $U_{Hare} = 2 + x(t)$. Andelen hjortspelare ökar om $U_{Hjort} > U_{Hare}$, och minskar om $U_{Hjort} < U_{Hare}$. Om vi jämför dessa nyttor så ser vi att hjortspelarna tjänar mer än harespelarna om och endast om $x(t) > 0,4$, med andra ord om och endast om andelen som spelar hjort är större än 40 procent. Med vår inlärningsmodell betyder det att andelen hjortspelare ökar om det redan är fler än 40 procent som spelar hjortstrategin och minskar om det är färre än 40 procent som gör det.

För att tydliggöra vad detta innebär kan vi illustrera med figur 3. Eftersom $x(t)$ står för andelen hjortspelare är $x(t)$ en siffra mellan noll och ett. Eftersom andelen harespelare är $1 - x(t)$ räcker det med att ange siffran $x(t)$ för att fullständigt beskriva i vilket tillstånd populationen befinner sig. Intervallet mellan noll och ett representerar därför de möjliga populations-tillstånden. När $x(t) = 0$ är alla individer harespelare, vilket betyder att hare-hare-jämvikten spelas varje gång ett par möts. När $x(t) = 1$ är istället alla individer hjortspelare, vilket betyder att hjort-hjort-jämvikten spelas varje gång ett par möts.

När $x(t) = 0,4$ så är 40 % av populationen hjortspelare och 60 % harespelare. Det innebär att om vi slumpmässigt drar två individer från populationen för att spela så är sannolikheterna för de olika utfallen (hjort-hjort, hjort-hare, hare-hjort, och hare-hare) precis desamma som om vi hade betraktat två spelare som var för sig valde att spela hjort med sannolikheten 0,4 och hare med sannolikheten 0,6. Det betyder alltså att en blandad jämvikt inte bara kan användas för att beskriva en situation där spelarna väljer handlingar slumpmässigt utan även en situation där det istället är spelarna som väljs slumpmässigt. Om man tycker att slumpmässigt valda handlingar verkar orimliga i vissa interaktioner så kan man alltså ändå ha nytta av idén om en blandad jämvikt när man betraktar populationen som helhet.

13 I själva verket var det teoretiska biologer som först studerade dessa ekvationer. Replikatorodynamiken introducerades av Taylor & Jonker (1978). Hofbauer & Sigmund (1988) beskriver utförligt dess biologiska tillämpning.

Figur 3. Inläring i koordinationsspel.



I den streckade linjen i figur 3 har skillnaden i nytta mellan hjort- och hare-alternativen ritats in som en funktion av $x(t)$. Denna skillnad är $U_{Hjort} - U_{Hare} = 6x(t) - (2 + x(t)) = 5x(t) - 2$. När denna skillnad är positiv är det bättre att spela hjort än hare, och då ökar andelen som spelar hjort, det vill säga att $x(t)$ ökar. Detta visas i figuren med pilar som pekar åt höger längs den horisontella axeln. När skillnaden $U_{Hjort} - U_{Hare}$ är negativ så minskar istället andelen som spelar hjort, det vill säga att $x(t)$ minskar, vilket indikeras av pilarna som pekar åt vänster längs den horisontella axeln.

Från figur 3 blir det tydligt att om populationen initialt befinner sig i en punkt till höger om $x(t) = 0,4$ så ökar $x(t)$ till dess att alla är hjortspelare. Från ett utgångsläge till vänster om $x(t) = 0,4$ så minskar istället $x(t)$ till dess att alla är harespelare. Det innebär att den blandade jämvikten är mycket instabil. Om populationen i utgångsläget befinner sig i $x(t) = 0,4$ så räcker det med en liten störning som för populationen bort från $x(t) = 0,4$ för att inlärningsprocessen obönhörligen ska föra populationen till någon av de rena jämvikterna.¹⁴ Det blir uppenbart hur inläring implicit koordinerar förväntningarna om andras beteende, vilket i detta spel leder till att beteende koordineras.

En liknelse kan tydliggöra vad som menas med stabilitet i sammanhanget. Föreställ dig en boll som ligger på toppen av en kulle. En liten vindpust flyttar bollen ett stycke från toppen, bollen börjar rulla nedför kullen och hamnar till

14 Det är rimligt att tro att strategier inte alltid uppdateras på så sätt att mer framgångsrika strategier väljs framför mindre framgångsrika. Om det finns en positiv sannolikhet för att val av mindre framgångsrika strategier inträffar så kan det hända, av ren slump, att en population som befunnit sig i en jämvikt under lång tid på ganska kort tid förflyttar sig till den andra jämvikten. Man kan då studera vilken jämvikt som är svårast att lämna och lättast att hamna i. En jämvikt som är svår att på detta vis lämna sägs vara stokastiskt stabil. Teorin om stokastisk stabilitet utvecklades av Young (1993) och Kandori m.fl. (1993).

slut långt från toppen. Vi säger att bollen befann sig i ett instabilt läge eftersom en liten förändring av dess utgångsläge ledde till att den fortsatte att röra sig bort från utgångsläget. Jämför detta med situationen för en boll som ligger i botten av en sänka. Om vi knuffar bollen lite grand åt något håll så kommer sluttningen i sänkan göra att bollen rullar tillbaka till utgångsläget när vi släpper den. Vi säger då att bollen befann sig i ett stabilt läge. Den blandade jämvikten är instabil som bollen på toppen av kullen – en liten förändring fortplantar sig och tar oss långt bort från utgångsläget. De båda rena jämvikterna är däremot stabila som bollen i botten av sänkan – efter en liten förändring kommer vi tillbaka till utgångsläget.

HÖK-DUVA

I hjortjaktsspelet fann vi att den blandade jämvikten är instabil och att de rena jämvikterna är stabila. Vi ska nu se att slutsatsen blir annorlunda i hök-duvaspelet. Låt $x(t)$ beteckna andelen hökspelare vid tidpunkten t , så att $x(t)$ är sannolikheten att möta en individ som spelar hök.

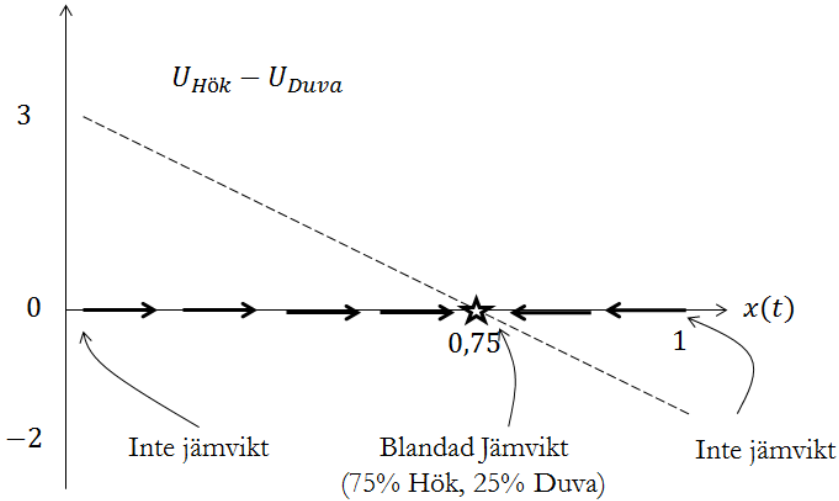
När $x(t) = 0,75$ beter sig populationen som i den blandade jämvikten. Om vi slumpmässigt drar två individer från populationen för att spela så är sannolikheterna för de olika utfallen (hök-hök, hök-duva, duva-hök, och duva-duva) samma när två spelare var för sig spelar hök med sannolikheten $0,75$ och duva med sannolikheten $0,25$. Liksom i koordinationsspelet ovan kan vi här tolka den blandade jämvikten som en situation där det inte är spelarna som väljer handlingar slumpmässigt utan istället spelarna som väljs slumpmässigt.

Det finns däremot inget värde på $x(t)$, alltså ingen blandning av hök- och duvspelare, som svarar mot de båda asymmetriska jämvikterna (duva-hök och hök-duva). Anledningen till detta är att dessa jämvikter förutsätter att man kan välja olika strategier beroende på om man är en kolumn- eller radspelare. Eftersom alla spelare i populationen slumpmässigt tilldelas roller som rad- eller kolumnspelare kan den asymmetriska jämvikten inte spelas. Som vi skall se nedan ändras detta om vi hade två populationer och drog radspelarna från den ena populationen och kolumnspelarna från den andra populationen.

Den förväntade nyttan av att spela hök är $U_{Hök} = 6 - 7x(t)$ och den förväntade nyttan av att spela duva är $U_{Duva} = 3 - 3x(t)$. Hökspelarna tjänar mer än duvspelarna, och ökar därmed som andel av befolkningen, om och endast om $x(t) < 0,75$. Det betyder att $x(t)$ växer om $x(t) < 0,75$ och minskar om $x(t) > 0,75$. Detta illustreras i figur 4. Där har skillnaden $U_{Hök} - U_{Duva}$ ritats in som en funktion av $x(t)$. När skillnaden är positiv så ökar $x(t)$, vilket indikeras av pilarna som pekar åt höger längs den horisontella axeln. När skillnaden istället är negativ så ökar $x(t)$, vilket indikeras av pilarna som pekar åt vänster längs den horisontella axeln. Vi ser att inlärningsprocessen för populationen mot den blandade jämvikten. Till skillnad från koordinationsspelet ovan är den blandade jämvikten här stabil. Om populationen ursprungligen befinner sig i punkten

$x(t) = 0,75$ och sedan störs lite grand, så att hamnar över eller under $0,75$, kommer inlärningsprocessen så småningom att föra tillbaka populationen till $x(t) = 0,75$.¹⁵

Figur 4. Inlärning i hök och duva-spelet.



HÖK-DUVA MED TVÅ POPULATIONER

En allmän lärdom av analysen ovan är att studiet av inlärning hjälper oss skilja på vilka jämvikter som det är rimligt och orimligt att förvänta sig. Detaljerna i inlärningsmodellen kan dock spela roll för slutsatserna. Hittills har alla individer som skall spela ett spel dragits från en och samma population. Vi skall nu se vad som sker om vi istället har två populationer och drar radspelarna från den ena populationen (radpopulationen) och kolumnspelarna från den andra populationen (kolumnpopulationen). Låt $x(t)$ beteckna andelen hökspelare i kolumnpopulationen vid tidpunkten t , så att $x(t)$ är sannolikheten för en radspelare att möta en individ som spelar hök. Låt $y(t)$ beteckna andelen hökspelare i radpopulationen vid tidpunkten t .

Vi modifierar tumregeln för inlärning på så sätt att en individ som funderar på att byta strategi jämför sig med en slumpvist vald annan individ från *samma population som individen själv*. Om den individen följer en strategi som ger en

15 Oprea m.fl. (2011) studerar hök-duva i ett experiment där försöksdeltagarna upprepade gånger får välja mellan hök och duva och belönas med pengar i enlighet med siffrorna i spelmatrisen, beroende på hur de andra försökspersonerna beter sig. Forskarna finner att populationen av försöksdeltagare så småningom kommer mycket nära den jämvikt som förutsågs av teorin.

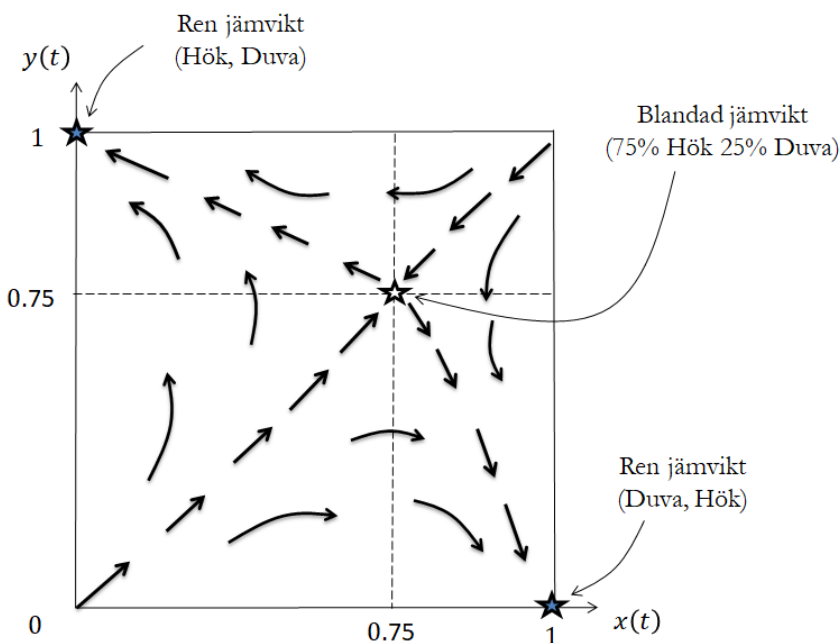
högre förväntad nytta så är den jämförande individen benägen att byta till den mer framgångsrika strategin.

Den förväntade nyttan av att spela hög för en radspelare är $U_{Hök} = 6 - 7x(t)$ och den förväntade nyttan av att spela duva är $U_{Duva} = 3 - 3x(t)$. Hökspelarna i radpopulationen tjänar mer än duvspelarna i radpopulationen, och ökar därmed som andel av radpopulation, om och endast om $x(t) < 0,75$. På motsvarande sätt tjänar hökspelarna i kolumnpopulationen mer än duvspelarna i kolumnpopulationen, och ökar därmed som andel av kolumnpopulation, om och endast om $y(t) < 0,75$.

Genom att gå igenom de fyra olika möjligheterna som ges av huruvida $x(t)$ eller $y(t)$ är större eller mindre än 0,75, och jämföra vilken strategi som ger mest i vardera populationen, kan vi skapa oss en bild av hur andelen hög- och duvspelare förändras inom endera populationen som en funktion av sammansättningen av den andra populationen. Ett sådant, så kallat fasdiagram, återfinns i figur 5.

Den horisontella axeln mäter andelen hökspelare i kolumnpopulationen och den vertikala axeln mäter andelen hökspelare i radpopulationen. De tjocka svarta pilarna indikerar hur populationsandelarna rör sig över tid. Om vi startar på diagonalen så förs vi mot den blandade jämvikten men om vi startar någon annanstans så leder inlärningsprocessen istället mot någon av de rena jämvikterna. Vilken av de rena jämvikterna vi hamnar i beror på vilken sida av diagonalen vi börjar. Till skillnad från i fallet med en enda population är den blandade jämvikten här instabil medan de båda rena jämvikterna är stabila.

Figur 5. Inläring i hög och duva-spelet när rad- och kolumnspelare kommer från var sin population.



KÖNSROLLER I "DE FRIVILLIGAS DILEMMA"

Låt oss nu något spekulativt tolka analysen med två populationer i det abstrakta hök-duva-spelet i termer av något så konkret som könsroller. Vi antar då att kvinnor imiterar, och jämför sig med, andra kvinnor, medan män imiterar, och jämför sig med, andra män.

Betrakta följande exempel. Två medarbetare delar arbetsplats. Kopiatorn har gått sönder och för att få den lagad måste någon göra en felanmälan via leverantörens krångliga webformulär. Det räcker med att en person gör felanmälan för att maskinen skall lagas. Vi kan generellt tänka på en grupp av två personer där det räcker att *en* person utför uppgiften för att problemet skall vara löst. Om båda anmäler sig frivilligt så singlar man slant om vem som skall utföra uppgiften. Var och en av medarbetarna värderar det tre olika utfallen enligt följande. Att själv utföra den tråkiga uppgiften ger nytta 2. Att den andre medarbetaren utför den tråkiga uppgiften ger nytta 4. Att ingen utför den tråkiga uppgiften ger nytta 1. Vi kan kalla spelet, som sammanfattas i figur 6, för *de frivilligas dilemma*. Notera att nyttan när båda erbjuder sig att hjälpa till är $(4 + 2)/2 = 3$ för var och en, eftersom man i detta fall singlar slant om vem som skall utföra arbetsuppgiften.

Figur 6. De frivilligas dilemma.

	Inte hjälpa till	Hjälpa till
Inte hjälpa till	1, 1	4, 2
Hjälpa till	2, 4	3, 3

Notera den strukturella likheten med hök-duva-spelet. Det finns två Nashjämvikter i rena strategier: en där radspelaren hjälper till och kolumnspelaren inte hjälper till, och en annan där kolumnspelaren hjälper till och radspelaren inte hjälper till. Den blandade jämvikten består i att båda spelare lägger exakt lika stor vikt vid att hjälpa till som att inte hjälpa till.¹⁶ Inläring, av den sort vi diskuterat, skulle i de frivilligas dilemma ge upphov till samma fasdiagram som för hök-duva-spelet i figur 5, med den enda skillnaden att den instabila, blandade jämvikten ligger vid $(x, y) = (1/2, 1/2)$.

Babcock m.fl. (2017) genomförde ett experiment som är snarlikt det jag beskrev som de frivilligas dilemma ovan.¹⁷ En skillnad var att grupperna bestod av tre medarbetare. En annan skillnad var att man kunde avvakta och se om de

16 Notera att $U_{\text{inte hjälpa}} = q + 4(1 - q) = 2q + 3(1 - q) = U_{\text{hjälpa}}$ implicerar $q = 1/2$.

17 Holm (2000) är en tidig studie med ett liknande tema.

andra skulle hjälpa till. Om ingen anmält sig frivilligt inom två minuter räknades det som att ingen hjälpte till.

Experimentdeltagarna var helt anonyma för varandra när de fattade sina beslut sittande framför datorskärmar i en stor sal. Innan man gick in i experimentalen kunde man dock se vilka andra som deltog i experimentet. I vissa sessioner var det en jämn könsfördelning medan det i andra ingick enbart män eller enbart kvinnor. Experimentet gick ut på att jämföra hur beteendet skilde sig beroende på denna könssammansättning. Man fann att män i grupper som bestod av enbart män var *lika* benägna att ställa upp och hjälpa till som kvinnor i grupper som bestod av enbart kvinnor. I könsblandade grupper var män däremot *mindre* benägna att ställa upp och hjälpa än kvinnor.

Det ligger nära till hands att tolka resultaten som att män och kvinnor tycker lika illa om att utföra uppgiften, men att både män och kvinnor förväntar sig att kvinnor i allmänhet hjälper till och att män i allmänhet låter bli. Dessa förväntningar om könsroller är självuppfyllande i den meningen att det är rationellt för kvinnor att ställa upp och rationellt för män att låta bli (beteendena i fråga utgör en Nashjämvikt).¹⁸

Analysen av inläring i de frivilligas dilemma (vilken som sagt är kvalitativt identisk med den för analysen av konfliktspelet hök-duva i figur 5) indikerar att könsrollerna kan vara mycket stabila och svåra att ändra. Låt oss anta att kvinnorna utgör kolumnpopulationen och männen radpopulationen. Jämvikten där kvinnor hjälper till och män låter bli motsvarar då den punkt i diagrammet där $(x, y) = (0, 1)$. För att inte halka tillbaka till denna jämvikt skulle man i ett enda svep behöva ändra beteendet hos en så stor del av männen och kvinnorna att man hamnade på en punkt under diagonalen (45-graderslinjen). Om man lyckades med detta skulle dock inlärningsdynamiken (i denna enkla modell) föra oss till en punkt där männen alltid hjälpte till och kvinnorna aldrig hjälpte till. Det enda som skulle kunna leda fram till ett stabilt jämförbart läge där både män och kvinnor turades om vore ifall män och kvinnor slutade bete sig som två separata populationer antas göra i modellen. Det vill säga, det som krävs är att kvinnor och män blir könsblinda så att båda könen imiterar, och jämför sig med, både kvinnor och män, när de uppdaterar sitt val av strategi.

18 Mer allmänt kan det vara fruktbart att tolka sociala normer som jämvikter i spel med flera jämvikter. När normen är en jämvikt så har spelarna inga incitament att bryta normen så länge de förväntar sig att andra följer den. Om detta säger Binmore & Samuelson (1994) följande: "In consequence, one should make no sharp distinction between *homo economicus* and *homo sociologicus*. In using a social norm in a situation to which it is well adapted, *homo sociologicus* behaves as though he were optimizing. Similarly, when optimizing, *homo economicus* behaves as though he were employing a social norm that is well adapted to his problem".

Inläring och upprepade spel

ETT SOCIALT DILEMMA: FÅNGARNAS DILEMMA

Det kanske mest kända exemplet från spelteorin är fångarnas dilemma. Spelet beskriver en situation där deltagarna skulle tjäna på att hjälpas åt, men där var och en av deltagarna föredrar att själva låta bli att bidra. Låt oss strunta i hur fångarnas dilemma fått sitt namn och istället tänka oss följande situation. Ask och Embla arbetar med ett projekt tillsammans. Båda har två handlingsalternativ, eller strategier, nämligen att anstränga sig eller att lata sig. Båda föredrar att någon anstränger sig så att projektet blir färdigt i tid. Men båda vill helst slippa anstränga sig och vill istället åka snålskjuts på den andres ansträngning. Problemet är att om ingen anstränger sig blir projektet inte klart i tid. Ask och Embla har därför motstridiga intressen. Vi kan sammanfatta dessa preferenser med matrisen nedan i figur 7.

Figur 7. Socialt dilemma: fångarnas dilemma.

	Anstränga sig	Lata sig
Anstränga sig	3, 3	0, 5
Lata sig	5, 0	1, 1

Oavsett vad Ask gör så når Embla en högre nytta genom att lata sig (eftersom 5 är större än 3 och 1 är större än 0). På samma vis är det bästa Ask kan göra att låta bli att anstränga sig, oavsett vad Embla gör. Spelet har därmed en enda Nashjämvikt, vilken består i att båda latar sig. Det hade varit bättre för båda spelarna om de båda valt att anstränga sig, eftersom de då fått 3 vardera istället för 1 vardera, vilket blir fallet om båda latar sig. Det sociala dilemman består i att det kollektivt fördelaktiga utfallet (att båda anstränger sig) inte är förenligt med att var och en av spelarna betar sig (individuellt) rationellt.¹⁹

Vad säger vår inlärningsmodell om fångarnas dilemma? Låt $x(t)$ vara andelen samarbetsspelare vid tidpunkten t . Den förväntade nyttan av att spela samarbete (anstränga sig) är $U_{\text{Samarbete}} = 3x(t)$ och den förväntade nyttan av att spela icke-samarbete (lata sig) är $U_{\text{Icke-samarbete}} = 4x(t) + 1$. De som inte samarbetar har därför högre förväntad nytta än de som samarbetar, oavsett vilket värde $x(t)$ tar. Det betyder att $x(t)$ alltid rör sig mot noll. Populationen konvergerar mot den enda jämvikten – ingen samarbetar.

19 Vi skulle kunna ha antagit att Ask och Embla är mera generösa och föredrar att städa om den andra personen städar, exempelvis för att de får dåligt samvete annars. Då hade vi kunnat byta ut siffran 5 mot en lägre siffra och få ett annat spel, med andra jämvikter. Men givet våra antaganden, som de sammanfattas i figur 7, är det ovedersägligen så att om spelarna är rationella så kommer de att välja att lata sig och det leder till ett utfall som är sämre än om båda hade ansträngt sig.

UPPREPADE SPEL

I beskrivningen av hur inläring går till antogs att individer matchas ihop med en ny motspelare i varje period. Eftersom befolkningen förutsätts vara stor så är sannolikheten liten för att två individer ska mötas snart igen. När individerna möts väldigt sällan kan de låta bli att bry sig om de framtida konsekvenserna av sina handlingar. Detta är avgörande för vår slutsats att samarbete inte kan överleva i fångarnas dilemma.

Låt oss nu undersöka vad som händer om individerna spelar ett spel flera gånger med samma motpart. Vi modellerar detta genom att anta att det, varje gång två spelare möts, finns möjlighet att de spelar fångarnas dilemma flera gånger med varandra. Mer exakt tänker vi oss att de två matchade individerna med säkerhet spelar fångarnas dilemma en gång. Därefter spelar de en andra gång med sannolikheten δ . Om de spelar fångarnas dilemma en andra gång så spelar de sedan en tredje gång med sannolikheten δ . Varje gång de spelat en omgång så är sannolikheten δ att de ska spela ytterligare en gång. Vi säger att de spelar ett obestämt upprepat spel, med upprepnings sannolikheten δ .

Det finns oändligt många sätt att betinga sitt val mellan samarbete och icke-samarbete på vad som hänt hittills i ett obestämt upprepat spel. För enkelhets skull begränsar vi oss till att undersöka två enkla tumregler, eller strategier, för det obestämt upprepade fångarnas dilemma-spelet.

- *Aldrig samarbete.* Som namnet anger väljer en individ som följer denna strategi aldrig att samarbete.
- *Samarbete tills vidare, sedan aldrig mer ("Grim trigger").* En individ som följer denna strategi börjar med att samarbete och fortsätter att samarbete så länge som alla andra hittills alltid samarbetat. Om någon inte samarbetar så byter man till att aldrig mera samarbete.²⁰

När två som aldrig samarbetar möts för att spela fångarnas dilemma med de nyttor som anges i figur 7 så tjänar de 1 i varje spelomgång. Den förväntade nyttan är därför $1 + \delta + \delta^2 + \delta^3 + \dots$. Man kan visa att detta är lika med $1/(1 - \delta)$. När två som spelar *samarbete tills vidare*-strategin möts samarbetar de i varje omgång de möts (eftersom de båda börjar med att samarbete och fortsätter därmed så länge ingen gjort något annat). Den förväntade nyttan är därför $3 + 3\delta + 3\delta^2 + 3\delta^3 + \dots = 3/(1 - \delta)$. När en som följer strategin *samarbete tills vidare* möter en som följer strategin *aldrig samarbete* så samarbetar *samarbete tills vidare*-spelaren endast i första perioden. Den förväntade nyttan för *samarbete tills vidare*-spelaren är därför $0 + \delta + \delta^2 + \delta^3 + \dots = \delta/(1 - \delta)$, och den förväntade nyttan för den andre är $5 + \delta + \delta^2 + \delta^3 + \dots = 5 + \delta/(1 - \delta)$. Vi

20 Resultaten skulle vara desamma om vi istället använde strategin tit-for-tat, berömd från Axelrods turnering.

sammanfattar detta i figur 8(a). Om vi sätter $\delta = 2/3$ så erhåller vi spelmatrisen i figur 8(b)

Figur 8. Upprepat fångarnas dilemma.

(a) Uppreningssannolikhet δ

	Samarbete tills vidare	Aldrig samarbete
Samarbete tills vidare	$\frac{3}{1-\delta}$, $\frac{3}{1-\delta}$	$\frac{\delta}{1-\delta}$, $5 + \frac{\delta}{1-\delta}$
Aldrig samarbete	$5 + \frac{\delta}{1-\delta}$, $\frac{\delta}{1-\delta}$	$\frac{3}{1-\delta}$, $\frac{3}{1-\delta}$

(b) Uppreningssannolikhet $\delta = 2/3$

	Samarbete tills vidare	Aldrig samarbete
Samarbete tills vidare	9, 9	2, 7
Aldrig samarbete	7, 2	3, 3

Om vi studerar matrisen i figur 8(b) närmare, är det lätt att se att den har en struktur som påminner om koordinationsspelet i figur 2. Det bästa man kan göra är att välja samma strategi som sin motspelare. Därför finns det två rena jämvikter i spelet, där man kan välja mellan att följa *samarbete tills vidare* eller *aldrig samarbete* i ett obestämt upprepat spel med uppreningssannolikheten $\delta = 2/3$. I den ena jämvikten spelas bara *samarbete tills vidare*, vilket betyder att alla samarbetar. I den andra samarbetar ingen. Om vi jämför med det icke-upprepade spelet, det vill säga det vanliga fångarnas dilemma, så ser vi att jämvikten utan samarbete visserligen finns kvar som en möjlighet, men att det på grund av uppreningen också tillkommit en möjlighet att samarbete. Detta sker genom att samarbete idag belönas med samarbete imorgon och att icke-samarbete idag straffas med icke-samarbete imorgon.²¹ Samarbetet förutsätter här ingen omtanke om motparten; genom att betinga framtida beteende på dagens beteende skapar spelarna ömsesidiga incitament att samarbete.

21 Denna insikt, att obestämd upprening skapar möjligheter till samarbete, går under benämningen folkteoremet i spelteorin, eftersom den var en del av spelteorins "folkvisdom" innan resultatet formaliserades. Den introducerades i den evolutionsbiologiska litteraturen av Trivers (1971). Studiet av evolution i det upprepade fångarnas dilemma-spelet stimulerades kanske särskilt av Axelrods (1984) datorturneringar. Man kan även studera scenarier där individer inte spelar upprepade fångarnas dilemma i par utan varje omgång möter en ny motpart i fångarnas dilemma. Även under dessa omständigheter kan evolution leda till samarbete, ifall spelarna får veta något om hur motparten brukar spela, så att spelarna har möjlighet att skapa ett visst rykte om sig. För analyser av detta scenario, se Kandori (1992), Nowak & Sigmund (1998) och Heller & Mohlin (2016).

Liksom i koordinationsspelet så är de båda rena jämvikterna i det upprepade fångarnas dilemma-spelet stabila. Vi kan tänka på det som två olika normer som kan råda i samhället. I den ena är alla optimistiska och förtroendefulla – de räknar med att motparten samarbetar och eftersom det är precis vad som sker så fortsätter alla att samarbeta och vara optimistiska. I den andra jämvikten är alla pessimistiska och misstänksamma – de utgår helt korrekt från att ingen samarbetar och har därför inga skäl att samarbeta.²² Vilken norm som blir rådande beror, enligt den enkla inlärningsmodell vi studerat, på utgångsläget. Om samhället till att börja med mestadels består av dem som spelar *samarbete tills vidare* så kommer de att ta över, men om samhället innehåller en tillräckligt stor andel som spelar *aldrig samarbete* så kommer istället dessa att ta över.²³

Inläring som inte leder till jämvikt

I alla spel vi har studerat hittills leder inlärningsprocessen till att en jämvikt spelas. I koordinationsspelet spelas visserligen inte den blandade jämvikten och i hök-duva-spelet inte de rena jämvikterna (när det bara finns en population), men någon av jämvikterna spelas i alla dessa spel. Man kan fråga sig om det alltid är så att den inlärningsprocess vi beskrivit leder till en jämvikt. Det vore ju bra om det vore så eftersom så mycket av ekonomisk teori förutsätter jämvikt. Tyvärr är det inte så.²⁴

Du har antagligen spelat sten, sax och påse som barn. Det är ett spel med två spelare där påse vinner mot sten som vinner över sax som i sin tur vinner över påse. Låt oss anta att en vinst ger nytta 1, en förlust ger nytta -1 och om det är oavgjort så får man nyttan 0. Figur 9 avbildar spelmatrisen för detta spel.

22 Det bör noteras att koordinationsproblemen grovt underskattas av att vi begränsar oss till två strategier i det upprepade spelet. Det finns som sagt oändligt många strategier i det upprepade fångarnas dilemma-spelet. Om vi betraktar sociala dilemman med fler än två handlingar, som svarar mot olika grader av samarbete, kanske med mindre symmetriska utfall, växer koordinationsproblemen ytterligare.

23 Litteraturen om evolution av samarbete, särskilt i fångarnas dilemma, är mycket omfattande. För en introduktion, se exempelvis Novak (2006). En översikt över den experimentella forskningen rörande inläring i det upprepade fångarnas dilemma-spelet ges av Dal Bó & Fréchette (2018).

24 Man kan fråga sig om det åtminstone är så att inläring leder till att dominerade strategier, eller iterativt dominerade strategier inte används. Inte heller detta är generellt sant. Se Hofbauer & Sandholm (2011) och Bernergård & Mohlin (2019) samt referenser däruti.

Figur 9. Sten, sax och påse.

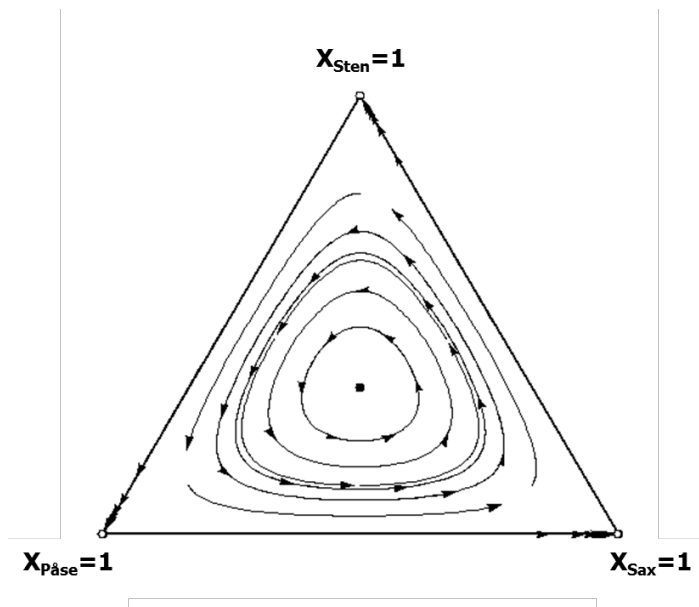
	Sten	Påse	Sax
Sten	0, 0	-1, 1	1, -1
Påse	1, -1	0, 0	-1, 1
Sax	-1, 1	1, -1	0, 0

Detta spel har en enda Nashjämvikt. I denna jämvikt spelar båda spelare alla sina tre strategier med lika stor sannolikhet. En spelare väljer alltså vart och ett av sina handlingsalternativ med sannolikhet $1/3$.

Vad händer om detta spel spelas i en population av individer, som anpassar sig med hjälp av den tumregel för inläring som vi arbetat med hittills? Man kan visa, att populationen aldrig kommer att röra sig mot jämvikt. Detta illustreras i figur 10. Varje punkt i triangeln står för en blandad strategi: en kombination av strategierna sten, påse och sax. I triangelns översta hörn spelar alla sten, i det nedre vänstra hörnet spelar alla påse och i det nedre högra hörnet spelar alla sax. Pilarna visar att populationen betar sig cykliskt. Vi ser att var vi än börjar så rör vi oss så småningom tillbaka mot ursprungspunkten.

Om de flesta spelar sten så tjänar de som istället spelar påse mest och växer som andel av befolkningen. När de flesta börjat spelar påse så tjänar de som spelar sax mest och växer som andel av befolkningen. När de flesta spelar sax så är det de som spelar sten som tjänar mest och på så vis återgår populationen till utgångsläget där de flesta spelade sten.²⁵

25 Inläring i sten, sax, och påse har studerats experimentellt av Cason m.fl. (2014). De finner stöd för att beteende cirkulerar på det sätt som förutsägs av teorin. Det här spelet verkar dessutom ha biologisk relevans: hanar av ödlearten *Uta stansburiana* ("side-blotched lizards") använder sig av tre reproduktionsstrategier som står i samma relation till varandra som strategierna i sten, sax och påse (Sinervo m.fl 1996). De olika hanarna har olika färg på halsen. Orangehalsade hanar kontrollerar stora territorier med flera honor. Gulhalsade hanar håller till i utkanten av de orange-halsades territorier och smyger in och parar sig med honor på de orange hanarnas territorier när tillfälle ges. Blåhalsade hanar vaktar en enskild hona. De kan försvara sig mot intrång från gulhalsade hanar men inte mot de orangehalsade. Om populationen främst består av orangehalsade hanar är det fördelaktigt att vara gulhalsad. Om populationen istället domineras av gulhalsade hanar är det fördelaktigt att vara blåhalsad. Om de blåhalsade dominerar är det fördelaktigt att vara orangehalsad (bilden kompliceras sedan ytterligare av att honorna har två olika strategier).

Figur 10. Inläring i sten, sax och påse.²⁶

En fallstudie

Fram till nu har vår diskussion huvudsakligen varit teoretisk, med vissa hänvisningar till laboratorie-experiment. Vi skall nu titta närmare på en studie av inläring utanför laboratoriet (Mohlin m.fl. 2020, se även Östling m.fl. 2011).²⁷

Under 49 dagar våren 2007 drev företaget Svenska Spel ett lotteri som de kallade Limbo. Något förenklat var spelet uppbyggt kring följande regler. Varje dag fick alla som ville delta. Deltagarna betalade en avgift och fick sedan välja ett nummer mellan 1 och 99 999. När alla hade gjort det identifierade man det enda – och lägsta – nummer som hade valts av endast en enda person. Vederbörande vann en stor summa pengar (i storleksordningen 100 000 kronor). Spelarna kände till dessa regler. I genomsnitt deltog mer än 50 000 spelare varje dag. Det går att bevisa att spelet har en unik jämvikt. I denna följer alla spelare samma blandade strategi, vilken uttrycks som den streckade linjen i figur 11(a) och 11(b). Denna blandade strategi lägger positiv sannolikhet på alla nummer mellan 0 och 99 999. Sannolikheten för att välja höga nummer är dock lägre

²⁶ This phase diagram was created using the software Dynamo (Sandholm & Dokumaci 2007).

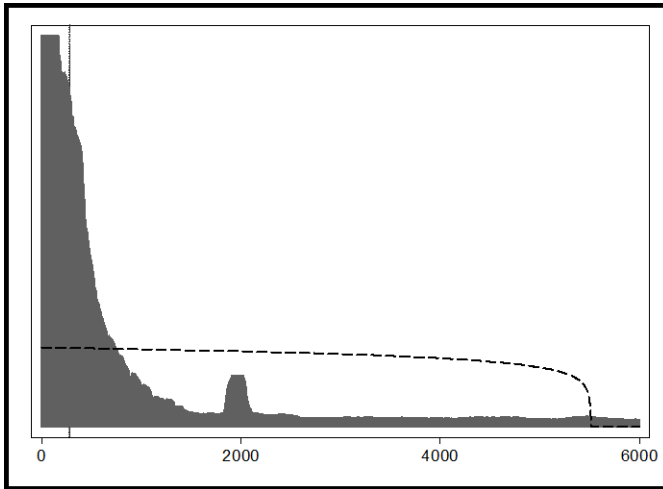
²⁷ Att studera inläring i spel utanför laboratoriet är svårt, bland annat eftersom man måste kunna avgränsa vad deltagarna uppfattar som det relevanta spelet. Därför är studier baserade på helt naturliga inläringssituationer ovanliga. Doraszelski m.fl. (2018) studerar dock den engelska energimarknaden och finner stöd för inlärningsdynamiker som är besläktade med dem vi diskuterat här.

än sannolikheten för att välja låga nummer. Sannolikheten för att välja de första 5 500 numren är avsevärt högre än sannolikheten för att välja högre nummer.

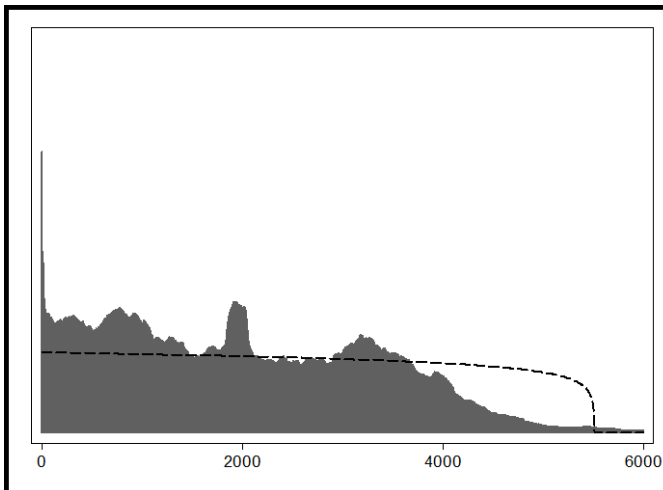
Det verkar orimligt att anta att spelarna skulle kunna resonera sig fram till vad jämvikten är. Och även om de hade kunnat det, så ligger det inte i deras intresse att bete sig i enlighet med jämvikten om de inte också tror, att alla andra kommer att bete sig på samma sätt. I figur 11(a) ser vi mycket riktigt att fördelningen av nummer som valdes den andra dagen (den grå massan i figur 11) av spelet avviker kraftigt från jämvikten. Men i figur 11(b) ser vi att förhållandet förändras avsevärt fram till den sista gången spelet genomfördes. Fördelningen av valda nummer under dag 49 ligger betydligt närmare jämvikten, även om det är en bit kvar till full jämvikt.

Figur 11. Inlärnin g och jämvikt i Limbo-spelet.

(a) Beteende dag 2



(b) Beteende dag 49 (sista dagen)

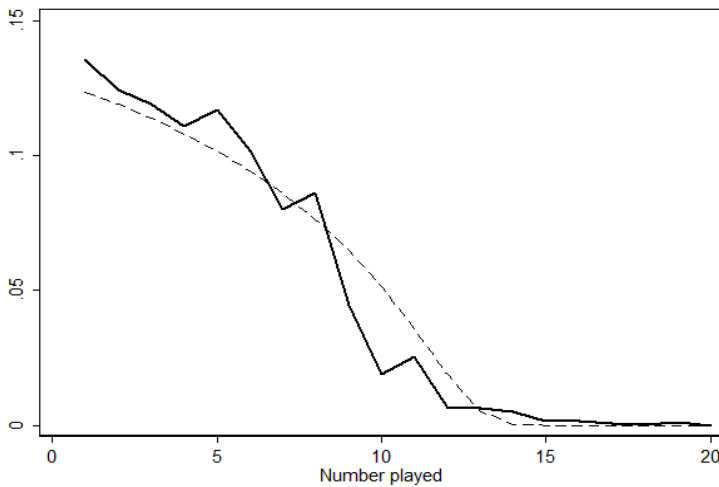


Hur går detta till? Spelarna har tillgång till information om vilka nummer som vunnit tidigare. Det verkar som att spelarna använder denna information som en indikation på vilka nummer som kan tänkas vinna i framtiden. De väljer nämligen nummer som ligger nära tidigare vinnare, och de är mer benägna att välja ett nummer, ju närmare de tidigare vinnarna detta nummer ligger.

För att undersöka samma spel under mera kontrollerade former, genomfördes ett experiment där studenter rekryterades och fick spela en version av Limbo-spelet på datorer 49 gånger i rad. Antalet nummer att välja mellan minskades från 99 999 till 99 och det genomsnittliga antalet deltagare minskade från över 50 000 per dag till i genomsnitt 26 stycken per omgång. Prissumman för varje omgång var 7 dollar (ungefär 50 kronor). Den enda information deltagarna hade var vilka nummer som hade vunnit i förgående spelomgångar.

Figur 12 jämför den teoretiska jämvikten (streckad linje) med det faktiska beteendet (svart linje) efter en period av inläring som motsvarar de första 20 omgångarna. Vi ser att inläring fört beteendet mycket nära jämvikten. Att experimentet använder färre nummer och ett mindre antal spelare, som utslutande fokuserar på just detta spel under en stund i en datorsal, verkar leda till snabbare inläring – det krävs färre omgångar av spelet för att beteendet skall hamna mycket nära jämviktstillståndet.

Figur 12. Experimentversionen av Limbo-spelet.



Komplikationer

EXTRAPOLERING MELLAN SPEL

I vår modell av inlärning har vi utgått från att individer från en population möts för att spela ett och samma spel gång på gång. De lär sig spela detta spel utan hänsyn till erfarenheter från andra spel. I verkligheten erfar vi sällan exakt samma typ av interaktion mer än en gång. Istället är vi delaktiga i ett otal mer eller mindre olika spel samtidigt. De innebär att våra möjligheter att genom inlärning hamna i jämvikt i ett enskilt spel beror på hur vi extrapolerar från erfarenheter i tillräckligt likartade spel. Även om spelteoretiker förstår detta så har det nästan inte alls studerats explicit.²⁸

Ett intressant undantag är Grimm & Mengel (2014). De låter experimentdeltagare spela antingen två eller sex olika spel i slumpvis ordning vid sammanlagt 100 tillfällen. Varje enskilt spel har en enda Nashjämvikt som det skulle vara lätt att konvergera mot, om experimentdeltagarna endast mötte en och samma strategi under de 100 tillfällena. Att hålla reda på två olika speltyper torde heller inte vara så svårt, men att däremot försöka överblicka sex olika spel skulle kunna vara kognitivt påfrestande. De sex olika spelen kunde emellertid klassificeras i tre par, där spelen i varje par hade samma strategiska struktur. Spel inom ett par kunde därför lämpligen användas för att lära sig det andra spelet i paret.

I experimentet visade det sig att en hög andel deltagare förmådde upptäcka att de sex olika spelen väsentligen kunde behandlas som tre olika kategorier av spel, och därmed lära sig att spela jämvikt. Detta förutsatte dock att deltagarna erhöll tillräckligt rik återkoppling om resultatet i tidigare perioder. När deltagarna fick mer begränsad återkoppling lärde de sig inte att spela jämvikt.²⁹

INTERAKTIONER UTAN MÖJLIGHET TILL INLÄRNING

Inte alla interaktioner ger deltagarna möjlighet att ackumulera erfarenhet som gör att de gradvis kan anpassa sitt beteende till dess att de eventuellt hamnar i en jämvikt. Många viktiga interaktioner beskrivs kanske till och med bäst som engångsföreteelser för de inblandade. Om jämviktsteorier endast är användbara för att beskriva beteende som grundats på erfarenhet och inlärning genom upprepade försök, vad skall vi använda för teori för strategiskt beteende innan inlärning ägt rum, eller i situationer där det saknas möjligheter att ackumulera tillräckliga erfarenheter att basera sina beslut på?

28 Förmodligen är huvudorsaken till att det inte studerats är att det är svårt. Dels är det svårt att bedöma vad som är en rimlig modell av hur erfarenheter extrapolerats (de teoretiska valmöjligheterna är mycket större än när man fokuserar på ett enskilt spel), dels är det svårt bestämma vad som är en realistisk klass av spel som gör inlärning "lagom" svårt.

29 Relaterat finner Mengel & Sciuuba (2014) att spelare lär sig fortare om de tidigare spelat ett spel med likartad strategisk struktur och lär sig långsammare om de tvärtom tidigare spelat ett spel med annorlunda (motsatt) strategisk struktur.

En teori som försöker göra reda för beteende innan spelare hunnit skaffa sig erfarenhet av ett spel är så kallad *level-k* teori. Teorin stipulerar att olika individer använder sig av olika tumregler som tillsammans bildar en hierarki. En individ på nivå 0 antas ofta välja likformigt slumpmässigt mellan sina strategier. En individ på nivå 1 betar sig som om den trodde att alla andra var på nivå 0, dvs en individ på nivå 1 väljer den strategi som är nyttomaximerande när motparten väljer likformigt slumpmässigt mellan sina strategier. En individ på nivå 2 betar sig som om den trodde att alla andra var på nivå 1, dvs som om alla andra trodde att alla andra var på nivå 0. Mer allmänt väljer en individ på nivå k en strategi som är nyttomaximerande, givet att alla andra betar sig som nivå $k-1$.³⁰

Avslutning

Vi har sett att inläring baserad på enkla tumregler i många viktiga spel leder till att en grupp individer så småningom betar sig i enlighet med teoretiska antaganden om jämvikt. Genom att studera de dynamiska processerna bakom kan vi avgöra vilka jämvikter som torde vara stabila och vilka som är instabila. Om ett spel har flera jämvikter så kan inläring och evolution ge en vink om vilken eller vilka jämvikter som det är troligast att man kommer att röra sig mot i det långa loppet. I vissa fall leder inläring och evolution dock inte till jämvikt. Det kan hända även om spelet har en enda jämvikt. Evolutionär spelteori och experimentella undersökningar av inläring i spel har därmed avsevärt förbättrat vår förmåga att bedöma hållbarheten i tillämpningar av klassisk spelteori.

Litteratur

- Axelrod, R., 1984. *The Evolution of Cooperation*. Basic Books.
- Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L., 2017. "Gender differences in accepting and receiving requests for tasks with low promotability", *American Economic Review*, 107(3), s. 714–47.
- Bernergård, A., & Mohlin, E., 2019. "Evolutionary Selection against Iteratively Weakly Dominated Strategies", *Games and Economic Behavior* 117 (Sept.), s. 82–97.
- Binmore, K. & Samuelson, L., 1994. "An economist's perspective on the evolution of norms", *Journal of Institutional and Theoretical Economics* 150 (1), s. 45–63.
- Björnerstedt, J. & Weibull, J., 1996. "Nash equilibrium and evolution by imitation", i K. J. Arrow, E. Colombatto, M. Perlman & C. Schmidt (red.), *The Rational Foundations of Economic Behaviour*. London: MacMillan.

30 *Level-k*-teorin introducerades av Nagel (1995) och Stahl & Wilson (1995). Camerer m.fl. (2004) utvecklade en besläktad teori kallad *cognitive hierarchy*. Crawford m.fl. (2013) är en grundlig genomgång av dessa modeller, såväl teori som evidens och tillämpningar. Man skulle kunna fråga sig vad som sker om fördelningen av nivåer i en population är föremål för inläring eller evolution, se Stahl (2000) och Mohlin (2012).

- Börgers, T. & Sarin, R., 1997. "Learning through reinforcement and replicator dynamics", *Journal of Economic Theory* 77(1), s. 1–14.
- Camerer, C., 2003. *Behavioral Game Theory*. Princeton: Princeton UP.
- Camerer, C., Ho, T-H., & Chong, J.K., 2004. "A Cognitive Hierarchy Model of Games", *Quarterly Journal of Economics* 119(3), s. 861–898.
- Cason, T. N., Friedman, D., & Hopkins, E., 2014. "Cycles and Instability in a Rock–Paper–Scissors Population Game: A Continuous Time Experiment", *Review of Economic Studies*, 81(1), s. 112–136.
- Crawford, V., Costa-Gomes, M., & Iriberri, N., 2013. "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications", *Journal of Economic Literature* 51 (March), s. 5–62.
- Dal Bó, P. & Fréchette, G. R., 2018. "On the determinants of cooperation in infinitely repeated games: A survey", *Journal of Economic Literature*, 56(1), s. 60–114.
- Doraszelski, U., Lewis, G., & Pakes, A., 2018. "Just starting out: Learning and equilibrium in a new market", *American Economic Review*, 108(3), s. 565–615.
- Fudenberg, D. & Levine, D. K., 1998. *The theory of learning in games*. Cambridge MA: MIT Press.
- Grimm, V. & Mengel, F., 2012. "An experiment on learning in a multiple games environment", *Journal of Economic Theory*, 147(6), s. 2220–2259.
- Heller, Y., & Mohlin, E., 2017. "Observations on cooperation", *The Review of Economic Studies*, 85(4), s. 2253–2282.
- Hofbauer, J. & Sandholm, W. H., 2011. "Survival of dominated strategies under evolutionary dynamics", *Theoretical Economics* 6(3), s. 341–377.
- Hofbauer, J. & Sigmund, K., 1988. *The Theory of Evolution and Dynamical Systems*. Cambridge: Cambridge University Press.
- Holm, H. J., 2000. "Gender-based focal points", *Games and Economic Behavior*, 32(2), s. 292–314.
- Kandori, M., 1992. "Social Norms and Community Enforcement", *Review of Economic Studies* 59, s. 61–80.
- Kandori, M., Mailath, G. & Rob, R., 1993. "Learning, Mutation and Long-Run Equilibria in Games", *Econometrica* 61, s. 29–56.
- Mengel, F., & Scitubba, E., 2014. "Extrapolation and structural similarity in games", *Economics Letters*, 125(3), s. 381–385.
- Mohlin, E., 2012. "Evolution of theories of mind", *Games and Economic Behavior*, 75(1), s. 299–318.
- Mohlin, E., Östling, R. & Wang, J. T.-Y., 2020. "Learning by Similarity-weighted Imitation in Games", *Games and Economic Behavior* 120 (March), s. 225–245.
- Nash, J., 1950. *Non-cooperative Games*, PhD Thesis, Princeton University.
- Nagel, N., 1995. "Unraveling in Guessing Games: An Experimental Study", *American Economic Review* 85 (5), s. 1313–1326.
- Nowak, M., & Sigmund, K., 1998. "Evolution of indirect reciprocity by image scoring", *Nature* 393, s. 573–577.
- Nowak, M., 2006. *Evolutionary Dynamics*. Cambridge MA: Harvard Belknap Press.
- Oprea, R., Henwood, K. & Friedman, D., 2011. "Separating the Hawks from the Doves: Evidence from continuous time laboratory games", *Journal of Economic Theory*, 146(6), s. 2206–2225.

- Sandholm, W. H., 2010. *Population games and evolutionary dynamics*. Cambridge MA: MIT Press.
- Sandholm, W. H. & Dokumaci, E., 2007. *Dynamo: Phase diagrams for evolutionary dynamics* (software suite), <http://www.ssc.wisc.edu/whs/dynamo>.
- Schlag, K. H., 1998. "Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits", *Journal of Economic Theory* 78(1), s. 130–156.
- Stahl, D. & Wilson, P., 1995. "On Players' Models of Other Players: Theory and Experimental Evidence", *Games and Economic Behavior* 10, s. 218–254.
- Stahl, D. O., 2000. "Rule learning in symmetric normal-form games: theory and evidence", *Games and Economic Behavior*, 32(1), s. 105–138.
- Sinervo, B. & Lively, C. M., 1996. "The rock-paper-scissors game and the evolution of alternative male strategies", *Nature* 380, s. 240–243.
- Taylor, P. D. & Jonker, L. B., 1978. "Evolutionary stable strategies and game dynamics", *Mathematical Biosciences* 40(1–2), s. 145–156.
- Trivers, R., 1971. "The evolution of reciprocal altruism", *Quarterly Review of Biology*, 46, s. 35–56.
- Weibull, J. W., 1995. *Evolutionary Game Theory*. Cambridge MA: MIT Press.
- Young, P., 1993. "The Evolution of Conventions", *Econometrica* 61, s. 57–84.
- Young, H. P., 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton NJ: Princeton University Press.
- Östling, R., Wang, J. T.-Y., Chou, E. Y. & Camerer, C. F., 2011. "Testing game theory in the field: Swedish LUPI lottery games", *American Economic Journal: Microeconomics* 3(3), s. 1–33.

Appendix: Biologisk evolution och social inläring

Vi antar en stor men ändlig population. Vid varje tidpunkt spelar varje individ en av de rena strategierna i ett givet spel. Antalet individer som spelar strategi i vid tid t är $p_i(t)$ och den totala populationens storlek är $p(t) = \sum_{i=1}^k p_i(t)$. Andelen som spelar strategi i är $x_i(t) = p_i(t)/p(t)$. Populationens tillstånd är en vektor $x(t) = (x_1(t), x_2(t), \dots, x_k(t))$. Notera att ett populationstillstånd svarar mot en blandad strategi $x(t) \in \Delta$. Låt e^i beteckna enhetsvektorn som motsvarar den "degenererade" blandade strategin som lägger sannolikheten ett på den rena strategin i . Förväntad nytta för en som spelar strategi i mot en slumpvist vald motspelare är $u(e^i, x)$. Den genomsnittliga förväntade nyttan i populationen är $u(x, x) = \sum_{i=1}^k x_i u(e^i, x)$.

BIOLOGISK EVOLUTION

Under en biologisk tolkning är varje individ "programmerad" att spela en av de rena strategierna. Vi antar ny att det vi tidigare kallat nytta nu representerar effekter på en individs "fitness" mätt som förväntad avkomma per tidsenhet. Varje avkomma ärver sin enda förälders strategi. Låt $\beta \geq 0$ stå för bakgrunds-"fitness" (oberoende av utfallet i spelet vi studerar). Om reproduktion äger rum kontinuerligt över tid så är fertiliteten (födelseakten) vid tidpunkt

t , för individer som är programmerade till den rena strategin i , $\beta + u(e^i, x(t))$. Låt mortaliteten $\delta \geq 0$ vara samma för alla individer. När $p(t)$ är stor nog, så kommer det genomsnittliga antalet avkomma för en individ som spelar strategi i att vara nära det förväntade antalet avkomma $\beta + u(e^i, x(t))$, och den faktiska andelen som dör kommer vara nära δ . Om vi använder punkter för att indikera derivata med avseende på tid så kan vi skriva ned den resulterande populationsdynamiken:

$$\dot{p}_i(t) = [\beta + u(e^i, e(t)) - \delta] p_i(t) = 3 - 3q$$

Notera att $p(t)x_i(t) = p_i(t)$ och derivera med avseende på tid på båda sidor.

$$\dot{p}(t)x_i(t) + p(t)\dot{x}_i(t) = \dot{p}_i(t)$$

Om vi flyttar om detta så får erhåller vi (efter lite arbete)

$$\dot{x}_i(t) = [\beta + u(e^i, x(t)) - \delta] x_i(t) - [\beta + u(x(t), x(t)) - \delta] x_i(t)$$

Om vi förenklar detta uttryck (och tar bort hänvisningen till tid) så får vi fram den så kallade replikatorodynamiken:

$$\dot{x}_i = [u(e^i, x) - u(x, x)]x^i$$

SOCIAL INLÄRNING

Under en social tolkning följer varje individ för närvarande en av de rena strategierna. Det vi tidigare kallade nytta bör nu tolkas som ett mått på individers observerbara "framgång" per tidsenhet. Framgång kan bestå i vadhelst spelarna strävar efter – exempelvis välstånd, status eller hälsa. När en aktör erhåller en möjlighet att revidera sin strategi (för enkelhetsskulle antar vi att de följer en Poissonfördelning med parameter 1 som är oberoende av individ och tidpunkt) så väljer hon slumpmässigt en annan individ att ha som sin (potentiella) förebild. Hon observerar förebildens strategi och nytta/framgång. Om hennes förebilds nytta är högre än hennes egen nytta så imiterar hon förebilden med en sannolikhet som står i proportion mot skillnaden i nytta mellan aktören och hennes förebild. Om aktören har en högre nytta än förebilden så byter hon inte strategi. Om en aktör som för närvarande använder strategi observerar en förebild som för närvarande använder strategi så kan vi uttrycka sannolikheten för att aktören byter till förebildens strategi som

$$[u(e^i, x(t)) - u(e^j, x(t))]_+.$$

Här indikerar det nedsänkta plustecknet att vi endast bryr oss om uttrycket inom hak-parenteserna ifall det är icke-negativt. Det följer att andelen av populationen som byter från strategi i vid tid t är

$$x_i(t) \sum_{j=1}^k x_j(t) [u(e^i, x(t)) - u(e^j, x(t))]_+,$$

och andelen av befolkningen som byter till vid tidpunkten är

$$\sum_{j=1}^k x_j(t) x_i(t) [u(e^i, x(t)) - u(e^j, x(t))]_+.$$

Om vi sätter samman dess uttryck, och tar bort hänvisning till tid, så erhåller vi

$$\begin{aligned} \dot{x}_i &= x_i \sum_{j=1}^k x_j [u(e^i, x) - u(e^j, x)]_+ \\ &\quad - x_i \sum_{j=1}^k x_j [u(e^j, x) - u(e^i, x)]_+ \\ &= x_i \sum_{j=1}^k x_j [u(e^i, x) - u(e^j, x)] \\ &= x_i [u(e^i, x) - u(x, x)] \end{aligned}$$

Återigen har vi härlett replikatorodynamiken.

Antagonistens roll

Vad Napoleon och KGB kan lära oss om beslutsfattandets konst inom militär ledning och underrättelseanalys

Tony Ingesson

Abstract

Professionalization is a growing trend in both military and intelligence organizations. However, unlike the traditional professions, both military decision-makers and intelligence analysts frequently find themselves locked in a struggle with an intelligent antagonist. One key aspect of professionalization, the teaching of structured, scientifically based knowledge and best practices, can lead to predictability. This creates a vulnerability that can be exploited by an antagonist. Using theories formulated by Clausewitz and Sun Tzu, together with examples from Napoleon's campaign in Northern Italy and the KGB's deception operation to conceal a mole in the CIA, it is shown how personal qualities, creativity and unpredictability rather than conformity are required to prevail against an intelligent, well-informed and capable antagonist. Thus, both the military decision-maker and the intelligence analyst have more in common with the chess player than the traditional professional.

Inledning

Inom såväl militärt beslutsfattande som underrättelseverksamhet har under senare tid funnits en ambition att professionalisera de yrkesverksamma, mer specifikt officerare respektive underrättelseanalytiker (se exempelvis Ledberg 2019 och Gentry 2016).¹ En del av denna ambition har kretsat kring det upplevda behovet av att i större utsträckning basera beslutsfattandet på vetenskapligt beprövade modeller, i likhet med hur evidens används

¹ Vissa militära organisationer, exempelvis svenska Försvarsmakten, går längre än bara officerare, och inkluderar i princip samtliga anställda i professionsbegreppet (se *Vår militära profession*, 2016).

inom det medicinska området (se exempelvis Wahlbergs kapitel i detta nummer). Förekomsten av en aktiv motståndare gör det dock extremt riskabelt att basera alla beslut på en vetenskaplig eller beprövad grund inom såväl militärt beslutsfattande som underrättelseanalys. Detta av samma anledning som en schackspelare inte kan uppnå elitnivå genom att enbart anamma beprövade strategier, eller en pokerspelare inte enbart kan spela på basis av sannolikhetskalkyler. Förekomsten av en aktiv motståndare, som kommer att utnyttja all dylik förutsägbarhet, innebär att beslutsfattandet oundvikligen måste vara en konstform lika väl som en mer vetenskapligt grundad verksamhet. Fenomenet kan vi benämna den *antagonistiska* dimensionen. Detta går bortom det som i medicinska sammanhang kallas *beprövad erfarenhet*, då det handlar om en förmåga att förstå, och i förlängningen idealt sett manipulera, en föränderlig, intelligent motståndare som aktivt eftersträvar samma sak. Således måste beslutsfattaren ständigt vara beredd att variera sitt angreppssätt genom att konstant läsa av sin motståndare. Detta fordrar, utöver förmågan att förstå, även ett mått av kreativitet.

DEN MILITÄRE BESLUTFATTAREN: EN IDEALTYP

För att kunna gå vidare med en analys behöver vi först klargöra vad som avses med begreppet "militär beslutsfattare". För detta kan vi använda oss av Webers (1921/1983, s. 7) definition av en idealtyp som "den rena typen" av en "vanlig företeelse". Denna renodling förtydligar och underlättar analysen, men läsaren måste hålla i åtanke att verkligheten endast ungefärligen överensstämmer med den idealtypiska konstruktionen. Vi kan för den kommande diskussionen utgå från följande idealtyp av den militäre beslutsfattaren: en person i någon form av chefsposition i en militär organisation, som direkt eller indirekt måste förhålla sig till någon form av motståndare. Detta förbiser de mer vardagliga sysslorna en militär chef också måste förhålla sig till, likväl som alla de militära chefer som har mer tekniska eller administrativa uppdrag. Om vi dock utgår från att en militär organisations huvudsakliga syfte är att skydda nationen mot olika typer av angrepp och hot från fientligt inställda aktörer, fångar denna idealtyp upp kärnan i det militära ledarskapet, medan i detta sammanhang mindre relevanta detaljer kan utelämnas. Detta passar även in på den svenska synen på den militära professionen, som lyfter fram att det mest centrala är Försvarsmaktens förmåga till väpnad strid (*Vår militära profession*, 2016).

DEN MILITÄRA PROFESSIONENS PROBLEM

Ett av de mest grundläggande problemen med den militära professionen som koncept, i relation till ovanstående idealtyp, ligger i själva definitionen av en traditionell profession. De kriterier Huntington (1964) tog upp i sitt tongivande verk *The Soldier and the State* var:

- en hög grad av generell och systematisk kunskap
- att primärt tjäna samhällliga intressen snarare än ett individuellt egenintresse
- en hög grad av självreglering genom internaliserade regelverk för yrkesetik, som beivras genom frivilliga sammanslutningar som drivs av de yrkesverksamma själva
- belöningar som främst är symboliska snarare än lukrativa

I en mer samtida publikation på professionstemat (se Brante 2011) kan vi skönja en annan uppsättning kriterier:

- det är en heltidssysselsättning av ett slag som inte är ett hantverk
- det etablerar ett monopol på arbetsmarknaden för experttjänster
- det är autonomt eller självreglerande (utan inblandning från lekmän eller staten)
- utbildningen är specialiserad, systematisk och akademisk
- behörigheten till yrket regleras genom specifika examina, titlar och motsvarande

Utöver ett antal problem som har att göra med det etiska regelverket och hur en militär profession skulle se på de olika personalkategorierna som ryms i en samtida militär organisation (se Ingesson 2016 och Ingesson 2018 för en mer detaljerad diskussion kring dessa aspekter) så finns det också en dimension som helt saknas i ovanstående listor på kriterier: antagonisten.

Till skillnad från en läkare eller advokat måste den militäre beslutsfattaren i regel förhålla sig till en antagonist, vilket särskiljer militärt beslutsfattande från de typer av verksamhet traditionella professioner är förknippade med.² Närvaron av en antagonist förändrar hela grunden för hur beslut kan bedömas och värderas. En läkare eller annan medlem av en traditionell profession kan försvara ett givet beslut med att det var det mest rationella i den givna situationen. En militär beslutsfattare kan däremot klandras för att ha agerat på samma sätt, då detta gjort denne förutsägbar och därmed skapat en öppning som en antagonist kan exploatera. Det som fordras går utöver det som kan formaliseras och formuleras i termer av regler och rutiner. Relationen till antagonisten är dessutom något den militäre beslutsfattaren har gemensamt med många yrkesverksamma kategorier inom underrättelsevärliden, vilket kommer att framgå av exemplen som presenteras nedan.

2 Advokaten kan förvisso sägas ha en egen antagonist i formen av åklagaren, men deras samspel är så reglerat av lagstiftningen och domaren/domarna att det inte kan liknas vid den tvekamp två militära beslutsfattare utkämpar.

Antagonistens roll

På samma sätt som en schackspelare kan studera spelöppningar, strategier och sannolikheter, kan en militär beslutsfattare studera taktik och ledarskap, medan underrättelseanalytikern kan studera metoder. Dessa färdigheter kan i olika mån underbyggas vetenskapligt. Exempelvis finns det gott om verk inom såväl krigsvetenskapens område (se Smedberg 2004, Van Creveld 1982, Fox & Burks 2019, m.fl.) som inom vetenskapligt förankrad metodutveckling för underrättelseanalys (se Heuer 1999, Coulthart 2017, m.fl.).³

DET MILITÄRA PERSPEKTIVET

Inom det krigsvetenskapliga området har mer eller mindre vetenskapligt förankrade rekommendationer och principer framförts i århundraden. Här finner vi exempelvis ett antal principer, såsom vikten av att sätta upp och hålla fast mål, behålla initiativet, eftersträva enkelhet, kraftsamla, säkerställa handlingsfrihet, osv. (se Smedberg 2004: 387). Det finns ofta en rationell och praktisk bas, byggd på såväl teoretisk grund som praktisk erfarenhet, bakom dessa idéer. Det ryska nederlaget vid Tannenberg 1914 kan således tillskrivas brister i *samordning* och *säkerhet*, medan tyskarnas framgångar vid samma slag kan förklaras med deras *kraftsamling* och *initiativ* (Smedberg 2004: 393).

Vissa numera klassiska koncept tar hänsyn till behovet av självständigt och kreativt tänkande, kanske mest renodlat i begreppet *uppdragstaktik*. Dess moderna tappning kan spåras tillbaka till sluttampen av första världskriget, då tyska *Stosstruppen* skördade framgångar på västfronten genom självständigt och flexibelt agerande i mindre grupper (Zetterling 2008: 17). Det var vid det laget alltför sent för Tyskland att vända kriget, men innovationen som lanserats gjorde ändå ett bestående intryck. Tyskarna själva fortsatte att experimentera med denna modell och införde sedan under mellankrigstiden den på bred front i hela tyska krigsmakten (Zetterling 2008: 19). Under andra världskriget fick hela världen bevittna hur effektiv denna ledningsmodell kunde vara, vilket inspirerade några andra nationer att anamma samma idé under kalla kriget, i synnerhet Sverige och Israel (se exempelvis *Arméreglemente del 2: Taktik* 1982 och Van Creveld 1985). I en del andra länder, exempelvis USA, blev det mer en läpparnas bekännelse där mer formaliserad och regelstyrd ledning fortsatt dominerade (se t.ex. Orsini 2019).

Uppdragstaktiken har blivit ett naturligt inslag i ledningsfilosofin hos ett antal länder idag, och dess syfte brukar beskrivas i termer av den grad av

3 Även spelteori har använts för att studera relationen mellan antagonister, men fokus ligger då ofta på nationalstatsnivå, snarare än den enskilde beslutsfattaren i en organisation, och kretsar mer kring förhandlingar och avskräckning än direkt konflikt (se exempelvis Schelling 1963). En viktig skillnad är den långsiktighet och storskalighet som präglar relationer mellan nationalstater, kontra de mer kortsiktiga och begränsade konflikter som utspelar sig mellan enskilda beslutsfattare längre ner i hierarkierna. Spelteori har heller inte fått samma genomslag inom utbildningen av beslutsfattare inom de berörda områdena, vilket gör den mindre relevant i det här sammanhanget.

flexibilitet den medför, vilket skapar möjligheter för militära chefer. Den amerikanska definitionen fokuserar exempelvis på den enskilde beslutsfattarens ansvar att uppmuntra initiativ, formulera ett övergripande syfte, och understödja sin personal i detta (*Field Manual 3-0*, 2017). I ett svenskt sammanhang brukar fokus likaledes ligga på chefens ansvar för att agera självständigt och kraftfullt. Här tas även upp att omdöme och självsäkerhet är centrala aspekter (*Militärstrategisk doktrin* 2016: 68).

Ett annat samtida exempel är *multi-attribute military decision-making*, ett begrepp som syftar på en modell för att prioritera bland ett antal olika alternativ och beslutskriterier (Fox & Burks 2019: 195). Det finns vidare enligt Fox och Burks fyra metoder för att uppnå detta syfte: *Data Envelopment Analysis* (DEA), *Simple Average Weighting* (SAW), *Analytical Hierarchy Process* (AHP), samt *Technique of Order Preference by Similarity to Ideal Solution* (TOPSIS). Utan att gå in på de mer tekniska detaljerna kan vi konstatera att detta tillvägagångssätt primärt bygger på matematisk analys av preferenshierarkier och dylikt.

Det som framgår av ovanstående kortfattade översikt är att klassisk krigsvetenskaplig metodutveckling i något högre grad tar hänsyn till den antagonistiska dimensionen i militärt beslutsfattande. Mer samtida perspektiv, och i synnerhet de som är associerade med en professionssträvan, tar endast i begränsad mån hänsyn till det faktum att en antagonist kan förväntas känna till samma kunskapsstoff. Om vi således anammar dessa principer och metoder slentrianmässigt blir vi *förutsägbara* och därmed skapar vi en sårbarhet som kan exploateras. Uppdragstaktiken ger oss en del av lösningen på det här problemet genom att betona vissa av de egenskaper som fordras hos en militär chef, och den grad av flexibilitet som kan krävas. Det är dock relativt lite som sägs om just antagonisten. Hur ska vi förhålla oss till denne? För att besvara denna fråga lämnar vi de verk som har mer vetenskapliga anspråk, och gräver istället djupare i krigsvetenskapens mer filosofiskt orienterade klassiska arbeten.

Antagonisten i militärt beslutsfattande

Historiskt sett har de klassiska studierna av militärt beslutsfattande betonat vikten av den enskilde beslutsfattarens förmåga att överlista sin motståndare. Clausewitz (1832/1989: 75) liknade ett krig vid en tvekamp, där två motsatta parter medelst våld försöker tvinga den andre till underkastelse.⁴ Han beskrev även det han kallade för *den militära genialitetens essens* (Clausewitz 1832/1989: 100). Enligt Clausewitz måste en god militär ledare besitta mod,

4 Här bör det poängteras att tvekan är tesen i Clausewitz dialektiska resonemang kring krigets natur, dess antites är det mer välkända påståendet att kriget är en fortsättning av politiken med andra medel. Både tesen och antitesen ingår följaktligen i syntesen. Se Bassford 2007 för en mer detaljerad diskussion om detta.

styrka och intellektuell spänst. De förstnämnda kriterierna är en nödvändighet för att kunna hantera krigets oundvikliga påfrestningar, medan det sistnämnda kriteriet behövs för att utveckla det omdöme som krävs för att kunna hantera den osäkerhet som Clausewitz menade skyler tre fjärdedelar av de faktorer på vilka beslut baseras.

Det intellekt Clausewitz efterlyste kännetecknas av två egenskaper: *coup d'oeil* och beslutsamhet. Det förstnämnda definierar han, i något poetiska ordalag, som förmågan att även i de mörkaste stunder hålla fast vid det inre ljus som leder till sanning (Clausewitz 1832/1989:101-102). Den andra egenskapen är modet att följa detta ljus varhelst det må leda (av Clausewitz benämnt *courage d'esprit*). Enligt Clausewitz reflekterar detta mod inte en militär befälhavares intellektuella förmåga, utan är snarare ett uttryck för ett specifikt temperament (*ibid.*).

Bägge dessa egenskaper kan förefalla svårgripbara och svepande, men deras relevans i relation till en antagonist framgår desto tydligare om vi jämför med exempelvis Ulfvings (2000) resonemang om den ryska doktrinen för militär vilseledning. Enligt denna syftar vilseledning till att exploatera motståndarens förutfattade meningar, och förmå denne att agera på ett sätt som gynnar den egna planen med hjälp av skapade illusioner. Här ser vi konkret vikten av att skilja på det sanna respektive falska. De egenskaper Clausewitz framhåller är viktiga inte bara för att undvika att aktivt låta sig vilseledas, utan även för att förhindra passivitet och förlust av initiativ som ett resultat av sådana dimridåer.

Enligt en annan, men betydligt äldre klassisk krigsvetenskaplig tänkare, Sun Tzu (2008: 28-29), verksam på 500-talet före Kristus, bör en befälhavare inte bara bör kunna genomskåda fientlig vilseledning, utan även kunna tillämpa samma metoder själv. Sun Tzu (2008: 15) menar även att en skicklig militär beslutsfattare bör kunna uppnå resultat utan att bruka våld.

Sammantaget är samspelet mellan den militäre beslutsfattaren och antagonisten centralt för bägge dessa tänkare. De egenskaper såväl Clausewitz som Sun Tzu efterlyser ligger bortom de vetenskapligt förankrade metodernas och praktikernas plan. Istället är det som fordras en kombination av talang och personliga egenskaper som närmast påminner om den skicklige schackspelarens. Vi ska härnäst se hur detta kan ta sig konkret uttryck i beslut fattade av en av världshistoriens mest välkända och framgångsrika fältherrar, Napoleon.

FALLSTUDIE: NAPOLEONS KRIGSLISTER

Det är inte utan anledning som Napoleon Bonaparte är en av de mest välkända militära beslutsfattarna i historien. Den kanske mest intressanta aspekten av hans framgångar är att de i hög grad var ett resultat av hans förmåga att överlista sin motståndare, snarare än någon form av teknologiskt eller numerärt övertag (även om en av de stora tillgångarna med hans *Grande Armée* var,

som namnet antyder, just dess storlek, vilket möjliggjordes av hans innovativa ledningsorganisation).

Ett tidigt exempel på Napoleons förmåga att utmanövrera sina motståndare inträffade 1796 under första koalitionskriget, när vapenstilleståndsavtalet vid Cherasco skulle undertecknas av Frankrike respektive kungariket Sardinien. I avtalet fanns en hemlig klausul som skulle ge franska trupper fri passage över en bro över floden Terdoppio i staden Valenza. Detta snappades upp omgående av Österrike-Ungerns spioner i det sardinska hovet, som tolkade det som att det fanns en risk att fransmännen planerade att använda bron för att inom en snar framtid anfälla Österrike-Ungern. Följaktligen placerade de ut trupper för att skydda denna passage. Därmed hade Österrike-Ungern åtagit sig att befästa inte mindre än tre olika möjliga anfallsriktningar. Detta var exakt vad Napoleon hade planerat från början. Den enda anledningen till att den aktuella hemliga klausulen existerade var för att lura Österrike-Ungern att tro att det var en del av en fransk anfallsplan (Rummage 2019a), då Napoleon av allt att döma var väl medveten om att informationen skulle läcka. Detta förstärkte Napoleon sedan med en mer konventionell form av vilseledning, då han samlade styrkor vid Valenza för att ge intryck av att något var på gång där (Boycott-Brown 2001: 293-294). Hans motståndare tvingades följaktligen att avdela trupper för att skydda mot en flodövergång som aldrig var tänkt att ske, och gav därmed Napoleon ett övertag som han utnyttjade till fullo genom att istället korsa floden vid Piacenza. Därmed hade han försatt motståndaren i en hopplös situation, och tvingade denne att retirera. Detta hade inte varit möjligt utan en god förståelse för både motståndaren Österrike-Ungern, och samspelet mellan motståndaren och det besegrade Sardinien.

En annan av Napoleons krigslistor var manipulationen av den unge kung Fredrik Vilhelm III av Preussen under andra koalitionskriget. Preussen försökte upprätthålla sin neutralitet under kriget, men Napoleon hade identifierat en öppning i Preussens historiska anspråk på Kurfurstendömet Braunschweig-Lüneburg. Detta innebar att det fanns en grund för konflikt mellan Preussen och Storbritannien, då den brittiske kungen Georg III också var kurfurste av Braunschweig-Lüneburg. Napoleon lade således fram ett ultimatum: om inte Preussen invaderade Braunschweig-Lüneburg så skulle Frankrike göra det istället (Rummage 2019b). Även om Preussen var ovilligt att inträda i kriget, dessutom som fiende till Storbritannien (vilket i princip skulle göra Preussen till Frankrikes allierade) så insåg Napoleon att Fredrik Vilhelm III var än mer ovillig att helt avstå från att kunna göra anspråk på Braunschweig-Lüneburg i framtiden. Alltså invaderade Fredrik Vilhelm III Braunschweig-Lüneburg, och Napoleon hade därmed slagit in en kil mellan Storbritannien och Preussen, som bara några år tidigare varit allierade.

Ovanstående exempel är bara två detaljer från Napoleons långa karriär, som var fylld av mängder av framgångsrika beslut, lika väl som ett antal

ödesdigra misstag. Icke desto mindre kan de fungera som en illustration över denna fältherres sällsynta förmåga att läsa av sin antagonist, och förutspå dennes reaktion. Detta kunde Napoleon dessutom exploatera till fullo, och kombinera med sina andra talanger, vilket torde vara en viktig del av förklaringen till hans anmärkningsvärda eftermäle. Napoleons motståndare var väl införstådda med krigsvetenskapens grundprinciper, men blev icke desto mindre utmanövrerade av den unge generalen. Napoleon besatt således den klarsynthet Clausewitz efterlyste, som både lät honom läsa av sin motståndare och förstå hur denne skulle manipuleras. Napoleon hade dessutom det mod som krävdes för att agera beslutsamt för att sätta dessa insikter i verket. Exemplet med Fredrik Vilhelm III visar att Napoleon även, i likhet med Sun Tzus idealiske befälhavare, kunde förmå andra att agera i enlighet med sina planer utan att tillgripa våld.

Antagonisten i underrättelsevärlden

Underrättelsevärlden kan närmast liknas vid en slags lågintensiv kamp, om än en i regel relativt oblodig sådan. I synnerhet inom de verksamheter som bedrivs av stater och syftar till att inhämta information om andra stater eller sådant som de ogärna vill dela med sig av (exempelvis militär förmåga, tekniska landvinningar eller motsvarande) finns en konfliktdynamik som inte är helt olik den Clausewitz menar var kännetecknande för kriget.

UNDERRÄTTELSEANALYTIKERN: EN IDEALTYP

Det finns många olika typer av underrättelseanalytiker, som är verksamma i ett antal olika sammanhang. Spannet kan sägas sträcka sig från den utrikesorienterade och säkerhetspolitiskt präglade underrättelseverksamheten som bedrivs i statens regi, till den kriminalunderrättelsetjänst som rättsvårdande myndigheter ägnar sig åt, och vidare till den typ av informationsinhämtning och analys av konkurrenter och marknader som vi kan finna hos större företag. För denna analys kommer fokus att ligga på den typ av underrättelseanalys som har en tydlig motståndare att förhålla sig till. Denna kan vara en nationalstat, en kriminell aktör, eller en konkurrent. Det de har gemensamt är att de alla fungerar som en antagonist, dvs. de kan förutsättas ha motsatta intressen och ett incitament att försvåra eller förhindra det egna analysarbetet.

UNDERRÄTTELSEANALYSENS VETENSKAPLIGA GRUND

Underrättelseanalysens vetenskapliga förankring har främst kretsat kring risken för att individuella eller gruppbaseade psykologiska mekanismer underminerar analysen av inhämtade underrättelser, och därmed leder till felaktiga

slutsatser.⁵ Ett klassiskt exempel på detta är den roll fenomenet *groupthink* tillskrivs i relation till Israels misslyckande med att förutse Yom Kippur-krigets utbrott 1973 (se exempelvis Bar-Joseph 2005). Ett annat är Heuers (1999) arbete med att utveckla nya analysmetoder för att reducera inverkan av såväl denna typ av gruppdynamik, som mer individuella faktorer. Detta har sedan bl.a. resulterat i utvecklingen av de s.k. *Structured Analytic Techniques* (SAT-teknikerna) som publicerats öppet (se *A Tradecraft Primer* 2009).

Liksom i kriget finns den ständigt närvarande antagonisten. Clausewitz idé om tvekampen ligger därmed nära till hands även här. Ändå är antagonisten liksom i mer samtida militär teoriutveckling ofta förpassad till en sekundär roll. Detta leder till samma grundproblem: i likhet med kriget kan den kvalificerade motståndaren antas försöka utnyttja förutsägbarhet och fantasilöshet för att vinna egna fördelar. Slutligen, beslutsamhet och förmågan att se "sanningen" spelar även här en viktig roll, som vi ska se nedan.

REFLEXIVE CONTROL

I "spelet" mellan underrättelseorganisationerna rörande spionage och kontra-spionage ser vi tydligt vikten av kreativitet och förmåga att läsa av och förstå den andre för att erhålla ett övertag. I sovjetisk/rysk doktrin kan detta ses i begreppet *Reflexive Control*, som utvecklades under kalla kriget. Detta innebär att man använder subtila instrument för att manipulera motståndarens besluts-situation genom desinformation, inducerad stress, hot och olika typer av påverkan (t.ex. genom att förhindra vissa beslutsvägar) för att skapa en situation där det beslut som förefaller mest rationellt för motståndaren sammanfaller med det utfall den egna sidan redan på förhand definierat som mest önskvärt (Thomas 2004). På så vis förmås motståndaren att, utan att inse det själv, agera i strid med sina egna intressen. *Reflexive Control* var välkänt som metod inom ryskt militär- och underrättelseväsende under kalla kriget (Hosaka 2019: 327). I följande fallstudie kan vi se hur detta kunde få ödesdigra följder för en motståndare som inte hade förmåga att genomskåda det.

FALLSTUDIE: SPELET KRING KGB:S MULLVAD

En god illustration av den antagonistiska dimensionen i underrättelsevärlden är spelet mellan CIA och KGB under 80-talet, i samband med att den sistnämnda organisationen försökte dölja att den hade en källa i CIA:s avdelning för kontraspionage riktat mot Sovjetunionen. CIA hade under senare delen av 70-talet och början av 80-talet byggt upp ett relativt stort nätverk av källor i Sovjetunionen och Östeuropa, men efter att CIA-tjänstemannen Aldrich Ames

5 Det finns dock även många andra ämnen som ryms inom underrättelseanalysens vetenskapliga dimension; för ett bredare urval av exempel, se *Statsvetenskaplig tidskrift*, vol. 3, 2020.

rekryterades av KGB 1985 gick det snabbt utför (Fischer 2011). Redan under sitt första år som KGB-spion avslöjade Ames tio av CIA:s och FBI:s mest värdefulla sovjetiska källor. Totalt under 1985–1986 röjdes tjugo operationer, vilket resulterade i ett antal gripanden och avrättningar. Framåt senare delen av 1986 hade CIA:s operationer i Sovjet kollapsat (*An Assessment...*:15–16).

Enligt en utredning som publicerades 1994 har ingen amerikansk spion någonsin åsamkat lika stor skada som Ames (*An Assessment...*:53). Det faktum att Ames kunde fortsätta verka inom CIA under nästan nio år innan han slutligen greps förklaras enligt utredningen främst av tillkortakommanden inom CIA. Utredarna medger dock också att KGB förtlöpande ägnade sig åt vilseledning i syfte att avleda uppmärksamheten från Ames. Ett exempel är från juni 1988, då KGB satte en sådan plan i verket, genom att information "läcktes" till CIA att fem av de avslöjade operationerna i själva verket röjts på grund av bristande *tradecraft* hos antingen CIA-tjänstemännen eller källorna själva.⁶

Detta villospår analyserades under inte mindre än två år, och förbrukade under denna tid avsevärda resurser som kunde ägnats åt att hitta den riktiga läckan, innan det slutligen kunde avfärdas (*An Assessment...*:32). Redan 1985 inledde KGB en annan vilseledningsoperation för att skydda Ames. I denna användes en dubbelagent (en källa som rekryterats av CIA men i själva verket arbetade för KGB) för att få CIA att tro att KGB utvecklat ett nytt kemiskt ämne, kallat *Metka*, som kunde användas för att "markera" CIA-personal och därmed följa deras spår i efterhand utan deras kännedom (Fischer 2011: 276).

Detta var, som all god vilseledning, baserat i viss del i sanning. Under 60-talet hade KGB experimenterat med ett markeringsämne, och använt det i viss utsträckning. 1985 hade det dock tjänat ut sin roll, och kunde därmed "offras" som en s.k. *throw-away* genom att KGB påstod att de vidareutvecklat ämnet för att göra det mer effektivt. Dubbelagenten tillhandahöll t.o.m. ett prov på ämnet, och CIA kunde därefter upptäcka det på flera viktiga platser. Det som inte framgick var att KGB förmodligen hade kontaminerat dessa platser med ämnet just för att CIA skulle upptäcka det (Fischer 2011: 276). Även denna avledning fungerade som avsett, och fick CIA att lägga tid och möda på att hantera ett icke-existerande problem, som även kunde användas som förklaring för de röjda operationerna.

Ovanstående exempel visar hur KGB på ett skickligt sätt kunde läsa av och manipulera CIA:s analytiker genom att anspela på deras förutfattade meningar, exempelvis deras benägenhet att tro på mer tekniska förklaringar snarare än att acceptera att det fanns en spion i deras innersta krets. Denna typ av manipulation, som fortgick under flera år i följd med hjälp av olika angreppssätt, fordrar

6 Begreppet *tradecraft* avser här de procedurer som källor och deras kontaktpersoner använder för att undvika upptäckt. Detta inkluderar bland annat s.k. "brevlådor" (*dead drops*) för att utväxla information, utrustning och kontanter; metoder för att upptäcka och undkomma övervakning; kommunikationsrutiner, osv.

inte bara kunskap om antagonisten, utan även ett visst mått av kreativitet. Det är ingen slump att sovjetisk doktrin understryker vikten av originalitet i planering och genomförande av vilseledningsoperationer (se Ulfving 2000: 65).

Slutsatser

Ovanstående exempel illustrerar den roll den antagonistiska relationen mellan beslutsfattare och motståndare spelar i såväl militärt beslutsfattande som i underrättelsesammanhang. I bägge fallen finns det mer eller mindre vetenskapligt förankrade metoder, rutiner och doktriner för att vägleda beslutsfattare. Samtidigt präglas den praktiska verksamheten i hög grad av samspelet mellan beslutsfattaren/analytikern och en aktiv antagonist som förutsätts ha tillgång till samma kunskapsbas. Således kan ingen av dessa verksamheter reduceras till evidensbaserade rutiner eller beprövad erfarenhet, då detta hade öppnat upp en möjlighet för antagonisten att exploatera den förutsägbarhet som följer av ett alltför regelstyrt uppträdande. Det var detta såväl Napoleon som KGB utnyttjade i exemplen ovan.

Utöver de två fallstudier som behandlats ovan kan vi även anta att samma dynamik uppträder i fler sammanhang. Detta gäller mängder av militära eller underrättelserelaterade scenarier, men även i andra situationer som kan betraktas som en tvekamp. Exempelvis relationen mellan företaget och konkurrenten, eller för all del mellan två utövare av kampsport under en match.

Vi landar återigen där vi började: med schackmetaforen. En skicklig schackspelare kan förutsättas ha läst på om klassiska spelöppningar och historiska partier. Samtidigt måste spelaren också sätta sin personliga prägel på sitt spel, för att på så vis kunna överlista motspelaren. Detta var nyckeln till Garry Kasparovs framgångar som schackspelare, då hans personliga stil med omfattande förberedelser och aggressiva spelöppningar skördade stora framgångar. Samtidigt blev detta också hans akilleshäla, när Vladimir Kramnik insåg hur Kasparovs spelstil kunde kontrastas med skraddarsydda strategier (Byrne 2000). Den personliga prägel skapar en oförutsägbarhet jämfört med de konventionella, allmänt kända ramarna, men samtidigt måste spelaren ständigt kunna förnya sig och anpassa sig efter varje individuell motståndare, för att upprätthålla oförutsägbarheten.

Det är talande att doktrinen om *Reflexive Control*, som visat sig vara så användbar för ryskt militär- och underrättelseväsende, i princip återfinns redan hos Napoleon. Både vid övergången av Po, och vid manipulationen av Fredrik Vilhelm III, visade Napoleon prov på en förmåga att forma sin motståndares beslutsramar på ett sätt som resulterade i utfall som gynnade Napoleon själv. I detta avseende har hans agerande mycket gemensamt med KGB:s vilseledning av CIA under 1980-talet. En motpart som resonerar konventionellt, i enlighet

med transparenta rationella ramar, hamnar i underläge gentemot den som har förmågan att tänka ”utanför boxen”.

Sammantaget belyser detta en i alltför hög grad förbisedd aspekt av såväl militärt beslutsfattande som underrättelseanalys: hur ska den enskilde befälhavaren eller analytikern förhålla sig till det faktum att dessa verksamheter så ofta kretsar kring ett samspel med en antagonist? Som ovanstående fallstudier visar så räcker det inte med att hålla sig till allmänna principer eller vetenskapligt förankrade metoder för att kunna förutse och kontra dennes drag. Det som i slutändan krävs är utöver detta en djupare förståelse för motståndaren, förmågan att läsa av densamme, samt tillräcklig klarsyn för att genomskåda försök till vilseledning. Den som dessutom har den kreativitet och originalitet som krävs för att själv lägga ut dimridåer och förvilla sin antagonist kan uppnå de verkligt stora framgångarna.

Referenser

- A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis*, 2009. Central Intelligence Agency. <https://www.cia.gov/library/publications/publications-rss-updates/tradecraft-primer-may-4-2009.html>.
- An Assessment of the Aldrich H. Ames Espionage Case and its Implications for U.S. Intelligence*, 1994. The Select Committee on Intelligence, United States Senate. Washington: U.S. Government Printing Office.
- Arméreglemente del 2: Taktik*, 1982. M7741-100611. Stockholm: Chefen för armén och Försvarets läromedelscentral.
- Bar-Joseph, Uri, 2005. *The Watchman Fell Asleep: the Surprise of Yom Kippur and its Sources*. Albany: SUNY Press.
- Bassford, Christopher, 2007. ”Teaching the Clausewitzian Trinity”, *Clausewitz.com*, <https://www.clausewitz.com/readings/Bassford/Trinity/TrinityTeachingNote.htm>.
- Boycott-Brown, Martin, 2001. *The Road to Rivoli: Napoleon's First Campaign*. London: Cassell & Co.
- Brante, Thomas, 2011. ”Professions as Science-Based Occupation”, *Professions & Professionalism* 1(1), s. 4-20.
- Byrne, Robert, 2000. ”CHESS; How Kramnik Kept Kasparov Off His Game”, *New York Times*, 5 november 2000. <https://www.nytimes.com/2000/11/05/nyregion/chess-how-kramnik-kept-kasparov-off-his-game.html>.
- Clausewitz, Carl von, 1832/1989. *On War*. Övers. Howard & Paret. Princeton: Princeton University Press.
- Coulthart, Stephen J., 2017. ”An Evidence-Based Evaluation of 12 Core Structured Analytic Techniques”, *International Journal of Intelligence and CounterIntelligence*, 30(2), s. 368-391.
- Field Manual 3-0: Operations*, 2017. Washington D.C.: Headquarters, Department of the Army.
- Fischer, Benjamin B., 2011. ”Spy Dust and Ghost Surveillance: How the KGB Spooked the CIA and Hid Aldrich Ames in Plain Sight”, *International Journal of Intelligence and CounterIntelligence*, 24(2), s. 268-306.

- Fox, William P. och Burks, Robert, 2019. *Applications of Operations Research and Management Science for Military Decision Making*. Cham: Springer Nature.
- Gentry, John A., 2016. "The 'Professionalization' of Intelligence Analysis: A Skeptical Perspective", *International Journal of Intelligence and CounterIntelligence*, 29(4), s. 643-676.
- Heuer, Richards J. Jr., 1999. *Psychology of Intelligence Analysis*. Washington D.C.: Center for the Study of Intelligence, Central Intelligence Agency.
- Hosaka, Sanshiro, 2019. "Putin the 'Peacemaker'? Russian Reflexive Control During the 2014 August Invasion of Ukraine", *Journal of Slavic Military Studies*, 32(3), s. 324-346.
- Huntington, Samuel, 1964. *The Soldier and the State: The Theory and Politics of Civil-Military Relations*. New York: Vintage Books.
- Ingesson, Tony, 2016. *The Politics of Combat: the Political and Strategic Impact of Tactical-Level Subcultures, 1939-1995*. Doktorsavhandling. Lund: Statsvetenskapliga institutionen och Samhällsvetenskapliga fakulteten.
- Ingesson, Tony, 2018. "When the Military Profession Isn't", i Mayfield, Tyrell & Finney, Nathan (red.). *Redefining the Modern Military: the Intersection of Profession and Ethics*. Annapolis: Naval Institute Press.
- Ledberg, Sofia K., 2019. *Officeren, staten och samhället: ett professionsperspektiv*. Lund: Nordic Academic Press.
- Militärstrategisk doktrin*, 2016. M7739-354028. Försvarsmakten. <https://www.forsvarsmakten.se/siteassets/4-om-myndigheten/dokumentfiler/doktriner/militarstrategisk-doktrin-2016-ny.pdf>.
- Orsini, Ryan, 2019. "U.S. Army Mission Command at a Crossroads", *The Strategy Bridge*, 29 oktober 2019, <https://thestategybridge.org/the-bridge/2019/10/29/us-army-mission-command-at-a-crossroads>.
- Rummage, E.M., 2019a. *Age of Napoleon: episode 27*. Podcast. <https://ageofnapoleon.com>.
- Rummage, E.M., 2019b. *Age of Napoleon: episode 63*. Podcast. <https://ageofnapoleon.com>.
- Schelling, Thomas C., 1963. *The Strategy of Conflict*. New York: Galaxy Books.
- Smedberg, Marco, 2004. *Krigföring: från Austerlitz till Bagdad*. Lund: Historiska Media.
- Sun Tzu, 2008. "The Art of War", övers. Lionel Giles, i *The Complete Art of War*. Radford: Wilder Publications.
- Thomas, Timothy L., 2004. "Russia's Reflexive Control Theory and the Military", *Journal of Slavic Military Studies* 17(2), s. 237-256.
- Ulfving, Lars, 2000. *Den stora maskeraden: Sovjetrysk militär vilseledning*. Stockholm: Förvarshögskolan.
- Van Creveld, Martin, 1982. *Fighting Power: German and U.S. Army Performance, 1939-1945*. Westport: Greenwood Press.
- Van Creveld, Martin, 1985. *Command in War*. Cambridge: Harvard University Press.
- Vår Militära Profession – Agerar när det krävs*, 2016. FM2015-1597:7. Försvarsmakten. <https://www.forsvarsmakten.se/siteassets/4-om-myndigheten/vart-uppdrag/underbilaga-1-1-fm2025-1597-7-fmsi-2015-a1-var-militara-profession-med-var-vardegrund.pdf>.
- Weber, Max, 1921/1983. *Ekonomi och samhälle: förståendesociologins grunder*. Vol. 1. Lund: Argos.

Zetterling, Niklas, 2008. *Blixtkrig: 1939-1941*. Stockholm: Månocket.

Översikter & meddelanden

Political Hedgehogs

The Geographical Sorting of Refugees in Sweden¹

Johan Wennström & Özge Öner

Abstract

In sharp contrast to how admitted refugees are being placed in comparable European countries, refugees in Sweden have been disproportionately placed in peripheral and rural areas with high unemployment and rapid native depopulation where the prospects for integration, both socially and economically, are poor. We explore and evaluate some potential reasons for this outcome. Factors such as an intimidating political and intellectual climate in favor of receiving large numbers of asylum seekers and immigrants and the economic support given by the central government to municipalities that accept refugees are not sufficient to understand the actions of rural local governments. Instead, we argue that Tetlock's fox/hedgehog model of predictive style may provide a useful approach for understanding the seemingly irrational actions of local politicians in rural and peripheral municipalities, and of other policymakers in other similar situations.

Introduction

In recent years, Sweden has received unusually large numbers of refugees. In 2015, the top year for refugee admission, 162,900 asylum seekers arrived in the country, which corresponds to 1.6 percent of the total population and exceeds the number of native births by almost 50 percent. In sharp contrast to how admitted refugees are being placed in comparable European countries,² refugees who

1 We are grateful for useful comments and suggestions from Magnus Henrekson and Johan Tralau. We also gratefully acknowledge financial support from the Catarina and Sven Hagströmer Foundation (Wennström) and *Jan Wallanders och Tom Hedelius stiftelse* (Öner).

2 For example, in Germany, refugees are concentrated in the most densely populated and urban areas of the country (Katz et al. 2016).

Johan Wennström (PhD Political Science) works at the Research Institute of Industrial Economics (IFN). Özge Öner (PhD Economics) works at the Department of Land Economy, University of Cambridge, and IFN.
E-mail: johan.wennstrom@ifn.se; 00263@cam.ac.uk

have been granted a residence permit in Sweden have been disproportionately placed in peripheral and rural areas with high unemployment and rapid native depopulation, where the prospects for integration, both socially and economically, are poor. Relative to their populations, peripheral and rural municipalities with declining populations have received more refugees than growing urban municipalities with expanding employment opportunities.

As we explore at length in an extended working paper version of this article that is available open-access (Wennström & Öner 2019), which presents the above-mentioned empirical results in detail, this pattern seems both counter-intuitive and counterproductive given that labor market integration is more likely to be achieved in larger and more diverse labor markets. In the current study, we explore and evaluate some potential reasons for why the pattern has manifested itself.

It can be partly explained by the fact that Sweden in international comparisons “has been ranked in the group with the politically and functionally strongest local government forms in Europe” (Wollmann 2004: 647). Thus, and in the absence of legislation forcing municipalities to receive large numbers of refugees per capita, metropolitan municipalities have simply been able to choose to accept fewer refugees. However, the question remains why governments of peripheral and rural municipalities would voluntarily accept more refugees per capita than governments of urban municipalities with better labor market prospects.

We first consider whether a structurally oriented rational choice theory can offer an exhaustive, or, at least, a proximate explanation. However, we argue that factors such as the disciplinary political and intellectual climate in favor of receiving large numbers of asylum seekers and immigrants and the economic support given by the central government to municipalities that accepted refugees—clearly considerations that rational agents would need to address in this context—are not sufficient for understanding the actions of rural local governments, particularly if the assumed rationality of politicians is interpreted in non-egotistic terms. Indeed, given that a vast literature on urbanization and social capital (reviewed in Wennström & Öner 2019) predicts that the chances of refugees being successfully integrated into villages and small towns are slim or nonexistent, it appears irrational for local politicians to accept disproportionate numbers of refugees. Thus, we have to turn from structural to psychological levels of analysis.

We maintain that the chosen course of action is best explained by the views of the politicians, which diverged sharply from the views of the general public on refugee reception. However, most existing theories within political science fail to offer an account of how ideas and other psychological forces can induce irrational behavior in politicians and human actors more generally: “Most political scientists continue to be environmental determinists, believing that

human behavior is entirely the product of environmental forces.” (Hibbing & Smith 2007: 6–7). In the current study, we focus on one promising attempt that we have encountered within social psychology: Tetlock’s (2005) fox/hedgehog model of predictive style, which shows that “hedgehogs” are poorer forecasters because they make bold predictions stemming from a single central vision while “foxes” perform better because they are more guarded and flexible in their thinking.

In line with, for example, Mitchell and Tetlock’s (2010) application of the fox/hedgehog model to judicial decision-making and Tirole’s (2017) application of the model to the economics profession, we argue that this framework can be applied in diverse settings and that it may provide a useful approach for understanding the actions of local politicians in rural and peripheral Swedish municipalities. This is an exploratory study, which does not aim to establish a causal relationship between quantifiable municipal characteristics and refugee placement. Our purpose here is to introduce our “Tetlockian” interpretation of the erratic policy of refugee placement building on observed empirical regularities, adding to attempts to go beyond the analytical confines of rational choice theory (e.g., Blyth 2003), which may have implications for the understanding of political behavior more broadly. In particular, the analysis may shed additional light on why many politicians embarking on major policy changes do not appear to think ahead and be able to predict negative consequences (Forstedt 2018), and often are reluctant to perform sophisticated policy assessments (Nilsson et al. 2008).

We proceed as follows. In the next section, we provide information about asylum immigration to Sweden and the terms under which refugees are placed in municipalities, as well as a brief summary of our empirical findings. We then offer our discussion of those findings.

Asylum Immigration and Its Geographical Distribution

Sweden has experienced a significant influx of refugees during the 2000s and 2010s. The first surge in asylum immigration during this period occurred between 2005 and 2008—peaking in 2006, when approximately 25,000 refugees were granted permission to stay—because of the large migration into Europe following the Iraq War and its aftermath. The second surge in asylum immigration occurred after 2011, when the “Arab Spring” sparked political turmoil across the Middle East, leading to civil war in Syria.

Until 2016, the number of asylum seekers continued to increase drastically, from 29,000 in 2011 to approximately 81,000 in 2014. In 2015, that figure was doubled when 162,900 asylum seekers arrived in Sweden. During the autumn of the top year 2015, more than 9,000 people applied for asylum each week.

The largest group was Syrian refugees. However, the growth came to a halt the following year after Sweden strengthened its border controls and enacted new temporary legislation in the summer of 2016 that makes it more difficult for asylum seekers to obtain a permanent residence permit and be reunited with their families in Sweden. Before the enactment of this law, the default was that Sweden granted permanent residence permits and that Syrian refugees were prioritized. An agreement on migration policy struck in 2011 between the center-right minority government of 2010–2014 and the Green Party also included the right to healthcare for undocumented immigrants and the right to elementary education for their children, which was a strong signal that Sweden encouraged immigration.

Per capita, Sweden has received more asylum seekers than any other EU member country, exceeding the reception in France and Germany by a factor of 4.5 in 2015. Such a staggering difference (see further Henrekson et al. 2019) raises the question of where to house the refugees who have been granted a residence permit in Sweden.

The process of placement is initiated once a refugee is permitted by the central government to stay permanently in Sweden. He or she is offered a place in a municipality unless he or she can arrange housing for him- or herself. It became mandatory for all municipalities to accept refugees based on prescribed annual quotas determined by the central government and relevant agencies in March 2016 but before that it was voluntary. No sanctions for noncompliance were imposed in conjunction with the new law.

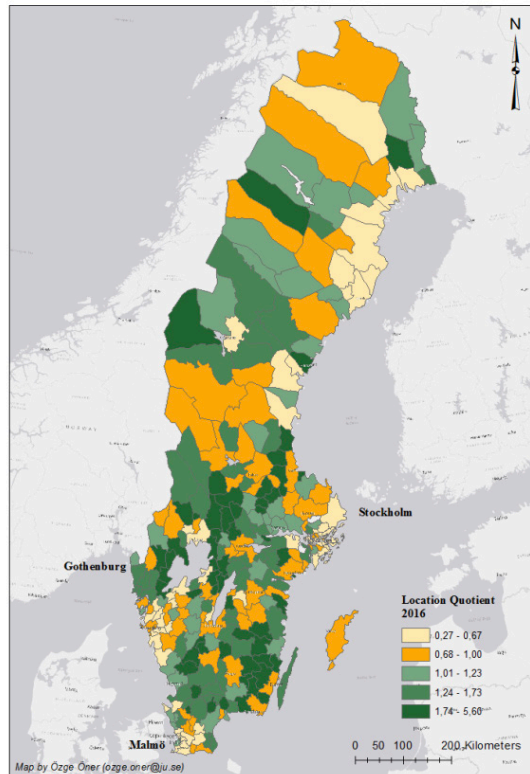
The municipalities are financially compensated by the central government for receiving refugees. The main grant is a flat-rate compensation for each new refugee arriving in the municipality of SEK 133,200 (USD 14,000), or roughly one-fourth of the average annual income (including mandatory social security) for full-time workers, which is intended to cover all direct costs incurred by the municipalities during the first two years. There are also certain forms of ex post compensation that municipalities can apply for, which mainly cover social assistance.

While municipalities will not profit from accepting these grants, they are intended to ensure that refugee reception does not end in a financial loss. However, as observed by a recent public inquiry (SOU 2018:22), the current flat-rate compensation does not necessarily cover actual costs incurred and applications for ex post compensations are slow to be granted, causing some municipalities to suffer financially.

Our empirical analysis shows that refugees who have been granted a residence permit are disproportionately placed in peripheral, depopulating municipalities in the vast northern part of Sweden and the industrial hinterland in the south. The principal metropolitan areas—Stockholm and Uppsala in the east, Gothenburg in the west, and Malmö and Lund in the south—receive

relatively few refugees (see FIG. 1). The relationship between population size and refugee reception first turned negative after 2006 and became even more pronounced after 2010, when Sweden experienced unprecedented levels of asylum immigration. In 2016, the municipalities with the most rapid population decline received on average double the national average of refugees per capita. There are even examples of depopulating municipalities that received four to five times the national average.³ Our analysis shows a similar negative relationship between the municipal employment rate and refugee reception, suggesting that an increase in population and/or the employment rate is associated with a decline in the relative reception of refugees. Finally, we observe a substantial positive association between higher levels of relative refugee reception in a municipality on the electoral support for the immigration-critical Sweden Democrats between the 2010 and 2014 national elections.⁴

Figure 1. Refugee concentration calculated by location quotients (LQ) in 2016.



- 3 At the same time, the variance within the group of depopulating municipalities is large, with several of the municipalities receiving refugees well below the national average; we come back to this point in the discussion.
- 4 The same effect is seen in Mehic’s (2019) study of refugee reception and the electoral support for the Sweden Democrats between the 2014 and 2018 elections.

Discussion

How can we explain the counterintuitive decision of rural and peripheral municipalities with declining populations to accept disproportionate numbers of refugees? In line with Shapiro's (2005) seminal discussion about "problem-driven" versus "theory-driven" research, we approach this question by asking whether an existing theory that has been widely applied to the study of political decision-making can shed light on the motivations of rural local governments and how, if at all, it is counterproductive in this case. If this theory fails to explain the actions of rural and peripheral municipalities, we ask whether there is an alternative explanation.

Our theoretical point of departure is rational choice. In its simplest form, rational choice theory assumes that individuals act consistently in line with their preferences. Thus, given individuals' desires and the information available to them, their actions must be assumed to be the best possible actions among all feasible alternatives (Hindmoor & Taylor 2015). Green and Shapiro (1996: 267), well-known skeptics of the universal applicability of rational choice theory, identified Taylor's (1996) "threefold requirements that the number of options be limited, their costs and benefits clear to the agents, and the stakes high" as reasonable conditions for when rational choice is likely to apply. It would not be unreasonable to assume that such conditions were present in our case and, therefore, that the actions of rural local governments are consistent with rational choice.

THE POLITICAL AND INTELLECTUAL CLIMATE

In the early 2000s, the two major left and right parties, the Social Democrats and the Moderate Party, constituted a restrictive axis in asylum and immigration issues. However, both parties gradually changed their stance and became as committed to generous immigration policies as the five smaller left- and right-wing parties. By 2010, the year in which the negative relationship between population growth and the rate of refugee reception became even more pronounced, all mainstream parties, from left to right, had officially embraced a liberal stance on asylum seekers and immigration in general.

Except for the Sweden Democrats, which entered parliament in that year's election, the collective left and right now also considered it inappropriate to question or criticize immigration to Sweden. Regardless of the individual characteristics of immigrants and the number of refugees arriving in the country, the mainstream parties all claimed that immigration represents a positive cultural and economic contribution to Sweden (e.g., Sanandaji 2017). Any politician, including cabinet ministers and members of parliament (MPs), who did not seem to toe this line in public was severely criticized, even by his or her party. Moreover, national borders were explicitly said to be undesirable

by several of the political parties on both sides of the left–right spectrum,⁵ and then Prime Minister Fredrik Reinfeldt of the Moderate Party (2006–2014) claimed that Sweden lacked a national culture of its own that is worthy of preserving.⁶

In addition, in the wider political debate, the consensus at the time was that it was an expression of racism to question immigration (see, e.g., the recollections of the former leader of the Liberal Party Lars Leijonborg 2018). As a prominent journalist (Hedenmo 2017: 6) later explained, “anyone who raised the issue of the size of immigration found him- or herself out in the cold.” Ekengren Oscarsson (2013) famously coined the term “opinion corridor” to describe this phenomenon of a quickly narrowing public discourse. An opinion poll documented that it had even affected private discussions. Individuals with a conservative or nationalist outlook were less likely to share their views on issues such as immigration outside a small circle of friends than individuals with a liberal or left-wing outlook (Santesson 2015).

Given the political and intellectual climate, which prohibited views on immigration that diverged from the norm, and the fact that both political blocks at the national level were committed to generous asylum and immigration policies, it would have been costly for local politicians in rural and peripheral municipalities to reject refugees. Certainly, all municipalities faced pressure from the central government to accept refugees, but rural municipalities with declining populations had plenty of vacant housing and thus lacked a credible reason for not accepting refugees. Therefore, it would have been more costly for smaller, peripheral municipalities to challenge the *Zeitgeist* by attempting to reject refugees.

Thus, governing local politicians had little to gain by not assenting to large-scale refugee reception in their own municipalities. At the same time, local politicians could contend that they were not personally responsible for the consequences given that their choices had, in effect, been highly constrained by the political and intellectual climate. Arguably, this climate made the reception of disproportionate numbers of refugees appear to be the rational and strategically appropriate option to most politicians.

Moreover, it is conceivable that some local politicians interpreted the economic support from the central government as “pork barrel” (Evans 2011) that would boost the local economy in the form of taxable income for any salaried staff who would have to be employed and large profits for property owners and other service providers, which, in turn, would be spent on local consumption.

5 See the political programs of the Center Party (Centerpartiet 2013), the Liberal Party (Liberalerna 2013), the Green Party (Miljöpartiet 2013), and the Moderate Party (Moderaterna 2011).

6 See “Reinfeldt: What is purely Swedish is barbarous” (Dagens Nyheter 2006).

To some degree, this perception could have alleviated whatever concerns local politicians might have had about accepting large groups of refugees.

THE QUESTION OF ELECTORAL SUPPORT

While these factors are highly relevant, they nevertheless overlook an even more significant issue, namely, that large-scale refugee reception is an unpopular policy among large voter groups. Since 1990, approximately half of the Swedish population has advocated that fewer refugees be accepted. During peaks in Sweden's refugee reception, this share climbed to above 50 percent and even above 60 percent while only between a quarter and a third has been in favor of accepting more refugees (Demker 2017).

Hence, local politicians faced the risk of eroding their electoral support by accepting disproportionate numbers of refugees. Indeed, research suggests that growing shares and regional concentrations of immigrants (including an increase in perceived exposure to immigrants) increase electoral support in both national and local elections for immigration-critical political parties and adversely affect electoral support for parties promoting liberal immigration and asylum policies (Belluci et al. 2019; Halla et al. 2017; Harmon 2018; Lubbers & Scheepers 2001; Otto & Steinhardt 2014; Rydgren & Ruth 2011). The effect seems to be strongest in regard to non-European immigrants with low education levels (Edo et al. 2019). Our empirical analysis suggests further evidence of the existence of such an effect.

Although rational choice theory has made considerable advances since Downs' (1957) claim that politicians are exclusively vote-maximizers, acknowledging that politicians have both policy-seeking and office-seeking motives, rational choice theorists still assume that politicians have "at least one eye upon the prevailing public mood" (Hindmoor & Taylor 2015, 69). Hence, we should expect some ideological divergence but not the flight from the position of large voter groups that took place in Sweden. For several years, only the Sweden Democrats represented the view on refugee immigration of roughly half the Swedish population. How can we reconcile this with rational choice theory?

One explanation that has often been suggested is that the mainstream parties did not realize that they had lost public support for their generous asylum and immigration policies (e.g., Leijonborg 2018). However, a more plausible explanation, given the stability of the public's views on refugee immigration over time, is that the growth of the Sweden Democrats hurt the traditional blocks to roughly the same extent. Indeed, in the party's breakthrough election in 2014, in which it obtained almost 13 percent of the popular vote, SD gained voter shares from both the left and the right (Oscarsson 2016). Hence, it could be argued that a prisoners' dilemma-type situation emerged in which it was deemed preferable by both the left and the right to stay the course rather

than risk being punished by their opponents for changing position on refugee immigration before the other political block had also done so. If this was the choice that the mainstream parties made at the national level, then it is plausible that local politicians were under great pressure not to take any initiatives on their own, and, arguably, it was rational for them to conform and not upset the cart.

Only in the autumn of 2015, when it was evident that the country had exceeded its capacity to receive more refugees, could both political blocks escape this dilemma by simultaneously switching positions, which they did. As Leijonborg (2018: 319), who was a cabinet minister in the successive center-right coalition governments of 2006–2014, recounted: “In the autumn of 2015, the pendulum in immigration policy swung back with full force. [...] Now Sweden’s refugee policy would meet the EU’s minimum standard, the bridge across the Oresund Strait would be closed and [Social Democratic] government ministers talked about chartered planes leaving Sweden with rejected refugees.”

Cast in game-theoretic terms, the breakdown in refugee reception in late 2015 escalated to a Schelling point (Schelling 1960) that provided the window of opportunity that allowed both the left and the right to reverse their previous stance without risking being branded as racist or inhuman by their opponents; a mutual policy shift was thus rendered possible without the need for explicit coordination. Hence, it could be argued that only at that point would it have been rational for local politicians to refuse to accept disproportionate numbers of refugees and that rational choice theory therefore holds as an explanation for the actions of rural and peripheral municipalities.

NON-EGOTISTIC RATIONALITY

However, as, for example, Lichbach and Zuckerman (1997: 24) have observed, rational choice theory “is not bound by the utility or wealth-maximizing assumptions that characterize economics” and “does not even require the assumption that individuals are self-interested.” The addition of non-egotistic considerations increases the complexity of the analysis but is not ruled out (see, e.g., Ostrom 1990). If we, in line with this observation, instead interpret rationality in politicians’ behavior as doing what is most beneficial for the people they represent, any interpretation in line with rational choice becomes more problematic.

As argued at length in Wennström and Öner (2019), refugees have little chance of integration into the local community and the labor market in rural and peripheral areas. Briefly summarized, this is due to the different economic functions of small communities and large cities and to the different kinds of social capital that exist in rural and metropolitan areas. Instead, a large influx of refugees who are unlikely to find work in rural and peripheral municipalities

risks exacerbating an already severe unemployment problem, fueling a growing sense of alienation between inhabitants of rural and urban areas, and creating an ethnic conflict over scarce resources between the native population and the refugees.

Previous research has also demonstrated that immigrants, including refugees, constitute a net cost to the public sector in Sweden (Ekberg 1999; 2009). This is mainly because the employment rate of immigrants, even those of working age, is substantially lower than that of the native population but also because employed immigrants on average have lower annual incomes than natives (Sanandaji 2017). These costs are likely to be felt even more keenly in rural and peripheral municipalities where employment opportunities are scarcer.

Given these factors and the non-egotistic interpretation of rationality we have suggested, it would seem that the actions of most rural local politicians are, in fact, irrational. While it is beyond the scope of this study to adequately account for why some peripheral municipalities with declining populations have not accepted large numbers of refugees, it is likely against this background that they have adopted a comparatively more moderate approach to refugee reception.

Certain northern municipalities with significant population decline but fewer refugees per capita, e.g., Härjedalen and Gällivare, can be compared to Buchanan's (1965) concept of "economic clubs," which are characterized by exclusivity and bonding social capital. Most plausibly, such municipalities have been intent on not disturbing the social order, and the fact that several of them have been governed by a single party (the Social Democrats) most of the time has likely made local politicians more immune to political pressure to receive more refugees. In comparison, several of the municipalities with declining populations that received the greatest numbers of refugees, e.g., Lessebo and Hylte in the south of Sweden, have been governed by intermittent coalitions of Social Democrats, Greens, and center-right parties.

FOXES AND HEDGEHOGS

Rational choice theory should therefore be considered a non-exhaustive explanation. What else might then account for local politicians' actions? One indicator could be the fact that there is an extreme divergence between the opinions of the general public and the opinions of elected politicians in regard to refugee reception. In recurring parliamentary surveys conducted between the years 1994 and 2010, only between 7 and 16 percent of elected MPs favored accepting fewer refugees (see Ekengren Oscarsson 2015). During the period 2002–2010, covering the surge of Middle Eastern refugees between the years 2005–2008, this share varied between 6 and 7 percent.

Although there are no comparable surveys of local politicians' opinions, this large divide between elite and popular opinion, which only widened as Sweden accepted more refugees, was likely also present at the local municipal level. Individual MPs in Sweden need to maintain a close relationship with the party organization in their home districts to be re-elected (Esaiasson et al. 1996), which makes it unlikely that there would be a large divergence in opinion between MPs and local politicians in their home base.

We can instead assume that there was a similar opinion divide over the issue of refugee reception at the local level and should perhaps explore the possibility that this contributed to the reception of disproportionate numbers of refugees in rural and peripheral municipalities. The most promising vehicle for explaining how ideas can induce such irrational behavior in politicians that we have encountered is Tetlock's (2005) model of predictive style.

Tetlock solicited thousands of predictions from several hundred political experts working in academia and government. He classified the experts along a continuum extending between "foxes" and "hedgehogs," a reference to Berlin's (1953) idea that writers and thinkers can be divided into these two categories, and found that foxes are considerably better forecasters than hedgehogs. This is because foxes have a more balanced style of thinking about the world. They are more tolerant of nuance, skeptical of claims that deep laws govern history, and they tend not to reject unpalatable truths to maintain "moral purity" (Tetlock 2005: 106). By contrast, hedgehogs believe in big ideas and governing principles, tend to maintain the same approach in all circumstances. As they do now know "when to apply the mental brakes" (Tetlock 2005: 103), hedgehogs are also more likely to be swept away by their rhetoric and are more prone to making extreme predictions of radical negative or positive change.

Tetlock's (2005) model provides a useful approach for understanding the actions of local politicians in rural and peripheral municipalities. Although determining what particular hedgehog hypothesis regarding refugee immigration may have swayed politicians falls outside the scope of the current study, we can offer a couple of competing suggestions.

One possibility could be a belief that it is always right to choose the option that seems morally good and generous, perhaps supported by "positive asymmetry"—a common way of seeing that "foregrounds or underscores only the best characteristics and potentials of people, places, objects, and events" (Cerulo 2006: 6). Given that different groups or "thought communities prioritize and attend to different categories of people, places, objects, and events" (Cerulo 2006: 12) depending on their goals and values, it is conceivable that the traditional left and right blocks, having converged toward a liberal stance on asylum seekers and immigration, saw only the best in the refugees. The mainstream parties may have received further inspiration from various experts and think-tank scholars who imposed on society and the public debate a

uniform body of allegedly grounded “knowledge” about immigration, which only described positive outcomes and soon became dogma (e.g., Norberg & Segerfeldt 2012).⁷

A different hedgehog hypothesis is that depopulating municipalities must do whatever they can to reverse the trend and increase their populations, which is a widely shared goal among local politicians in Sweden (Syssner 2014). Declining populations are seen as embarrassing and a political failure since “growth for a long time has been the norm in local and regional development policy” (Syssner 2014: 39). The central government has, under various political leaderships, also advocated the idea that all parts of the country should be competitive and prosperous. Given that immigration has often been proposed as a panacea for an aging population in Sweden (Sanandaji 2017), it does not seem unlikely that a large influx of refugees may have offered a perceived short-cut to local politicians to reverse a negative population trend in depopulating rural and peripheral municipalities. As in the alternative scenario, there was support to be had from experts and think-tank scholars who advocated large-scale immigration as a solution to depopulation in rural areas (e.g., Bergström 2014).

In any case, administering Tetlock’s (2005) fox/hedgehog test to politicians could provide a good starting point in future research to determine the motivations of rural municipal authorities in Sweden, and of other policymakers in other similar situations who have failed to look ahead.

References

- Belluci, Davide, Conzo, Pierluigi, & Zotti, Roberto, 2019. “Perceived Immigration and Voting Behavior”. Turin: University of Turin. <https://econpapers.repec.org/paper/utodipeco/201915.htm>.
- Bergström, Andreas, 2014. “Utan invandring avfolkas Sverige: en studie om kommunernas demografiska utveckling”, FORES Policy Paper 2014:3. Stockholm: FORES.
- Berlin, Isaiah, 1953. *The Hedgehog and the Fox: An Essay on Tolstoy’s View of History*. London: Weidenfeld & Nicolson.
- Blyth, Mark, 2003. “Structures Do Not Come with an Instruction Sheet: Interests, Ideas, and Progress in Political Science”, *Perspectives on Politics* 1 (4), pp. 695–706.
- Buchanan, James M., 1965. “An Economic Theory of Clubs”, *Economica* 32 (125), pp. 1–14.
- Centerpartiet, 2013. *Centerpartiets idéprogram*. Stockholm: Centerpartiet.
- Cerulo, Karen A, 2006. *Never Saw It Coming: Cultural Challenges to Envisioning the Worst*. Chicago, IL: University of Chicago Press.
- Dagens Nyheter, 2006. “Reinfeldt: Det ursvenska är blott barbari”, 15/11. Available at: <https://www.dn.se/nyheter/politik/reinfeldt-det-ursvenska-ar-blott-barbari/>.

7 How such processes work is described in Koppl’s (2018) work on “expert failure.” For example, experts with dissenting views on immigration had little incentive to offer their evidence, as such views were perceived as morally inferior.

- Demker, Marie, 2017. "Ökat motstånd mot flyktingmottagning och invandrades religionsfrihet", pp. 475–488 in Andersson, Ulrika, Ohlsson, Jonas, Ekengren Oscarsson, Henrik & Oskarson, Maria (eds.), *Larmar och gör sig till*. Göteborg: Göteborgs universitet, SOM-institutet.
- Downs, Anthony, 1957. *An Economic Theory of Democracy*. New York, NY: Harper & Row.
- Edo, Anthony, Giesing, Yvonne, Öztunc, Jonathan, & Poutvaara, Panu, 2019. "Immigration and Electoral Support for the Far-Left and the Far-Right", *European Economic Review* 115 (June), pp. 99–143.
- Ekberg, Jan, 1999. "Immigration and the Public Sector: Income Effects for the Native Population in Sweden", *Journal of Population Economics* 12 (3), pp. 411–430.
- Ekberg, Jan, 2009. *Invandringen och de offentliga finanserna. Rapport till Expertgruppen för studier i offentlig ekonomi 2009:3*. Stockholm: Fritzes.
- Ekengren Oscarsson, Henrik, 2013. "Väljare är inga dumbommar". Available at: <http://www.henrikoscarsson.com/2013/12/valjare-ar-inga-dumbommar.html>.
- Ekengren Oscarsson, Henrik, 2015. "Om åsiktsrepresentation". Available at: <http://www.henrikoscarsson.com/2015/02/om-asiktsrepresentation.html>.
- Esaiasson, Peter, Holmberg, Sören, & Mayers, Peter, 1996. *Representation From Above: Members of Parliament and Representative Democracy in Sweden*. Aldershot: Dartmouth Publishing.
- Evans, Diana, 2011. "Pork Barrel Politics", pp. 315–339 in Edwards III, George C., Lee, Frances E. & Schickler, Eric (eds.), *The Oxford Handbook of the American Congress*. Oxford: Oxford University Press.
- Forsstedt, Sara, 2018. *Tänk efter före! En ESO-rapport om samhällsekonomiska konsekvensanalyser. Rapport till Expertgruppen för studier i offentlig ekonomi 2018:5*. Stockholm: Ministry of Finance.
- Green, Donald P. & Shapiro, Ian, 1996. "Pathologies Revisited: Reflections on Our Critics", pp. 235–276 in Friedman, Jeffrey (ed.), *The Rational Choice Controversy: Economic Models of Politics Reconsidered*. New Haven, CT & London: Yale University Press.
- Halla, Martin, Wagner, Alexander F. & Zweimüller, Josef, 2017. "Immigration and Voting for the Far Right", *Journal of the European Economic Association* 15 (6), pp. 1341–1385.
- Harmon, Nikolaj A., 2018. "Immigration, Ethnic Diversity, and Political Outcomes: Evidence from Denmark", *Scandinavian Journal of Economics* 120 (4), pp. 1043–1074.
- Hedenmo, Anna, 2017. "Förtroendekris", *Expressen*, 8/6, p. 6.
- Henrekson, Magnus, Öner, Özge, & Sanandaji, Tino, 2019. "The Refugee Crisis and the Reinvigoration of the Nation State: Does the European Union Have a Common Refugee Policy?". IFN Working Paper No. 1265. Stockholm: Research Institute of Industrial Economics (IFN).
- Hibbing, John R. & Smith, Kevin B., 2007. "The Biology of Political Behavior: An Introduction", *The Annals of the American Academy of Political and Social Sciences* 614, pp. 6–14.
- Hindmoor, Andrew & Taylor, Brad, 2015. *Rational Choice*. London: Palgrave.
- Katz, Bruce, Noring, Luise & Garrelts, Nantke, 2016. *Cities and Refugees – The German Experience*. Washington, DC: Brookings Institute.
- Koppl, Roger, 2018. *Expert Failure*. Cambridge: Cambridge University Press.

- Leijonborg, Lars, 2018. *Kris och framgång: mitt halvsekel i politiken*. Stockholm: Ekerlids förlag.
- Liberalerna, 2013. *Frihet i globaliseringens tid – Liberalernas partiprogram*. Stockholm: Liberalerna.
- Lichbach, Mark Irving & Zuckerman, Alan S., 1997. *Comparative Politics: Rationality, Culture, and Structure*. Cambridge: Cambridge University Press.
- Lubbers, Marcel & Scheepers, Peer, 2001. "Explaining the Trend in Extreme Right-Wing Voting: Germany 1989–1998", *European Sociological Review* 17 (4), pp. 431–449.
- Mehic, Adrian, 2019. "Invandring och väljarstödet för Sverigedemokraterna – lärdomar från flyktingkrisen 2015", *Ekonomisk Debatt* 47, pp. 6–16.
- Miljöpartiet, 2013. *Partiprogram*. Stockholm: Miljöpartiet.
- Mitchell, Gregory & Tetlock, Philip E., 2010. "Cognitive Style and Judging", pp. 279–284 in Klein, David & Mitchell, Gregory (eds.), *The Psychology of Judicial Decision Making*. Oxford: Oxford University Press.
- Moderaterna, 2011. *Moderaternas idéprogram. Ansvar för hela Sverige*. Stockholm: Moderaterna.
- Nilsson, Måns, Jordan, Andrew, Turnpenny, John, Hertin, Julia, Nykvist, Björn & Russel, Duncan, 2008. "The Use and Non-Use of Policy Appraisal Tools in Public Policy Making: An Analysis of Three European Countries and the European Union", *Policy Sciences* 41 (4), pp. 335–355.
- Norberg, Johan & Segerfeldt, Fredrik, 2012. *Migrationens kraft: Därför behöver vi öppna gränser*. Stocksund: Hydra.
- Oscarsson, Henrik, 2016. *Flytande väljare*. Stockholm: Statistiska centralbyrån.
- Ostrom, Elinor, 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Otto, Alkis & Steinhardt, Max, 2014. "Immigration and Election Outcomes – Evidence from City Districts in Hamburg", *Regional Science and Urban Economics* 45 (C), pp. 67–89.
- Rydgren, Jens & Ruth, Patrick, 2011. "Voting for the Radical Right in Swedish Municipalities: Social Marginality and Ethnic Competition?", *Scandinavian Political Studies* 34 (3), pp. 202–225.
- Sanandaji, Tino, 2017. *Massutmaning: ekonomisk politik mot utanförskap och antisocialt beteende*. Stockholm: Kuhzad Media.
- Santesson, Peter, 2015. "Vem vågar prata om sina åsikter?". Available at: <https://demoskop.se/vem-vagar-prata/>.
- Schelling, Thomas, 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Shapiro, Ian, 2005. *The Flight from Reality in the Human Sciences*. Princeton, NJ: Princeton University Press.
- SOU 2018:22. *Ett ordnat mottagande – gemensamt ansvar för snabb etablering eller återvändande. Betänkande av Mottagandeutredningen*. Stockholm: Elanders Sverige AB.
- Syssner, Joseфина, 2014. *Politik för kommuner som krymper*. Norrköping: Centrum för kommunstrategiska studier.
- Taylor, Michael, 1996. "When Rationality Fails", pp. 223–234 in Friedman, Jeffrey (ed.), *The Rational Choice Controversy: Economic Models of Politics Reconsidered*. New Haven, CT & London: Yale University Press.

- Tetlock, Philip E., 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.
- Tirole, Jean, 2017. *Economics for the Common Good*. Princeton, NJ: Princeton University Press.
- Wennström, Johan & Öner, Özge, 2019. "Political Hedgehogs: The Geographical Sorting of Refugees in Sweden". IFN Working Paper No. 1266. Stockholm: Research Institute of Industrial Economics (IFN). <https://www.ifn.se/eng/publications/wp/2019/1266>.
- Wollmann, Hellmut, 2004. "Local Government Reforms in Great Britain, Sweden, Germany and France: Between Multi-Function and Single Purpose Organisations", *Local Government Studies* 30 (4), pp. 639–665.

Demokrati med radvalg og fondsvalg

Knut Midgaard

Det følger logisk av begrepet 'demokrati' at der hvor vi har et demokratisk styre, må borgerne, eller medlemmene, kunne tilkjenne sitt standpunkt i saker som er oppe til diskusjon eller avgjørelse; og for det tilfelle at det ikke rår full enighet, må det eksistere en metode til å utlede en gruppepreferanse eller en beslutning på grunnlag av individenes preferanser.

Det er en faglig utfordring – for statsvitere, filosofer, økonomer, sosiologer og andre – å klarlegge hvilke egenskaper en slik metode kan ha – og vurdere hvilke den bør ha. Etter Kenneth J. Arrows bok *Social choice and individual values* fra 1951 er det blitt skrevet mye om dette emnet. Det drøftes også, på en meget fortjenestfull måte, i den bok som her skal omtales, som først kom ut i 2003, og som ble utgitt på nytt i 2019: islendingen Bjørn S. Stefánssons *Demokrati med radvalg og fondsvalg*. Stefánsson legger vekt på at en voteringsmetode ikke bør avskrekke fra å legge frem nyanserte alternativer, men snarere legge til rette for det. En voteringsmetode bør videre være elastisk også på den måten at den gir individene frihet til å gi mye eller lite informasjon om sine preferanser. Den bør videre utnytte den informasjon individene har gitt, så godt som mulig, og utledningen av resultatet bør være presis og teknisk uproblematisk. Det er også et ønskemål at en kan unngå spill av tid og andre ressurser ved at det må votes flere ganger i samme sak. Endelig bør ikke ordningen gi sterke incentiver til å stemme uoppriktig.

Stefánsson sier mer om ønskelige egenskaper ved en voteringsmetode, men jeg stopper her. Mitt poeng er at Stefánsson gir oss et godt grunnlag for å vurdere de metoder som boken behandler: radvalg og fondsvalg. Dette skjer for en vesentlig del gjennom sammenligning med – og til dels kritikk av – andre metoder.

Imidlertid ligger bokens faglige verdi først og fremst i selve utarbeidelsen av de to voteringsmetodene, radvalg og fondsvalg, og i kombinasjonen av to ting: drøfting av metodenes egenskaper, og praktisk utnyttelse av dem. La meg utdype dette.

Knut Midgaard är professor emeritus i statsvitenskap ved Universitetet i Oslo.
E-post: knut.midgaard@st.vuio.no

Radvalg er en metode som forutsetter at hvert individ rangerer de foreliggende alternativer. I en gitt rangering får et gitt alternativ like mange poeng som det har alternativer under seg på listen; og så summeres alternativenes poengtall. Metoden kalles vanligvis Bordas metode etter det franske akademi-medlem 1700-tallsmatematikeren Jean-Charles de Borda, men allerede den store teolog og filosof Nikolaus Cusanus anbefalte en slik metode på 1400-tallet. Stefánsson føyer seg altså inn i en stor sammenheng, og han har selv pekt på det som er å lære av rangeringsmetodene til praktikere som sjakkspillere. Det verdifulle i Stefánssons bidrag ligger for det første i det forhold at hans metode, radvalg, ikke forutsetter fullstendige og strenge preferanser; Stefánsson har funnet frem til en velbegrunnet måte å ta vare på alle alternativer av preferanserelasjoner på – mer eller mindre fullstendige preferanser, med større eller mindre konsentrasjon om hva man ønsker – eller ikke ønsker. For det annet er det svært verdifullt at Stefánsson har pekt på ulike bruksmåter, således både sondering av folks preferanser, opinionsundersøkelser og bruk av metoden i faktiske avgjørelser. Og ikke minst: Han har vist betydelig innsikt, fantasi og energi når det gjelder å legge opp til – og gjennomføre – praktiske anvendelser på høyst forskjellige områder.

La meg så si et par ord om fondsvalgmetoden, som kan brukes hvor det skal voteres over flere saker, kanskje en lang rekke saker. Denne metoden er et virkelig originalt bidrag, et virkelig nybrottsarbeid. Jeg vet bare om én person som har vært inne på en lignende metode, den amerikanske sosiologen James Coleman. Som Stefánsson nevner i boken, unnfanget Coleman sin idé om politiske penger om lag samtidig med at ideen om fondsvalg var utformet på Island. Men Coleman utviklet ikke metoden til praktisk bruk slik Stefánsson har gjort. Slik jeg husker det fra et seminar i Norge hvor Coleman var med, fikk han sin idé fra den amerikanske kongressen: hvordan kongressmedlemmene forhandler om gjensidig støtte i viktige voteringer, slik at A gir sin støtte til B i en sak av stor betydning for B mens B på tilsvarende måte gir A sin støtte i en sak av stor betydning for A, osv. Stefánsson har konstruert en metode som muliggjør noe tilsvarende, men uten at det forutsettes eksplisitte forhandlinger: Man disponerer i utgangspunktet over et fond av stemmer, og så kan man satse ekstra sterkt – og nyansert – i saker som betyr mye for en, med den følge, dersom en lykkes, at en har mindre stemmekraft å sette inn i saker av mindre betydning.

Også her utmerker Stefánsson seg med sin evne til å se mulighetene for interessante og nyttige anvendelser. Mer spesielt har han vist at fondsvalg like som radvalg kan brukes til nyttige sonderinger og opinionsundersøkelser. Jeg er imponert over bredden i Stefánssons anvendelser, og jeg er imponert over Stefánssons evne til å gå djervt og sikkert inn i sentrale – og samtidig sammenfattede – sakskomplekser.

Bjørn Stefánsson er en skikkelse som det ikke finnes maken til på området gruppeavgjørelser og voteringsteori. Han er en dyktig teoretiker, som på

en klargjørende måte har diskutert Arrows umulighetsteorem med Arrow selv. Men ikke alle dyktige teoretikere har nyskapende og praktisk evne. Det har Stefánsson. Han er oppfinner og ingeniør, og han har et sosialt og politisk blikk som medfører at han ser ulike og meget interessante muligheter for anvendelse – i ulike sammenhenger og på ulike nivåer.

Dette viser seg ikke minst i to bøker som følger opp Stefánssons første – hans «Radvalg og fondsvalg: Status i 2018. Artikler og kommentarer» og hans «On Arrow's possibility theorem».

I sammenheng med sitt sosiale og politiske blikk har Stefánsson også en klar bevissthet om de verdier demokratiet kan og bør ta vare på og utvikle. Han er opptatt av hvordan man kan motvirke tilløp til flertallstyranni, og han er opptatt av hvordan man kan motvirke tillukning og polarisering, og i stedet bidra til prosesser preget av gjensidig respekt og en åpent prøvende holdning som muliggjør kreativitet og velavveide avgjørelser.

Jeg har lurt på om denne type originalitet er noe som lettere utvikles i et lite land med særegne tradisjoner og en særegen natur enn i tunge, kompetitive lærdomssentre. Jeg er fristet til å nevne noe den store norske matematiker Atle Selberg ved Princeton Institute for Advanced Study en gang sa om bakgrunnen til de mest betydelige og originale norske matematikere: De hadde for en stor del sin bakgrunn i norske fjordbygder.

Nok om det! Nå er det et spørsmål hvor Stefánssons ideer kan ha mulighet for å bli utprøvd på en tilfredsstillende måte. Det ligger vel muligheter mange steder og i mange miljøer – kanskje noen steder fordi man der sliter med vanskelige konflikter og dårlige metoder for votering og konfliktløsning, og kanskje andre steder fordi det der er et godt samarbeidsklima.

Det er en observasjon flere har gjort: at det er mye konservatisme i vårt forhold til voteringsmetoder – og mer generelt: aggregeringsmetoder. Vi er sterkt tilbøyelige til å tro at de metodene vi har arvet i vårt eget land, er de som må brukes – dette uavhengig av hvilke konkrete metoder det er tale om. Jeg vil tro at det vil gi grunnlag for mer velbegrunnede valg av metoder – og dermed kanskje også bedre beslutningsprosesser – om det ble gitt mer rom for forsøk med alternative voteringsmetoder og aggregeringsmetoder. Mer spesielt håper jeg det kan bli gitt rom for videre utprøving og anvendelse av radvalg og fondsvalg. Teoretiske analyser og overveielser bør gå hånd i hånd med praktisk utprøving, slik det har skjedd på Island – på Stefánssons initiativ og under hans ledelse.

Litteratur

Björn S. Stefánsson, 2019. *Demokrati med radvalg og fondsvalg*. Reykjavik: Demokratisenteret.

Björn S. Stefánsson, 2019. *Radvalg og fondsvalg: status i 2018, artikler og kommentarer siden 2003*. Reykjavik: Demokratisenteret.

Björn S. Stefánsson, 2018. *On Arrow's possibility theorem*. Reykjavik: Demokratsisenteret.

Björn S. Stefánsson, 2013. *Democracy with sequential choice and fund voting*. Reykjavik: Demokratsisenteret.

Björn S. Stefánsson, 2018. *Sequential choice and fund voting—Status in 2018: Articles and comments since 2003*. Reykjavik: Demokratsisenteret.

For kjøp, gå inn på: https://www.amazon.com/gp/product/Bo7S5DDRGX/ref=dbs_a_def_rwt_bibl_vppi_i1.

American Exceptionalism

Supreme Court och det historiska argumentet

Kjell Å Modéer

Inledning

Den 18 september 2020 avled domaren i den amerikanska *Supreme Court* Ruth Bader Ginsburg, RBG, 87 år gammal. Hon hade då suttit i domstolen i 26 år; rekordet har Oliver Wendell Holmes som dog 93 år gammal. Han hade varit domare i 30 år. RBG dog på sin post. Trots sjukdom upprätthöll hon sin plats i domstolen med ”full steam”. När meddelandet om hennes död kablades ut mitt i presidentvalskampanjen utlöstes en uppmärksammas politisk aktivitet. Av de nio domarna på domarbänken hade rättighetsföreträdaren RBG i många år varit den liberala falangens *doyenne*. De sista tjugo åren av sitt liv var hon en närmast ikonisk förebild för många jurister – särskilt för en yngre generation engagerade för mänskliga rättigheter.

RBG hade som sin sista önskan uttryckt, att hennes efterträdare inte skulle utses förrän efter valet, men såväl president Trump som majoritetsledaren i senaten Mitch McConnell var av en annan uppfattning. Medan hon fortfarande låg på lit de parade i Statuary Hall i kongressbyggnaden nominerade president Trump hennes efterträdare, Amy Coney Barrett, en 48 årig domare vid appellationsdomstolen för 7:e distriktet i Chicago, katolik och sjubarnsmamma. Hon hade varit assistent, *law clerk*, hos den konservativa domaren Antonin Scalia – också han katolik – och som textualist hade hon gjort sig känd för att följa hans konservativa rättsskola.

Efter att Scalia oväntat avlidit i februari 2016 nominerade president Obama chefsdomaren vid appellationsdomstolen i Washington D.C., Merrick Garland som hans efterträdare. Garland är känd i både republikanska och demokratiska läger som en skicklig och erfaren domare. Men senatens republikanska majoritet med McConnell som talesman blockerade hans kandidatur. Den ansåg att Obama var ”*lame duck*” och att nomineringen av Scalias efterträdare skulle ske under den president, som skulle väljas i november samma år. Därigenom föll Garlands nominering. Det blev i stället president Trump, som när han tillträtt våren 2017 nominerade och utsåg domaren vid appellationsdomstolen i tionde distriktet i Denver, Neil Gorsuch – också han en dokumenterad

företrädare för Scalias konservativa tolkning av författningens ”ursprungliga mening” (Gorsuch 2019: 108 ff).

I juli 2018 fick Trump möjlighet att också nominera efterträdaren till domaren Anthony Kennedy, som 82 år gammal önskade lämna sin plats i domstolen. Han uttryckte önskemål om att till hans efterträdare skulle nomineras en annan pålitlig konservativ domare, den femtiofemåriga domaren från appellationsdomstolen i Washington, Brett Kavanaugh (Marcus 2019). Denne nominerades av president Trump och efter infekterade senatsförhör godkändes han med siffrorna 50-48 och utsågs i oktober 2018.

Därigenom hade Trump genom sina nomineringar (och med den republikanska senatens tillstyrkan) – och genom att domarna sitter på livstid – sannolikt garanterat en konservativ majoritet i domstolen för lång tid framåt.

Under den nuvarande konservative chefsdomaren John Roberts (utsedd av Bush) har domstolen i flera kontroversiella ärenden – till Trumps stora förtret – flera gånger röstat med den liberala falangen. Mest uppmärksammat är när han 2012 anslöt sig till den liberala gruppen och därigenom räddade Obamas sjukvårdsreform, *Affordable Care Act*, Obamacare. I juli 2020 röstade Roberts emot en inskränkning i aborträtten (*June Medical Services v. Russo*) och även i det mål som rörde offentliggörande av uppgifter rörande president Trumps skatteåterbäringar (*Trump v. Vance*) gick Roberts emot Trump. Senast, bara några dagar efter presidentvalet 2020, röstade både Roberts och Kavanaugh i ett nytt mål mot ett upphävande av Obamacare (*Oral Argument 10.11.2020 No 19-840, California et al v. Texas et al.*).

Med presidentvalet för dörren blev frågan om efterträdare till Ruth Bader Ginsburg således en ytterst viktig och brådskande politisk angelägenhet för presidenten. Det gällde för honom att utse hennes efterträdare före valdagen den 3 november 2020. Med en garanterad 6 – 3 sammansättning av ledamöterna i domstolen skulle Trump för lång tid framåt kunna ge domstolen en konservativ majoritet. För honom blev det därför en triumfartad framgång att utnämna Amy Coney Barrett och bryta den syn på domstolen som reformorienterad som den i princip haft sedan Earl Warrens dagar. I den extrema situation som uppstått efter valet, där Trump utgått från att valet skulle avgöras till hans fördel i ”den högsta rättsinstansen”, har emellertid de många republikanska stämningarna om valfusk fallit platt till marken.

Domstolen och lagprövningen

Tillsättningsärendet efter RBG har visat hur infekterat och politiserat det amerikanska rättssystemet är. När Tage Erlander var på officiellt besök i USA 1952 besökte han *Supreme Court* i Washington D.C. (Erlander 1974: 228; Modéer 2007: 345 ff). Det blev en avskräckande upplevelse för honom. Han mötte en författning som historiskt bygger på Montesquieu’s maktodelningsprincip och

vars tillämpning utvecklats med dynamiska *checks and balances* mellan den verkställande, lagstiftande och dömande grenen av statens funktioner.

För Tage Erlander var det artfrämmande att domstolar skulle få agera politiskt och få pröva politiska myndigheters, regeringens och riksdagens beslut. En sådan lagprövning, *judicial review*, hade enligt honom ingen plats i den svenska rättsordningen. Den amerikanska modellen innebar en politisering av domstolsväsendet. Samtidigt kom den svenska rättskulturen under efterkrigstiden alltmer att präglas av angloamerikanska influenser. Kort efter svenska regeringsformens ikraftträdande 1975 infördes 1979 i Sverige inte bara en rättighetskatalog utan också – om än i begränsad omfattning – lagprövning av normer (Regeringsformen 11:14 och 12:10).

I dag, 70 år senare, har de svenska domstolarna i hög grad – och främst genom medlemskapet i den Europeiska Unionen och dess domstolar – påverkats av maktdelningsläran och dess institutioner. Rättskulturen i de amerikanska domstolarna har i dag också en annan inriktning än den då hade.

Författningens djupstruktur

Rättsväsendet i USA har djupa historiska rötter. Det regleras främst genom författningen 1787 och dess tillägg samt domarlagen (*Judiciary act*) från 1789. Montesquieus maktdelningslära med principen *Equal justice under law* finns inhugget i arkitraven på framsidan av *Supreme Court*, som också genom sin praxis utvecklat den starka ställning domstolen har i den amerikanska rättskulturen. Genom rättsfallet *Marbury v. Madison* 1801 påtog sig domstolen rätten att granska om en lag var i överensstämmelse med den amerikanska författningen (Marcus 1986). Denna av domstolen självpåtagna rätt att ogiltigförklara kongressens och presidentens lagstiftning har sedan dess gett domstolen dess starka autonoma ställning.

Domstolens avgöranden har historiskt också fått djupgående politiska konsekvenser. I avgörandet *Dred Scott v. Sandford* 1857 tillät domstolen slaveri och förvägrade de amerikaner som var av afro-amerikanskt ursprung medborgarskap – och därigenom också rättshandlingsförmåga i de federala domstolarna. Det beslutet blev direkt avgörande för inbördeskrigets utbrott 1861. Också genom *Plessy v. Ferguson* (1897) bekräftade domstolen att i det då framväxande moderna samhället stred inte segregationslagstiftningen mot författningen. Därigenom etablerades principen ”åtskilda men lika”, *Separate but Equal*, som legitimerade den segregationspolitik som var förhärskande fram till mitten av 1900-talet.

Under 1900-talets första decennier var den progressiva erans klassiska rättssyn förhärskande. President Franklin D. Roosevelt's jämlikhetspräglade *New Deal*-lagstiftning kom däremot på kontrakurs med de ”nio gubbarna” i *Supreme Court*. När denna moderna lagstiftning ogiltigförklarades av

domstolen i rättsfall efter rättsfall, svarade presidenten i mars 1937 med sin sk *court packing* proposition, som föreslog att antalet ledamöter i domstolen skulle ökas från nio till tretton. Med fyra nya liberala domare hade omröstningarna kunnat förändras, men kritiken mot förslaget blev massiv och Roosevelt drog tillbaka förslaget (Lash 1988: 291 ff). När ledamöterna i *Supreme Court* efter hand började dra sig tillbaka, såg Roosevelt sin chans att nominera unga "New Deal Lawyers" till domstolen. Under Roosevelts och Harry Trumans presidentskap utsågs sammanlagt tolv domare, vilka för lång tid framåt garanterade en liberal tolkning av författningen.¹ Cass Sunstein har framhållit uppfattningen att det var på 1930-talet som domstolen kapitulerade för Roosevelt och hans *New Deal*. Från att tidigare ha sett som sin uppgift att tolka den skrivna författningen, "den verkliga författningen" – "*the real Constitution*" – började domstolen spåra ur och anpassa sig till den moderna statens regleringslagstiftning (Sunstein 2005: 3 f).

Earl Warren, tidigare republikansk guvernör i Kalifornien, utsågs 1953 till domstolens nye ordförande. Hans ordförandetid (1953 – 1969) präglades av reformtänkande. Han såg författningen som ett levande dokument (*a living constitution*), som kontinuerligt anpassades till sin samhälleliga kontext. Warren inledde med att skriva domen i det rättsfall, som fortfarande framstår som det mest betydande för 1900-talets rättighetsdebatt, *Brown v. Board of Education* (1954). I *Brown* kom domstolen att i ett enigt avgörande förbjuda separata skolor för färgade och vita. En viktig kontext till detta avgörande var Gunnar Myrdals omdebatterade bok *An American Dilemma* (1944), som med sociologisk metod avslöjade rasproblemen i USA. 1950-talets intensiva debatt om medborgliga rättigheter resulterade i *Civil rights act* 1964 under Lyndon Johnson (Kalman 2017: 17). "Den levande författningen" kom under Warrens ordförandeskap till uttryck genom ett antal kontroversiella avgöranden, exempelvis om fri abort 1973 (*Roe v. Wade*), rättighetsfrågor och jämställdhet (*affirmative action*). Redan 1968 hade Nixon fokuserat sin kritik mot Warren Court, och när han utsett William Rehnquist till ledamot 1972 slog också pendeln tillbaka. Sedan dess har också republikanernas kritik av "den levande författningen" varit en i hög grad levande diskurs (Kalman 2017: 313).

Under efterkrigstiden, och särskilt från Earl Warrens tid, blev således domstolens politiska framtoning betydande. Domstolens aktivism ställdes mot kongressens lagstiftningsuppgift och både domare och juristprofessorer uttalade sitt stöd för en traditionell och mer inskränkt tillämpning av lagprövningsinstitutet (Gunther 1994: 654 ff). Domstolens praxis sattes under lupp. Plazan framför domstolsbyggnaden i Washington blev arena för ständiga protestmarscher och demonstrationer.

Allt större intresse fokuserades på de enskilda ledamöternas rättsliga

1 <https://www.senate.gov/legislative/nominations/SupremeCourtNominations1789present.htm>

position, deras *judicial philosophy*. Domarens främsta uppgift är ju att tolka författningen/lagen – inte att stifta lagar. Detta är fortfarande ett återkommande mantra i utfrågningen av domarkandidater till federala domartjänster, och det blev mycket tydligt i senatsförhören med appellationsdomaren Robert Bork, som 1987 nominerades av president Reagan. Senaten dominerades vid denna tid av demokraterna, och de avlog nomineringen (Kalman 2017: 300). Bork föll således på sin konstitutionella doktrin. Han ifrågasatte *Brown* och ansåg som originalist att domstolen skulle tolka författningstexten och endast låta sig vägledas av grundlagsfädernas idéer och åsikter. När författningens 14:e tillägg antogs 1868 uttalade lagstiftaren ingen avsikt att undanröja segregeringen mellan vita och svarta, och därför hade domstolen 1954 enligt Bork inte heller någon laglig grund för sitt ställningstagande (Hoffer m fl. 2007: 411).

Det historiska argumentet

Alltsedan 1980-talet har domstolens ledamöter alltmer använt sig av en historisk argumentation i sina avgöranden. I konstitutionell rättshistorisk teori skiljer Robert W. Gordon mellan en statisk, dynamisk eller kritisk historisk argumentation (Gordon 2002: 181 ff).

Det statiska historiska argumentet är i dag starkt framträdande i *Supreme Court's* avgöranden. Detta argument utgår från att det endast är författningens text som domarna ska utgå ifrån. Det var Antonin Scalia (1938–2016) som i slutet av 1980-talet blev den domare som kom att symbolisera denna bokstavstolkning som kallas *originalism*. Scalia ansåg konsekvent att det var författningens ursprungliga mening, *original meaning*, som skulle beaktas. Han framhöll att domaren skulle referera till den ursprungliga meningen av det värdeladdade språk som tillämpades när författningen ratifierades i slutet av 1700-talet (Scalia 2017: 180 ff). Denna bokstavstroga tolkning, s k *textualism*, dominerade både hans och den konservativa majoritetens rättssyn (Scalia & Garner 2012; Calabresi 2007).

Den statiska tolkningen ställs emot den som bygger på en mera modern, liberal och samtida, – *dynamisk* – rättighets-(policy)tolkning. Det är den vi har mött i Earl Warrens domstol och i begreppet *the living constitution*. Den innebär att tolkningen av författningen ska anpassas till samtiden och dess förhållanden. När en part hävdade en bokstavstroga författningstolkning brukade Warren ställa frågan: ”Men är det rättvist?” – *Is it fair?* (Choper 2007: 357 ff). Denna rättvisetolkning omfattade Ruth Bader Ginsburg i hög grad i sin rättssyn (Sherron de Hart 2018). Därtill kommer det kritiska argumentet, som exempelvis Earl Warren och domstolen gav uttryck för 1954 i rättsfallet *Brown v. Board of Education*. Den rättsliga argumentationen var i detta fall kritisk till de tidigare nämnda argumenten, vilka på skilda sätt bygger på historiska tolkningar av författningen.

Samhällsteater i rättens tempel

Supreme Court har sedan 1930-talet sina lokaler på Capitol Hill i ett magnifikt marmortempel, som inrymmer alla de arkitektoniska symboler som de nyantika klassicistiska idealen företrädde. Tempelmetaforen representerar en egen rättslig, närmast sakral kultur (Semonche 1998; Modéer 2004: 922 ff). Flera av domstolens ledamöter vägrade att flytta in i detta domstolspalats, när det stod färdigt 1935. Byggnaden har en teatral inscenering och i den närmast sakrala sessionssalen agerar domarna i sina individuella roller, som också kan följas i de offentliga protokollen från de muntliga förhandlingarna. I varje mål har parterna och deras advokater ca 30 minuter på sig vardera att framhålla målets fakta och problem. Domstolens nio ledamöter får därvid tillfälle ställa frågor beträffande det skriftliga material (*briefs*), som parterna ingivit till domstolen. Förhandlingarna är ofta både livfulla och intellektuellt stimulerande. Domarna ställer frågor efter anciennitet, och dialogerna mellan advokaterna och domarna är skarpa, ofta klagörande och inte sällan raljanta. Antonin Scalia var som senior ledamot ofta den som brutalt fullbordade dessa verbala ”tjurfäktningar”. Avslutningsvis ställer domstolens ordförande sina frågor och summerar. Vid en session där jag var närvarande i november 2015 började domarnas frågor hagla redan efter någon minut. Parternas advokater hade inte en chans att utveckla och fullborda sina väl förberedda presentationer (*Supreme Court*, 13-1339 *Spokeo v. Robins*, November 2, 2015).

Domarnas interna tankesmedjor

Varje domare i *Supreme Court* omger sig i sina ämbetsrum med både sekreterare och tre – fyra juridiska assistenter, s k *law clerks*, som förbereder målen, gör rättsliga utredningar och även skriver domarens utlåtanden i domstolens rättsliga avgöranden. Varje domare väljer själv sina assistenter. När Brett Kavanaugh tillträdde sin tjänst 2018 tillsatte han endast kvinnliga assistenter (Marcus 2019: 373). Antonin Scalia har berättat hur informellt han brukade nominera sina *clerks*. Han tog kontakt med några av sina vänner vid de främsta juridiska fakulteterna på Harvard, Columbia, Chicago och bad dem föreslå kandidater. Sedan kallade han ner de utvalda till Washington för intervjuer. Alltså ett mycket informellt och flexibelt förfarande. Ruth Bader Ginsburgs selektion var än mer familjär. Det var under många år hennes man, Martin Ginsburg, professor i skatterätt på *Georgetown School of Law*, som gick igenom ansökningarna och rangordnade de kandidater som han ansåg skulle inbjudas till intervju (Peppers & Ward 2012: 392). Också en av hennes tidigare assistenter, tillika dekan vid *Columbia Law School* David Schizer, hade också normalt förslagsrätt till en av tjänsterna (Peppers & Ward 2012: 392).

Att tjänstgöra som *law clerk* för en domare i *Supreme Court* eller någon av de federala appellationsdomstolarna är en mycket viktig merit för unga

amerikanska jurister och avgörande för deras fortsatta karriärer. Domarnas mentorsförhållanden till sina *law clerks* är föremål för en rikhaltig anekdotsflora (Peppers & Ward 2012). Assistenterna präglas givetvis av domarens rättsyn, vilket inte minst Amy Coney Barrett framhållit i förhållande till Scalia. Dessa tjänstgöringar följs därför med stort intresse av jurister i karriären.

Professor Dick Howard vid University of Virginia, som tjänstgjorde som *law clerk* för domaren Hugo L. Black 1962 – 1964, återkommer gärna till den bildning som Black förmedlade till sina *law clerks*. Black utgick från att de var väl bevandrade inte bara i juridisk litteratur utan att de också tog chansen att använda hans välsorterade privata skönlitterära bibliotek (Howard 1967: 1068).

Nominering och utseende av ledamöter

Vid nomineringarna till den moderna *Supreme Court* har presidenten varit noga med att välja domare med en domarteori (*judicial theory*) som överensstämmer med presidenten och hans parti. I 85 % av nomineringarna till *Supreme Court* har presidenten valt en kandidat från sitt eget parti. Detta kom särskilt till uttryck genom Franklin D. Roosevelts alla nomineringar från och med domaren Hugo Black. Under den progressiva perioden hade domarna representerat konservativa klassiska ideal, vilka inte överensstämde med den moderna tidens reformideal. Domarna kallades i demokraternas folkmun för ”De nio gubbarna”, *The Nine Old Men* (Pearson & Allen 1937). Roosevelt nominerade därför unga övertygade ”*New Dealers*”, som för lång tid framåt skulle kunna garantera upprätthållandet av presidentens plattform (Frankfurter, Black, Jackson & Rutledge).

Den republikanske presidenten Dwight D. Eisenhower nominerade och utsåg två domare i *Supreme Court*, Earl Warren och William J. Brennan. När Eisenhower efter sin presidenttid (1965) fick frågan om han hade gjort några misstag som president, svarade han: ”Jag har gjort två misstag, och de sitter båda i *Supreme Court*”. Domarnas roll som oberoende domare överskuggas således inte alltid av deras politiska profil, och domarnas rättsyn är inte alltid så ortodox som presidenterna tror. Både Sandra Day O’Connor och Anthony Kennedy ansågs vara sk *swing voters* och därigenom bildade de majoritet i s k 5-4 avgöranden. Som nämnts har också den nuvarande *chief justice* under senare år flera gånger tagit på sig rollen som *swing voter*.

Amy Coney Barrett den femte kvinnliga domaren

Den första kvinnliga domaren i *Supreme Court* var Sandra Day O’Connor. Hon utnämndes av Ronald Reagan 1980. Även om hon var en konservativ jurist var hon under sina senare år i domstolen en *swing voter* (Thomas 2019: 383). Den andra kvinnliga domaren var Ruth Bader Ginsburg; hon utsågs av Clinton 1993

(Ginsburg 2016: 165 ff). När Sandra Day O'Connor drog sig tillbaka från domstolen 2005 blev Ginsburg ensam som kvinna i domstolen och kände sig under några år isolerad (Ginsburg 2016: 76). Men under president Obama utsågs två kvinnliga domare, Sonia Sotomayor med puertoricansk bakgrund (2009) (Sotomayor 2013), och riksadvokaten (*Solicitor general*) Elena Kagan (2010). Båda anslöt sig till domstolens liberala falang.

Med Amy Coney Barrett tillförs *Supreme Court* tvärtom en pålitlig textuallist. Av hennes välmeriterade fyra nyutnämnda *clerks* har alla hämtats från de främsta advokatbyråerna i USA; tre av dem har varit *clerks* för konservativa domare i *Supreme Court*.

Politiseringen av domarnomineringarna

Under president Trump har domarutnämningarna varit uttalat politiska. Det framgår i hög grad av senatsutfrågningarna av kandidaterna. Att de nominerade juristerna söker distansera sig från den politiska etiketteringen visade sig beträffande Brett Kavanaugh, som introducerades av en framstående och erfaren advokat vid *Supreme Court*, Lisa Blatt. Hon markerade, att hon är en liberal jurist, att hon hade tjänstgjort som *law clerk* hos Ginsburg, och personligen hade önskat att alla nio domarna hade samma rättssyn som sin "hjärte". Blatt framhöll, att Kavanaugh har alla de kvaliteter, som kvalificerade honom att vara domare i *Supreme Court*. Hennes introduktion belyser den autonomi och integritet som juristernas kultur i *Supreme Court* förutsätts företräda.²

Ett annat belysande exempel på domarnas förmåga att skilja mellan sak och person är den djupa vänskap som förelåg mellan den mest konservativa domaren Scalia, och den mest liberala, Ginsburg. I domstolen företrädde de den statiska respektive den dynamiska historiska argumentationen, men de var personliga vänner och umgicks *en famille* (Carmon & Knizhnik 2015: 115 ff; Ginsburg 2016: 38ff).

Avvikande meningar – Dissents

Avgörandena i *Supreme Court* författas av en företrädare för majoriteten. Domare med avvikande mening redovisar sitt ställningstagande särskilt i en *dissens*. I den ständigt pågående författningsrättsliga diskursen är sådana avvikande meningar viktiga, inte sällan mer betydelsefulla än majoritetens dom. Ruth Bader Ginsburg tillhörde inte sällan minoriteten. Uppmärksammat blev ett mål 2007 rörande kvinnors rätt till lika lön med män (*Ledbetter v. Goodyear*), där Ginsburg skrev för minoriteten och krävde explicit ändrad lagstiftning.

² <https://www.c-span.org/video/?c4747282/user-clip-attorney-lisa-blatt-introduces-judge-brett-kavanaugh-senate-judiciary-committee>

Dissensen riktade sig alltså till den lagstiftande kongressen, och med en demokratisk kongress under president Obama korrigerades domstolsavgörandet genom *Lilly Ledbetter Fair Pay Act 2009* (Carmon & Knizhnik 2015: 136 ff.).

Utblick och sammanfattning

Den franske politologen Alexis de Tocqueville konstaterade redan 1840 att ”det finns knappt någon politisk fråga i Förenta Staterna som inte förr eller senare blir en domstolsfråga” (Tocqueville 1840, Hoffer 2007: 2 f). Påpekandet är fortfarande högaktuellt i dag, när frågor rörande poströster i det amerikanska presidentvalet gång efter annan avgörs av *Supreme Court* och det i en tid då RBG:s konservativa efterträdare faller avgörandet. En dryg månad efter hennes bortgång är *Supreme Court* åter fulltalig. Det är 45 år sedan tillsättningsprocessen gick så fort (John Paul Stevens 1975). Den 27 oktober 2020, en dryg vecka före valet, förestavade domaren Clarence Thomas eden för Ginsburgs efterträdare, Amy Coney Barrett.

Redan domartillsättningarna under Lyndon Johnson och Richard Nixon hade visat på utvecklingen mot en konservativ *Supreme Court* (Kalman 2017). Det republikanska partiet har under presidentvalen de senaste trettio åren haft tillsättningen av konservativa domare i högsta domstolen högt på sin agenda (Marcus 2019: 6). Republikanska presidenter har kritiserat de liberala domarna och deras tolkning av ”den levande författningen”. År 1982 grundades också *Federalist Society* som en ideell organisation till förmån för en textuell och originalistisk tolkning av författningen. Med Ronald Reagans utnämning av William Rehnquist som ordförande och Antonin Scalia som ledamot 1986 förstärktes domstolens konservativa falang, som sedan dess bildat majoritet. Men sedan O’Connor, Kennedy och senast John Roberts visat sig inte vara fullt ut lojala med de konservativa värderingarna blev tillsättningen av Scalias efterträdare särskilt viktig för Trumpadministrationen.

President Trump har under sin presidentperiod fått möjlighet att nominera tre domare med en konservativ rättssyn till *Supreme Court*. Genom dessa har domstolen kantrat över från vänster till höger. Den liberala flank, som sedan Earl Warrens dagar varit så betydelsefull för domstolens reformavgöranden har nu (till synes för lång tid framöver) ersatts av en konservativ rättssyn.

Samtidigt föreligger stor risk för att politiseringen kommer att rubba förtroendet för domstolen, vilket på ett förödande sätt ägde rum genom avgörandet *Bush v. Gore* (2000) (Zelden 2010). Det tog åtskilliga år innan förtroendet var återställt och därmed också respekten för institutionerna och författningens *checks and balances*. Upprätthållandet av domarnas självständighet och förtroendet för deras opartiskhet blir allt viktigare i vår tid, där politiken alltmer judikaliseras (Tate & Vallinder 1997).

Inte bara domarna i *Supreme Court* utan hela den federala domarkåren har under Trump för lång tid framåt fått en helt ny konservativ prägel. Den

republikanska senaten hade under Obama försvårat och fördröjt domartillsättningarna. Det har resulterat i att drygt femtio domare enbart vid appellationsdomstolarna har nominerats och tillsatts av Trump. Denna strategi är som visats inte ny, men den kommer att för årtionden framåt att prägla den dömande makten i USA.

Källor och litteratur

- Calabresi, Steven G. (red.), 2007. *Originalism: A quarter-century of debate*. Washington D.C.: Regnery Publishing.
- Carmon, Irin & Knizhnik, Shana, 2015. *Notorious RBG: The Life and Times of Ruth Bader Ginsburg*. New York: William Morrow Publishers.
- Choper, Jesse, 2007. "Earl Warren – A Law Clerk's Memory of the Man and the Court", i N. Scheiber, Harry (red.). *Earl Warren and the Warren Court: The Legacy in American and Foreign Law*. Lanham: Lexington Books.
- Erlander, Tage, 1974. *1949–1954*. Stockholm: Tidens förlag.
- Ginsburg, Ruth Bader, 2016. *My own Words* (With Mary Hartnett and Wende W. Williams). New York: Simon & Schuster.
- Gordon, Robert W., 2002. "The Historical Argument in American Legal Culture", i Modéer, Kjell Å. (red.), *Rättshistoria i förändring*. Stockholm: Rättshistoriska studier.
- Gorsuch, Neil, 2019. *A Republic. If You Can Keep It*, [med Jane Nitze och David Feder]. New York: Crown Feder.
- Gunther, Gerald, 1994. *Learned Hand: The Man and the Judge*. New York: Knopf.
- Hoffer, Peter Charles, Hull Hoffer, William James & Hull, N. E. H., 2007. *The Supreme Court: An Essential History*. Lawrence: University Press of Kansas.
- Howard, Dick, 1967. "Mr. Justice Black: The Negro Protest Movement and the Rule of Law", *Virginia Law Review*, 53.
- Kalman, Laura, 2017. *The Long Reach of the Sixties: LBJ, Nixon, and the Making of the Contemporary Supreme Court*. New York, Oxford: Oxford University Press.
- Lash, Joseph P., 1988. *Dealers and Dreamers: A New Look at the New Deal*. New York: Doubleday.
- Marcus, Maeva, 1986. "Marbury v. Madison: John Marshall's Selective Use of History," *University of Wisconsin Law Review*, 2.
- Marcus, Ruth, 2019. *Supreme Ambition: Brett Kavanaugh and the Conservative Takeover*. New York: Simon and Schuster.
- Modéer, Kjell Å., 2004. "Det heliga rummet" – Domstolsbyggnadernas roll i det civila samhället. *Svensk Juristtidning* 922 ff.
- Modéer, Kjell Å., 2007. "An American Dilemma and the Scandinavian Dream", i Scheiber, Harry N. (red.), *Earl Warren and the Warren Court: The Legacy in American and Foreign Law*. Lanham: Lexington Books.
- Pearson, Drew & Allen, Robert S., 1937. *The Nine Old Men*. Garden City/New York: Doubleday, Doran & Co. Inc.
- Peppers, Todd C. & Ward, Artemus, 2012. *In Chambers: Stories of Supreme Court Law Clerks and Their Justices*. Charlottesville and London: University of Virginia Press.

- Scalia, Antonin, 2017, *Scalia Speaks: Reflections on Law, Faith and Life Well Lived*, [red. Christopher J Scalia & Edward Whelan]. New York: Crown Forum.
- Scalia, Antonin & Garner, Bryan A., 2012. *Reading Law: The Interpretation of Legal Texts*. St. Paul: Thomson West.
- Semonche, John E., 1998. *Keeping the Faith: A Cultural History of the U.S. Supreme Court*. Lanham: Rowman & Littlefield Publishers.
- Sherron de Hart, Jane, 2018. *Ruth Bader Ginsburg: A Life*. New York: Knopf.
- Sotomayor, Sonia, 2013. *My Beloved World*. New York: Vintage Books/Random.
- Sunstein, Cass R., 2005. *Radicals in Robes: Why Extreme Right-Wing Courts Are Wrong for America*. New York: Basic Books.
- Tate, C. Neal & Vallinder, Torbjörn, 1997. *The Global Expansion of Judicial Power*. New York: New York University Press.
- Thomas, Evan, 2019. *First: Sandra Day O'Connor. An Intimate Portrait of the First Woman Supreme Court Justice*. New York: Random House.
- de Tocqueville, Alexis, 1840. *De la démocratie en Amérique*. London: Saunders and Otley.
- Zelden, Charles L., 2010. *Bush v. Gore: Exposing the Hidden Crisis in American Democracy*. Lawrence: University Press of Kansas.

INTERNET

US Senate:

<https://www.senate.gov/legislative/nominations/SupremeCourtNominations1789present.htm>.

<https://www.c-span.org/video/?c4747282/user-clip-attorney-lisa-blatt-introduces-judge-brett-kavanaugh-senate-judiciary-committee>.

US Supreme Court:

Supreme Court oral argument in *Spokeo v. Robins*, November 2, 2015. https://www.supremecourt.gov/oral_arguments/argument_transcripts/2015/13-1339_j5fl.pdf.

FÖRVALTNINGSPOLITISKA BETRAKTELSE¹

Skapar FN:s hållbarhetsmål ordning på den globala politiska nivån?

Lars Niklasson

Will the UN Sustainable Development Goals contribute to a more coherent global level of politics?

The UN Sustainable Development Goals are a set of policy ambitions agreed upon by all the member states of the UN. They deserve to be taken seriously as a unanimously agreed framework for global politics. However, they raise a number of questions. One concerns the democratic legitimacy and the weak involvement of the public in these decisions. Another set of questions relate to the design and implementation of the SDGs, which are partly contradictory and open to rival interpretations. This relates to all the main themes, i.e. environment vs growth and social development, as well as to the role of good governance. The latter seems to be of great importance for raising taxes and making rational decisions on fossil fuels etc. At the same time, the SDGs could provide a general framework for the reform of global governance. International organizations often have unclear, limited and yet overlapping mandates. Such reform could be pursued through the EU as a global actor.

Bör man ta FN:s hållbarhetsmål – även kallade Agenda 2030 – på allvar? Målen har blivit en självklarhet bland bistånds- och miljöorganisationer. Den svenska regeringen har utsett en nationell samordnare och EU-kommissionens nya ordförande Ursula von der Leyen har gett alla sina kommissionärer i uppdrag

¹ Statsvetenskaplig tidskrift inleder med denna översikt en serie som vi väljer att kalla "Förvaltningspolitiska betraktelser". Här inryms både samtidskritik, debatt och analyser av förvaltningspolitiska problem och utmaningar, nationella som internationella. Välkommen med ett bidrag!

Lars Niklasson är verksam vid Sieps och Linköpings universitet.
E-post: Lars.niklasson@sieps.se

att verka för målen. Bland den övriga allmänheten finns ett stort intresse att arbeta med målen, särskilt de som har med miljöfrågor att göra. För en statsvetare, däremot, infinner sig ett antal frågor om vad detta är för slags politik, vart målen leder och varför just dessa mål har upphöjts till global norm. I det sistnämnda ligger frågor om hur det gick till när målen arbetades fram, men även frågor om de värderingar som uttrycks; varför just dessa värderingar? Och på vilka grunder blir vi som medborgare bundna att följa ett beslut i FN? En del av detta kommer jag att analysera i det följande.

Hållbarhetsmålen ("Sustainable Development Goals", SDG) antogs av FN i september 2015 (United Nations 2015a). Det var en stor politisk framgång att alla länder enade sig om en gemensam uppsättning mål som ska uppfyllas till år 2030. Målen följer efter de så kallade Millenniemålen (MDG), som gällde för fattiga länder under åren 2000-2015, men har vidgats till att gälla hela området "hållbar utveckling" och för alla länder. Det finns en förväntan om att varje land ska utarbeta en nationell strategi som är anpassad efter de specifika förutsättningarna. Däremot finns det inga redskap som FN kan använda för att tvinga länder att göra på ett visst sätt. Grupptricket är det enda påtryckningsmedlet att använda mot ovilliga regeringar. Samtidigt finns det internationella organisationer som har vissa befogenheter, vilket väcker frågor om hur dessa kan samordnas med varandra. Reducerar FN:s hållbarhetsmål komplexiteten på den globala politiska nivån och skapar något slags ordning för de inblandade aktörerna, eller uppstår det nya otydligheter? De underliggande politiska konfliktlinjerna försvinner inte med ett beslut i enighet. Istället borde forskningen diskutera vilken typ av politik som kommer till uttryck i målen och vilka politiska ståndpunkter som inte kommer till uttryck (jfr Niklasson 2019 kap 5).

Målen har stora likheter med Parisavtalet för klimatet (United Nations 2015b). Även där finns mål – att den globala uppvärmningen ska stanna vid 1,5 grader – men inga gemensamma strategier eller tvångsmedel. Varje land ska själv utforma sin strategi. Parisavtalet antogs inom ramen för FN:s klimatkonvention i december 2015 och har en koppling till hållbarhetsmålen såtillvida att de senare målen hänvisar till Parisavtalet. Därför är det rimligt att säga att de två hänger ihop och att de är designade på ett likartat sätt, vilket gör att de står inför likartade utmaningar. Det som gör hållbarhetsmålen särskilt intressanta är att de är bredare och handlar om att lösa två stora utmaningar som riskerar att kräva motstridiga åtgärder, att rädda miljön och att samtidigt utrota fattigdomen i världen.

För en statsvetare utgör hållbarhetsmålen en intressant blandning av politiskt ansvarstagande och önsketänkande. Medan politiker talar om målen som om de har ett entydigt innehåll, vill forskaren framhålla komplexiteten och behovet av fördjupad analys. Det handlar både om nationella och internationella politiska processer samt relationen mellan dessa båda. Några av de relevanta frågorna berörs här.

Syftet med den här artikeln är att ställa några vanliga statsvetarfrågor om hållbarhetsmålen; vad är målen, varför just dessa och hur ska de genomföras? En sådan fråga handlar om att beskriva målen, att se målen som en global politik eller åtminstone ett slags gemensam global ram, en "governance-arkitektur" (Borás & Radaelli 2011). Vad är speciellt med globala mål och vad är speciellt med dessa mål? Går de att uppnå? Kan vi säga något om implementeringen redan innan vi vet hur det har gått? Mitt fokus är således i första hand förvaltningspolitiskt, att förstå vad som förväntas göras, av vem och hur det kommer att gå. Jag kommer avslutningsvis att beröra frågan om hur EU skulle kunna ta en större roll i det globala genomförandet av målen, utifrån antagandet att EU har en ambition att påverka den globala politiken.

Jag kommer att utgå från att vi bör ta hållbarhetsmålen på allvar, dvs att vi accepterar FN:s ståndpunkt som rimlig, även om det finns skäl att kritiskt granska de värderingar som FN ger uttryck för. Jag ska återkomma till hur dessa värderingar kan beskrivas.

Mina frågor är främst instrumentella och handlar om vad som bör göras för att uppnå målen, givet att vi accepterar dem. Först kommer jag dock att beröra den mer fundamentala frågan om vilken grund målen vilar på. Blir vi som medborgare bundna av ett beslut i FN? Finns det någon koppling till den enskilde medborgaren som gör det rimligt att lyda globala beslut?

En annan fråga är den större om hur den globala politiken fungerar i stort och hur den borde fungera. Just nu ifrågasätts världsordningen av USA och Kina. Tidigare var det globaliseringsmotståndare som protesterade mot Världshandelsorganisationen (WTO) och mot de ekonomiska toppmötena i Davos. Samtidigt är det många som vill se mer global politik för att hantera miljöproblemen och för att hjälpa fattiga och förtryckta människor ta sig ur sin situation. Hållbarhetsmålen är ett steg i den riktningen, men med viktiga frågor obesvarade. Här finns med andra ord många "vanliga" statsvetarfrågor att ställa om en ovanlig politik.

Forskningen om hållbarhetsmålen är starkt växande och knyter an till stora frågor om hållbar utveckling i allmänhet och sådant som staters styrförmåga och styrelsekick ("good governance", "statebuilding" osv.). Även om målen gäller för alla länder, så är det några som har mer fundamentala problem med genomförandet av de åtgärder som krävs. Mitt intresse är särskilt riktat mot länder som har stora problem med genomförandet.

Forskardebatten om målen började redan för tio år sedan, som en diskussion om lärdomar inför en ny politik 2015 (se t ex Fukuda-Parr 2017 eller Kanie & Biermann 2017). Det finns en växande litteratur om hur förhandlingarna gick till, med detaljer om var man hade svårt att bli eniga (t ex Kamau, Chasek & O'Connor 2018). Det är ännu få som studerar hur genomförandet är tänkt att gå till och hur det kan utvecklas (Niklasson 2019). De studier som gjorts av genomförandet av millenniemålen år 2000–2015 är begränsade i olika

avseenden. Däremot finns det generell litteratur av typen ”vad fungerar för att hjälpa de fattigaste i Afrika?” (Lawson, Ado-Kofie & Hulme 2017), som dock inte tydligt kopplar lärdomarna till FN:s mål.

Mitt bidrag är att sätta in målen i en diskussion om implementeringen, särskilt behovet att lära av erfarenheter och att utveckla insatserna som görs. Då hamnar sådant som val av strategi, genomförandeorganisation och utvärderingsmetoder i fokus (se t ex Knill & Tosun 2012). Det behövs mer av vetenskaplig fördjupning kring de ”best practices” som samlas in. Statsvetenskaplig forskning inom områdena internationell och jämförande politik erbjuder verktygslådor som FN och andra borde använda bättre. Ett exempel är att genom jämförande fallstudier förstå orsakssambanden i det nationella genomförandet av målen. Ett annat exempel är att genom analys av den politiska styrningen av internationella organisationer lägga en grund för reformering och konsolidering av den globala styrningen (”global governance”) för att understödja arbetet med att uppfylla målen.

I den här artikeln ger jag en övergripande beskrivning av problemen på den globala nivån att styra i riktning mot målen. Utgångspunkten är att målen inte bara vänder sig till nationella regeringar, utan även ger en samlad bild av övergripande globala prioriteringar. Det väcker frågan om FN:s mål kan skapa ordning i det som många beskriver som en ”global röra” av överlappande regleringar och organisationer (Slaughter 2004). En växande forskning inom bland annat miljöpolitiken pekar på problemet med fragmentering och konflikter mellan olika regelverk på den globala nivån (van Asselt 2014). Det väcker också följdfrågan om EU, som en viktig global aktör, kan bidra till att reducera komplexiteten på den globala nivån.

Demokratisk förankring?

Om vi börjar med den stora frågan om politikens demokratiska förankring, så tycks den vara tunn. FN är ett slags samarbetsorganisation där ländernas regeringar gör gemensamma uttalanden och fattar vissa bindande beslut. Vi har således indirekt, genom regeringen, varit med om att fatta beslutet om hållbarhetsmålen. Detta är ett vanligt sätt att rättfärdiga beslut i internationella organisationer, men ju mer omvälvande beslut de fattar, desto mer ställs det krav på en starkare koppling till medborgarna. En vanlig kritik är att indirekt representation skapar ett demokratiskt underskott, t ex i EU (Hix & Høyland 2011:132ff, jfr Keohane 2015, Bäckstrand 2006). De internationella frågorna försvinner ofta i den allmänna nationella valdebatten. Och varför ska vi bara företrädas av regeringen? Borde inte oppositionen delta i det internationella samarbetet? Den typen av invändningar har lett till att EU, som också är en sorts internationell organisation, har utvecklat ett direktvalt parlament som bidrar till den folkliga förankringen. Priset är dock en mer komplicerad konstitutionell

modell. Exemplet EU visar en väg att utveckla FN och skapa starkare förankring för de globala målen. Det finns debattörer som vill se ett världsparlament som komplement eller ersättning för FN (Leinen & Bummel 2018).

De globala målen hämtar snarare sin legitimitet ur det folkliga engagemanget i de berörda frågorna. En stor mängd konsultationer genomfördes för att fånga in synpunkter ”nerifrån”, samtidigt som flera processer organiserades för att förhandla ”uppifrån” (Kamau, Chasek & O’Connor 2018). Miljörörelsen är stark, liksom engagemanget för de fattiga och förtryckta människorna i världen. Däremot finns det ingen metod att väga dessa och andra frågor mot varandra. Hållbarhetsmålen innehåller ett antal kompletterande ambitioner och uttalanden om t ex styrelseskick som bidrar till en mer avvägd helhet (nedan). Målen utgör en ”lagom” blandning av ambitioner som kan accepteras av de flesta medborgare.

Det finns två principiella ståndpunkter som ger ytterligare tyngd åt hållbarhetsmålen. Den ena är att det finns gemensamma problem som kräver i stort sett enigt agerande. Klimathotet är en sådan fråga. Spridningen av dödliga virus är en annan. Med ekonomernas terminologi finns det globala kollektiva nyttigheter som vi alla vill ha men inte kan nå om vi inte skapar en samarbetsordning (Barrett 2007). Det andra argumentet är en filosofisk ståndpunkt som säger att politiken bör styras av globala (”kosmopolitiska”) principer snarare än nationella (Føllesdal 2012). I vår tid är det dock motsatsen, nationalismen, som blivit en dominerande princip. Donald Trump och andra ifrågasätter nyttan av samarbete och betonar nationalstatens intressen. Även bland miljödebattörer har det hävdats att nationalismen är en nödvändig drivkraft för förändring och uppoffringar (Lieven 2020).

Det paradoxala är att behovet av gemensamma globala lösningar tycks vara större än någonsin, samtidigt som möjligheterna att uppnå dem är mindre än tidigare. Hållbarhetsmålen är ett slags bästa möjliga kompromiss. De utgör ett samarbete mellan nationalstater, ett exempel på flernivåstyrning (multilevel governance), där organisationer på flera politiska nivåer samarbetar (och konkurrerar) om gemensamma ambitioner (Zürn 2010). Samarbetet om målen borde följas av ett samarbete om att reformera de internationella organisationerna för att skapa starkare gemensamma regelverk i linje med hållbarhetsmålen. Troligen kommer vi att få se krav på hårdare åtgärder och mer global styrning efter 2030, när hållbarhetsmålen löper ut. Särskilt miljörörelsen kommer att kräva åtgärder för att uppnå klimatmålen.

Som statsvetare blir man nyfiken när alla länder gör upp om gemensamma mål. Uppfattar vi målen lika? Går de att uppfylla samtidigt, eller måste det göras avvägningar? Politiska kompromisser uppnås ofta till priset av otydlighet; vad är det egentligen man har kommit överens om? Kommer enigheten att bestå eller finns det särintressen som kommer att driva sina perspektiv i genomförandet? Det är lätt att se målen som en kompromiss mellan länders

regeringar och även mellan de politiska ståndpunkter som dessa regeringar har. Mycket av ståndpunkterna är ”mainstream”, men det finns inslag som sticker ut. Exempelvis är stödet för politisk demokrati svagt, för att texten skulle accepteras av regeringar som inte är demokratiska.

Som statsvetare ser man att det finns fler konfliktlinjer i politiken än mellan nationalstater. En sådan konfliktlinje går mellan politikområden, även inom nationella regeringar. Miljöpolitikens företrädare tävlar med företrädarna för bistånd, handel och näringspolitik om att uttolka vilka gemensamma strategier som behövs. Departementen gör upp med sina likar, mot sina rivaler, på den europeiska och den internationella nivån. Dragkampen går vidare i tolkningarna av hållbarhetsmålen. På engelska kallas de rivaliserande grupperna ”epistemic communities” eller med modern jargong ”silos”. På svenska säger vi ”stuprör” eller ”järntrianglar” om konkurrerande politikområden (Haas 1992, jfr Collinson 1999; Sullivan & Skelcher 2002). Här finns ett behov att forska om konflikter i genomförandet av hållbarhetsmålen.

Vilka är målen?

Hållbarhetsmålen följer i spåren av de stora internationella mötena som diskuterade miljöfrågor och utvecklingsfrågor, bland annat i Stockholm 1972 och i Rio 1992. I Sverige har miljöfrågorna fått störst uppmärksamhet, men målen handlar om teman som delvis är motstridiga och som måste fås att fungera tillsammans. Målen gäller i alla länder. Däremot är problemen olika och möjligheterna att lösa dem ser olika ut.

Målen har fyra teman. Tre av dessa ingår i begreppet hållbar utveckling – ekonomisk, social och miljömässig hållbarhet – medan det fjärde temat handlar om förutsättningar för de övriga målen genomförande. Kategorierna överlappar och målen är avsedda att genomföras tillsammans, men det är lättare att beskriva dem i fyra grupper.

Figur 1. FN:s hållbarhetsmål grupperade i fyra teman.

Område	Mål
Miljö	Nr 12: Hållbar konsumtion och produktion Nr 13: Bekämpa klimatförändringarna Nr 14: Hav och marina resurser Nr 15: Ekosystem och biologisk mångfald
Socialt	Nr 2: Ingen hunger Nr 3: God hälsa och välbefinnande Nr 4: God utbildning för alla Nr 5: Jämställdhet Nr 6: Rent vatten och sanitet för alla Nr 7: Hållbar energi för alla Nr 11: Hållbara städer och samhällen
Ekonomi	Nr 1: Ingen fattigdom Nr 8: Anständiga arbetsvillkor och ekonomisk tillväxt Nr 9: Hållbar industri, innovationer och infrastruktur Nr 10: Minskad ojämlikhet
Förutsättningar (styrelseskick mm)	Nr 16: Fredliga och inkluderande samhällen Nr 17: Genomförande och globalt partnerskap

Hållbarhetsmålen som fokuserar på miljön är inte helt tydliga, utan hänvisar till Parisavtalet och andra internationella överenskommelser. I praktiken är det inom ramen för Parisavtalet som det globala klimatarbetet sker. I arbetet med hållbarhetsmålen ska varje land hantera alla fyra teman, det vill säga samtidigt som länderna verkar för miljö- och klimatmålen ska de säkerställa ekonomisk, social och politisk utveckling. Målen anger bara i liten utsträckning hur denna blandning ska uppnås.

Bakom målen för miljön och klimatet finns en analys av vad som behöver göras för att begränsa den globala uppvärmningen, sammanfattad i den så kallade Stern-rapporten (Stern 2006). I korthet handlar det om att minska användningen av fossila bränslen, vilket kräver utveckling av nya teknologier och särskilda åtgärder för att inte snabbväxande utvecklingsländer som Kina och Indien ska välja att bygga upp energisystem baserade på relativt billig olja och kol. Mer kontroversiella vägval handlar om personlig livsstil kontra fokus på de stora utsläppen. Hållbarhetsmålen är försiktigt formulerade om vilka åtgärder som bör prioriteras.

De sociala målen är tydligare än miljömålen. De handlar om att utrota fattigdom, hunger och ohälsa. Sjukvård och utbildning är två stora insatsområden, liksom utbyggnad av grundläggande infrastruktur (vatten, avlopp, elektricitet, vägar med mera). Det finns vidare ett antal mål som handlar om ekonomisk tillväxt, bland annat för att finansiera de sociala åtgärderna. Dessa mål handlar om sådant som effektivare jordbruk, industrialisering, lönearbete, ekonomisk infrastruktur och om att sluta diskriminera de fattiga. Målen gäller framför allt fattiga länder som ännu inte har en utvecklad ekonomi.

Det finns anledning till optimism eftersom fattigdomen i världen halverats sedan 1990, en nedgång från två miljarder till en miljard extremt fattiga människor, samtidigt som jordens befolkning ökat från sex till sju miljarder. FN har drivit ståndpunkten att de fattigaste människorna behöver starthjälp så att de kan förbättra sin egen situation. Exempelvis skulle investeringar i vägar och elektricitet hjälpa familj jordbruken att öka sin avkastning i ett första utvecklingssteg (UN Millennium Project 2005).

De två sista målen handlar om grundläggande förutsättningar för att uppnå de övriga målen. Mål nr 16 handlar om gott styrelseskick ("good governance"), bland annat att den nationella politiken måste fungera enligt rättsstatens och demokratins principer, även om inte ordet demokrati används i texten. Världsbanken och andra biståndsorganisationer har drivit ståndpunkten att korruption och andra problem måste bekämpas för att se till att omvärldens stöd används effektivt.

Mål nr 17 handlar om internationellt stöd till fattiga länder bland annat i form av investeringar, kunskapsöverföring (forskning och utveckling) och handel. Detta är saker som de rika länderna ska ställa upp med, samtidigt som de fattiga länderna ska finansiera en allt större del av sin välfärd genom skatter. Dessutom sägs att de fattiga länderna bör få större inflytande i internationella organisationer.

En implicit orsakskedja

Det finns många uppfattningar om hur målen bör tolkas. Flera av dem lyfter fram att det gäller att hitta synergier och samband mellan målen som underlättar genomförandet (se till exempel Kanie & Biermann 2017). Frågan är om målen kan översättas i nationella handlingsplaner, som dessutom tar hänsyn till de lokala förutsättningarna, det vill säga lägesbilden och de instrument som står till regeringarnas förfogande. Detta är en viktig fråga eftersom ett politiskt handlingsprogram måste klara ett rationalitetstest, att insatserna verkligen kan leda till målet. Statsvetare framhåller att det ofta är stor skillnad mellan välformulerade planer och hur genomförandet av politiska program går till i praktiken (Hill & Hupe 2002).

Min tolkning är att målen kan läsas som en implicit orsakskedja, vilket innebär att några mål är medel för att uppnå andra mål (se Niklasson 2019). Den orsakskedja som tonar fram ligger i linje med modern forskning i statsvetenskap och institutionell ekonomi om ekonomisk och social utveckling (jfr Lindvall & Rothstein 2010). Kort sammanfattat handlar det om att en välfungerande politik (statsapparat) skapar förutsättningar för ekonomisk utveckling och för regleringar som skyddar miljön. Den ekonomiska utvecklingen skapar resurser som kan beskattas – om medborgarna har förtroende för politiken – för att finansiera den sociala välfärden. Många specifika frågor återstår att

reda ut, men nyckeln till utvecklingen är medborgarnas vilja att underordna sig gemensamma regler, det vill säga att regeringen och statsapparaten uppfattas som legitima.

Figur 2. Den implicita orsakskedjan (Niklasson 2019).



Det goda statsskicket är fundamentalt eftersom det är det som lägger grunden för välfungerande regler, skatter och politiska strategier. De ekonomiska målen ligger också tidigt i orsakskedjan eftersom det är tillväxten som ska finansiera den utbyggda välfärden och den tekniska utvecklingen. Det finns en tanke om att varje land på sikt ska generera egna resurser ("domestic revenue mobilization", på engelska) för att hantera sina sociala problem, utan att detta går ut över miljön. Regleringar och politiska strategier har betydelse både för uppbyggnaden av välfärden och för övergången till ett fossilfritt samhälle och annat som leder till en god miljö.

Omvärlden ska framför allt bidra med investeringar och kunskap för teknisk utveckling (mål nr 17). Tillspetsat är det en teknikoptimistisk vision som uttrycks, där alla länder ska få möjlighet att utvecklas som länderna i Europa gjorde i början av 1900-talet och där ny teknik gör det möjligt att minska miljöförstöringarna. Målen nämner inte omfördelning eller normer om "global rättvisa". Snarare ger de uttryck för en liberal syn, att ekonomin skapar nya resurser som löser sociala problem över tid, med en viss politisk styrning. Kritikerna hävdar att det krävs mer radikala åtgärder för att bryta de fattiga ländernas beroende av de rika länderna.

Bakom tankegången ligger bland annat iakttagelsen att två av de tre länder där de flesta extremt fattiga individerna lever, är länder med stark ekonomisk utveckling. De två länderna är Indien och Kina, med Nigeria mellan sig (UN 2015c). Detta pekar både på ett ökat nationellt ansvar för fattigdomsbekämpning och på en förhoppning om att de ekonomiska framstegen i Asien ska följas av framsteg i övriga världen. En nyckel till utvecklingen i Asien var välfungerande stater ("developmental states") med en stor del av det som kallas "good governance". Många av de fattigaste länderna är istället "fragile states" eller "failed states" som domineras av klaner och krigsherrar. Det är med andra ord mycket fundamentala samhällsproblem som måste hanteras i länder som Afghanistan, Irak, Libyen och Somalia.

Det bör sägas att det finns andra uppfattningar om alla de här frågorna, både i forskningen och i politiken. Bland annat gäller det grundantagandet i FN:s hållbarhetsmål, att tillväxt och fattigdomsbekämpning kan kombineras med ambitionerna att förbättra klimatet och miljön. Även synen på det goda styrelseskicket har ifrågasatts, där det hävdas att åtminstone delar av ett gott styrelseskick är ett resultat av tillväxt, snarare än en förutsättning för tillväxt, eftersom vissa diktaturer har lyckats sätta igång en nationell utvecklingsprocess. En ståndpunkt är att det räcker med "tillräckligt gott styrelseskick" ("good enough governance"; Grindle 2004).

En implikation av den här analysen är att hållbarhetsmålen behöver diskuteras samlat, så att både målkonflikter och synergier (orsakssamband) kan uppmärksammas. FN:s mål är mer komplexa och mer avancerade än de framställs. De pekar mot ett behov av en samlad syn på särskilt miljöpolitik, handelspolitik och biståndspolitik, där handel och bistånd (demokratistöd med mera) lägger en grund för fattiga länders välfärdspolitik och miljöpolitik.

På den globala nivån ställer målen krav på en mer utvecklad multilateral ordning.

En annan implikation är att större fokus bör läggas på att lösa grundläggande problem med de nationella styrelseskicken, eftersom dessa problem begränsar effekten av övriga insatser. Korruption, diktaturer och klansamhällen borde motarbetas av det globala samfundet med större kraft. Detta krockar dock med normen om att alla stater är suveräna (Sørensen 2011). Genom FN har vissa inskränkningar i den nationella suveräniteten kunnat göras (Alter 2014, Acharya 2018, jfr Wittor 2016). Mer borde göras.

Kritiken mot målen

Ett övergripande problem är att målen ofta är försiktigt formulerade. Ett annat problem är att det inte finns några sanktionsmöjligheter inom ramen för överenskommelsen om målen. Det handlar (underförstått) om gruppträck som ska få ländernas regeringar att leva upp till sina åtaganden. Dock utgör målen ett gemensamt ställningstagande av jordens länder, vilket skapar en viss tyngd och förhoppning om att de kommer att efterlevas. I många fall bygger målen på tidigare överenskommelser, vilket gör att det kan finnas ytterligare påtryckningsmöjligheter genom internationella organisationer.

Hållbarhetsmålen har ifrågasatts för att de fokuserar just på mål och lämnar öppet för olika tolkningar av hur målen ska uppnås (Kanie & Biermann 2017). I några fall uttrycker dock hållbarhetsmålen strategier (medel), framför allt vad gäller tillväxt och finansieringen av de sociala insatserna. Målen är ofta okontroversiella, medan valet av medel är kontroversiellt, både bland forskare

och bland nationella politiker. Traditionellt har debatten om biståndspolitiken handlat just om vilka strategier som är verkningsfulla (Williams 2012). Den liberala synen som kommer till uttryck i FN:s hållbarhetsmål omfattas inte av alla biståndspolitiska debattörer.

Ett problem med fokus på mål är att de riskerar att tränga ut diskussionen om vad som leder till att målen uppnås. Istället för att fokusera på indikatorer (lägesbilder) borde analysen handla om att förstå vad olika länder gör och vad det är som påverkar utfallet (orsakssambanden). Då kan erfarenheterna överföras till fler länder. Med andra ord borde FN bygga upp ett lärande över tid, med fokus på vad som fungerar i olika situationer.

En annan kritik är att enigheten i FN om målen döljer att det finns både oklarheter och alternativa vägval. Målen anger på flera ställen att tillväxt måste förenas med hänsyn till miljön, men säger inte mycket om hur det ska gå till. Målen pekar på att det är den tekniska utvecklingen som ska lösa detta i kombination med en handlingskraftig nationell politik. Här skulle analysen av genomförandet kunna tillföra intressanta data om vad som fungerar. I åtminstone ett fall är målet uppseendeväckande svagt: subventioner till fossila bränslen ska ses över, inte avskaffas.

Ett annat sätt att uttrycka dilemmat är att målen handlar om stora samhälls-transformationer som inte uppenbart låter sig förenas. Klimatomställningen (att leva fossilfritt) får stort utrymme i den allmänna debatten, men den ekonomiska utvecklingen i fattiga länder är också en stor transformation med konsekvenser för alla länder i en ekonomiskt integrerad värld, till exempel att vi alla blir mer lika i våra livsstilar och kan sträva efter större samarbete och större geografisk rörlighet.

Sammanfattningsvis ger hållbarhetsmålen ett slags målbild för den globala politiken, dock utan att ändra något i tidigare globala överenskommelser. Samtidigt är det svårt att implementera hållbarhetsmålen utan att göra en närmare analys av vilka åtgärder som leder dit. Det behövs en kvalificerad diskussion om vad som fungerar i olika situationer och hur målkonflikter bör hanteras. Annars är risken stor att hållbarhetsmålen blir symbolpolitik, där frågor rycks ur sitt sammanhang och åtgärder väljs för att de låter bra här och nu.

Nationellt genomförande

FN lägger ansvaret för genomförandet på de nationella regeringarna. De uppmanas att redovisa sina erfarenheter i form av ”voluntary national reviews” vid de årliga uppföljningsmötena i New York. Regeringarna förväntas även redovisa sitt arbete med klimatfrågorna inom ramen för uppföljningen av Parisavtalet. Det finns en risk att de två procedurerna kräver olika redovisning och olika åtgärder.

Mycket av det intressanta arbetet sker i rådgivningen till fattiga länder om vad de bör göra. Bland andra FN:s utvecklingsprogram (UNDP) ger sådant stöd. Här finns en risk att råden blir generella ("uppifrån") och allt för lite anpassade till den specifika situationen. Man skulle önska mer av en forskarstrategi som jämför strategier "nerifrån" och ser vart de leder och varför. I Afrika sker ett parallellt arbete inom den Afrikanska unionen, som driver en gemensam övergripande strategi kallad Afrika 2063, med ett 50-års-perspektiv från beslutet om strategin 2013. Förhoppningsvis bidrar det till ett lärande mellan länderna om hur olika situationer kan hanteras, men risken för symbolpolitik är stor.

Lärandet från egna och andras erfarenheter är centralt. Här kan statsvetare bidra med metoder eftersom frågan om att utveckla strategier har likheter med frågan om hur demokrati utvecklas. I de stora frågorna om modernisering, industrialisering och demokratisering har länder valt olika vägar. En rimlig inställning är att det inte finns en väg som passar alla. Utvecklingen bör snarare ses som ett komplext system, där flera faktorer kan påverka varandra på olika sätt. Ändå vill man som forskare veta vilka strategier som leder till goda resultat i olika kontexter, för att driva analysen framåt.

Frågan om "governance" illustrerar några av problemen. Det finns ett stöd i mål 16 att ta styrelseskicket (rättsstat och demokrati) på allvar. Vissa forskare skulle säga att det är en helt avgörande fråga för att övriga insatser ska bli effektiva (Norris 2012). Korruption och frånvaro av rättsstat snedvrider insatserna och minskar resurserna. Ändå förs det debatt mellan forskare om i vilken ordning åtgärderna behöver genomföras. Några invänder att det goda styrelseskicket kan vara ett resultat av tillväxt, likaväl som en förutsättning för den samma (Kelsall & Booth 2010, se Niklasson 2019). Som forskare vill man fördjupa analysen, för att inte ge förenklade råd till de nationella regeringarna. Ingen vet med säkerhet om ekonomisk utveckling leder till goda politiska system eller tvärtom.

Den multilaterala ordningen

Internationella organisationer spelar viktiga roller i den globala politiken, men ifrågasätts från olika håll. Merparten av de frågor som tas upp i FN:s hållbarhetsmål är reglerade internationellt och föremål för agerande genom internationella organisationer. Därför är det relevant att lyfta fram dessa organisationer och hur hållbarhetsmålen påverkar dem. Kan hållbarhetsmålen – trots sina brister – reducera komplexiteten på den globala politiska nivån av organisationer och regleringar? En av de intressanta sakerna med målen är att det finns en global politisk uppslutning kring dem. I stort sett alla jordens länder är eniga om en agenda till år 2030. Det tycks finnas ett närmast unikt tillfälle

att åtgärda några av de problem med den internationella ordningen som har diskuterats under lång tid. Samtidigt visar konflikterna kring de internationella organisationerna att samsynen om de globala frågorna inte är så djup som man kan förledas att tro om man bara studerar framväxten av målen.

I det följande ger jag en översiktlig bild av några av de berörda organisationerna och några av deras grundläggande problem, som enskilda organisationer och gemensamt, som ett samlat system för global styrning. Ambitionen är att visa vilka typer av problem som behöver hanteras, snarare än hur det kan göras i de enskilda fallen. Forskningen om hur målen påverkar de internationella organisationerna är mycket begränsad, men det finns studier av de tidigare målens effekter på styrningen som jag återkommer till.

Varje internationell organisation är föremål för en omfattande debatt och många av dem har reformerats, om än i små steg. Några har bytt inriktning sedan de inrättades, till exempel OECD och Internationella valutafonden, IMF. Organisationerna har olika roller i den globala politiken. Några är till stor del förhandlingsmaskinerier. Andra reglerar på egen hand. Vissa löser konflikter. Andra ger bidrag och lån. Jag hänför organisationerna till olika teman, även om de ibland har uppdrag som sträcker sig över flera områden (se till exempel Weiss & Wilkinson 2014 för en översikt).

På miljöområdet sker de viktiga överenskommelserna inom ramen för FN:s ramkonvention om klimatförändringar, "klimatkonventionen" (*United Nations Framework Convention on Climate Change, UNFCCC*) och dess årliga möten (*Conference of the Parties, COP*), men även inom konventioner och organisationer som reglerar naturresurser (haven, skogar mm) och energiproduktion. En viktig organisation är den internationella energimyndigheten (*International Energy Agency, IEA*) och myndigheten för förnyelsebar energi (*International Renewable Energy Agency, IRENA*). Därtill har FN en organisation för miljöfrågor (*United Nations Environment Program, UNEP*). En viktig kategori av genomförandeorganisationer är de globala fonder som finansierar investeringar i ny teknik och omställning, exempelvis *Green Climate Fund, GCF*, och *Clean Technology Fund, CTF*.

Parisavtalet och dess föregångare inom UNFCCC (bland annat Kyoto-avtalet) anses vara svaga i meningen att det sällan finns tvingande instrument att använda mot staterna. Konventioner om haven och hotade djurarter med mera är mer tvingande. En uppenbar fråga är om regelverken borde vara mer tvingande, om de nationella insatserna borde styras hårdare av en global organisation, en *World Environment Organization* (som inte har kunnat skapas). Utifrån perspektivet att globala regleringar behövs för att samordna nationell politik, ter sig en sådan organisation önskvärd.

Den svaga regleringen kompenseras i någon mån av att det finns ett starkt folkligt engagemang i miljöfrågorna. Ländernas regeringar förväntas vara lojala med Parisavtalet och genomföra stora förändringar. På den globala nivån

handlar en diskussion om huruvida de industrialiserade länderna ska bära en större del av den gemensamma bördan än de länder som är i början av sin industrialiseringsprocess. En sådan formell uppdelning gjordes i Kyoto-avtalet, men togs bort i Parisavtalet.

Energipolitiken har en nyckelroll i regleringen av fossila bränslen, som är en huvudfråga i klimatpolitiken. Energipolitiken är ett område som är känt för sin splittrade organisation. IEA uppfattas som en svag organisation, vilket bidrog till inrättandet av en ny myndighet för förnyelsebar energi, IRENA. Ingen organisation äger hela energipolitiken.

Figur 3. Några av de viktigaste internationella organisationerna.

Frågor	Exempel på organisationer
Miljö	UNFCCC IEA IRENA UNEP
Ekonomi	WTO ILO IMF UNCTAD UNIDO G7, G20
Socialt	WHO UNAIDS UNESCO UNDP FAO
Styrelseskick	ICC HRC UNSC

På det ekonomiska området är det framför allt Världshandelsorganisationen (*World Trade Organization, WTO*) som skapar gemensamma spelregler för ekonomin genom sina handelsregler. Miljöregler och arbetsrättsregler kan ses som regelverk för att upprätthålla rättvis konkurrens. Arbetsrätten regleras genom den internationell arbetsorganisationen (*International Labour Organization, ILO*). Internationella valutafonden (*International Monetary Fund, IMF*) ger lån och reglerar vissa makroekonomiska frågor. G7 och G20 är fora för de 7 respektive 20 ledande länderna att diskutera övergripande frågor om ekonomin med mera. Organisationen för ekonomiskt samarbete och utveckling (*Organization for Economic Co-operation and Development, OECD*) är en bred tankesmedja och inspiratör för sina medlemsländer.

WTO är intressant som en relativt stark organisation med möjlighet att utfärda sanktioner som upprätthåller den globala handeln och anger spelregler

för ekonomisk utveckling. Dock har förhandlingsmaskineriet gått i stå på grund av konflikter om hur nya avtal bör se ut. Dessutom vill inte USA tillsätta nya domare i WTO:s tvistelösningsorgan, vilket leder till en blockering. Kina bryter mot de nuvarande reglerna genom dolda subventioner till företag.

Handelspolitiken har ett nära samband med andra politikområden som reglerar aspekter av hur marknader fungerar. Gemensamma miljöregler kan i någon mån skapas genom de miljöpolitiska överenskommelserna. Internationella arbetsorganisationen (ILO) arbetar med gemensamma regler på arbetsrättens område, till exempel för att begränsa barnarbete. Organisationen anses mycket svagare än WTO, vilket innebär att det globala samarbetet är mindre utvecklat.

Världsbanken (*World Bank, WB*) – som egentligen består av fem separata organisationer – ger bidrag och lån och har stort inflytande som ledande aktör inom den globala biståndspolitik. Regionala utvecklingsbanker lånar ut medel för investeringar. FN har särskilda organisationer för handelsutveckling (*United Nations Conference on Trade and Development, UNCTAD*) och industriell utveckling (*United Nations Industrial Development Organization, UNIDO*) som stöder utvecklingsländerna.

På det sociala området finns Världshälsoorganisationen (*World Health Organization, WHO*) och diverse specifika organisationer för AIDS och vaccineringar mm. Inom utbildningssektorn är FN:s organisation för utbildning, vetenskap och kultur (*United Nations Educational, Scientific and Cultural Organization, UNESCO*) ett viktigt forum. FN:s livsmedels- och jordbruksorganisation (*Food and Agriculture Organization, FAO*) jobbar bland annat med effektivisering av jordbruk. Det bredaste mandatet för utvecklingsfrågor har FN:s organisation för utveckling (*United Nations Development Program, UNDP*).

De sociala frågorna och frågorna om det goda styrelseskicket är framför allt föremål för biståndspolitiska insatser. Fredsbevarande insatser sker främst i FN:s regi, men det har visat sig svårt på många håll i världen att skapa varaktig fred som leder till utveckling av rättsstat och demokrati. Samtidigt kritiserar Världsbanken och IMF för att ha drivit "good governance" ("gott styrelseskick") för hårt på bekostnad av ett fattigdomsfokus (jfr Grindle 2004). Hållbarhetsmålen mjukar upp detta, även om "good governance" är en viktig förutsättning för de övriga åtgärder som behöver vidtas. Både Världsbanken och IMF kritiserar för att vara för starkt dominerade av USA och av de europeiska länderna.

En annan kritik handlar om att bistånd kan skapa beroende och leda till korruption. FN:s hållbarhetsmål framhåller ländernas eget finansieringsansvar för sin välfärd, vilket kan ses som ett sätt att komma ur beroendet av rika länder. FN:s konventioner om mänskliga rättigheter och mot korruption har en viss betydelse men saknar starka instrument att ställa dåliga politiska ledare till svars.

Länders styrelseskick regleras mycket knappt och då främst på juridisk väg, genom konventioner om mänskliga rättigheter (Den allmänna förklaringen om mänskliga rättigheter, *United Nations Convention on Human Rights*), mot korruption (*United Nations Convention on Corruption*) och flera domstolar för att ställa krigsförbrytare till svars, bland annat den internationella brottmålsdomstolen (*International Criminal Court, ICC*). Inom FN finns ett råd för mänskliga rättigheter (*Human Rights Council, HRC*). Inom området fredsbevarande insatser är det FN:s säkerhetsråd som har ett mandat att agera (*United Nations Security Council, UNSC*). FN-stadgan reviderades nyligen så att det nu finns en skyldighet för alla regeringar att skydda sin befolkning (*Responsibility to Protect, R2P*). Däremot anses det inte finnas en juridisk skyldighet att styra länder på ett demokratiskt sätt (Wittor 2016).

Jag kommer att fokusera på de internationella organisationerna, men det bör noteras att det sker en utveckling på det juridiska området som särskilt är intressant för frågorna om ett gott styrelseskick. Karen Alter pekar på att internationella domstolar gradvis har fått ett utökat mandat att blanda sig i länders inre angelägenheter. Dessutom tillåter flera domstolar att enskilda väcker talan, vilket ökar mängden fall där stater ställs till svars (Alter 2014). Här finns en intressant parallell mellan den globala och den europeiska integrationen, där EU-domstolen påverkar utvecklingen.

Problemen med organisationerna

Ett problem med både organisationer och regelverk är att de ofta är svaga i förhållande till de förväntningar som hållbarhetsmålen väcker, vilket i sin tur har att göra med att det är svårt att uppnå enighet om organisationernas uppdrag. Det är också svårt för organisationerna att få tillräcklig finansiering och att uppfylla andra grundläggande förutsättningar. Samtidigt skapar organisationernas regler och insatser spår som påverkar den fortsatta händelseutvecklingen.

Det förs debatter om varje enskild organisations effektivitet i förhållande till sina uppdrag. Här är frågan mer begränsad, om organisationerna gör tillräckligt för att uppnå hållbarhetsmålen. Möjliga åtgärder för att stärka organisationerna kan handla om att vässa styrningen, uppdragen eller resurserna, vilket jag kommenterar i det följande.

För en organisation kan det vara rationellt att göra parallella aktiviteter som tillfredsställer olika målgrupper men som kanske inte är konsistenta med varandra (Engwall 2006). Trängd mellan oförenliga krav kan det vara lättast för en organisation att ge upp ambitionen om konsistens i sitt handlande. Det är inte heller givet att en organisations mål är att få genomslag för sina åsikter. En försiktigare strategi är att undvika risker och att inte framstå som ansvarig för misslyckanden (att undvika "blame"; jfr Hood 2011), särskilt om organisationen är ifrågasatt.

En internationell debatt handlar om hur styrelserna är sammansatta i några av de viktigaste organisationerna. En kritik handlar om att fattiga länder är underrepresenterade. En variant av detta är att tillväxtekonomier som Kina och Indien inte släpps fram att spela en större roll. Det gäller bland annat IMF, där USA inte vill minska sin andel av rösterna. I diskussionen om styrelserna föreslås både att EU-länderna ska samordna sig bättre och att andra ska få utrymme. När det gäller FN finns en diskussion om att fler världsdelar borde vara representerade i säkerhetsrådet. Före Brexit diskuterades möjligheten att Frankrike och Storbritannien skulle avstå sina platser till förmån för en gemensam europeisk plats. Ett annat alternativ är att lägga till platser för stora länder i andra världsdelar. I WTO är medlemskapet bara öppet för dem som lever upp till vissa krav. Det betyder även att organisationen skulle kunna utesluta medlemmar som inte lever upp till kraven.

En annan fråga är om organisationerna borde ha starkare mandat (uppdrag) att självständigt besluta och genomdriva nya regler. Det skulle innebära mer av överstatliga befogenheter och mindre av mellanstatligt samarbete. I den nuvarande ordningen är staternas suveränitet okränkbar, det vill säga det är bara staterna själva som kan ge organisationer överstatliga befogenheter. Alla berörda stater måste samtycka till nya regler. Särskilt på miljöområdet hörs röster som vill öka möjligheterna att tvinga länder att följa gemensamma regler som beslutas av en majoritet eller i vissa fall av en upplöst elit (Eckersley 2004).

Figur 4. Två dimensioner av handlingsutrymmet.

	Hård styrning	Begränsad styrning
Begränsat mandat	1	2
Stort mandat	3	4

De två frågorna om extern styrning och mandatets styrka kan grovt sammanfattas i fyra möjliga kombinationer. Styrningen från huvudmännens (staternas) sida kan variera mellan hård och begränsad, det vill säga röra sig mellan vänster och höger i figuren. Mandatet att driva igenom beslut kan variera mellan begränsat (obefintligt) och omfattande, det vill säga röra sig uppifrån och ner i figuren. Merparten av organisationerna har begränsade mandat. Ett sätt att göra den globala nivån starkare är då att ge organisationerna större mandat, att göra de mellanstatliga organen mer överstatliga. Staterna skulle kunna välja att ändra den nuvarande situationen från huvudsakligen ruta 1 till ruta 4 för att skapa en stark och professionell global styrning.

En lättare fråga är om organisationerna har resurser i form av pengar och personal som gör det möjligt att lösa deras uppgifter. Världshälsoorganisationen (WHO) används som exempel på en organisation som hämmas av både en

stark decentralisering och låga anslag, vilket gör att ledningen blir begränsad i sitt val av strategier (Hanrieder 2015). Huvudmännen utsätter organisationen för en hård styrning vad gäller organisationsform och finansiering. Dessa får anses vara avgörande för organisationens manöverutrymme.

Världshandelsorganisationen (WTO) har använts som exempel på en organisation som hamnat i svårigheter för att den lyckats med sitt uppdrag. En tolkning är att handelspolitiken har låsts fast därför att stora utvecklingsländer som Kina och Indien har fått en bättre förhandlingsposition. Anledningen till att de har fått en bättre position är att frihandeln, det vill säga WTO:s regler, har bidragit till ländernas utveckling (Hopewell 2016). Framgången har med andra ord skapat problem för WTO.

Kring 2010 fördes en debatt om att reformera flera av de internationella organisationerna. USA under president Obama (2009–16) var intresserade att driva fram åtminstone vissa förändringar av FN, IMF och WTO. Det diskuterades även att skapa en världsmiljöorganisation (*World Environment Organization, WEO*), men det var ett förslag som inte fick gehör bland de tilltänkta medlemsstaterna. Brookings Institution var en arena för den amerikanska debatten (Bradford & Linn 2007), där flera universitet bidrog med förslag till reformer, bland annat genom *The Princeton Project on National Security* (Ikenberry & Slaughter 2006).

Missnöjet med de internationella organisationerna har sammantaget lett till ett svagare stöd för den multilaterala ordningen. Det har sagts att världen är på väg från multilateralism till något som skulle kunna kallas minilateralism, där staterna bara gör upp om ett slags minsta gemensamma intresse (Brummer 2014). På handelsområdet sluts regionala och bilaterala avtal mellan intresserade länder när det inte går att göra upp med alla länder inom WTO. Handelspolitiken består nu av ett lapptäcke av överlappande begränsade avtal och en fråga är om detta gör det lättare eller svårare att senare göra nya överenskommelser inom WTO.

Både Parisavtalet och hållbarhetsmålen kan tolkas som exempel på minilateralism. Visserligen är i stort sett alla länder med, men det finns inga tvångsmedel i överenskommelserna. Möjligen kan man kalla det för en symbolisk multilateralism eftersom det yttligt betraktat bevarar bilden av multilaterala överenskommelser.

Problemen med helheten

Det är inte bara de enskilda organisationerna som står inför utmaningar. En annan typ av problem gäller helheten av organisationer. Ett sådant problem är att det saknas organisationer med uppdrag att ingripa mot vissa av problemen som hållbarhetsmålen strävar efter att åtgärda, till exempel nationell korruption, trots att det finns en FN-konvention som förbjuder korruption. Det finns internationella domstolar för krigsförbrytelser men inte för korruption.

Doktrinen om en "responsibility to protect" är inte tillräcklig för att legitimerade åtgärder av FN:s säkerhetsråd. Ett annat exempel är att det finns ett flertal konventioner och domstolsavgöranden som pekar på värdet av ett demokratiskt styrelseskick, men dessa kan ändå inte anses utgöra en förpliktelse för stater eller en rättighet för individer att leva i en demokratisk stat (Wittor 2016). Därmed finns ingen grund för tvångsmakt från FN eller någon annan organisation.

Ett vanligare problem är motsatsen, att det finns organisationer med överlappande mandat (Slaughter 2004). Det kan i sig vara ineffektivt, men allvarligare är om konflikter kan hanteras på flera ställen och därmed de berörda får möjlighet att välja var det passar dem bäst att ta upp frågorna. Särskilt inom den globala miljöpolitiken diskuteras problemet med överlappande och konkurrerande regelverk (van Asselt 2014). Ett exempel är det tidigare nämnda, att miljöfrågor kan hanteras inom WTO eller inom klimatkonventionen (UNFCCC). På engelska kallas det "venue shopping", när det finns möjlighet att välja arena och därmed att spela ut organisationer mot varandra.

En vanlig beskrivning är att politiken på den globala nivån är fragmenterad, med många små organisationer som konkurrerar med varandra. Uttrycket "regimkomplex" används för att beskriva en grupp av regelverk och organisationer som påverkar ett visst område och som bör analyseras som delar i ett system eller ett nätverk (van Asselt 2014). En poäng med begreppet regimkomplex är att zooma ut för att se den större helheten av organisationer och hur den fungerar i hanteringen av en viss grupp av frågor. I ett sådant regimkomplex kan ingå offentliga och privata organisationer med olika typer av roller (Mattli & Woods 2009). Organisationerna kan vara konkurrenter och samarbetspartners samtidigt.

En komplicerande omständighet är att stater ibland skapar nya organisationer istället för att reformera dåligt fungerande organisationer (Morse & Keohane 2014). Effekten blir att mängden organisationer ökar och att de nya konkurrerar med de gamla. Dessutom använder organisationer ofta en annan organisation som utförare, på ett sätt som kallas orkestrering. Om en organisation har svårt att själv genomdriva regler kan den anlita en annan organisation med större förtroende eller handlingsfrihet (Abbott et al 2015). Resultatet blir ett svåröverskådligt och svårstyrt nät av relationer.

Ett grundläggande problem är att organisationerna är smalare än hållbarhetsmålen. Det som tas upp i målen är ofta uppdelat på flera organisationer, vilket kan vara ett problem när målkonflikter ska hanteras. Detta är sedan länge uppmärksammat när det gäller balansen mellan tillväxt och miljö, där WTO är en av få organisationer som har i uppdrag att hantera båda, dock med tyngdpunkt på handelsperspektivet. Det skulle kunna vara bra att organisationerna är specialiserade och föreslår delvis motstridiga åtgärder, om det finns någon annan som kan göra en helhetsbedömning och fatta

balanserade beslut, ungefär som en nationell regering förväntas göra i förhållande till sina myndigheter. Det finns dock ingen sådan "världsregering". Istället är det upp till staterna och EU att koordinera och balansera styrningen av organisationerna.

Situationen med överlappande ansvar över politikområden och mellan aktörer på flera politiska nivåer (lokal, regional, nationell och europeisk) kallas inom Europa för "flernivåstyrning", där aktörer som väljs och finansieras av medborgarna på olika politiska nivåer är inblandade i genomförandet av politiska beslut (Hooghe & Marks 2001; Zürn 2010). Ett traditionellt exempel i Sverige är relationen mellan staten och kommunerna, där kommunerna visserligen regleras av staten, men samtidigt är självständiga aktörer med instrument och roller som överlappar med statliga myndigheters uppdrag. Exempelvis krävs det kommunala detaljplaner för byggande av nya järnvägar. Med EU och de internationella organisationerna har ytterligare aktörer med överlappande uppdrag och finansiering tillkommit i Sverige.

Begreppet flernivåstyrning är en neutral beskrivning av ett system där aktörerna är sammanflätade på olika vis. Det pekar mot en otydlig organisation i nätverksform och ett utrymme för påverkan där aktörernas strategier och föreställningar har betydelse för utfallet. Begreppet lanserades för att beskriva relationen mellan EU och medlemsstaterna på ett neutralt sätt, utan att använda begreppet "federalism" som beskrivning av ett system med sammanflätade politiska nivåer (Hooghe & Marks 2001). Traditionellt skiljer statsvetare mellan enhetsstater och federala stater, men EU är inte en stat. Det globala samfundet är inte heller en stat, men ändå ett slags politisk ordning, där aktörerna på olika nivåer är sammanflätade (Zürn 2010).

I det här sammanhanget blir FN:s hållbarhetsmål intressanta som exempel på en samlad styrning av komplexa frågor som behöver hanteras tillsammans för att vägas mot varandra. Dock är erfarenheten från de tidigare millenniemålen (*Millennium Development Goals, MDG*) att målen blev ett slags dubbelkommando vid sidan av den ordinarie statliga styrningen av de berörda internationella organisationerna, t ex Internationella valutafonden, IMF (Gutner 2010). Det blev oklart för IMF hur de globala målen förhöll sig till de egna instruktionerna. Med andra ord ökade oklarheterna snarare än minskade. Det korta svaret på frågan i rubriken är således att det finns en risk att målen inte bidrar till att skapa ordning på den globala nivån. Samtidigt finns det många exempel på att organisationer arbetar med den gemensamma agendan. Exempelvis är flera stora organisationer involverade i att stödja uppbyggnaden av moderna skattesystem som kan vinna medborgarnas förtroende.

Slutsatser

Sammantaget innebär den här beskrivningen av de internationella organisationerna att den multilaterala ordningen står inför stora utmaningar. Den ordning som har vuxit fram är inte effektiv för att hantera stora och delvis motstridiga utmaningar, särskilt när en del av utmaningarna handlar om nationella förhållanden (exempelvis länders styrelseskick). Det är önskvärt att se över enskilda organisationer så att uppdragen blir breda och ändamålsenliga. Mandatet (handlingskraften) behöver ofta förstärkas, samtidigt som styrningen och huvudmannaskapet behöver bli tydligare och bygga på ett bredare deltagande av jordens länder och invånare. Former för helhetssyn behöver utvecklas, vilket innefattar diskussion om hur målkonflikter mellan politikområden ska hanteras. I förlängningen pekar detta i riktning mot mer utvecklade former för global styrning, t ex utifrån dagens FN, G20 eller WTO.

Svaret på frågan om hållbarhetsmålen bidrar till större ordning på den globala nivån blir då att målen å ena sidan utgör en övergripande strategi, vilket ger ett slags ordning. Å andra sidan riskerar målen att skapa större oordning i styrningen av de internationella organisationerna. Vägen ur dilemmat skulle kunna vara att fortsätta reformera organisationerna, gärna med hållbarhetsmålen som utgångspunkt. För att detta ska hända krävs det politisk vilja från de stora aktörernas sida. Ur svenskt perspektiv är EU en naturlig plattform för ett sådant globalt reformarbete.

Vad kan EU göra?

EU jobbar sedan länge med de frågor som berörs av hållbarhetsmålen, framför allt genom miljöpolitiken, biståndspolitiken och handelspolitiken. Flera andra politikområden berörs. Generellt ligger EU:s ståndpunkter i linje med FN:s hållbarhetsmål, men det är naturligtvis en intressant fråga att undersöka vilka instrument EU använder, vilka deras effekter är samt hur stor den faktiska överensstämmelsen med FN:s mål är.

Sverige och andra länder i Europa har förbundit sig att gå före och visa möjligheter att kombinera tillväxt med miljöhänsyn. Europa kan visa upp modeller för teknikutveckling, inklusive investeringar och företagsklimat, samt för reglering av miljön.

EU är en tung aktör både inom den globala miljöpolitiken och den globala biståndspolitiken. EU har höga ambitioner om att påverka politiken på dessa områden men har även goda möjligheter att föra en diskussion internt och externt om målkonflikter mellan de berörda politikområdena. Den breda överblicken och erfarenheten att föra dialog mellan politikområden kan vara en tillgång i det globala sammanhanget.

Vidare är EU en stor ekonomisk aktör och har på olika sätt öppnat sina marknader för export från fattiga länder. Samtidigt förs det en debatt om att vi

begränsar importen av jordbruksprodukter. Handelspolitiken är central för EU och för uppbyggnaden av globala ekonomiska regelverk.

Sverige och EU är stora aktörer när det gäller fredsbevarande insatser, t ex genom konflikthantering i andra delar av världen. Det samma gäller insatser för att utveckla andra länders styrelseskick (rättsstat och demokrati) och skattesystem. Detta arbete sker både genom biståndspolitiken och genom utvecklingen av internationella regelverk och domstolar.

Mest intressant är kanske att EU kan bidra till en bättre analys av de insatser som görs genom att höja nivån för sina utvärderingar av biståndspolitiken. Hållbarhetsmålen kan sammantagna betraktas som en bred utvecklingspolitik, där det är intressant att studera hur olika länder hanterar både de enkla och de komplicerade och motstridiga målen. EU kan ta ett större ansvar för ett sådant lärande. Det skulle dessutom bidra till forskningen om miljöpolitik och välfärdsmodeller i utvecklingsländer (t ex Death 2016 respektive Rudra 2008).

De nya kommissionärerna har inskrivet i sina uppdrag att de ska verka för hållbarhetsmålen inom sina områden. Liknande ställningstaganden finns i EU:s globala strategi från 2016 (Europeiska unionen 2016). Men det är många steg som återstår om EU ska omsätta detta i ett samlat internationellt agerande. Målkonflikterna inom målen och mellan berörda politikområden måste hanteras på ett konstruktivt sätt. Detta talar för att analys och diskussion bör vara en central del i det fortsatta arbetet med hållbarhetsmålen.

Just nu verkar EU vara den mest samarbetsorienterade aktören på den globala politiska nivån. EU har uttalat en vilja att verka för genomförandet av de globala målen, både i Europa och i övriga världen. Det verkar ligga inom politiskt räckhåll att EU skulle kunna återropa de globala målen i sin strävan att skapa ordning i de gemensamma regelverken. Sammantaget verkar det finnas ett folkligt stöd för att hantera utmaningarna mer kraftfullt och mer samlat, både på den nationella och den internationella nivån. Samtidigt drar sig USA och andra länder tillbaka från det globala samarbetet. Troligen kommer vi snart att få se en debatt om hur målen och styrningen bör reformeras efter 2030, när hållbarhetsmålen i sin nuvarande form löper ut. Debatten om den globala ordningen har redan startat.

Referenser

- Abbott, Kenneth W., Genschel, Philipp, Snidal, Duncan & Zangl, Bernhard (red.), 2015. *International Organizations as Orchestrators*. Cambridge: Cambridge University Press.
- Acharya, Amitav, 2018. *Constructing global order. Agency and change in world politics*. Cambridge: Cambridge University Press.
- Alter, Karen J., 2014. *The New Terrain of International Law. Courts, Politics, Rights*. Princeton: Princeton University Press.

- Barrett, Scott, 2007. *Why Cooperate? The Incentive to Supply Global Public Goods*. Oxford: Oxford University Press.
- Biermann, Rafael & Koops, Joachim A. (red.), 2017. *The Palgrave Handbook of Inter-Organizational Relations in World Politics*. London: Palgrave Macmillan.
- Borás, Susana & Radaelli, Claudio Maria, 2011. "The Politics of Governance Architectures: Creation, Change and Effects of the EU Lisbon Strategy", *Journal of European Public Policy*, 18(4).
- Bradford, Colin I. & Linn, Johannes F., (red.), 2007. *Global Governance Reform. Breaking the Stalemate*. Washington, D.C.: Brookings Institution Press.
- Brunner, Chris, 2014. *Minilateralism. How trade alliances, soft law and financial engineering are redefining economic statecraft*. Cambridge: Cambridge University Press.
- Bäckstrand, Karin, 2006. "Democratizing global environmental governance? Stakeholder democracy after the World summit on sustainable development", *European Journal of International Relations*, 12(4), s. 467-498.
- Collinson, Sarah, 1999. "'Issue-systems', 'multi-level games' and the analysis of the analysis of the EU's external commercial and associated policies: a research agenda", *Journal of European Public Policy*, 6(2), s. 206-24.
- Death, Carl, 2016. *The Green State in Africa*. New Haven: Yale University Press.
- Eckersley, Robyn, 2004. *The Green State: Rethinking Democracy and Sovereignty*. Cambridge: The MIT Press.
- Engwall, Lars, 2006. "Global enterprises in fields of governance", s. 161-179 i Djelic, Marie-Laure & Sahlin-Andersson, Kerstin (red.), *Transnational Governance. Institutional Dynamics of Regulation*. Cambridge: Cambridge University Press.
- Europeiska unionen, 2016. *Shared Vision, Common Action: A Stronger Europe, A Global Strategy for the European Union's Foreign and Security Policy*. Bryssel.
- Follesdal, Andreas, 2012. "Cosmopolitan democracy. Neither a category mistake nor a categorical imperative", s. 96-114 i Archibugi, Daniele, Koenig-Archibugi, Mathias & Marchetti & Raffaele (red), *Global Democracy. Normative and Empirical Perspectives*. Cambridge: Cambridge University Press.
- Fukuda-Parr, Sakiko, 2017. *Millennium Development Goals. Ideas, Interests and Influence*. Abingdon: Routledge.
- Grindle, Merilee S., 2004. "Good Enough Governance: Poverty Reduction and Reform in Developing Countries", *Governance: An International Journal of Policy, Administration, and Institutions*, 17(4), s. 525-548.
- Gutner, Tamar, 2010. "When 'doing good' does not: The IMF and the Millennium Development Goals", s. 266-291 i Avant, Deborah D., Finnemore, Martha & Sell, Susan K. (red.), *Who governs the globe?*. Cambridge: Cambridge University Press.
- Haas, Peter M., 1992. "Introduction: Epistemic Communities and International Policy Coordination", *International Organization*, 46(1), s. 1-35.
- Hanrieder, Tine, 2015. *International Organization in Time. Fragmentation and Reform*. Oxford: Oxford University Press.
- Hill, Michael & Hupe, Peter 2002. *Implementing Public Policy*. London: SAGE.
- Hix, Simon & Høyland, Bjørn, 2011. *The political system of the European Union*. Houndmills: Palgrave Macmillan.
- Hood, Christopher, 2011. *The Blame Game. Spin, Bureaucracy, and Self-Preservation in Government*. Princeton: Princeton University Press.

- Hooghe, Liesbet & Marks, Gary, 2001. *Multi-Level Governance and European Integration*. Lanham: Rowman & Littlefield Publishers, Inc.
- Hopewell, Kristen, 2016. *Breaking the WTO. How emerging powers disrupted the neoliberal project*. Stanford: Stanford University Press.
- Ikenberry, John & Slaughter, Anne-Marie, 2006. *Forging a World of Liberty under Law: U.S. National Security in the 21st Century*. Final report of the Princeton Project on National Security, The Princeton Project Papers, The Woodrow Wilson School of Public and International Affairs. Princeton: Princeton University.
- Kamau, Macharia, Chasek, Pamela & O'Connor, David, 2018. *Transforming Multilateral Diplomacy. The Inside Story of the Sustainable Development Goals*. Abingdon: Routledge.
- Kanie, Norichika & Biermann, Frank (red.), 2017. *Governing through goals. Sustainable Development Goals as Governance Innovation*. Cambridge, Mass: The MIT Press.
- Kelsall, Tim & Booth, David et al, 2010. "Developmental patrimonialism? Questioning the orthodoxy on political governance and economic progress in Africa", Working Paper no 9, Africa Power and Politics Programme (APPP). London: Overseas Development Institute, ODI.
- Keohane, Robert, 2015. "Nominal democracy? Prospects for democratic global governance", *International Journal of Constitutional Law*, 13(2), s. 343-353.
- Knill, Christoph & Tosun, Jale, 2012. *Public Policy. A New Introduction*. Houndmills: Palgrave Macmillan.
- Lawson, David, Ado-Kofie, Lawrence & Hulme, David (red.), 2017. *What Works for Africa's Poorest? Programmes and Policies for the Extreme Poor*. Burton on Dunsmore: Practical Action Publishing.
- Leinen, Jo & Bummel, Andreas, 2018. *A World Parliament. Governance and Democracy in the 21st Century*. Berlin: Democracy without borders.
- Lieven, Anatol, 2020. *Climate Change and the Nation State. The Case for Nationalism in a Warming World*. Allen Lane.
- Lindvall, Johannes & Rothstein, Bo, 2010. *Vägar till välstånd. Sambandet mellan demokrati och marknadsekonomi*. Stockholm: SNS Förlag.
- Mattli, Walter & Woods, Ngaire (red.), 2009. *The politics of global regulation*. Princeton: Princeton University Press.
- Morse, Julia C. & Keohane, Robert O., 2014. "Contested Multilateralism", *Review of International Organization*, 9:385-412.
- Niklasson, Lars, 2019. *Improving the Sustainable Development Goals. Strategies and the Governance Challenge*. Abingdon: Routledge.
- Norris, Pippa, 2012. *Making Democratic Governance Work. How Regimes Shape Prosperity, Welfare, and Peace*. Cambridge: Cambridge University Press.
- Rudra, Nita, 2008. *Globalization and the Race to the Bottom in Developing Countries. Who really gets hurt?* Cambridge: Cambridge University Press.
- Slaughter, Anne-Marie, 2004. *A New World Order*. Princeton: Princeton University Press.
- Sorensen, Georg, 2011. *A liberal world order in crisis. Choosing between imposition and restraint*. Ithaca: Cornell University Press.
- Stern, Nicholas, 2006. *The Economics of Climate Change. The Stern Review*. Cambridge: Cambridge University Press.
- Sullivan, Helen & Skelcher, Chris, 2002. *Working Across Boundaries. Collaboration in Public Services*. Houndmills: Palgrave Macmillan.

- United Nations, 2015a. *Transforming our world: The 2030 agenda for sustainable development*. Resolution adopted by the General Assembly on 25 September 2015, A/Res/70/1.
- United Nations, 2015b. *Adoption of the Paris Agreement. Proposal by the President, Framework Convention on Climate Change, Conference of the Parties, Twenty-first session, FCCC/CP/2015/L.9/Rev.1*.
- United Nations, 2015c. *The Millennium Development Goals Report 2015*. New York
- United Nations Millennium Project, 2005. *Investing in Development. A Practical Plan to Achieve the Millennium Development Goals*. New York.
- van Asselt, Harro (red.), 2014. *The Fragmentation of Global Climate Governance. Consequences and Management of Regime Interactions*. Cheltenham: Edward Elgar.
- Weiss, Thomas & Wilkinson, Rorden (red.), 2014. *International Organization and Global Governance*. Abingdon: Routledge.
- Williams, David, 2012. *International Development and Global Politics. History, theory and practice*. Abingdon: Routledge.
- Wittor, Linda, 2016. *Democracy as an International Obligation of States and Right of the People*. Frankfurt am Main: Peter Lang.
- Zürn, Michael, 2010. "Global governance as multi-level governance" i Enderlein, Henrik, Wälti, Sonja & Zürn, Michael (red.), *Handbook on Multi-level Governance*. Cheltenham: Edward Elgar.

Varför ska vi läsa klassikerna?

En essäserie om politiskt tänkande

Som läsare har vi inte sällan formats av vissa specifika böcker som har fått stor inverkan på hur vi ser på världen och historien – och ibland påverkat även vår personliga identitet. I överförd mening kan detsamma sägas om samhällen och civilisationer: vissa författare och verk har haft en sådan djupgående inverkan på så många människor, antingen i ett specifikt historiskt ögonblick eller över längre tid, att de bidragit till att forma själva samhället och generationer av dess medlemmar.

De klassiska politiska tänkarna är av det skälet värda att återvända till gång efter annan. Vi i redaktionen har därför bett ett antal skribenter att göra just det: att återvända till tänkare eller specifika verk som har varit betydelsefulla inom ett visst ämne eller för världen i stort och att reflektera över det historiska sammanhanget, den idéhistoriska betydelsen och deras relevans för vår tid.

Med denna serie av essäer – vår alldeles egna klassikerserie – hoppas vi kunna erbjuda läsaren spännande ingångar och uppslag, och uppmuntran att själv slå sig ner i soffan eller läsfåtöljen med en klassiker.

Tredje i serien är Jon Wittrock som skriver om Carl Schmitts *Der Begriff des Politischen* från 1932.

En kontroversiell klassiker

Jon Wittrock

Carl Schmitts kompakta essä *Der Begriff des Politischen*, eller *Det politiska som begrepp*, utkom 1932 och har blivit berömd för sin kärnfulla huvudtes: att *det politiska*, den farliga, kokande kärna som genomströmmar mänskligt samliv, inklusive men inte uteslutande det vi kallar "politik", karakteriseras av skillnaden mellan *vän* och *fiende*. "Varje religiös, moralisk, ekonomisk, etnisk eller annan motsättning", skriver Schmitt, "förvandlas till en politisk motsättning om den är stark nog att dela upp människor i vänner respektive fiender."¹

En provocerande tes. En farlig tes. 1932 är som bekant inte vilket årtal som helst. Bättre blir det inte av att Schmitt kom att legitimera Hitlers maktövertagande. Att Schmitt hos många fått dåligt rykte överraskar alltså inte. Likväl hävdade Schmitt själv efter kriget att han aldrig varit övertygad nazist – men att han var en tänkare som fann inspiration hos reaktionära teoretiker och som hyste grava misstankar gällande demokratins förmåga att upprätthålla fred och stabilitet står bortom rimligt tvivel.

Debatten kring Schmitts roll i Nazityskland handlar inte om huruvida Schmitt var ett oskyldigt offer som motvilligt tvingades stödja den nya regimen eller inte – han stöttade öppet och entusiastiskt Tredje Riket och hade inget emot att engagera sig i antisemitiska attacker på judiska tänkare och deras verk. Frågan är snarare i vilken utsträckning Schmitt faktiskt höll med nazisterna, eller om han snarare "bara" var en fascistoid reaktionär, eller kanske till och med en konservativ kristen tänkare, vars antisemitism var grundad i katolicism snarare än rasbiologi. Kanske var Schmitt ytterst en auktoritär opportunist som inte alltid var en lysande ränskmidare? Nazisternas maktövertagande kom hur som helst att flytta Schmitt längre från, inte närmare, maktens och den intellektuella debattens centrum och inom den vittförgrenade nazistiska rörelsen fanns många som var nog så misstänksamma mot Schmitt, som bara en kort tid före Weimarrepublikens fall försökt bistå dess sista ledare med råd.

Efter kriget blev Schmitt en perifer figur och för många helt enkelt *persona non grata*, men hans intellektuella inflytande levde kvar och har vuxit kraftigt de senaste decennierna. Schmitt har blivit en postum gigant och även de som

1 Carl Schmitt, *Det politiska som begrepp* (Göteborg: Daidalos: 2010), s. 56.

ogillar både hans tänkande och hans politiska gärningar kan erkänna att han ofta är en skarp och klarsynt kritiker av den liberala demokratins dilemma.

Vad är då så speciellt med *Der Begriff des Politischen* och den tes texten framför? Schmitt skriver inte om privata, utan om offentliga fiender. Det handlar inte om personkonflikter, utan om konflikter mellan grupper av människor, som ytterst kan utmyнна i dödligt våld: "Konkurrenter eller motståndare i allmänhet", skriver Schmitt, "är alltså inga fiender, inte heller den personliga antagonisten som man känner motvilja eller hat mot."² Med andra ord behöver "En politisk fiende", enligt Schmitt, "inte vara moraliskt ond eller estetiskt ful. Han behöver inte uppträda som ekonomisk konkurrent, utan det kan rentav vara fördelaktigt att göra affärer med honom."³

En kritiker kan förstås invända att det politiska inte alls primärt är en fråga om vän eller fiende, utan snarare, som Aristoteles menar, om samordnat handlande, om förmågan att i samspel med andra bygga samhällen. Eller kanske mer allmänt en fråga om makt och motstånd. Eller så kanske kritikern kan invända att det politiska varken är ett objekt eller en organism, utan bara ett mångtydigt begrepp, som saknar en entydig kärna.

Ett annat problem är att Schmitt verkar blanda samman olika former av fiendskap: å ena sidan är den politiska fienden, som vi sett, enligt Schmitt en *offentlig* fiende, å andra sidan skriver han att "Var och en måste själv avgöra om det faktum att motparten i den aktuella konflikten är annorlunda innebär en negation av den egna existensformen och därför måste avvärjas eller bekämpas för att bevara det egna sättet att leva."⁴ Det verkar direkt motsäga idén om den offentliga fienden, vars status "var och en" knappast kan avgöra. Det är också oklart i vilken utsträckning fienden hotar vår fysiska existens, våra politiska och andra institutioner, eller i vidare mening vår "existensform", vårt "sätt att leva" – den politiska fienden är, skriver Schmitt, "helt enkelt den andre, främlingen, och det ligger i hans natur att på ett särskilt eftertryckligt sätt vara något existentiellt annorlunda och främmande, vilket i extremfall kan leda till konflikter."⁵ Måste offentliga fiender alltid "på ett särskilt eftertryckligt sätt vara något existentiellt annorlunda och främmande"? Är det inte snarare så att just den *politiska* fienden kan vara den existentiellt närmaste, medan den politiska vännen kan vara en främling?

Likväl har många funnit Schmitts ansats fruktbar. Betoningen av det politiska som någonting som kan bryta fram egentligen var som helst, medför en blick som inte nödvändigtvis binder *politiska* frågor till det vi i dagligt tal kallar *politik* – och de intressantaste frågorna behöver inte vara de som debatteras.

2 Ibid. 48.

3 Ibid. 46.

4 Ibid. 47.

5 Ibid. 46.

Centrala frågor kan tvärtom *avpolitiseras* – ännu ett Schmittianskt begrepp som senare teoretiker funnit fruktbart.

Frågor kan avpolitiseras – föras bort från de politiska konflikternas agendor och arenor – på flera olika sätt. Vissa frågor kan helt enkelt beskrivas som rent ”tekniska”, eller av en sådan art att det finns givna lösningar som svarar mot ett ”gemensamt intresse”.⁶ Vi kan också se mer djupgående processer där vissa begrepp, normer och institutioner politiseras och avpolitiseras över tid. Den liberala demokratin, till exempel, som förvandlats från skällsord till självklart ideal i stora delar av världen. Eller marknadsekonomin, invandringen, miljön, etc. På så sätt kan ”centrala områden”, påpekar Schmitt i en text som publiceras i anslutning till *Det politiska som begrepp*, upprättas och bli till generella utgångspunkter för alla problem som diskuteras – om vi bara löser problemen på den centrala domänen, löser sig andra problem av sig självt.⁷

Schmitt presenterar en skissartad bild över en historisk utveckling i det tidigmoderna och moderna Europa, där sådana centrala områden avlöser varandra. Det viktiga i Schmitts analys är inte hur historiskt korrekt eller detaljerad den är, utan att han väcker själva frågan om hur sådana områden bildas och hur ”de problem som finns på andra områden” kommer att ”lösas på det centrala områdets villkor. De reduceras till sekundära problem som påstås lösa sig automatiskt så fort man lyckats lösa problemen på det centrala området.”⁸

Om vi bara åtgärdar den socioekonomiska ojämlikheten, kommer vi inte längre ha problem rotade i migration och mångkultur. Eller om vi bara främjar tillväxten genom att sänka skatterna, kommer kreativa entreprenörer lösa de globala miljöproblemen. Om vi bara följer den korrekta tolkningen av vissa kanoniska texter – ideologiska, religiösa – så kommer...Etc.

Sådana centrala områden avlöser inte varandra i en entydig historisk rörelse: ”Tvärtom”, menar Schmitt, ”existerar alltid flera olika, redan passerade stadier parallellt med varandra. Personer från samma tid och samma land, från samma familj rentav, lever sida vid sida men i olika stadier.”⁹ Även om de allra flesta accepterar exempelvis den liberala demokratin, det sekulära samhället, eller marknadsekonomin, finns likväl grupper som ivrigt debatterar vilket mer eller mindre radikalt eller esoteriskt alternativ som ska efterträda dessa etablerade ordningar.

Vissa frågor avpolitiseras alltså eftersom svaren på dem ses som givna. Andra, däremot, kan försvinna från agendan helt enkelt för att de allra flesta saknar intresse för dem, även om de tidigare varit helt centrala för offentliga

6 Ibid. 41.

7 Carl Schmitt, ”Neutraliseringens och avpolitiseringsens tidsålder”, i *Det politiska som begrepp* (Göteborg: Daidalos: 2010), s. 97-111.

8 Ibid. 103.

9 Ibid. 99.

konflikter. Exempelvis är det relativt få människor i dagens Sverige som är ivrigt engagerade i de politiska följderna av relationerna mellan katolska och protestantiska kyrkor och teologiska tolkningar; något som under vissa historiska epoker naturligtvis varit en oerhört central fråga.

Den moderna politiska historien kan ur detta perspektiv betraktas som processer av politisering och avpolitisering. Kring de centrala domäner som politiseras öppnar sig sedan konfliktlinjer – i parlamentariska debatter, i media, kring köksbord och på nätföra – kring vilka politiska partier kan gruppera sig. Politiska konflikter berör inte minst gränserna för det politiska.

Schmitts teori om det politiska är alltså inte så endimensionell som det först kan verka. Han beskriver i själva verket flera olika sorters rörelser. För det första offentligt proklamerade uppdelningar i vänner och fiender, upprättandet av dikotomier i ett försök att rikta våld i vissa riktningar. Den domän där politiska ledare, men även exempelvis terrorister och gerillakrigare, deklarerar vem som är vän och vem som är fiende. Sådana förklaringar kan alltså upprätthålla men också gå på tvärs med etablerade uppdelningar. För det andra rörelser av politisering och avpolitisering, där vissa frågor lyfts upp på agendan och blir föremål för offentliga konflikter, medan andra sjunker undan, försvinner ur det offentliga ljuset och undandras ifrågasättande och konflikt.

Der Begriff des Politischen förblir ett centralt verk i Schmitts gärning, som både samlar temata han sysselsatt sig med innan han skrev den och pekar framåt mot sådant han kom att utveckla under kommande år och decennier. Schmitts verk bildar tillsammans ett vittförgrenat nät av rottrådar, som letat sig in i andra tänkares verk och pekar från hans tid mot vår egen och framtiden. Hans idéer om den politiska teologin ekar i samtiden, hans konfliktorienterade syn på det politiska har inspirerat exempelvis Chantal Mouffe i hennes egen kritik och analys av liberalism och populism, hans syn på suveränitet och undantagstillstånd ligger till grund för en väldig debatt, bland annat med utgångspunkt i Giorgio Agambens verk, hans tankar kring partisaner och stora rum pekar fram mot senare ansatser som exempelvis Mary Kaldors begrepp om nya krig och hans beskrivning av den politiska romantikens ”immanenta occasionalism” påminner inte så lite om exempelvis Zygmunt Baumans kritik av den pointillistiska tidens fragmenterade politik.

Schmitts skrifter under Weimartiden behandlar ämnen som är brännande aktuella i vår egen tid. Den starka suveränen som upphäver rättsordningen för att bevara den, folksamlingens acklamation eller avstyrkande utrop som demokratin ”urfenomen”, demokratin som primärt en fråga om identiteten mellan styrande och styrda, snarare än vissa institutionella mekanismer, eller folkets tänkta position ovanför, inuti och vid sidan av den konstitutionella ordningen, där många ser en fara i hur kanalerna från folkvilja till politiska beslut tunnas ut eller övertrumfas av andra kanaler för inflytande – allt detta är temata som

Schmitt berörde på tjugo- och trettitalen, och som nu används för att förstå vår egen samtid och dess dilemman.

Givetvis har många analyserat Donald Trumps entré på den politiska arenan och efterföljande och för många chockartade valseger med hänvisning till Schmitts tankar kring den starka suveräna makten och idéer om demokratin som fundamentalt en fråga om identiteten mellan styrande och styrda, där en demokratisk diktator kan förverkliga folkviljan utan att köra fast i långdragna förhandlingar eller urblekta kompromisser. Även Xi Jinpings mer nationalistiska och auktoritära inriktning i Kina har både analyserats och legitimerats med hjälp av Carl Schmitts begreppsvärld. Schmitt förblir, på gott och ont, aktuell i vår egen tid.

Som ingång till Schmitts provocerande, tankeväckande verk fungerar *Det politiska som begrepp* fortfarande utmärkt, men läst i isolering riskerar den att tolkas på ett alltför enahanda sätt, särskilt om vi inte uppmärksammar dess interna motsättningar och oklarheter.

KLASSIKERN: Carl Schmitt, *Der Begriff des Politischen, mit einer Rede über das Zeitalter der Neutralisierungen und Entpolitisierungen, neu herausgegeben von Carl Schmitt*. München und Leipzig: Duncker und Humblot, 1932; översatt av Svenja Hums som *Det politiska som begrepp*. Göteborg: Daidalos, 2010.

Statsvetenskapliga förbundet

Förbundsredaktör: Katarina Roos

Swepsa 2020

av Katarina Roos

Denna gång är temat för förbundssidorna Swepsa 2020. Det blir även andra aktuella rapporter med relevans för förbundet.

SWEPSA 2020 I DIGITALT FORMAT

Ett återkommande inslag på förbundssidorna i *Statsvetenskaplig tidskrift* är en kort rapport från Swepsa. Det var första gången som Mittuniversitet stod värd för arrangemanget och som konferensen genomfördes med hjälp av videosamtal. Efter att såväl ECPR som NOPSA ställts in, var det extra välkommet att Swepsa kunde genomföras, om än i digitalt format. De flesta arbetsgrupper kunde genomföras som planerat. Däremot fanns inte några gemensamma programpunkter utöver själva årsmötesförhandlingarna.

Inför konferensen hade elva arbetsgrupper förannonserats: 1) *Gender, Politics and Theory*; 2) *Environmental Politics*; 3) *Political Behavior and parties (Swepop)*; 4) *Challenges to democracy, answers from political theory?*; 5) *Governance and Public Administration*; 6) *European Studies and International Relations*; 7) *Migration and integration – structures, policies and actors*; 8) *Critical perspectives on power and politics*; 9) *Political Communication*; 10) *Risk Assessment, Risk Analysis, and Crisis Management*; resp. 11) *Comparative Politics*. Det blev dock vissa ändringar i programmet och slutligen fullföljde åtta arbetsgrupper. Antalet papper som behandlades i dessa grupper varierade från ett till tolv. Vid föregående års Swepsa-konferens presenterades omkring 90 papper. I år blev

det drygt 50 uppsatser. Minskningen får sannolikt tillskrivas formatet och den osäkerhet som präglar många arbetsituationer under den pågående pandemin.

RAPPORT FRÅN ARBETSGRUPPERNA

På grund av för få anmälda papper blev grupp nio och elva inställda och i arbetsgrupp fyra presenterades endast ett papper. En av de inställda grupperna, *Political Communication*, med fokus på relationen mellan medborgare, media och politik, var en nykomling i Swepsa-sammanhang. Kajsa Falasca som skulle ha lett den tillsammans med Lars Nord, båda Mittuniversitetet, berättar att hon tror att det berodde just på att arbetsgruppen är ny och inte hunnit göra sig känd. Förhoppningen är dock att kunna genomföra en arbetsgrupp vid nästa års konferens. En annan nykomling i listan över arbetsgrupper var *Risk Assessment, Risk Analysis, and Crisis Management*. Fem papper presenterades av lika många närvarande författare. Två av artiklarna var samförfattade, så det totala antalet författare var nio. Deltagarna i arbetsgruppen kom från Försvarshögskolan, Uppsala universitet, Örebro universitet, Utrikespolitiska institutet samt värduniversitetet. Två teman kunde urskiljas, Covid-19 och klimatförändringar, och ett papper behandlade crisis management. Pär Olausson, Mittuniversitetet, som ledde arbetsgruppen, berättar att alla deltagare i gruppen uttryckte en förhoppning om att den ska bli återkommande på Swepsas årsmöten.

Grupp ett och åtta slogs samman till en gemensam arbetsgrupp som leddes av Sara Nyhlén, Mittuniversitet, och Andreas Öjehag, Karlstads universitet. Nio papper presenterades med totalt 13 författare. Arbetsgruppen

Environmental Politics leddes av Stefan Linde, Mittuniversitetet, och Gregory Poeltzer, Luleå tekniska universitet. Nio papper presenterades, varav flera var samförfattade, så det totala antalet författare som bidragit var femton. De papper som behandlades handlade om naturresursförvaltning och miljöbeteenden, gröna stater och internationell miljöpolitik, samt hållbar utveckling och klimatanpassning. Gruppen hade en bred representation, med deltagare från Göteborgs universitet, Karlstads universitet, Luleå tekniska universitet, Lunds universitet, Mittuniversitetet, Stockholms universitet, Umeå universitet, Örebro universitet, samt Utrikespolitiska institutet.

Liksom vid föregående års upplaga av Swepsa var arbetsgruppen *Politiskt beteende och partier (Swepop)* den största. Den leddes av Niklas Bolin och Stefan Dahlberg, Mittuniversitetet, och totalt behandlades tolv uppsatser. Av dessa var fem ensamförfattade medan övriga hade 2–4 författare. Totalt var 20 personer involverade som författare. Deltagarna i arbetsgruppen representerade ett antal svenska lärosäten: Göteborgs universitet, Karlstads universitet, Mittuniversitetet, Umeå universitet samt Uppsala universitet. Gruppen hade även internationella deltagare. En uppsatsförfattare kom från Helsingfors universitet och en från University College of Dublin. Arbetsgruppen hade möten samtliga tre dagar och arbetsspråket var engelska.

Gruppen *Governance and Public Administration* leddes av Evangelia Petridou, Mittuniversitetet, och Petra Svensson, Högskolan i Halmstad. Totalt presenterades sju papper varav tre var samförfattade, så det totala antalet författare var 12. Det var stor spännvidd bland de papper som presenterades. Två av dem behandlade styrning av svensk skola, men ur helt olika perspektiv. De andra papperen behandlade representativ byråkrati, överklaganden av beslut, nätverk inom äldreomsorg, implementering samt medborgardialoger.

Arbetsgruppen *European Studies and International Relations* leddes av Sofie

Blombäck, Mittuniversitetet. Två av de för-anmälda papperen drogs tillbaka kort inpå konferensen, vilket resulterade i att totalt elva papper presenterades. Flera av papperen var samförfattade, men sammantaget 18 författare. Totalt deltog 17 personer aktivt i arbetsgruppens arbete, men inte under alla tre dagar. En grupp av papper kom från ett projekt som utvärderar Sveriges medlemskap i FN:s säkerhetsråd. Flera papper kombinerade arbetsgruppens två teman, Europapolitik och internationella relationer. Exempelvis fokuserades EU:s utrikes- och säkerhetspolitik. Sofie Blombäck berättar att de valde att bjuda in även medförfattare och andra som hade intresse av att diskutera enskilda papper. Det digitala formatet möjliggjorde detta, men konsekvensen blev att författarna inte kunde närvara hela tiden eftersom de hade undervisning och andra inbokade aktiviteter.

Arbetsgruppen *Migration and integration – structures, policies and actors* leddes av Gustav Lidén, Mittuniversitetet. Totalt inkom sex papper som sammantaget nio författare stod bakom. I arbetsgruppen deltog utöver sex av författarna, ordföranden och ytterligare en deltagare. Deltagarna representerade många av landets statsvetenskapliga miljöer, inklusive Göteborgs universitet, Linnéuniversitetet, Mälardalens högskola, Stockholms universitet, Umeå universitet, Uppsala universitet samt värduniversitetet. Arbetsgruppen präglades därtill av ett visst tvärvetenskapligt inslag genom en deltagare från ämnesområdet folkhälsa, som tematiskt anslöt väl till gruppens inriktning. Flera papper berörde gemensamma teman. Ett sådant innefattade studier av lokal integrationspolitik, såväl utifrån ett institutionellt perspektiv som ett individcentrerat sådant. Ett annat berörde geografiska analyser av frågor kopplade till tillit och utveckling av bostadsområden, särskilt i relation till individer med utländsk bakgrund. I arbetsgruppen diskuteras därtill frågor om anhöriginvandring och hälsoaspekter inom ramen för integrationspolitiken. De deltagande bidragen uppvisade

stor metodologisk bredd, vilket i allmänhet är en styrka inom detta forskningsområde. Gustav Lidén berättar att diskussionerna förlöpte väl och i konstruktiv anda, det digitala formatet till trots.

ÅRSMÖTESFÖRHANDLINGARNA

Statsvetenskapliga förbundets årsmöte hölls den 1 oktober i anslutning till Swepsa och programpunkten bjöd på sedvanliga årsmötesförhandlingar. Såväl ordförande Patrik Hall som sekreterare/kassör Magnus Erlandsson, både Malmö universitet, omvaldes på sina poster. Vice ordförande Annica Sandström, Luleå tekniska universitet omvaldes likaså. I övrigt blev det små förändringar i styrelsen. Niklas Bolin (MIUN), Johan Eriksson (SH), Gustav Sundqvist (MDH), Annica Fredén (KU), Magnus Hagevi (LNU), Kjell Engelbrekt (FHS), Martin Karlsson (ORU), Sara Svensson (HH), Khalid Khayati (LIU), Anna Michalski (UU) och Henrik Oskarsson Ekengren (GU) omvaldes som ledamöter, liksom doktorandrepresentanten Fredrik Carlsson (LIU). Tre nya ledamöter valdes in i styrelsen: Ulrika Mörth (SU), Dahlia Muhktar-Landgren (LU) och Therese Bjärstig (UMU). De ersätter Maria Jansson (SU), Douglas Brommesson (LU) och Katarina Roos (UMU), som lämnar styrelsen.

Styrelsen har liksom tidigare två allmänrepresentanter. Thomas Larue, sekretariatschef vid Riksdagens utredningsenhet, omvaldes. Peter Santesson, opinionschef Demoskop, lämnade styrelsen och ersattes av Tove Danne-stam. Tove är disputerad vid Lunds universitet och är nu chef för Omvärld och samverkan på Myndigheten för tillgängliga medier.

Styrelsen hade föreslagit att Statsvetenskapliga förbundets årsmöte 2021 ska arrangeras av Mittuniversitetet eftersom 2020-års upplaga blev en Coronaanpassad minimivariant. Förslaget bifölls av årsmötet. Det innebär också att Örebro universitet, som står på tur att arrangera enligt det rullande schemat, får arrangera årsmötet ett år senare än planerat, 2022. Årsmötet fattade även beslut om rutinen för valberedningens arbete. I syfte

att få kollegial behandling av nomineringar till styrelsen vid varje lärosäte, samt för att få till stånd en mer varierad representation, ska valberedningen även fortsättningsvis kontakta ämnesansvariga vid medlemslärosätena i början av valberedningsprocessen och då uppmana varje lärosäte att föreslå två kandidater, en man och en kvinna. Det kan noteras att könsfördelningen i den nuvarande styrelsen är jämn men inte i valberedningen, där sex av totalt åtta ledamöter är män.

FÖRBUNDETS PRIS FÖR BÄSTA UPPSATS I STATSVETENSKAPLIG TIDSKRIFT 2019

Vid årsmötet meddelades även förbundets pris för bästa uppsats i *Statsvetenskaplig tidskrift* som denna gång tilldelas Emma Ricknell för uppsatsen ”’Is Sweden Finally Waking Up?’ Debating the 2018 Swedish National Election on 4chan”. Emma är doktorand vid Institutionen för statsvetenskap vid Linnéuniversitetet. I den prisbelönta artikeln görs en innehållsanalys av attityder och beskrivningar av politiska partier i diskussionstrådar i ett engelskspråkigt, anonymt sub-forum på plattformen 4chan. Syftet är att bidra till ökad kunskap om hur politiska positioner formeras i forum av den här typen och vilken roll de spelar i komplexa information-s och kommunikationsflöden. Uppsatsen finns att läsa i sin helhet i nummer 3 av *Statsvetenskaplig tidskrift* 2019.

APROPÅ PRISER – NYINSTIFTAT PRIS FÖR BÄSTA UPPSATS OM RIKSDAGEN

Riksdagen har instiftat ett uppsatspris och 2020 är första gången som det delas ut. Uppsatsen kan vara på grund- eller avancerad nivå och ska ha fått högsta betyg. Ämnet är fritt, men uppsatsen ska ha en tydlig koppling till riksdagen. Det är institutionen som nominerar och uppsatsen ska vara examinerad under en viss tidsperiod för att kunna komma ifråga. Thomas Larue, som är allmänrepresentant i förbundets styrelse, berättade vid styrelsemötet den 23 november att ett antal uppsatser nominerats och att processen för att

utse vinnaren pågår. Det är riksdagsdirektören som fattar beslut på rekommendation av en priskommitté som består av talmannen, de vice talmännen och tjänstepersoner. Mer information och besked om vem som tilldelas 2020 års uppsatspris annonseras på riksdagens webbplats. Intentionen är att detta ska bli ett årligt återkommande pris, vilket kan vara värt att ha i åtanke när terminsslutet nu närmar sig och det är dags att examinera uppsatser.

AVRUNDNING

Som nämndes i föregående nummer sammanställde Leif Johansson, Lunds universitet, en lista över alla svenska statsvetenskapliga avhandlingar från 1890 till 2009 inför förbundets 50-års jubileum. Leif har nu utökat den till utgången av 2019. Den uppdaterade listan finns i sin helhet på förbundets hemsida, under fliken Doktorander, <https://swepsa.org/doktorander/avhandlingar.html>. En länk finns även bland inläggen på Statsvetenskapliga förbundets facebook-sida.

Den nyvalda styrelsen hade sitt första styrelsemöte den 23 november, 2020. Den beslutade då att adjungera den nuvarande förbundsredaktören för förbundssidorna i Statsvetenskaplig tidskrift till styrelsen. Det är viktigt att redaktören har kännedom om styrelsens arbete för att kunna återrapportera och lyfta viktiga frågor som styrelsen behandlat här på förbundssidorna.

Då var det så dags att avrunda, såväl förbundssidorna som året 2020, som för alltid kommer att associeras med coronapandemin. Ett år vars professionella sfär präglats av social distansering, hemarbete, fjärrundervisning, inställda konferenser och uppskjutna datainsamlingar. Nya ord och uttryck har letat sig in i och integrerats i det professionella vardagspråket såsom "muta" och "unmuta" och "joinande" av "breakoutrooms". Visst finns det också fördelar med att vi har lärt oss att undervisa, samtala och konferera i digitala mötesrum. Helt plötsligt är det möjligt att delta i intressanta seminarier och

workshops världen över, och att bjuda in kollegor till samtal har aldrig varit enklare. Vi ska dock inte underskatta det sociala mötet, också som drivkraft till att anmäla sig till en konferens. En konferens är inte bara presentationer av och diskussioner kring papper, utan också minglande och nätverkande under bensträckare, kaffepauser och middagar. Under de pauserna knyts intressen samman, som lägger grunden för framtida forskningssamarbeten – och vänskap. De sociala relationerna fungerar som nödvändigt kitt och smörjmedel även i de professionella relationerna. Innebörden i det förstår vi kanske fullt ut först nu, när möjligheterna till social interaktion har begränsats så kraftigt.

Häromsistens stod jag i begrepp att anmäla mig till en årligen återkommande konferens som jag sällan missar. Plötsligt insåg jag att konferensdeltagandet denna gång var liktydigt med att tillbringa två och en halv dag hemma framför min dataskärm. Inga goda middagar tillsammans med nya bekantskaper och gamla vänner, utan mikrovärmda makaroner med mig själv som enda sällskap. Än värre, den sista konferensdagen är alltid en lördag. Den dagen skulle jag ha sällskap av något önskat slag, eftersom jag skulle tvingas dela vardagsrummet – den enda platsen i hemmet där uppkopplingen är tillräckligt stabil för Zoomande – med min lördagslediga familj. De yngre familjemedlemmarna har förvisso gjort det till något av en konstform, att spana in kollegorna på skärmen utan att själva synas, men scenariot kändes ändå inte särdeles lockande. Jag klickade snabbt bort det halvt ifyllda digitala anmälningsformuläret till konferensen. Det kommer förhoppningsvis fler chanser. Låt oss hoppas att vi kan träffas och arbeta tillsammans på nätverksmöten, workshops och internationella konferenser i den fysiska världen snart igen, och att vi under Mittuniversitetets värdskap får träffas på plats på Swepsa hösten 2021.

Katarina Roos är verksam vid Statsvetenskapliga institutionen, Umeå universitet.
E-post: katarina.roos@umu.se

Litteraturgranskningar

Litteraturredaktör: Björn Östbring

De motade inte Adolf i grind

Benny Carlson

Barth, Rüdiger & Friedrichs, Hauke, 2019. *Dödgrävorna: Den sista vintern i Weimar-republiken*. Stockholm: Weyler förlag.

Vuillard, Éric, 2019. *Dagordningen*. Stockholm: Lind & Co.

Bouverie, Tim, 2020. *Appeasing Hitler: Chamberlain, Churchill and the Road to War*. London: Vintage.

Gilly, Pierre, 2019. *Europa mot katastrofen: Förspellet till andra världskriget*. Stockholm: Verbal förlag.

Historien uppvisar många exempel på politiska psyko- eller sociopater som ser sin chans i situationer av social och ekonomisk oro och som inte skyr några medel för att samla, utvidga och missbruka sin makt. Problemet för politiska ledare som utövar makt med måtta i länder under mer stabila förhållanden liknar det på skolgården: Hur ska de normalt funtade barnen kunna sätta mobbaren och hans medlöpare på plats? Att be snällt fungerar inte. Fyra böcker om bemötandet av Hitler och nazismen tydliggör farorna med att underskatta abnorma ledare och rörelser och tro att de går att kontrollera eller kompromissa med.

*

Rüdiger Barths och Hauke Friedrichs *Dödgrävorna* utspelar sig under ett par månader vintern 1932-33 i ett Tyskland förlamat av ekonomisk kris och politiska motsättningar. I riksdagsvalet i början av november har nazisterna fått 33 och kommunisterna 17 procent av rösterna. Franz von Papen är rikskansler och regerar med nödförordningar utfärdade av den 85-årige rikspresidenten Paul von Hindenburg. Demokratien är i fritt fall.

Barth och Friedrichs, historiker och journalister, berättar att de fick uppslaget till boken när de tittade på "House of Cards". De insåg att kampen om

Benny Carlson är professor emeritus vid Ekonomisk-historiska institutionen, Ekonomihögskolan vid Lunds universitet.

E-post: benny.carlson@ekh.lu.se

makten under Weimarrepublikens sista dagar var lika utstuderad som i den amerikanska TV-serien. I det här fallet överträffar verkligheten dikten med råge.

Den process som ledde fram till *Machtübernahme* är i stora drag välkänd men här får man följa den ”i realtid” under 75 dygn från 17 november 1932, när von Papen avgår som rikskansler, och fram till dess att Hitler – efter en serie förvecklingar in i det sista – svär ämbetseden som rikskansler vid lunchtid måndagen den 30 januari 1933. *Dödgrävarna* ingår i en trend som går ut på att skriva mikrohistoria för en kort period baserad på samtida primärkällor (protokoll, dagböcker, brev, tidningsartiklar). Andra aktuella och lyckade exempel i den vägen på tysk grund är Volker Weidermanns *München 1918* om den bayerska rådsrepubliken, Tobias Straumanns *1931* om regeringen Brüning:s desperata kamp för att undvika en finanskris samt Oliver Hilmes *Berlin 1936* som utspelar sig under Berlinolympiaden. När de stora dragen är klarlagda och måttligt intressanta att upprepa blir denna typ av historieskrivning mycket givande i och med att man kommer ett antal aktörer in på livet och genom dem kan återuppleva tidens stämningar och bedömningar under ovisshet om framtiden. I stället för historieskrivningar som blickar bakåt för att berätta och förklara blir man förflyttad i tiden och börjar nästan hoppas att slutet ska bli ett annat än det (man vet att det) blev.

Dödgrävarna försätter passande nog läsaren i begravningsstämning. Hur kunde detta för alla inblandade förvirrade och nervslitande drama – med huvudaktörer som den kompromisslöse Hitler, den ålderstigne Hindenburg, den intrigante von Papen, den maktsgugne Alfred Hugenberg och framför allt ”den sociale generalen” Kurt von Schleicher, som under större delen av de 75 dagarna (2 december till 28 januari) försökte lotsa Tyskland genom krisen – sluta så olyckligt? Hur kunde den traditionella ekonomiska och politiska eliten bli så orkes- och omdömeslös att den lät makten glida över i händerna på en äventyrare som Hitler, särskilt i en situation när nazistpartiets popularitet var dalande? Hur kunde man tro att det skulle gå att kontrollera en fanatiker som förfogade över en armé av brunklädda bråkstakar? Det är svårt att hitta motsvarande tillfällen i historien när några aktörers målmedvetenhet respektive fumlighet har fått lika ödesdigra konsekvenser. Ryska revolutionen är väl det exempel som ligger närmast i modern tid.

*

Éric Vuillards lilla konstverk *Dagordningen* börjar där *Dödgrävarna* slutar. Måndagen den 20 februari 1933 möter två dussin tyska industrialister med Gustav Krupp i spetsen Göring och Hitler, vilka utber sig om finansiering av valrörelsen inför riksdagsvalet den 5 mars. Göring lovar att det ska bli sista valet på hundra år om nazisterna får majoritet och Hitler lovar ”stoppa det kommunistiska hotet, avskaffa fackföreningarna och låta varje arbetsgivare bli führer för sitt eget företag” varpå den blivande riksbankschefen Hjalmar Schacht kommer

till saken: "Och nu, mina herrar, fram med plånboken" (s. 17-18). Innehållet i berättelsen är inte märkvärdigt men skildringen av den förtätade stämningen är sådan att läsaren tror sig kunna höra en knappål falla. Här handlar det inte så mycket om naivitet som om en investering, som, om Tyskland vunnit kriget, skulle ha visat sig oerhört lönsam inte bara i det korta perspektivet (slavarbete, krigsmateriel, gas) utan även i det långa. "Mötet den 20 februari 1933, som för oss skulle kunna framstå som ett unikt ögonblick i företagshistorien, en oerhörd eftergift till nazisterna, var för Krupp, Opel och Siemens inget annat än ett ganska vanligt mellanspel i affärslivet, en trivial penninginsamling", skriver Vuillard (s. 18).

I ett kapitel rubricerat "En artighetsvisit" återges ett exempel på brittisk sorglustig sorglöshet: Lord Halifax jagar och dinerar i november 1937 med den på en gång jovialiske och brutale Göring. Hur mycket ser han? "Ja, han kan inte ha undgått den skrämmande kärnan under den degiga och uppsvällda masken", tror Vuillard (s. 24). Halifax reser till Berghof och möts av Hitler, som han för ett ögonblick misstar för en betjänt. Vuillard kan hålla sig för skratt (s. 24): "Den brittiske aristokraten och diplomaten som står bakom sin lilla rad av anfäder, döva som tromboner, korkade som fårskallar, blinda som hönor, lämnar mig helt kall."

En politiker som säger "ja" inför Hitlers skrämselfaktik är den österrikiske kanslern Kurt von Schuschnigg. Han har i och för sig just inget annat val eftersom tyskarna tänker genomföra *Anschluss* oavsett. Schuschnigg framstår inte desto mindre som en ömklig figur eftersom han tidigare sagt "modigt nej" till pressfrihet, socialdemokrati, strejkrätt, parlamentsval och andra partier än sitt eget. I slutfasen avbokat han inför Hitlers hot en folkomröstning om Österrikes självständighet. Vuillards (s. 61) dom är hård: Då "vände sig den orubblige Schuschnigg, nejsägaren, negationens diktator, mot Tyskland, och gav med kvävd röst, rött tryne och fuktiga ögon ifrån sig ett kraftlöst 'ja'".

*

Episoden när Halifax misstar Hitler för en betjänt figurerar även i Tim Bouveries *Appeasing Hitler*, "en scen som kunde ha kommit från P. G. Wodehouse" (s. 142). Bouveries historik är till skillnad från Vuillards konventionellt berättad, en gedigen redogörelse för den brittiska eftergiftspolitikerna mot Hitler från maktövertagandet 1933 till slutet på "låtsaskriget" 1940, en politik som bedrevs av Stanley Baldwin och Neville Chamberlain, konservativa premiärministrar 1935-37 respektive 1937-40. De båda herrarna begrep aldrig med vilken ful fisk de hade att göra men medan Baldwin var allmänt loj och oföretagsam var Chamberlain en energisk optimist som in i det sista trodde att det skulle gå att "göra affärer" med Hitler. Chamberlain var inte, som man tidigare lätt kunnat få intryck av, en svag och obeslutsam karaktär utan en mycket målmedveten politiker, beredd att förhandla in i kaklet. Som Anthony Eden, utrikesminister 1935-38, anförtror sin dagbok: "Jag är rädd att det grundläggande problemet

är att N. tror att han är en man som har till uppgift att komma överens med diktatorerna” (s. 159). En annan samtida observatör liknade premiärministerns ansträngningar vid att ”klia en krokodil på huvudet för att få den att spinna” (s. 220). Chamberlain var naturligtvis inte ensam om dessa ansträngningar utan hade vid sin sida eftergiftsmän som Lord Halifax, utrikesminister 1938-40, och Nevile Henderson, ambassadör i Berlin 1937-39, för att nämna de mest kända.

Chamberlain underskattade Hitler grovt. Efter sitt första möte med Führern beskrev han denne som någon ”man aldrig skulle lägga märke till i en folksamling utan skulle ta för den husmålare han en gång var” (s. 250). Samtidigt trodde han – liksom Hindenburg, von Papen och Hugenberg gjort några år tidigare – att Hitler var en man som stod vid sitt ord. Hitler å sin sida hade naturligtvis bara förakt till övers för såväl Chamberlain som borgerliga hedersbegrepp.

Eftergiftspolitiken byggde emellertid inte bara på underskattning utan också på svaghet. Storbritanniens utgångsläge var inte det bästa. Hotat av Japan i Fjärran Östern, av Italien i Medelhavet och av Tyskland på den europeiska kontinenten, med dåligt samvete efter Versaillesfreden, starka inhemska pacifistiska strömningar efter första världskriget och ett nedrustat försvar som gav militärledningen ”ett rykte som de värsta defaitisterna i landet” (s. 88), gick Chamberlains strategi ut på att till varje pris undvika krig. Att inte heller fransmännen var beredda att sätta hårt mot hårt gjorde inte saken bättre.

Hitlers initiala planer på att inkludera alla befolkningar av tyskt ursprung i Tredje riket betraktades som rimliga. Inmarscherna i Rhenlandet och Österrike accepterades utan större knot. Först med Hitlers krav på att inkorporera den tyska minoriteten i Tjeckoslovakien ställdes saken på sin spets eftersom landet var allierat med Frankrike. Efter tre flygresor till Tyskland för att överlägga med Hitler kunde Chamberlain 30 september 1938 återvända från München i triumf och utropa ”fred i vår tid” och, till råga på allt, trots sveket mot tjeckerna, ”ärofylld fred”. Han blev för en kort tid så populär att han erbjöds ett lantställe i Frankrike, till vilket han tackade nej med en för omväxlings skull humoristisk glimt i ögat: han hade inga territoriella ambitioner i Europa.

När Hitler ett halvår senare marscherade in i Prag gick fredsdrömmarna upp i rök. Det stod klart att man hade med en motståndare att göra för vilken internationella avtal inte var värda papperet de var skrivna på. (Symptomatiskt nog hade bläckhornet varit tomt när Hitler skulle sätta sin namnteckning under Münchenöverenskommelsen.) Storbritannien och Frankrike ingick försvarsallians med Polen och när Hitler 1 september 1939 angrep landet hedrade de sitt åtagande. Den 10 maj 1940 tvingades Chamberlain avgå som premiärminister och ersattes av den man som hela tiden bekämpat eftergiftspolitiken, Winston Churchill.

Bouveries slutsatser går ut på att Storbritannien och Frankrike på ett tidigt stadium borde ha synat Tysklands och Italiens bluffar – inmarschen i Rhenlandet respektive angreppet på Abessinien – och försökt bygga breda allianser,

även om de tilltänkta parterna, det isolationistiska USA och det kommunistiska Sovjetunionen, inte var lätta att ha att göra med. Chamberlain gjorde emellertid aldrig några seriösa försök. Han behandlade USA med "kyligt förakt" (s. 421) och såg Sovjet som ett närmast värre hot än Nazityskland; de förhandlingar som fördes med ryssarna sköttes otroligt illa. I stället ställde han sitt hopp till att Mussolini skulle utöva ett modererande inflytande på Hitler. Det var som att sätta räven att vakta vargen. Det argument som eftergiftspolitikens försvarare har anfört, att Storbritannien köpte tid för att rusta upp sitt försvar, ger Bouverie inte mycket för eftersom tyskarna rustade upp mer än britterna mellan München och krigsutbrottet.

*

I Pierre Gillys *Europa mot katastrofen* figurerar många av de aktörer och episoder som befolkar de tre tidigare nämnda böckerna men här insatta i ett längre perspektiv från vapenvilan i november 1918 till krigsutbrottet 1939: *Machtübernahme, Anschluss*, Münchenöverenskommelsen och mycket mer. Även här tar Halifax Hitler för en betjänt, även här liknar Chamberlain Hitler vid en husmålare, även här (s. 197) gör Schuschnigg en slät figur: "Så länge motståndarna var demokratiska eller moderata värnade han alltid sin makt och Österrikes självständighet, men ställd inför Hitler kapitulerade han fort."

Trots det långa svepet genom mellankrigstiden blir framställningen ingen ytlig upprepning av allmängods utan bjuder, med hjälp av rikliga citat ur dagstidningar, politiska tal, dagböcker, memoarer och till och med erfarenheter ur den egna familjehistorien, på en överraskande mängd detaljer. Som gammal ekonomihistoriker kan man för övrigt med nöje notera att Dudley Dillards klassiker om *Västeuropas och Förenata staternas ekonomiska historia*, som i många år var inkörspport till ämnet, kommer till användning några gånger.

Halva boken behandlar eftergiftspolitiken. Gillys och Bouveries berättelser överlappar alltså till stor del varann men där Bouverie ställer britterna i förgrunden lägger Gilly, tack vare sin svensk-franska bakgrund, större vikt vid fransmännens agerande. Frankrike led av svårartad politisk polarisering och hamnade under eftergifternas gång i ett allt svårare läge (s. 284): "När Tjecko-slovakien försvann som militärmakt kunde trettio tyska divisioner sättas in mot Frankrike. 1939 blev läget ännu värre. Italien och Tyskland ingick den så kallade Stålpakten. Och i Spanien hade fascisterna segrat. Det innebar att Frankrike vid ett eventuellt krig hotades från tre håll." Dessutom förklarade sig Belgien, som varit integrerat i den franska försvarsplanen, neutralt. Konseljpresidenten Édouard Daladier var inte lika naiv som Chamberlain utan lär (enligt Jean-Paul Sartre) vid hemkomsten efter Münchenöverenskommelsen, när han möttes av jublande folkmassor, ha mumlat: "Förbannade kräk!"

Gilly har ett livfullt språk som triumferar inte minst när han introducerar huvudaktörerna. Mussolini beskrivs (s. 66) som "en robust men kortvuxen man med enorm haka, stor mun och stora, bruna, rörliga ögon som ibland såg ut

som om de skulle poppa ur sina hålor” och Frankrikes konseljpresident Aristide Briand (s. 74) som ”en man med ovårdat hår, vildvuxen mustasch, säckiga kläder och päronformad kropp”. Den franska Folkfrontens ledare Léon Blum sägs (s. 146) ha haft ”en vana att vrida och vända på politiska frågor så att alla sidor blev granskade, vilket ibland kunde ge intrycket att han själv varken visste ut eller in”. Dramatiken i det stora omväxlar med komiken i det lilla. Enligt Hitlers fotograf Heinrich Hoffman snubblade Hitler flera gånger på ett isbjörns-huvud i det hotell i Linz där han övernattade efter sitt triumfatoriska intåg i Österrike. Man kan inte, trots att Hitler var nykterist, låta bli att associera till TV-sketchen med grevinnan och betjänten.

*

Vilka lärdomar kan vi dra ur dessa böcker? Som jag ser det är sensmoralen, för den som vill leva i demokrati och frihet, i princip enkel (men svår att efterleva i praktiken):

Externt gäller det att bilda demokratiska allianser mot politiska psyko/sociopater och inte försöka blidka dem. Dessa typer är relativt lätta att identifiera eftersom de inte ens låtsas (även om de håller låtsasval) respektera demokrati, yttrande- och pressfrihet och mänskliga rättigheter. Problemet är att de ofta har en betydande del av folket i sina respektive länder, den genom historien ständigt skrånande pöbeln, bakom sig. Verkligt demokratiska ledare kan inte rida på en pöbel utan måste med rationella argument motivera sina befolkningar att stå emot lögn, hot, utpressning och andra standardvapen i de auktoritära regimernas arsenal.

Internt gäller det att å ena sidan att hålla högt i tak, å andra sidan hålla efter extremister som försöker använda demokratin som dörrmatta till sina (mar) drömsamhällen. Det gäller att dra gränsen mellan dem som angriper missförhållanden i syfte att åstadkomma samhällsförbättringar och dem som är ute efter att vända spelet över ända. Kruxet ligger i att skilja agnarna från vetet, att identifiera dem som spelar det demokratiska spelet på ett sluttande plan.

Det förefaller som om många nationer för närvarande befinner sig på det sluttande planet. Värst av allt är naturligtvis att USA, som så länge varit demokratins mäktiga utpost, på senare år har befunnit sig i ett tillstånd som för tankarna till den sena Weimarrepubliken.

REFERENSER

- Hilmes, O., 2019. *Berlin 1936: Sexton dagar i augusti*. Stockholm: Lind & Co.
- Straumann, T., 2019. *1931: Debt, Crisis and the Rise of Hitler*. Oxford: Oxford University Press.
- Weidermann, V., 2020. *München 1918: När författare och drömmare tog makten*. Stockholm: Lind & Co.

Gerring, John & Veenendaal, Wouter, 2020.
Population and Politics: The Impact of Scale.
Cambridge: Cambridge University Press.

Anmälan av Carsten Anckar

Antagandet att politiska enheters fysiska karaktäristika har betydelse för politikens utformning är egentligen lika gammalt som statsvetenskapen. Den mest uppmärksammade frågeställningen i sammanhanget berör relationen mellan storlek och demokrati. Den klassiska synen, framförd av Platon (1960) och Aristoteles (1991) och senare av Rousseau (1900) och Montesquieu (1944) samt i modern tid av Arend Lijphart (1977: 65), gjorde gällande att små enheter utgör en bättre grogrund för demokrati än stora. Den främsta orsaken här till antogs vara att litenhet var förknippad med homogenitet, som i sin tur skapade interdependens och samarbetsincitament mellan medborgarna. Den motsatta synen företrädde i första hand av den amerikanske grundlagsfadern James Madison, som i *The Federalist Papers* betonade att små enheter innebär en risk för majoritetstyranni i den meningen att en bestående majoritet kunde dominera totalt över en bestående minoritet. I stora politiska enheter, som präglades av heterogenitet och diversifiering, var denna risk mindre eftersom olika politiska frågeställningar skapade olika majoritetskonstellationer (Hamilton, Madison & Jay: 1961).

Storlek har dock inte endast konsekvenser för i vilken mån stater demokratiseras eller ej. Det finns starka skäl att anta att storlek har effekter på flera områden inom politiken. Det banbrytande verket inom forskningsområdet har länge utgjorts av Robert Dahls och Edward Tuftes *Size and Democracy* (1973). I sagda bok lade författarna fram teoretiska argument för att påvisa effekter av storlek på en rad politikområden. Överlag kan konstateras att Dahls och Tuftes verk var rikt på teori, men magert på empiri. Därmed utgjorde det en idealisk grund för fortsatt forskning, och i ljuset därav är det förvånande att området länge förblev outforskat. Enstaka studier om storlekens effekter på politiken publicerades förvisso sporadiskt och det har numera även blivit vanligt att inkludera befolkningsmängd bland kontrollvariablerna i regressionsanalyser som har som syfte att förklara graden av demokratisering. Fortfarande, konstaterar Gerring och Veenendaal (2020: xx), gäller dock att befolkningsmängd "plays a supporting role, but hardly ever a starring role in the work of anthropologists, economists, political scientists, and sociologists".

Gerrings och Veenendaals verk kan ses som en sen, men välbehövlig uppdatering av Dahls och Tuftes *Size and Democracy*. Ambitionen är att kartlägga

befolkningsmängdens effekter på en rad olika politikområden. Detta sker separat i de olika kapitlen, som alla inleds med en teoretisk diskussion som belyser relationen mellan storlek och ifrågavarande fenomen för att därefter gå över till att empiriskt testa sambandet. Upplägget gör det lätt för läsaren att tillgodogöra sig innehållet i boken. De erhållna resultaten redovisas därefter i tabeller, som också redogör för resultat som uppnåts i tidigare studier av motsvarande frågeställning. I ett avslutande kapitel tillhandahåller författarna därtill i tabellform en sammanställning över samtliga resultat som uppnåts i studien.

För att testa variabelrelationerna tillämpas statistiska analyser. I regel används data som finns tillgängligt i globala databaser såsom *Varieties of Democracy*, *Quality of Government*, *Multilevel Elections Archive* och *Global Leadership Project*. För de minsta staternas vidkommande kompletteras dessa uppgifter med nära tvåhundra intervjuer av lokala experter och beslutsfattare som en av författarna (Veenendaal) utfört. Detta är viktigt att påpeka, eftersom globala databaser alltjämt tenderar att sakna uppgifter om de allra minsta länderna. Författarna tillämpar genomgående en flernivåapproach; vissa variabelrelationer undersöks genom att studera enheter inom länder, t.ex. delstater och kommuner, medan andra frågeställningar besvaras genom länderjämförelser.

Den första relationen som avhandlas berör relationen mellan storlek och det vidsträckta begreppet *cohesion* (social sammanhållning). I denna del består nästan hela det redovisade materialet av resultat från externa studier. Författarna noterar bl.a. att befolkningsmängd är positivt relaterad till flera (men långtifrån alla) mått på etnokulturell heterogenitet, att känslan av social samhörighet minskar med ökad storlek och att graden av kriminalitet tenderar att vara högre i stora enheter än i små.

Två kapitel tar upp frågan om storlek och representation. Först avhandlas frågan i vilken utsträckning parlaments och regeringars storlek kan föras tillbaka på befolkningsmängd. Frågan kan tyckas något trivial; föga förvånande finner författarna att storlek och parlamentsstorlek går hand i hand samt att storlek även förklarar regeringsstorlek, dock i mindre omfattning. Författarna noterar vidare att kontakten mellan väljare och valda är mer livlig i små enheter än i stora. Mer intressanta är de fynd som visar att etniska minoriteter är bättre representerade i mindre enheter än i stora och att förtroendet för politiker minskar i takt med att befolkningsmängden ökar.

Författarna diskuterar därefter huruvida storlek förmår fånga upp skillnader mellan partikularism och universalism. De noterar härvidlag bl.a. att storlek varierar positivt med graden av institutionalisering av partier men negativt med företeelser som röstköp och den relativa storleken på kampanjbidrag. Nästa fråga som behandlas gäller huruvida storlek påverkar politiskt deltagande och här är resultaten entydiga: i takt med att storleken ökar minskar deltagandet. Mönstret gäller på samtliga nivåer, från individ till stat, och oberoende av om deltagande mäts i form av valdeltagande, medlemskap i partier eller med andra mått.

Kapitlet om relationen mellan storlek och valkonkurrens är högtintressant. Grundantagandet är enkelt; partisystemet antas fragmenteras i takt med att storleken ökar. Den teoretiska diskussionen om den kausala mekanismen mellan storlek och partisystem är dock något ytlig. I synnerhet saknas en diskussion om storlekströsklar och valsysteem, som i Dahls och Tuftes ovannämnda verk behandlades utförligt. Dahl och Tuftes antog nämligen att storlek kunde antas inverka på partisystemets fragmentering uttryckligen i väldigt små enheter som tillämpade proportionella valsysteem. Alltnog, resultaten är i enlighet med antagandena; stora enheter har mer fragmenterade partisystem än små och detta gäller såväl mellan som inom länder. Särskilt intressant i sammanhanget är den jämförelse av lokalval i Sverige, Storbritannien och Brasilien, som tyder på att storlekens effekter på partisystemets utformning inte påverkas av valsysteem. Resultatet kunde med fördel ha diskuterats mer ingående.

I det följande kapitlet diskuteras ledarsuccession. Här konstaterar författarna bl.a. att politiska ledare sitter en kortare tid vid makten i stora stater än i små. Därtill konstateras att monarker är mer frekvent förekommande i små stater än i stora. Det efterföljande kapitlet behandlar graden av professionalism i statsapparaten. Här finner man bl.a. stöd för antagandet att meritokratisk rekrytering utmärker stora enheter snarare än små samt att tjänstemännen i stora enheter har en högre utbildningsnivå (men inte lön) än sina kolleger i små enheter.

Efter detta studeras relationen mellan storlek och makt. Grundtanken är att en ökad befolkningens mängd ger större ekonomisk, militär och kulturell makt. Resultaten är härvidlag otvetydiga; samtliga aspekter på makt samvarierar positivt med befolkningens mängd. Däremot finner författarna inte något entydigt stöd för att storlek hänger samman med konfliktintensitet; om ett samband mellan företeelserna överhuvudtaget existerar, är sambandet kurvlinjärt snarare än linjärt.

I ett avslutande kapitel behandlas storlekens betydelse för en rad övriga faktorer. Det är något förvånande att man i detta kapitel även behandlar den i historisk mening mest centrala frågan, nämligen om demokratin har en bättre grogrund i små enheter än i stora. Frågeställningen hade tveklöst förtjänat ett eget kapitel, men måhända är en av orsakerna till att så litet utrymme ges frågeställningen att temat behandlats mer ingående av en av författarna för ett par år sedan i en kvalitativt upplagd studie omfattande världens minsta stater (Corbett & Veenendaal 2018). Alltnog, Gerring och Veenendaal använder flera mått på demokrati och resultaten visar, i motsats till de flesta tidigare utförda studierna på området, att något samband mellan befolkningens mängd och demokrati ej går att finna. Däremot noterar de att östater har en högre grad av demokrati än fastlandsstater.

Gerrings och Veenendaals bok utgör ett efterlängtat och välbehövligt tillskott till litteraturen om fysiska faktorerens relevans för politikens utformning.

Ambitionen är att göra en sammanhållen bedömning av storlekens betydelse för en rad politikområden. Den skiljer sig från de flesta andra statsvetenskapliga studier genom att den använder sig av en enda oberoende variabel och ett stort antal beroende variabler. Såsom författarna noterar (s.16) är ett dylikt upplägg utmanande eftersom författare idag knappast kan förväntas besitta expertis på väldigt många av statskunskapens forskningsområden. Jag menar dock att slutresultatet är lyckat, låt vara att resonemangen om de kausala relationerna mellan storlek och de beroende variablerna ibland blir något ytliga.

I pedagogiskt avseende är boken exemplarisk i den meningen att den innehåller tabellariska summeringar av de forskningsinsatser som tidigare gjorts på området. Härvidlag redogörs tydligt för vilka undersökningsenheter och vilken forskningsdesign som använts. Detta gör det möjligt för läsaren att bilda sig en uppfattning om i vilken utsträckning de erhållna resultaten bekräftar eller vederlägger resultat som erhållits i andra sammanhang. Att detta är av särskild betydelse i en studie där vi diskuterar effekter av en och samma oberoende variabel på väldigt många beroende variabler säger sig självt.

Ytterligare är det värt att konstatera att bokens flernivåstruktur tillför mycket; då resultaten av studier i vilka man använt sig av olika systemnivåer jämförs, framträder storlekens effekter ofta mer tydligt än om man använt sig av analysenheter på samma nivå. Samtidigt kommer man inte ifrån att en (alltför) stor del av de inomstatliga studierna handlar om ett och samma land, nämligen USA. Onekligen hade resultatets tillförlitlighet ökat om resultat från andra länder redovisats i större omfattning.

Det bör dock påpekas att Gerrings och Veenendaals verk endast handlar om den ena centrala storleksdimensionen, nämligen befolkningmängden. Den andra storleksdimensionen, arealen, åsidosätts från såväl de teoretiska resonemangen som de empiriska analyserna. Detta är måhända förstäligt med tanke på att boken redan behandlar ett stort antal områden. Därtill är det svårt att isolera befolkningmängdens inverkan från arealens, eftersom bägge dimensionerna samvarierar mycket starkt. Samtidigt står det klart att utelämnandet av arealen lämnar många frågor öppna, inte minst med tanke på att det finns empiriska belegg som tyder på att arealen möjligtvis har en starkare koppling till graden av demokrati än befolkningmängden (Anckar 2008: 439-440.).

Alltnog, det är helt klart att boken utgör ett viktigt bidrag till den empiriska teoribildningen kring relationen mellan storlek och politik. Samtidigt måste man konstatera att den läsare som hade hoppats att boken skulle ge ett definitivt svar på frågan om litenhet gynnar eller missgynnar demokrati sannolikt blir besviken; efter att ha suttit i överläggning i 2500 år har juryn ännu inte levererat ett utslag.

REFERENSER

- Anckar, Carsten, 2008. "Size, Islandness, and Democracy", *International Political Science Review* 29, s. 433-459.
- Aristoteles, 1991. *Politiikka*, A.M. Anttila (övers.). Jyväskylä: Gaudeamus.
- Corbett, Jack & Veenendaal, Wouter, 2018. *Democracy in Small States: Persisting against all All Odds*. Oxford: Oxford University Press.
- Dahl, Robert & Tuft, Edward, 1973. *Size and Democracy*. Stanford: Stanford University Press.
- Hamilton, Alexander, Madison, James & Jay, John, 1961. *The Federalist Papers*. New York: New American Library.
- Lijphart, Arend, 1977. *Democracy in Plural Societies: A Comparative Exploration*. New Haven: Yale University Press.
- Montesquieu, 1944. *De l'esprit des lois. Texte établi avec une introduction des notes et des variantes par Gonzague Truc*. Paris: Flammarion.
- Platon, 1960. "The Laws", i Taylor, Alfred Edward (red.), *The Laws*. London: Dent.
- Rousseau, Jean-Jacques, 1900. *Du contrat social. Ou Principes du droit politique*. Paris: Flammarion.

Stubhaug, Arild, 2019. *Stein Rokkan – fra periferi til sentrum*. Bergen: Vigmostad & Bjørke.

Anmälan av Ulf Lindström

I repris av Max Webers tonårsfasoner, att skriva uppsats som läraren erkänner sig inkompetent att bedöma, skrev Stein som fjortonåring sin analys av John Stuart Mills *On Liberty*. Till skillnad från Berlin fanns dock ingen lokal professor att inkalla som censor. Man skriver inte uppsats om Mill när man är 14 år i Narvik 1936. Då blir man, i bästa fall, lämnad ensam på skolgården under rasterna "med nevene plantet i bukselommene," kanske som Ibsens "sterkeste mann i verden." Ifjol utgavs biografien om en av Nordens få samhällsvetare av internationell tyngd: Stein Rokkan (1921-1979).

Nationalstaten har fortsatt sitt register av ytterligheter. Den var den gemenskap i och för demokrati om vilken Europas februarirevolution 1848 utkämpades, en "kritisk vattendelare" i Rokkans arbeten. Danmarks fredligt avskaffade envælde, ersatt av en grundlag året därpå, inspirerade demokratins spridning i övriga Norden. Samtidigt havererade enandet av *Kleindeutschland*, som fick

följdverkningar nedåt i historien som Kejsarriket och ohyggligheter som Tredje riket och DDR.

Rokkans teorier om demokrati och partiväsendet har framväxten av den europeiska nationalstaten som underlag för empiriskt prövbar forskning. Rudolf Kjelléns arbeten, om staten som livsform och historiska uppgift, var Rokkan främmande. Herbert Tingsten, käck student vid den institution där Kjellén till 1922 odlade sin millenarism, liknade professorns värv vid nattgroggsdebatt. För Rokkan var statsbyggandets utmaning gränslös och, efter 1848, demokratins tid inne: att övervinna trösklar mot undersåtars medborgarskap och deltagande i alla typer av jurisdiktioner.

Rokkans liv och insats förkortades av dålig hälsa, synnerligen olägligt när statsvetenskapen och sociologin inbjöds att studera demokratins betingelser bortom Västeuropa, särskilt i Östeuropa efter 1989. Över-, multi- och undernationella processer, som direktval till EU:s parlament, etnisk mobilisering och tema som "regionernas Europa" och "europeisering av partier," har uppslag att hämta i hans ofullbordade projekt *Economy, Territory, Identity*.

Unge Steins läshunger i unga år kunde bara ha kurerats med den grymma rutinen att sända pojkar till college som tolvåringar. Det hade tvingat honom till annan sysselsättning också, kanske forellfiske, samt gett honom ro från en "irettesettende, arrogant og bedrevitende" far. Nu fanns i *malmbyen* och järnvägens nordliga utpost Narvik ingen akademi utöver stadsbibliotek och morbror Arntzens abonnemang på *Dagbladet*. Det gick därefter, två överhoppade årsklasser på gymnasiet och ändå högsta betyg i alla ämnen i studentexamen.

Arild Stubhaug, med matematiken som sin egen ämnesbakgrund, gör med biografien den insats som förutsätter distans till samtidens samhällsvetenskap. Stein Rokkan var ingen institutionell produkt. Han skulle själv värka fram nya institutioner som frukt av dyrt inhämtad kunskap. Flykt undan eget skrivande (ofärdig doktorsavhandling om Humes kontraktsteori) blev till den akademiska republikens gagn.

Hans med Seymour Martin Lipset redigerade antologi *Party Systems and Voter Alignments* (1967) är som enskild volym ännu oöverträffad i originalitet om partiväsendets egenskaper och väljarnas beteende. I den finns bland annat Rokkans iakttagelse att partier inte uppenbarar sig för väljarna som nya, *de novo*, vid varje val. Partier är övermättade aktörer, komposit av historia och, ja, sprickfärdigt arvsbelastade. Med reservation för pilsnerpartier föds alla partier än idag ur frön i sprickor, ur samhällets klyftor som inte kan överbyggas av och i partier.

Livskraften i Rokkans *nachlass* är också de funderingar som det väcker. För Rokkan var också en generös pedagog med ett fantasifullt bruk av kartor, diagram, pilar och pregnanta nyckeltermer. Han var inte avvisande ens till fyndiga one-liners. "Röster räknas men resurser avgör" (1964) var hans sätt att fånga korporativismen under dess storhetstid i Norden när LO, arbetsgivarföreningen,

bondeorganisationerna och andra särintressen ännu hade tyngd i departementens korridorer att få valresultatets förväntningar överprövade.

Vad, vid sidan om fransk filosofi i originalspråk, läste han som student vid Universitetet i Oslo, dit han och skolkompisen Terje kom 1939 för att överleva på andrasorteringens frukt och grönsaker? Kort svar: ännu mer, ofta och överallt fransk filosofi, men snart med inspridda texter i sociologi och statsfilosofi. De böcker som han nu "borde ha" upptäckt – om hans väg framåt gått rakt genom institutioner – var Tingstens *Political Behaviour* (1936) och Heberles *From Democracy to Nazism* (1945), kanske rent av André Siegfrieds, *Tableau politique de la France* (1913), lärdomsgenealogiskt startpunkten i genren.

I de, som senare i Sven Rydenfelts *Kommunismen i Sverige* (fast 1954 fanns inte datamaskinen), finns de tankefigurer som förutskickar politisk ekologi, masspolitikens uttryck för samspelet mellan de i sediment och kluster ruvande föreställningarna och den miljö de nu lever i. Datamaskinen, att "köra" istället för att handräkna de tester av teorier som kunde vittna om avfrostning av frön i det politiska landskapets sprickor, blev mirakelverktyg. Så exempelvis kunde utgången av Norges folkomröstning 1972 om medlemskap i EU skönjas redan i sambanden bakom resultatet i folkomröstningen om alkoholförbud 1919. Osagt om konsekvensen, men om teorin var Rokkan nöjd: än en gång hade periferin talat emot centrum.

Arne Næss, utnämnd till professor i filosofi i Oslo 1939, var navet i det sällskap av akademiska ynglingar som Rokkan tillhörde under ockupationsåren. Næss vidtalades 1948 att delta i ett UNESCO-projekt i Paris om demokrati som ideal och realitet. Rokkan, som i mellantiden hade inhämtat politisk idéhistoria, anslöt sig som assistent. Att bidra till demokratins återväxt var vällovligt, i en omgivning som dessutom erbjöd nya forskningsmetoder. Likt Næss vid vunen professur var Rokkan 28 år gammal när han levererade slutrapporten! Bland de som noterade det var Alva Myrdal. (Som "near-miss" må räknas utblivna möten med maken Gunnar och unge Steins sårbarhet inför en docering om förtjänster i Rudolf Kjelléns geopolitik.)

Åren 1949–1955 kryssar han fram i en karg skärgård, fattig i institutionell trygghet. Norge var underetablerat av lärosäten. Universitetet, som Oslos länge skulle omtalas i singularis, var värdskap för lärostolar. Fördelen var att ingen statsvetenskap var barrikaderad bakom unken statsrätt. Forskning som kollektiv verksamhet vid fristående institut fick nu sin *Gründerzeit*, "tillsatt endast pengar." För samhällsvetenskapen blev träförädlingens mecenat familjen Rinde lyckan. Institut för Samfunnsforskning, ISF, blev mötesplats också för Rokkan. Därifrån kunde han söka nya vikariat, nya stipendier också utomlands, fortsätta att bygga nätverk. "For meg personlig kom denne stilen til å virke frigjørende, ja forløsende." Anslag från tidens ymnighetshorn, amerikanska stiftelser, finansierade bland annat intervjuundersökning om skollärares roll i demokratisk fostran.

Amerikansk samhällsvetenskap, med exiltysk accent och urval av intervju-personer som metod, blev *de rigueur* efter 1945. Väljarundersökningarna blev statsvetenskapens mediala signum. Den första amerikanska 1948 skulle få sina nordiska uppföljare, den norska motsvarigheten vid valet till Stortinget 1957. Då bytte sju procent parti sedan föregående val, idag är partitytarna 35 procent. Därifrån härstammar en ikonisk tabell i valforskningen. Kollega Henry Valen, i samförstånd med Rokkan om väljarnas personliga arv, ”kontrollerade” partival med fars - dock ej mors - partisympati.

När resultaten från väljarundersökningen 1957 fanns som manuskript i (aldrig utgivna) samlingsverket *Valg i Norge* var Rokkan igen rastlös. Att lyfta forskningen in längs det kontextuella spåret, den politiska ekologi som ska avsätta boken tillsammans med Seymour Martin Lipset, hastade.

För biografiska syften har dedikationen i den boken övertolkats, ”To our fathers, defenders of the periphery.” Lipset, ett år yngre än Stein, var son till rysk-judisk flyktning undan tsaren. Som yngling i New York City var ”Marty” vänsteraktiv, trotskist. Rokkans familj är en egen historia. Hans far, Georg, avsatt redaktör i Arbeiderpartiets *Fremover*, skollärare, brandman, esperantopionjär, åkeriägare, dömdes efter kriget för landsförräderi till två månaders fängelse. Han hade haft oregelbundna uppdrag som tolk åt ockupationsmakten och för den utfört logistiska tjänster med åkerifirman. Åtminstone för ett år hade han betalat medlemsavgiften i Quislings Nasjonal Samling. Det lockar gärna till övertolkning. Stubhaug håller emot, så gott det går.

Det var 93.000 som åtalades, 46.000 fälldes i vad som blev Europas mest svepande rättsuppgörelse. Deras avkommor provocerades inte i proportionell grad att genom just akademisk förtjänst återvinna namnets respekt. Landsförräderi är traumatiskt – Stein bad om och fick sin brittiska frus löfte om tystnad om krigsminnen i familjekretsen, själv höll han kontakten med fadern pro forma – men är inte ensamt differentia specifika för familjemedlemmars livslopp.

Rokkan var en av flera samhällsvetare som växte upp i periferin när det fortfarande var långt till centrum, för honom ett dygn till Oslo med tåg via Boden och Stockholm. Hans egenhet blev alltså inte den lärde lokalpatriotens. Hans far hade mer än väl förbrukat familjens kvot med kvädet ”statens krossande knytnäve” mot Nord-Norges väl, en bok utgiven 1932 på eget förlag. Att jämföra förhållandet, likheter och skillnader, mellan centra och utkanter i alla statsformationer var mer angeläget. För nya självständiga stater i Afrika nämnde Rokkan omgående språken som konfliktladdad tröskel i nationsbyggandet.

Det var i Bergen han etablerade sin bas, 1958 som forskningsledare vid ett fristående institut (CMI), från 1966 som professor vid stadens universitet. Där var hans engagemang för etableringen av NSD (Norsk Senter for forskningsdata) osvikligt. Kommundata, tidiga opinionsundersökningar, officiell statistik, ”rådata,” skulle arkiveras i användarvänligt format. Ännu var tron på kumulativitet, att träget flytta forskningsfronten, i samhällsvetenskapen stark.

Som omedelbar avkastning åtnjöt forskare och studenter i Bergen trafiken av gästföreläsare av "gamla skolans" yppersta rang: Erik Allardt, Reinhard Bendix, Robert Dahl, Martin Lipset, Shmuel Eisenstadt, Juan Linz. Inramningen av gästföreläsningen var en annan förr, när ett för gästen protokollärt program påbjöds innan denne dagen därpå mötte fullsatt andäktigt auditorium. (Föreläsarna under Rokkans värdskap minns kanske annat, som den disträ trakteringen. Immanuel Wallerstein bjöds efter sin globala utsvävning på *raspeballe*, jfr pitepalt, på närbelägna café Ugla, galonets och bruna perstorpsplattans tempel.) Idag räcker ett streck med rödpenna i universitetets budgetutkast och traditionen läggs som webinar på YouTube, den interaktiva taffeln till Starbucks. Minnesföreläsningar, som de årliga till Stein Rokkan själv, är få universitet förunnade.

Rokkans umgänge med sina studenter var otvunget, uppmuntrande och krävande. Tidsfrister sattes, studenters ursäkter om andra åligganden möttes av "då jag var i din ålder..." Om han där avbröt sig i insikt om den jämförande metodens fälla, "conceptual stretching," är okänt. Idag kan studenter antagna vid lärosäten i världsklass appellera inte bara till lärares medlidande utan också till inskränkningar av lärares *tankar* inför studenters – känslor! Om det är Rokkan tidlös, här i krönika i *Arbeiderbladet* 1947 [sic!]: "...vi trenger en vitenskap, ikke et konglomerat av vage kjensler, av uklare forestillinger, av tåketete spekulasjoner og av ordtak fulle av uutrunnelig visdom."

Från just universitet i världsklass fick Rokkan inom loppet av fyra veckor våren 1968 erbjudanden om professurer. Stanford i det soliga Kalifornien, därefter Harvard och Yale, erbjöd svindlande lön och förmåner! Endast läsare utan förlamande, fast med nykter förståelse för, kärlek till Bergen kan undgå att förvånas av hans lojalitet med hemlandet. Här är Stubhaug sparsam med sin utredning om "Rokkans Val". Hans patriotism erkändes i media först efter att han 1979 utnämnts till kommendör av St. Olavs Orden. Vid det laget hade han invalts i sex utländska vetenskapsakademier!

Stubhaug loggför annars tålmodigt sitt subjekt; de möten Rokkan deltog i, de ansökningar han skrev i arbetet med att bygga institutioner för bärkraftig forskning. Mellan 1970 och 1977 åtog han sig presidentskapen för UNESCOs samhällsvetenskapliga råd, ISSC, för statsvetenskapens och sociologins internationella organisationer (IPSA, vice-president ISA), arbetet att organisera den europeiska och nordiska statsvetenskapens samarbete. Redaktörskap för antologier, introduktör i tidskrifters temanummer, han tackade alltför sällan nej. Kort innan hans frånfälle hade han utkast färdigt till forskningsprogram om u-länder och i-länder och hur de, nytänkande, *ömsesidigt* kunde utväxla erfarenheter av stats- och nationsbyggande. Och ännu var det länge till mobiltelefon, e-post och internet.

Forskningspolitiskt, i en epok av storpolitik, är Rokkan teflonmänniska. Hans arbete fortskred utan störning av politikens dagordning, som

Vietnamkriget och EU-politik. En sällsynt förtvivlan finner dock vägen till hans penna, om fragmenteringen av sociologin, 1974: "the discipline wanted to pose as the queen, it ended up as the clown of the sciences of man." Ändå, betänkt vad som skulle komma, finns i Stubhaugs bok inte en enda referens till att Rokkan som president/ordförare tvingades mäkla i konflikter som idag har upphov i intersektionalismens måttligt komiska minfält!

Rokkan kunde aldrig partimärkas. Han röstade inte ens. Om hans sympatier och antipatier, höger som vänster, finns magra bevis. Hans ungdomliga uppskattning av Karl Poppers *Open Society and its Enemies* blev livslång, att döma av en senare schism som fick Rokkan att som fakultetsopponent röja sitt credo. Det jämförelsevis starka gensvaret som marxism-leninismen fick bland studenterna vid landets då fyra universitet hade marginell betydelse för den rokkanska statsvetenskapen. Weber, Marx, Durkheim, Parsons, det fanns för honom som för andra något hos alla, men inte allt hos en.

Stubhaugs biografi har sin styrka i dokumentationen och enstaka borrhande utredningar. Hans återhållsamhet i tolkningarna av Rokkans fyra egna vägval är berömvärd och timid på samma gång. Före "nej" till Stanford undvek Rokkan upprepade gånger att fångas i Alva Myrdals talangjakt för produktion av framtida ledarskap i FN-apparaten. Det tredje valet, sökandet efter insikt i de långa linjerna och djupa avlagringarna bakom politiskt handlande, innebar att han valde bort aktörsnära teorier som lånade från ekonomivetenskapen, "rational" och senare "public choice." Det viktigaste valet bestod han inte, till det var han för snäll: att bara tacka "nej" när kolleger igen sände meddelandet "We need a leadership."

Rokkans produktion av färdiga, väl definierade och avgränsade, artiklar och monografier var till antalet få. "Publikationer räknas men originalitet avgör." Ännu kastar de långa men bleknande institutionella skuggor in i politikforskningens tid, när esoteriska blomstrar medan nationalstaten återigen tar sig allt större och bullrigare utrymme.

Ritchie, Stuart, 2020. *Science Fictions: Exposing Fraud, Bias, Negligence and Hype in Science*. London: The Bodley Head.

Anmälan av Anders Sundell

Sjuåringen tyckte att det var tråkigt att göra läxor, så jag satte ihop ett litet belöningssystem. Varje dag hon gör en läxa får hon ett klistermärke att sätta upp i en kalender. Det fungerade: nu gör hon dem gärna. Men bara en åt gången, mer vore slöseri. Bättre att sprida ut dem över dagarna för att maximera antalet klistermärken.

Incitament har effekt, på små såväl som stora. Det blir uppenbart när man läser Stuart Richies *Science Fictions: Exposing Fraud, Bias, Negligence and Hype in Science*, en genomgång av hur de olika problem som räknas upp i titeln fördärvar vetenskapen.

Den bakomliggande orsaken är enligt Ritchie vetenskapssamhällets skeva incitament. Vetenskapen har professionaliserats och expanderat, med en ökande produktion av både forskare och vetenskapliga publikationer. Konkurrensen om tjänster och medel är hård, vilket ökar trycket på att samla karriärpoäng. Målet flyttas därmed, från sökande efter kunskap till sökande efter publikationer. Man kan säga att det viktiga inte är att resultatet är vetenskapligt, utan att det är *vetenskapslikt*.

Konsekvenserna är lika förutsägbara som tragiska. Den vetenskapliga revolutionen och den efterföljande tekniska utvecklingen är orsaken bakom det enorma uppsving i levnadsstandard mänskligheten upplevt de senaste århundradena. Förutsättningslöst kunskapssökande, hypoteser grundade i konsekventa och ofta matematiska teorier, prövade i väldokumenterade experiment, ledde till en ständigt växande kunskapsbank.

Nuförtiden växer dock kunskapen inte i samma takt som antalet publikationer. Studier från en mängd olika vetenskapsfält visar att en stor andel "forskningsresultat" inte visar sig hålla för närmare granskning, och inte kan replikeras av andra. Denna replikationskris har slagit hårdast mot psykologin (Ritchies eget ämne), men naturvetenskapen är inte undantagen – en uppmärksammas genomgång visade att 47 av 53 viktiga resultat inom cancerforskning inte gick att återskapa (Begley och Ellis 2012). Att statsvetenskapen än så länge inte drabbats i samma utsträckning beror antagligen mindre på att våra studier skulle hålla högre kvalitet än det faktum att de är svårare att replikera, eftersom de oftast bygger på observationsdata. Och även om boken i princip bara tar upp kvantitativ forskning tror jag inte heller att kvalitativa forskare skulle vara

mindre benägna att vinkla sina tolkningar för att passa in i ett mer övertygande narrativ. Det är bara svårare att kontrollera.

Ritchie inleder boken med att prata om fusk, och använder pinsamt nog för oss svenskar Macchiariniskandalen som främsta exempel. Och som statsvetare blir det extra genant att ett annat exempel är Lacour (och Green)-skandalen. Efter att en fältstudie misslyckats förfalskade statsvetardoktoranden Michael Lacour data, och fick tillsammans med Donald Green en artikel publicerad i *Science*. En enorm mediauppmärksamhet och ett jobb på Princeton följde. Bedrägeriet avslöjades av två andra doktorander, David Broockman och Joshua Kalla. När de själva senare faktiskt utförde en liknande studie (Broockman och Kalla 2016) visade det sig, ironiskt nog, att resultaten i princip var desamma som rapporterats av Lacour och Green!

En stor förtjänst med boken är emellertid att den inte fastnar i de här exemplen, hur sensationella de än må vara. Medvetet och utstuderat fusk är allvarligt, men inte det största problemet vetenskapen står inför. Istället är det antagligen mer vardagligt fusk och slarv som ligger bakom de flesta studier med opålitliga resultatet. I många fall handlar det troligen om att forskare lurar sig själva.

Drivkraften är även här att få fram publikationer, och eftersom tidskrifter tyvärr är mer benägna att publicera samband än ickesamband leder det till en jakt på statistisk signifikans. Så kallad "p-hacking" innebär att man mixtrar med de oräkneliga parametrar som finns i alla statistiska undersökningar – operationalisering av variabler, val av kontrollvariabler, estimeringsstrategi, hantering av outliers, interaktioner, med mera – för att komma på rätt sida av den magiska gränsen $p=0,05$. En kartläggning av p-värden i prestigetidskrifterna *American Political Science Review* och *American Journal of Political Science* visade till exempel att det är osannolikt få p-värden precis över gränsen, och osannolikt många precis under (Gerber och Malhotra 2008). Ett statistiskt mysterium, men inte alls oväntat för den som vet någonting om forsknings- och publiceringsprocessen. Mönstret är troligen resultatet av flera faktorer: forskarnas val i analysen, huruvida man färdigställer en studie eller inte, och tidskriftens beslut att publicera eller inte när ett manuskript väl skickats in.

Fenomenet p-hacking illustrerar väl målförskjutningen i modern vetenskap. $0,05$ är inte en naturligt magisk gräns. Sannolikheten att ett statistiskt samband är orsakat av slump i urvalet är i princip lika stor när p är $0,049$ som när p är $0,051$. Den stora skillnaden är att man i det förra fallet får sätta ut en stjärna vid koefficienten, och rätten att beteckna sambandet som *signifikant*. Värdet är symboliskt, inte substantiellt.

Till skillnad från andra böcker på liknande tema (till exempel *Return to Meaning – A Social Science with Something to Say* av Mats Alvesson, Yiannis Gabriel och Roland Paulsen 2017) är *Science Fictions* dock aldrig svepande i sin kritik. Ritchie utgår genomgående från empiriska studier, ofta metastudier,

som belägger problemen. Diskussionerna är också nyanserade, och visar på allvaret utan att slå över i sensationalism.

Man kan inte heller misstänka Ritchies för att klanka ner på ett system som behandlat honom illa. Sedan disputationen 2013 har han samlat ihop över 5000 citeringar på Google Scholar. Författaren klarar sig uppenbarligen bra i den dysfunktionella vetenskapsvärld som beskrivs. Boken ger dessutom inte uttryck för någon nostalgi över "den gamla goda tiden". Lösningen är inte att gå tillbaka till ett gentlemannaforskarideal, där kvalitetsgaranten är forskarens fina ideal och genuina kunskapssträvan. Medan Alvesson, Gabriel och Paulsen (2017) vill lösgöra sig från världsligheter som källhänvisningar och uppvärdera intellektualism vill Ritchie snarast gå åt andra hållet, mot större systematik, transparens och kontroll. Ett karriärsystem utan några mätbara kriterier för framgång skulle lämna stort utrymme för godtycke och nepotism.

Ett utmärkt exempel på vad som kan göras för att höja kvaliteten är krav på förregistrering av studier och tillhandahållande av material. Ritchie visar till exempel hur innehållet förändrades i en medicinsk tidskrift efter att förregistrering infördes. Resultaten i publicerade studier blev mindre sensationella, med många fler nollresultat, men ger antagligen en mycket mer representativ bild av forskningsfältet.

Men ökade krav på forskare bör också åtföljas av motsvarande skärpning från tidskriftshåll. En väg framåt kan vara modellen där artiklar antas på grundval av frågeställning, teori och metod. Resultaten avgörs i efterhand, och publiceras oavsett utfall.

Och även om replikationskrisen gett nödvändig krisinsikt ska vi inte tro att allt som publiceras är falskt. En nytt och ambitiöst replikationsförsök visade att resultat i studier med god metod – exempelvis hög statistisk styrka, förregistrering och stor transparens – också kunde återskapas i förväntad utsträckning (Protzko et al 2020).

Lösningen, tror jag, är att ta vetenskapen på större allvar, inte mindre. Vardagsfusk och slarv är mer lockande när man inte tror att någon kommer att använda ens forskningsresultat (om de ens blir lästa). När så är fallet blir vetenskapen till en hobby, eller ett spel, där det viktigaste är att vinna inom spelets ramar – få signifikanta resultat, fina publikationer – snarare än att närma sig en riktig sanning.

Vetenskapen ska i teorin vara mer rigorös och krävande än det allmänna tyckandet. I en krönika kan man skriva vad man tror, medan man i en vetenskaplig text behöver bevis. Men tanken är givetvis att även forskaren ska tro på sina teser – och jag är inte säker på att så alltid är fallet. Risken finns att det formella systemet med peer review slår ut det egna omdömet, att man intalar sig att resultat är rimliga bara de klarar hinderbanan som vetenskapssamhället satt upp. Problemet är att tidskrifternas granskare bara läser den färdiga texten. De ser inte hela processen, och kan därför inte åläggas hela ansvaret.

Om man tänker sig att forskningens slutmål inte är publikation, utan att kunskapen ska komma ut och användas i samhället, borde det bli mindre lockande att skicka in resultat som man själv tvivlar på. Att tvingas registrera sina studier på förhand är ett steg i rätt riktning, eftersom det underlättar kollegial granskning. Men en starkare koppling mellan forskning och undervisning och tydligare krav på att kommunicera resultat till allmänheten skulle också hjälpa. Forskare tvingas då att tydligare stå för sina resultat, inför kollegor, studenter och allmänheten. Gemene man kanske inte kan skilja på en Regression Discontinuity Design och en Time Series Cross Section, men har å andra sidan mindre tolerans för akademiskt navelskåderi. Om det nu är sanningen vi gör anspråk på att avtäcka i forskningen borde den rimligen hålla för kritisk granskning från alla håll.

För den som har följt debatten om replikationskrisen någorlunda erbjuder Stuart Ritchies bok egentligen inget nytt, men är ändå en gedigen och nyanserad översikt från en som vill vetenskapen väl. Den borde vara obligatorisk läsning på alla forskarutbildningar. Vetenskapen är för viktig för att reduceras till en jakt på klistermärken.

REFERENSER

- Alvesson, Mats, Yiannis Gabriel & Roland Paulsen, 2017. *Return to Meaning. A Social Science with Something to Say*. Oxford: Oxford University Press.
- Begley, Glenn & Lee Ellis, 2012. "Raise standards for preclinical cancer research", *Nature*, 483: 531-533.
- Broockman, David & Joshua Kalla, 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing", *Science*, 352: 220-224.
- Gerber, Alan & Neil Malhotra, 2008. "Do statistical reporting standards affect what is published? Publication bias in two leading political science journals", *Quarterly Journal of Political Science*, 3: 313-326.
- Protzko, John, Jon Krosnick, Leif D. Nelson, Brian A. Nosek, Jordan Axt, Matthew Berent, Nick Buttrick, Matthew DeBell, Charles Ebersole, Sebastian Lundmark, Bo MacInnis, Michael O'Donnell, Hannah Perfecto, James Pustejovsky, Scott Roeder, Jan Walleczek, Jonathan Schooler, 2020. "High Replicability of Newly-Discovered Social-behavioral Findings is Achievable".

Prenumerera på Statsvetenskaplig tidskrift!
www.statsvetenskapligtidskrift.org