

Utvärdering – en balansakt

Rolf Sandahl

Inledning

Utvärdering i Sverige

När utvärdering startade i Sverige kan inte exakt anges. I artikeln *Utvärdering i det svenska politiska systemet* (Vedung, Furubo och Sandahl 2000) ges tre alternativa berättelser. Berättelse ett påstår att utvärdering uppstod i slutet av 1960-talet utifrån uppfattningen att politik och offentlig förvaltning kunde bedrivas mer rationalistiskt och vetenskapligt. Användandet av samhällsvetenskapliga metoder skulle göra styrningen välgrundad, förnuftig och effektiv. Idéerna var hämtade från USA där president Lyndon B Johnson hade introducerat ett antal välfärdsreformer sammanfattade i begreppet Great Society. Utvärdering av dessa reformer var ett obligatorium.

I Sverige manifesterades dessa idéer i en offentlig utredning, Programbudgetutredningen (SOU 1967:11). I utredningen introducerades begrepp som produktivitet, effektivitet, prestationer, effekter, måluppfyllelse etc. begrepp som fortsättningsvis inte bara var avsett för företagsvärlden utan även för den offentliga sektorn. I en försöksverksamhet på ett 30-tal myndigheter skulle, vid sidan av resursförbrukning, även myndigheternas resultat redovisas (Sandahl 1999). Försöket kan betraktas som en tidig introduktion av det som senare kom att kallas New Public Management.

Berättelse två hävdar att utvärdering i Sverige startade ungefär 15 år tidigare inom utbildningssektorn, närmare bestämt i samband med de stora skolreformerna. Grundskolereformen 1950 startade med fältförsök i vissa skoldistrikt, medan det gamla skolsystemet bibehölls i andra distrikt för att möjliggöra kontrollgrupper. Utvärderingarna utfördes av de pedagogiska institutionerna vid Stockholms och Göteborgs universitet. Termen utvärdering användes dock inte i dessa undersökningar, utan man talade om studier, utprovning eller praktisk övning. Syftet var att ge saklig grund för succesiva revisioner av skolans allmänna läroplaner. Skolområdet var, enligt berättelse två, vid denna tidpunkt det enda område där en mer systematisk utvärderingsansats tillämpades.

Berättelse tre knyts till Sveriges unika och tidiga utredningssystem, Sveriges Offentliga Utredningar (SOU). Inom ramen för dessa utredningar fanns naturligtvis en stor kunskap om hur exempelvis statliga insatser skulle kunna bedömas. Frågor om kausalitet hade ju diskuterats länge exempelvis bland filosofer. I en kommittérapport från 1934 (SOU 1934:1, s. 8) skrev Gunnar Myrdal följande:

En utredning rörande verkningarna av en viss finansiell reglering måste nämligen gälla *skillnaden* mellan det ekonomiska händelseförlopp, som utlöser sig vid denna reglering, och det förlopp, som utlöser sig vid en annan reglering. Denna andra reglering (jämförelsenormen) måste därför ävenledes bestämmas. I en annan mening än såsom en skillnad mellan två förlopp vid alternativa finansiella regleringar är allt tal om verkningar utan bestämd innebörd

Vilken av berättelserna som bäst beskriver när utvärdering startade i Sverige vet vi inte, berättelserna beskriver snarare olika skeenden i utvecklingen, från utvärdering i enskilda

utvärderingar (berättelse tre) till mer systematiska utvärderingsupplägg inom ett enskilt sakområde (berättelse två) och slutligen till att omfatta hela den offentliga sektorn (berättelse ett).

Utvärdering

Likaväl som det finns olika berättelser när utvärdering tog fart i Sverige, finns det olika definitioner av utvärdering. Populära begrepp, dit utvärdering hör, får gärna karaktären av ”semantiska magneter”, det vill säga många typer av aktiviteter rubriceras som utvärdering (Weiss 1998, s. 3). Risken med en alltför vid definition kan jämföras med det Aron Wildavskys (Wildavsky 1973, s.127) gett uttryck för när det gäller planering, det vill säga ”*if planning is everything, maybe it’s nothing*”, också skulle kunna appliceras på begreppet utvärdering.

Även om det finns många definitioner av utvärdering, är det emellertid några grundelement som de flesta utvärderare ändå är överens om bör ingå i en definition av utvärdering. Låt oss ta statsvetaren Evert Vedungs definition av utvärdering som utgångspunkt (Vedung 2009, s. 22):

Noggrann efterhandsbedömning i efterhand av utfall, slutprestationer, förvaltning och beslutsinnehåll samt organisering av offentlig verksamhet, vilken tänkes spela en roll i praktiska beslutssituationer.

Utgångspunkten i denna definition är offentlig förvaltning, men utvärdering är naturligtvis inte reserverat för offentlig verksamhet. Alla är inte heller överens om att utvärdering är en efterhandsbedömning, det vill säga *ex post*. Lika ofta förekommer begreppet *ex ante* utvärdering. Ur definitionen ovan är det emellertid tre specifika faktorer eller nyckelord som kännetecknar och som anses viktiga i en utvärdering, nämligen *systematisk* eller noggrann insamling av data, *värdering* eller bedömning *och användning*.¹

Noggrannhet

Att noggrannhet anses viktigt inom utvärderingsområdet kan vi utläsa av titeln på en av de mer tongivande böckerna inom utvärderingsområdet, nämligen ”*Evaluation*” med undertiteln ”*A Systematic Approach*” (Rossi m. fl. 2004). Huvudbudskapet är att en utvärdering ska vara underbyggd av empirisk kunskap och analyserad enligt de metodmässiga regler som gäller, dvs. vara rättvisande (tillförlitlighet, relevans etc.) och transparent (visa hur underlaget är insamlat, urval, referenser etc.). Prefixet ”ut” i utvärdering är en förstärkning av begreppet precis som i orden uttömmande, utdöd, utbränd etc. ”Ut” betyder fullt ut eller just noggrann (Vedung 1998, s. 23).

Men *hur* noggrann ska en utvärdering eller hur noggrann *kan* den vara. Kring begreppet utvärdering har noggrannhet eller systematik ofta satts lika med eller anknutits till forskningsmetoder. ”... *program evaluation is the application of systematic research methods to the assessment of program design, implementation, and effectiveness*” uttalade sig Eleanor Chelimsky (Chelimsky 1985, s.7) redan i mitten av 1980-talet, då ansvarig för metodavdelningen på General Accounting Office (GAO) i USA. Carol Weiss, även hon en av pionjärerna på

¹ I boken *Evaluation Roots* (Alkin ed. 2013) symboliseras utvärderingsområdet med ett träd där trädets tre grenar benämns Use, Methods och Valuing.

utvärderingsområdet, kallade en av sina första böcker (Weiss 1972, s.6) för ”*Evaluation research*”. Hon menade att alla de metoder som användes inom forskning som upplägg, mätmetoder och analyser också gällde för utvärdering. Det som skiljde utvärdering från forskning var inte metoderna eller objekten i sig utan syftet. Utvärderingar är till för att användas till skillnad mot t.ex. grundforskning vars främsta syfte är generera kunskap, där forskaren dessutom ”äger” problemställningarna vilket inte alltid är fallet när det gäller utvärdering (Weiss 1998, s. 15 ff).

Just syftet med utvärdering gör, enligt en annan av de tidigare utvärderingsforskarna, Lee Cronbach, att utvärderingar inte alltid kan hålla den absolut högsta vetenskapliga standarden i praktiska situationer. Det kanske inte ens är lika viktigt eller önskvärt. Utvärdering skiljer sig från vetenskaplig forskning, enligt Cronbach (Rossi 2004, s. 20 och 23) eftersom ”... *evaluation is more art than science and should be oriented toward meeting needs of program decisionmakers and stakeholders.*” Forskning syftar således till att nå forskningsstandard medan utvärdering syftar till att uppnå användbar information och ”... *to inform action.* ”

Den så kallade ”evidensdebatten” under senare år kan i viss mån ses som en diskussion om noggrannhet och vetenskaplig standard. Debatten har till stor del handlat om huruvida det finns ett bästa upplägg (design) för effektmätning som i alla sammanhang ger mer tillförlitlig och användbar kunskap än andra metoder (Bohlin & Sager 2011, s. 13). Mer precist har debatten handlat om huruvida det så kallade randomiserade experimentella upplägget (Randomized Control Trials, RCT) förtjänar en sådan status. Inom det medicinska området har det otvivelaktigt en sådan status. Vissa debattörer har menat att detta upplägg är bättre än andra även för det samhällsvetenskapliga området, medan andra har framhållit att valet av upplägg måste bero på vilka frågor som ska besvaras.

Värdering

Värdefrågor är nog den del av utvärderingsbegreppet som berörts minst. Det kan tyckas underligt med tanke på att värdering är en del av begreppet utvärdering och att utvärdering, enligt en av de mer tongivande personerna inom utvärderingsområdet, Michael Scriven, bör ses ”...as the science of valuing” (Shadish m. fl. 1991, s. 74). En delförklaring, enligt Ernest House (House 2001, s. 313), är att värderingsfrågan upplevs som komplex: “*The fact – value dichotomy is a particularly difficult problem for evaluation since values lie at the heart of evaluation. I doubt that anything in the field has caused more trouble than this belief.*”

Ernest House menar att fakta och värden i många fall inte är två ytterligheter på en skala utan ofta förekommer samtidigt. Det som kan uppfattas som rena fakta är inte alltid värdeneutralt. Att ta upp vissa fakta men utesluta andra är i högsta grad en värdering. Att säga, enligt House, att George Washington var USA:s första president är rena fakta. Men att säga att George Washington var USA:s första president men också att han var man, vit och hade slavar på sin gård är visserligen också rena fakta, men görs utifrån syftet att appellera till vissa värderingar.

Att utvärderare använder eller värderar verktyg olika beroende på t.ex. disciplinär bakgrund är välkänt och skulle kunna illustreras med följande logiska lek (Smith 1994, s. 224):

“Tell me who the evaluator is and I will tell you what the design will be. Tell me the design and I will predict the results, and therefore: Tell me the evaluator and I predict the results!”

Skiljelinjen i diskussionen går i princip mellan dem som anser att utvärderaren enbart ska utgå från exempelvis intressenternas, brukarnas, politikernas, experternas värden, *deskriptiv värdering*, och de som anser att man bör utgå utvärderarens egna värderingar, *preskriptiv värdering*. Men även bland dem som förespråkar den deskriptiva utgångspunkten menar att det samtidigt inte finns något hinder emot att utvärderaren utöver att bedöma om en viss målsättning är uppfylld (enligt någons önskemål) samtidigt kan säga att detta också lett till andra kanske negativa följder eller bieffekter för vissa grupper.

Ove Karlsson Vestman menar, med hänvisning till statsvetaren Bo Badersten (Karlsson Vestman 2011, s. 51 f), att kunskap aldrig kan vara värderingsfri, vilket innebär att distinktionen mellan ”bör” och ”är” torde falla,” men att utvärderarens uppgift ändå är att förhålla sig till är och bör.

Diskussionen om värderingar i utvärderingssammanhang har således handlat om *vems* värderingar som ska gälla men också mot vilka standards eller måttstockar som resultatet ska bedömas emot, det vill säga var ska ribban läggas för att nivån ska anses som godtagbar, icke godtagbar, utmärkt eller liknande.

Offentligt publicerade utvärderingar är ett viktigt underlag för att möjliggöra en debatt och ett nödvändigt inslag i en demokrati. Men målsättningen med utvärderingar kan föras ytterligare ett steg. Ernest House & Kenneth Howe (House & Howe 1999), som förespråkar en deltagande (deliberativ) demokrati, menar att det i en utvärdering inte bara är viktigt att så många som möjligt, det vill säga intressenter, brukare med flera kommer till tals utan särskilt de grupper som i andra sammanhang har svårt att få sina röster hörda.

Användning

“Is Anybody There? Does Anybody Care?” Frågar sig Carol Weiss redan i artikel från slutet av 1980-talet (Weiss 1988, s. 5 ff). Frågan ställdes utifrån insikten om att beslutsfattare eller mottagare inte alltid reagerade utifrån den rationalistiska modell som förväntades av dem. Insikten var helt enkelt den att politiska beslut har sin egen eller en annan rationalitet och mer komplex än att bara utgå ifrån ett utvärderingsunderlag. Dels kanske utvärderingsunderlaget bara var *ett* av faktaunderlagen, dels att politiska beslut i grunden är normativa ställningstaganden. Politikernas syn på vad som är viktigt kan helt skilja sig ifrån utvärderarens.

En utvärdering kan dock, enligt Weiss, ha en annan användning är den synligt *instrumentella* användningen, som utvärderaren gärna ser, utan en *upplysande* eller *konceptuell* användning. Hur denna upplysning kommer politikerna till del behöver inte vara genom att själv läsa utvärderingen, med tanke på all den information som passerar politikernas bord, utan via

massmedia, personal, intressegrupper etc. Samtidigt kan vi konstatera att den som exempelvis läser den årliga budgetpropositionen inte behöver läsa länge för att se att regeringen refererar till olika utvärderingar eller ofta uppföljningar i sin argumentation för att påvisa ett gott resultat i syfte att äska medel hos riskdagen för fortsatta satsningar.

Frågan om användning har diskuterats mycket inom utvärderingsområdet under senare år. Förklaringen är helt enkelt att användningsfrågan är central i utvärdering. Få eller ingen gör en utvärdering av ren nyfikenhet utan en utvärdering är nästan undantagslöst en beställning från en uppdragsgivare med bestämda syften. I en utvärdering finns därför alltid en användningsaspekt, oavsett om eller på vilket sätt utvärderingsresultatet kommer att användas.

Den utvärderingsforskare som kanske starkast påverkat debatten inom användningsområdet är Michael Quinn Patton genom sin bok *”Utilization-focused evaluation”* (Patton 1997, s. 20):

Utilized-Focused Evaluation begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done, from beginning to end, will effect use. Nor is use an abstraction. Use concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the focus is on intended use by intended users.

Vem ska då bestämma vad som ska utvärderas och hur det ska göras. För Michael Quinn Patton är det helt enkelt den som utvärderingen är avsedd för (Patton 2000, s. 427):

Many decisions must be made in any evaluation. The purpose of the evaluation must be determined. Concrete evaluative criteria for judging program success will usually have to be established. Methods will have to be selected and timelessly agreed on. All of these are important issues in any evaluation. The question is: Who will decide these issues? The utilization-focused answer is: primary intended users of the evaluation.

Utvärdering kan sägas vara såväl en produkt, exempelvis ett skrivet dokument, som en process, det vill säga själva framtagandet av produkten. För Patton är processen mycket viktigare än själva produkten: *”...what happens from the very beginning of a study will determine its eventual impact long before a final report is produced.”* (Patton 1997, s. 20). I stället för att fundera på tänkbara användare ska man, enligt honom, arbeta med dem som är de primära användarna. Eftersom utvärdering inte är värderingsfritt är deras värderingar avgörande. Han menar att *”... evaluation use is too important to be left to evaluators.”* (Patton 1997, s. 21). Idén med det som kallas för följeforskning är att följa och vara behjälplig i själva utvärderingsprocessen.

Vi ser att det kan finnas en viss konflikt mellan användningsfrågan, det vill säga huruvida användaren har monopol på vad ska utvärderas och hur det ska ske och värdefrågan, det vill säga hur oberoende en utvärderare ska vara i förhållande till beställaren eller användaren av utvärderingen. Hur ofta hör vi inte i den offentliga politiken att det ska tillsättas en oberoende utredning eller utvärdering. Vad innebär det i praktiken? Ove Karlsson Vestmans syn att

utvärdering är en svår konst understryks ytterligare av att utvärderaren hela tiden måste ta ställning till och avväga noggrannhet, värderingar och användning mot varandra.

Utmaningar

Förutom nyss nämnda mer övergripande överväganden kan även en rad praktiska problem eller överväganden uppstå. Vi kan illustrera detta med en av Riksrevisionens rapporter där de granskat hur statligt stöd till innovationer och företagande fungerar. Myndigheten konstaterar sammanfattningsvis (Riksrevisionen 2016, s. 9):

Riksrevisionens bedömning är att regeringen inte har en tillräckligt samlad och tydlig bild av vare sig stöden eller deras effekter. Stöden utvärderas kontinuerligt av forskare och utredare, men den kritik som framförs tas inte alltid till vara för lärande och utveckling. Vidare saknas användbara data som gör att stöden går att följa upp på ett ändamålsenligt sätt, samtidigt som målen för de enskilda programmen formuleras så att de är svåra att följa upp. Utan användbara data och uppföljningsbara mål är det svårt att utvärdera resultat av stöden. Trots brist på överblick över stöden och deras effekter kan det vara så att stöden skapar nytta. Men för att öka tilltron till systemet och för att kunna styra insatserna dit där de gör mest nytta bör systemet utvecklas. Det bör dock påpekas att det fragmenterade systemet samt de oklara effekterna av stöden inte innebär att myndigheter eller program inte kan fungera väl.

Riksrevisionen tar således upp en rad olika problem. Målsättningarna är oklara, stödets effekter är osäkra, överblick av de olika stöden liksom data för uppföljning och utvärdering saknas. Situationen är inte knappast unik för detta sakområde. Rekommendationerna är också ungefär desamma för alla typer av offentliga insatser "... mål, syfte, metod, data och val av variabler bör klargöras för utvärderaren innan programmen påbörjas."

Hur ska eller kan man hantera denna problematik som vare sig är ny eller kommer att försvinna? Det finns flera olika delar i denna problematik. Vi kan dela in problematiken i mål, utvärderingsbarhet, kausalitet och kunskapsintresse.

Mål

När det gäller oklara eller vida målsättningar utgör, som sagts tidigare, utvärdering en del av den politiska processen. Att målsättningar kan vara vida eller oklara i politiken beror på att man kanske vare sig vill eller kan vara särskilt precisa. Det innebär att tolkningsutrymmet blir stort och det får utvärderare eller tjänstemän ofta leva med. Det kan inte lösas på något annat sätt än att uppdragsgivare och uppdragstagare innan uppdraget startar har en dialog huruvida det är möjligt att precisera målsättningen eller målsättningarna. Utvärderingsforskaren Michael Scriven lanserade för länge sedan uttrycket Goal-Free Evaluation, det vill säga att utvärderaren ska bortse från målsättningarna och utvärdera alla de typer av effekter som insatsen haft (Scriven 1973). Rent praktiskt blir det svårt för en utvärderare att inte utgå från målsättningarna men ansatsen i sig betonar att också andra typer av effekter bör beaktas, det vill säga såväl positiva som negativa bieffekter. Att veta eller att ställa frågor om eventuella andra typer av effekter kräver emellertid goda kunskaper om sakområdet.

En kontinuerlig dialog mellan uppdragsgivare och uppdragstagare under arbetets gång, oavsett hur precisa målsättningarna är, kan förhindra att förväntningarna på utvärderingens resultat blir orealistiska.

Utvärderingsbarhet

Lika viktigt som att diskutera målsättningen för en insats är att fundera över vilka förutsättningar som det finns för att faktiskt ta reda på det som uppdragsgivaren vill ha reda på. I en komplex verklighet där politiken används för att lösa vissa frågor och där det inom varje enskilt politikområde finns ett stort antal politiska insatser, måste man faktiskt ställa sig frågan om den insats som man avser att mäta går att mäta. Insatsen kanske är så invävd i tidigare insatser eller att den verklighet som ska påverkas också påverkas av andra faktorer. Detta är naturligtvis något som politikerna själva är mycket medvetna om (Proposition 2011/12:1, s. 53):

Samtidigt som jobbskatteavdraget infördes gjordes vidare flera andra regelförändringar som kan ha påverkat arbetsutbudet och sysselsättningen, t.ex. förändringar i arbetslöshetsförsäkringen, sjukförsäkringen och socialavgifterna, vilket innebär att det är svårt att identifiera den isolerade effekten av jobbskatteavdraget.

Man kan inför en utvärdering ställa ett antal frågor som kan påverka utvärderingsbarheten exempelvis: Hur omfattande är insatsen? Är det en av flera insatser inom området eller är insatsen unik? Vilka data finns tillgängliga? Vilken typ av kunskap sökes? Vilka resurser är avsatta för utvärderingen? När kan man förvänta sig att se effekter och i så fall vilka?

Ju längre i tid man väntar med att mäta desto fler faktorer kommer förmodligen att påverka utfallet och det blir svårare att isolera den insats man är intresserad av att mäta. Att vänta för länge innebär också att det tar lång tid innan man får reda på om insatsen är effektiv eller inte. Ett sätt att förhindra detta är att tänka i effektkedjetermer. Ett effektkedjeresonemang är i första ledet en tankekonstruktion där vi försöker få klarhet i vad som möjligtvis skulle kunna hända i de olika leden i en effektkedja, det vill säga en effekt i kedjan är en förutsättning för nästa effekt i effektkedjan.

Fördelen med att tänka i termer av effektkedjor är dels att man måste fundera över vilka led eller effekter som måste inträffa för att nästa led ska inträffa dels fundera över varför det kommer att ske, det vill säga själva programteorin. Det innebär att man kan fundera över vilka effekter som skulle mätas innan den slutliga effekten inträffar, det vill säga man behöver inte invänta en eventuell slutlig effekt. Om det visar sig att en tidigare effekt i en effektkedja inte inträffar kan man ju avsluta en insats alternativt justera insatsen. Om det visar sig att en tidigare effekt inträffar i en effektkedja kan man på motsvarande sätt säga att det verkar som om man är på rätt väg. Slutsatsen är helt enkelt att en mätning av effekter i ett tidigare led i en effektkedja innebär att man inte alltid behöver ”vänta” på att mäta den slutliga effekten.

En effektkedja liksom programteorin innebär att man i förväg funderar över hur en insats är tänkt att fungera men också i utvärderingsledet har en bättre uppfattning om vad som ska mätas och när det ska ske. Att i förväg fundera på hur en effektkedja skulle kunna se ut innebär inte att man därmed har "tänkt rätt". Men att tänka i effektkedjor löser inte kausalitetsfrågan, det vill säga vi måste fortfarande mäta varje led i kedjan för att se om insatsen ger effekter. Vad menas då med kausalitet?

Kausalitet

Kausalitetsfrågan är något som diskuterats länge och intensivt inom många ämnesområden, förmodligen livligast bland filosofer. Det finns flera olika kausalitetsteorier även om de inte är helt oberoende av varandra (Se exempelvis Sandahl & Petersson 2016). David Humes klassiska teori om kausalitet som ett *lagbundet förhållande* (Hume 1739) innebär att en viss typ av orsak alltid följs av en vis typ av effekt. Inom naturvetenskapen kan detta tänkande sägas vara besläktat med formulerandet av kontext-oberoende naturlagar. John Mackies resonemang om *INUS-villkor* (Insufficient but Necessary part of a complex of conditions which is Unnecessary but Sufficient), kan ses som en vidareutveckling av Humes teori och innebär att det inte räcker med att en viss typ av händelse inträffar för att en annan typ av händelse alltid ska följa. Istället menar Mackie att det alltid är flera olika villkor som samverkar och därigenom producerar en effekt (Mackie 1965). Endast tillsammans är de tillräckliga för att effekten ska uppkomma. Exempelvis räcker det inte med hög värme för att en eldsvåda ska utbryta. Det krävs också, bland annat, brännbart material och syre.

Kausalitetsbegreppet förknippas ibland också med *sannolikhetsbegreppet*. Det finns helt enkelt alltför många möjliga kombinationer av INUS-villkor för att vi ska kunna observera deterministiska samband. Kausalitet skulle då kunna ses som förändrad sannolikhet. Vad innebär detta? Humes teori innebär att en orsak förändrar sannolikheten för en effekt från 0 till 1. En teori om kausalitet i termer av sannolikhet innebär att en orsak förändrar sannolikheten för en effekt, men inte nödvändigtvis från 0 till 1. I sådana fall finns inga deterministiska lagar att upptäcka, utan endast statistiska lagar. Sådana teorier har visat sig fruktbara inom den del av naturvetenskapen som benämns kvantfysik, men inom samhällsvetenskapen har de visat sig problematiska.

Filosofen Nancy Cartwright (Cartwright 1979) menar att kausalitet inte kan definieras enbart i termer av sannolikhet. Ett exempel på en sådan invändning är att en faktor X kan ha både en direkt och en indirekt effekt på en annan faktor Y. Om den ena av dessa båda effekter är positiv och den andra är negativ, så kan de ta ut varandra. Cartwright menar därför att kausalitet måste definieras dels i statistiska termer, dels i termer av kausala strukturer. Det är alltså viktigt att skilja på direkta effekter och nettoeffekter. Filosoferna är idag överens om att statistiska förhållanden endast kan fånga kausalitet om vi konstanthåller *samtliga* faktorer som påverkar den beroende variabeln, vilket i praktiken givetvis är omöjligt. Kausalitet kan därför inte definieras enbart i termer av sannolikhet och tidsföljd.

Ibland är vi intresserade av att förstå varför en samhällsförändring uppkommer eller varför en insats leder till att människor gör vissa val snarare än andra. Teorier om kausalitet som *verkande mekanismer* kan då vara ett alternativ. Inom samhällsvetenskaperna kan mekanismerna kopplas till individer och deras beteenden. Det kan exempelvis handla om hur människors tolkningar av intryck, intentioner, känslor och föreställningar om framtiden samverkar och på så vis producerar val och beteendeförändringar. Och det är sådana mekanismer som förklarar varför insatser leder till effekter. Om vi exempelvis noterar att en informationskampanj tycks leda till minskad rattonykterhet, så kan detta kanske förklaras med mekanismer som omfattar att människors värderingar kring rattonykterhet har förändrats.

Inom sociologin finns en omfattande litteratur om så kallade sociala mekanismer, som kan förstås som idealtyper för interaktion. Framträdande författarnamn har varit exempelvis Robert Merton och James Coleman och under senare år exempelvis Peter Hedström och Richard Swedberg. Merton använder en självuppfyllande profetia inom banksektorn som exempel. Ryktet om att en ekonomisk kris är förestående gör att vissa sparare börjar ta ut sina pengar. Det ger signaler till andra att göra samma sak, vilket i slutänden leder till en bankkris. Ett sådant händelseförlopp kan förstås som en social mekanism.

Andra filosofer har menat att vi kan förklara vad kausalitet är med hjälp av termen *manipulerbarhet*. Manipulerbarhetsteori förknippas kanske främst med 1900-talsfilosofen Georg Henrik von Wright (von Wright 1971). Grundtanken är att huruvida ett förhållande mellan två faktorer är kausalt beror på huruvida en manipulering av orsaken skulle leda till en förändring av effekten. von Wright likställer manipulering med mänskligt handlande. Om vi skulle förändra den förmodade orsaken genom mänskligt handlande och om då även den förmodade effekten skulle förändras, så är förhållandet mellan dessa båda händelser kausalt. Manipulerbarhetsteorin innebär emellertid inte att en handling måste utföras för att ett förhållande ska kunna sägas vara kausalt. Det räcker med ett tankeexperiment som stödjer tanken att *om* orsaken manipulerades, så skulle effekten förändras.

Med utgångspunkt i manipulerbarhetsteorin menade von Wright också att samhällsvetare kan förklara människors handlingar med hänvisning till deras intentioner och föreställningar. Detta benämns en *ändamålsförklaring*. Det centrala elementet i en ändamålsförklaring är alltså handlingars avsedda effekter, inte deras faktiska effekter. En person som springer till tåget har för avsikt att hinna med tåget men det innebär inte att personen i fråga faktiskt hinner med tåget. .

En nära relaterad teori är den så kallade *kontrafakticitetsteorin*. Förespråkare för kontrafakticitetsteori menar att kausalitet kan definieras i termer av kontrafaktiskt beroende. Dessa filosofer menar alltså att kontrafaktiskt beroende är ett mer grundläggande och intuitivt begrepp än kausalitet. Effekten av exempelvis en insats är enligt detta synsätt skillnaden mellan den faktiska utvecklingen och den kontrafaktiska utvecklingen, exempelvis skillnaden i

bensinkonsumtion mellan en verklighet där regeringen höjer bensinskatten och en verklighet där regeringen inte höjer bensinskatten. Men vi har bara en verklighet. Vi kan naturligtvis aldrig vara säkra på vad som skulle ha hänt utan den insatta åtgärden. Effektmätningens uppgift blir därför enligt detta synsätt att så långt möjligt försöka konstruera denna tänkta utveckling.

När denna teori lanserades av David Lewis i början av 1970-talet (Lewis 1973 a och b) var den banbrytande, eftersom Humes teori om kausalitet som ett lagbundet förhållande ditintills hade varit norm. I ett kontrafaktiskt kausalitetsresonemang ligger intresset istället primärt på ett tankeexperiment om den aktuella situationen, inte på om förhållandet mellan två faktorer är uttryck för ett regelbundet förhållande, såsom en strikt lag eller en statistisk lag.

Den kanske mest omdiskuterade invändningen mot kontrafakticitetsteori rör så kallad *preemptive causation*, som på svenska skulle kunna översättas med företrädesorsaker. Förekomsten av företrädesorsaker är en allvarlig invändning mot kontrafakticitetsteorin och den har stora praktiska implikationer för den som intresserar sig för effekterna av offentliga insatser. Beakta följande exempel. Två lätt överförfriskade personer (Anders och Bertil) bestämmer sig för att krossa grannens fönsterruta. De tar var sin sten i handen och bestämmer att Anders ska kasta sin sten först. Om Anders missar så ska Bertil kasta sin sten. Anders träffar och rutan går sönder. Det tycks uppenbart att orsaken till att fönsterrutan gick sönder är att Anders kastade sin sten. Men kan vi i detta läge dra slutsatsen att om Anders hade missat så hade fönsterrutan inte gått sönder? Svaret måste bli nej, eftersom Bertil i sådana fall hade kastat sin sten. Vi vet inte hur det hade gått. Anta att om Anders hade missat så hade Bertil träffat. Då hade fönsterrutan gått sönder även om Anders hade missat. Det kausala påståendet kan alltså vara sant trots att effekten inte är kontrafaktiskt beroende av orsaken. I detta fall, menar kritikerna, vilseleder oss kontrafakticitetsteorin att förkasta ett kausalt samband som icke-kausalt. Och den kan vilseleda oss även i fall där det inte är uppenbart om ett förhållande är kausalt eller inte.

Vi kan konstatera att det i dag finns ett antal olika teorier om kausalitet och att det finns argument både för och mot dem alla. Var och en av dessa fem teorier om kausalitet fångar någon aspekt av kausalitet, som vi har en intuitiv förståelse för, men ingen teori fångar alla aspekter. Att dessa teorier fångar olika aspekter kan tolkas och värderas på olika vis. Dessutom kan vissa kausalitetsteorier sägas vara direkt beroende av varandra. Exempelvis bygger manipulerbarhetsteori på kontrafakticitetsteori, eftersom innebörden är att kärnan i kausalitetsbegreppet är att *om* vi genomförde en manipulering av orsaken så skulle effekten förändras, vilket är ett kontrafaktiskt resonemang. I von Wrights version av manipulerbarhetsteorin finns inget krav på att resonemangen ska förankras i kunskap om lagbundenheter. von Wrights teori kan på så vis sägas påminna om David Lewis tankar om kausalitet som kontrafaktiskt beroende, som definierar kausalitet utan hänvisning till generella förhållanden.

Det finns även kopplingar mellan begreppen kausalitet, sannolikhet och kontrafakticitet, men det är en kontroversiell fråga hur dessa begrepp exakt förhåller sig till varandra. Nancy Cartwright har varit kritisk mot metoder som syftar till att analysera korrelationer i termer av kontrafakticitet och menar att analyser av vilka kausala förhållanden det är som ger upphov till en viss korrelation bör utgå från en teori om kausalitet som håller sig till det som faktiskt inträffar i världen. Samtidigt har David Lewis valt att modifiera sin version av kontrafakticitetsteorin genom att inkorporera sannolikhetsresonemang.

I praktiken

Upplägg

Om vi antar att det finns förutsättningar för att exempelvis utvärdera effekterna av en insats och vi vet vilka effektvariabler som är aktuella, vilka val står man då inför? Ett val gäller vilket upplägg som ska väljas. Här har, som nämnts ovan, den så kallade evidensdebatten varit intensiv där evidens nästan likställts med randomiserade experiment och där den kontrafaktiska kausalitetsteorin dominerat. Shadish, Cook och Campbell menar att en effekt är en kontrafaktisk skillnad i utfall och att en central uppgift för forskning som försöker spåra kausala samband, är att avbilda det kontrafaktiska utfallet (Shadish m. fl. 2002, s. 5). Det experimentella upplägget är därför, enligt Shadish, Cook och Campbell, det bästa upplägget för att spåra kausala samband (Shadish m. fl. 2002, s. 7):

Hence experiments are well-suited to studying causal relationships. No other scientific method regularly matches the characteristics of causal relationships so well.

I praktiken är det emellertid inte alltid möjligt att tillämpa alla upplägg. Exempelvis måste experimentet väljas innan den utvärderade insatsen genomförs för att vi ska kunna randomisera till experiment- och kontrollgrupp. Andra gånger kan vi av praktiska skäl inte välja alla möjliga upplägg, kanske beroende på databrist, tidsbrist eller att resurserna helt enkelt är begränsade. Vi måste därför välja upplägg som vi kanske, idealt sett, inte skulle ha valt eller avstå från en situation där vi helst skulle vilja kombinera flera typer av upplägg.

Eftersom det finns argument både för och mot dessa olika, delvis konkurrerande, teorier om kausalitet, så finns det också starka argument både för och mot olika upplägg för att fånga kausalitet. De olika uppläggen framstår därmed snarare som komplement än som alternativ. Vi menar kort sagt att det inte kan finnas något bästa upplägg för effektmätning eftersom det inte finns någon bästa teori om kausalitet. Därför är inte heller olika tillvägagångssätt nödvändigtvis olika vetenskapliga eller ”rigorösa”. Det är viktigt att vara medveten om detta när man beställer, genomför, eller tolkar resultaten av en orsaks- eller effektanalys.

Däremot kan givetvis ett visst tillvägagångssätt sägas vara lämpligt och kanske till och med det mest lämpliga i ett visst fall, givet ett visst kunskapsläge och ett visst kunskapsintresse. Exempelvis bygger kombinationen av experiment och statistiska metoder på att vi har hypoteser

att pröva. Om vi har tydliga hypoteser och inte behöver förklara varför en insats ger eller inte ger avsedda effekter, så kan ett experiment som använder statistiska metoder för att spåra skillnader på en målvariabel vara ett gott alternativ. Detta framhålls av bland andra Shadish, Cook och Campbell – en författartrio som har varit mycket inflytelserik på utvärderingsområdet (Shadish m. fl. 2002, s. 457):

The purpose of experiments is not to completely explain some phenomenon; it is to identify whether a particular variable or small set of variables makes a marginal difference in some outcome over and above all other forces affecting that outcomes.

Om det är viktigt att kunna bidra till att förbättra studerade insatser och vi har lite kunskap att utgå från – det vill säga våra hypoteser är relativt osäkra – så kan det vara viktigt att förklara varför en viss insats har eller inte har genererat avsedda effekter. Ett experiment som kanske låter oss veta att effekter inte har uppkommit i avsedd omfattning kan då vara otillräckligt. Att i stället, utifrån hypoteser eller teorier, studera de mekanismer som skulle kunna ge upphov till effekter kan vara värdefullt och kan genomföras exempelvis med någon typ av fallstudie. Det kan också vara viktigt att studera skillnader i resultat mellan olika delar av målgruppen för att på så sätt kunna bidra till en förbättrad målgruppsanpassning. Ibland kanske vår kunskap är för liten för att vi ska kunna formulera tydliga hypoteser. Då kanske ett induktivt tillvägagångssätt är lämpligast. Det är alltså viktigt att i varje enskilt fall utgå från en precis fråga som ska besvaras: Räcker det att få veta exempelvis att en insats fungerar för en genomsnittlig mottagare, eller behöver man också veta varför eller varför inte avsedda effekter uppkommer hos olika typer av mottagare? Det räcker inte att fråga efter effekter – vi måste fråga efter en viss aspekt av kausalitet.

Vi kan också konstatera att de olika teorierna om kausalitet ibland uppträder i kombination, i vissa fall är beroende av varandra och att det i regel inte finns något som hindrar att de kombineras för att skapa en fullödig förståelse för vad kausalitet innebär. På motsvarande vis finns det många gånger inget som hindrar att olika upplägg kombineras. Det är exempelvis fullt möjligt att genomföra ett experiment, göra statistiska analyser och följa processerna från insats till målvariabel med iakttagande av mekanismer och INUS-villkor.

Intern och extern validitet

Debatten om kausalitet och upplägg har också handlat om att man i valet av upplägg många gånger måste välja mellan intern och extern validitet. Intern validitet är, enligt Shadish, Cook & Campbell, ett mått på sanningshalten i ett påstående om kausalitet i ett undersökt fall, exempelvis ett påstående om att ett visst program har gett vissa effekter för mottagarna.

We use the term *internal validity* to refer to instances about whether observed co-variation between A and B reflects a causal relationship from A to B in the form in which the variables were manipulated or measured (Shadish m. fl. 2002, s. 53).

Intern validitet gäller därmed ett påstående om kausalitet i en given kontext (Shadish m. fl. 2002, s. 54):

[I]nternal validity is about whether a complex and inevitably multivariate treatment package caused a difference in some variable-as-it-was-measured within the particular setting, time frames, and kinds of units that were sampled in a study.

Extern validitet är istället ett mått på sanningshalten i ett påstående om generaliserbarheten av en sådan fallspecifik slutsats om kausalitet. Det kan exempelvis röra sig om huruvida de fastställda effekterna av ett program skulle ha blivit desamma om insatsen hade rört en annan grupp av brukare (Shadish m. fl. 2002, s.83).

External validity concerns inferences about the extent to which a causal relationship holds over variations in persons, settings, treatments, and outcomes.

Huvudsyftet med att genomföra en utvärdering, enligt exempelvis Lee Cronbach, bör vara just att studera variation, det vill säga vilka satsningar som fungerar för vilka mottagare. Han uttryckte detta så att extern validitet borde prioriteras och att ” [...] internal validity is only of secondary concern to the evaluator” (Cronbach 1982, s. 112). Cronbach var alltså inte intresserad av kausala förhållanden som har inträffat i enskilda fall utan av generella kausala förhållanden eftersom endast generella kausala förhållanden kan spela en roll i det politiska beslutsfattandet. Och generaliseringar förutsätter kunskap om vilka delar av en målgrupp som påverkas på vilka vis, menar Cronbach.

I grund och botten handlar emellertid all politik om att manipulera samhällsutvecklingen och att ta ansvar för att vissa insatser valdes, snarare än andra. Att fråga vad som skulle ha hänt om en insats *inte* hade genomförts, det vill säga om samhället inte hade manipulerats, är också att fråga efter en kontrafaktisk skillnad.

Vilket upplägg som är att föredra måste avgöras från fall till fall och detta val måste göras beroende dels av vad som är praktiskt möjligt, dels av vilket kunskapsintresse vi har, det vill säga vilken fråga som ska besvaras. Ibland är vi bara intresserade av skillnaden i utfall, andra gånger av verkande mekanismer, i ytterligare andra fall av båda dessa aspekter. Kunskapsintresset avgörs av vad kunskapen ska användas till.

Vi frågade oss inledningsvis också hur vi ska se på att de praktiska förutsättningarna för analyser av orsaker och effekter onekligen många gånger inte är de bästa. Är det så att politiken borde anpassas så att insatser blir utvärderingsbara under rimliga förutsättningar? Eller är det snarare så att metodvalen borde anpassas så att de kan möta politikens kunskapsbehov? Frågan är alltså vilket förhållandet bör vara mellan beställare och utförare av analyser av orsaker och effekter.

Dessutom konstaterar vi att tillvägagångssätt, som ofta beskrivs som alternativ, rör olika nivåer av analysprocessen och därför med fördel kan kombineras. Exempelvis är det fullt möjligt att genomföra ett kontrollgruppsupplägg och välja att spåra verkande mekanismer i både experiment- och kontrollgrupp. Andra gånger kan det vara viktigt att kombinera ”hårda” och ”mjuka” tillvägagångssätt för att uppnå triangulering, något som framhölls av bland andra Donald Campbell (Shadish m.fl. 1991, s. 132). Det finns alltså flera olika goda skäl att betrakta olika upplägg som komplement för att bygga upp kunskap inom ett sakområde.

Vi vill också, som Donald Campbell, förespråka ett stort mått av pragmatism: Om upplägget är tillräckligt eller inte för att garantera intern validitet beror på hur bra kontroll man har på den grupp som undersöks, det vill säga hur lätt det är att falsifiera alternativa hypoteser (Campbell 1979). Det handlar om i vilken utsträckning insatsen kan isoleras från andra tänkbara orsaker. Om vi håller en syra i ett provrör, trycker på bromsen när vi kör, slår oss på tummen med en hammare eller bränner oss på en tändsticka, så kan vi med stor säkerhet säga vad orsaken är. Oavsett om vi intresserar oss för kontrafakticitet eller för någon annan aspekt av kausalitet så behöver vi ingen kontrollgrupp i dessa fall. Även om vi många gånger är hjälpta av att ha en kontrollgrupp för att kunna resonera om kontrafakticitet så är det, enligt exempelvis biståndsforskaren Howard White, (White 2010, s. 157) inte alltid nödvändigt.

I enlighet med Lee Cronbach bör man vid varje utvärdering utgå från en diskussion av de olika för- och nackdelar som olika upplägg har givet ett visst kunskapsintresse – det vill säga givet att vi vill fånga vissa aspekter av kausalitet snarare än andra (Cronbach 1982, s. 27).

Kunskapsintresse

Dessutom är det självklart att ideologi, politisk taktik och många andra faktorer alltid kommer att påverka när och hur insatser utformas och genomförs. Politiker kan inte styras med analyser av orsaker eller effekter, men kanske påverkas till viss del. Vi menar liksom många andra att det är frågan och inte metoden som bör vara utgångspunkten för en undersökning. Samhällsproblem har många gånger ett stort antal olika orsaker och vi kan vara intresserade av att utreda vissa hypoteser snarare än andra. Andra gånger har vi mindre förförståelse och måste ställa frågan om orsakerna till ett problem mer förutsättningslöst. Om man ställer sig frågan varför elever inte går ut med fullständiga betyg måste man kanske göra en orsaksanalys av möjliga påverkansvariabler som har med hemmiljö, lärarstöd, kamratskap och andra faktorer att göra. Om vi istället har genomfört en insats och vill ha kunskap om dess effekter, så kan vi även då många gånger välja att fokusera på vissa kausala förhållanden snarare än andra. Många insatser är avsedda att påverka flera olika aktörer och kanske i flera led. Då kan det vara viktigt att avgöra vilka förhållanden vi framför allt behöver kunskap om.

För att kunna ställa rätt frågor om orsaker och effekter i relation till en viss insats är det viktigt att ha en tydlig bild av hur insatsen är *tänkt* att fungera, det vill säga att förstå intentionerna. En sådan teoretisk grund om hur en insats är tänkt att ge avsedda effekter är det vi ovan benämnt

programteori, verksamhetslogik eller förändringsteori. Den något snusförnuftiga devisen att tänka efter före ger oss ett bra utgångsläge när det gäller att sedan utvärdera dessa insatser. I ett samverkansprojekt mellan polisen och Brottsförebyggande rådet, som gällde polisens brottsförebyggande arbete, menade man att just en sådan teoretisk grund saknades, vilket försvårade utvärdering (Brottsförebyggande rådet 2011, s.19):

Det polisiärt brottsförebyggande arbetet i Sverige vilar inte på någon tydlig teoretisk grund, vilket bland annat får till följd att möjligheterna att på ett tillfredställande sätt utvärdera verksamheten försvåras.

Inför en utvärdering är det viktigt att ta ställning till vilka olika aspekter av kausalitet man kan efterfråga kunskap om. Genom att fånga mekanismer kan vi förklara exempelvis *varför* människor omsätter eller inte omsätter den kunskap de fått genom en kurs eller en informationskampanj, en kunskap som är värdefull då vi behöver lära för att förbättra pågående eller kommande satsningar.

Ett annat kunskapsintresse som kan vara relevant är hur olika faktorer samverkar för att producera en effekt. Vissa filosofer har betonat att det inte är meningsfullt att diskutera orsaker utan att tänka i termer av samverkande faktorer. Detta stämmer väl överens med vår intuition att många olika faktorer kan påverka ett och samma samhällsfenomen eller leda fram till en viss händelse som väcker särskilt intresse. Sådan kunskap är viktig inte minst för att få veta vilken typ av insats som fungerar för vilken typ av mottagare. I en översikt av arbetsmarknadspolitikens effekter på sysselsättning och arbetslöshet konstaterade författarna (Forslund & Vikström 2011, s. 57) just att denna typ av kunskap är viktig:

Ett genomgående tema i översikten har varit att vi för program efter program vet relativt lite om vilken typ av arbetslösa som gynnas av en specifik insats. Sådan kunskap är viktig eftersom det är möjligt att ett visst program fungerar bra för en viss typ av arbetslösa men sämre för andra. Exempelvis skulle det kunna vara så att ungdomar gynnas av andra program än de äldre arbetslösa. En näraliggande fråga är hur de olika programmen ska utformas. Hur lång ska exempelvis arbetsmarknadsutbildningen vara? Vilken är den optimala längden och subventionsgraden för nystartsjobb och andra typer av subventionerade anställningar, och vilken typ av förmedlingsaktivitet är viktigast? Återigen kan vi konstatera att den nuvarande litteraturen är väldigt begränsad.

För att förstå ett skeende är mekanismer ett centralt element för att förklara varför en viss effekt inträffar. Sociologen Jon Elster (Elster 2007, s.36) menar att mekanismer är den svarta lådan som ger upphov till relationen mellan orsak och effekt:

[...] the cogs and wheels that have brought the relationship [between policies and outcomes] into existence.

Enligt Andrew Sayer (Danemark m.fl. 2003, s. 95) är en mekanism det som ”får något att hända”, som ”producerar”, ”genererar”, ”skapar”, ”bestämmer”, ”möjliggör”, eller ”leder till” en effekt. Inom samhällsvetenskaperna kan mekanismerna kopplas till individer och deras beteenden (Elster 1998, s.47). Den amerikanska utvärderingsforskaren Carol Weiss menar att vi måste söka mekanismer för att förstå hur offentliga insatser påverkar människors föreställningar och val. Annars kan vi inte förstå *varför* en insats leder till eller inte leder till vissa effekter. Weiss (Weiss 1997, s.46) menar att mekanismerna inte är någon del av själva insatsen, utan människors intryck av insatsen och omvandlingen av dessa till att respondera på ett visst vis.

För Ray Pawson (Pawson & Tilley 1997, s. 58) är det viktigt att se mekanismer som delar i ett större socialt sammanhang. Han uttrycker detta så att ” [...] causal outcomes follow from mechanisms acting in contexts”. Olika kontexter ger individerna olika stora möjligheter att agera på ett vis som omformar de sociala mekanismerna. En insats som ger avsedd effekt i ett visst socialt sammanhang behöver således inte ge avsedd effekt i ett annat socialt sammanhang (Pawson 2006, s.94):

The traditional “does it work” question transforms instantly into at least five different lines of inquiry: *what* is it about this kind of intervention that works, for *whom*, in what *circumstances*, in what *respects* and *why*?

Vi måste helt enkelt kunna beskriva kopplingen mellan orsak och effekt. Annars har vi inte klargjort *varför* orsaken leder till effekten (Elster 1983, s.25). Tar vi inte reda på dessa mekanismer, så förblir det kausala förhållandet ett mysterium menar Pawson och Tilley (Pawson & Tilley 1997, s.66). Det är inte själva övervakningskameran på en parkeringsplats som får presumtiva brottslingar att avstå från inbrott, enligt Pawson och Tilley. Det är vetenskapen om övervakningskameran som gör att den presumtive brottslingen börjar fundera över sitt beteende och det är denna omvandling av intryck till respons som är mekanismen.

Vi talar ofta om tekniska metoder för att försöka mäta, i det här fallet, effekter av olika insatser. Val av upplägg blir emellertid intressant först när frågor om utvärderingsbarhet och kunskapsfrågorna har besvarats. Kanske kommer man fram till att det inte går att utvärdera eller att det inte känns meningsfullt. Utvärderingar som inte känns meningsfulla att genomföra ska inte heller genomföras bara för att den finns en idé om allting ska utvärderas.

En fråga som måste ställas i utvärderingssammanhang är vilken tid det finns hos beställare, exempelvis hos Regeringskansliet, att lägga ned den tid som behövs för att diskutera utvärderingsupplägg, kunskapsfrågor etc. Eftersom denna tid oftast är knapp ställer det höga krav på utvärderaren att presentera möjligheter och svårigheter med att utföra utvärdering för att just uppdragsgivare och uppdragsmottagare inte ska bli besvikna. Svårigheterna kanske uppstår när uppdragsgivaren känner sig tvungen att göra en utvärdering och exempelvis en konsult har svårt

att avstyra en utvärdering eftersom den utgör ett levebröd. Vågar någon part säga nej, det vill säga att en utvärdering inte är meningsfull?

Det finns inte bara ett upplägg utan flera möjligheter. En enskild utvärderare kan inte hantera alla möjliga upplägg och alltid ställa rätt frågor. Vad kan man begära av en utvärderare och vilken kunskap ska en utvärderare ha? Metodfältet är stort och ingen behärskar alla samhällsvetenskapliga metoder och ingen behärskar ett helt sakområde. Att ställa rätt frågor kräver både en utvärderingsteoretisk kunskap och sakkunskap. Därför är det extra viktigt att föra en dialog mellan uppdragsgivare och uppdragstagare liksom att förankra metodansats och beställare hos metodexperter och sakområdesexperter om inte uppdragsgivare och uppdragstagare besitter dessa kunskaper. Utvärdering är, sammanfattningsvis, en "svår konst" en balansakt mellan noggrannhet, värderingar och användning.

Referenser

Alkin, Marvin C. (ed.) (2013). *Evaluation Roots. A Wider Perspective of Theorists' Views and Influences*, 2 uppl. Thousand Oaks: Sage Publications.

Brottsförebyggande rådet (2011). *Polisens strategiska brottsförebyggande arbete – återrapporering av ett regeringsuppdrag*. Dnr Brå 2011/0029.

Bohlin, Ingemar & Sager, Morten (red.) (2011). *Evidensens många ansikten. Evidensbaserad praktik i praktiken*. Lund: Arkiv förlag.

Campbell, Donald T. (1979). "Degrees of freedom" and the case study, i Cook, Thomas D. & Reichardt, Charles S., *Qualitative and Quantitative Methods in Evaluation Research*, s.49-67. Thousand Oaks: Sage Publications.

Cartwright, Nancy (1979) "Causal laws and effective strategies", *Noûs*, vol. 13 (4), s. 419-437. Special Issue on Counterfactuals and Laws.

Cartwright, Nancy (2007). *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press.

Chelmsky, Eleanor (ed.) (1985). *Program Evaluation: Patterns and Directions*. American Society for Public Administration. Washington D.C.

Cronbach, Lee J. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass Inc.

Danemark, Berth, Ekström, Mats, Jacobsen, Liselotte & Karlsson, Jan Ch. (2003). *Att förklara samhället*. 2 uppl. Lund: Studentlitteratur.

Elster, Jon (1983). *Explaining Technical Change. A Case Study in the Philosophy of Science*. Cambridge: Cambridge University Press och Universitetsförlaget.

Elster, Jon (1998). "A plea for mechanisms", i Hedström, Peter & Swedberg, Richard (red.), *Social Mechanisms. An Analytical Approach to Social Theory*, s. 45-73. Cambridge: Cambridge University Press.

Elster, Jon (2007). *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Forslund, Anders & Vikström, Johan (2011). *Arbetsmarknadspolitikens effekter på sysselsättning och arbetslöshet – en översikt*. Institutet för arbetsmarknads- och utbildningspolitisk utvärdering. Rapport 2011:7.

House, Ernest, R. & Howe, Kenneth, R. (1999). *Values in Evaluation and Social Research*. London: Sage.

House, Ernest, R. (2001). "Unfinished Business: Causes and Values", *American Journal of Evaluation*, vol. 22, no. 3, s. 309-315.

Hume, David (1739). *A Treatise of Human Nature*. The Online Library of Liberty.

Karlsson Westman, Ove (2011). *Utvärderandets konst*. Lund: Studentlitteratur.

Lewis, David (1973 a). "Causation", *Journal of Philosophy*, vol. 70, s. 556-567.

Lewis, David (1973 b). *Counterfactuals*. Oxford: Blackwell.

Mackie, John L. (1965). "Causes and conditions", *American Philosophical Quarterly*, vol.2, nr.4, s. 245-264.

Patton, Michael Quinn (1997). *Utilization – focused evaluation*. 3 uppl. Thousand Oaks: Sage Publications.

Patton, Michael Q. (2000). "Utilization-focused evaluation", i Stufflebeam, Daniel, L. & Madaus, George, F. & Kellaghan, Thomas (eds.) *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, 2 uppl., s. 425-438. New York: Kluwer Academic Publishers

Pawson, Ray & Tilley, Nick (1997). *Realistic Evaluation*. London: Sage Publications.

Pawson, Ray (2006). *Evidence-based Policy. A Realistic Perspective*. London: Sage Publications.

Proposition 2011/12:1. Budgetpropositionen för 2102.

Riksrevisionen (2016). *Statliga stöd till innovation och företagande*. RIR 2016:22.

Rossi, Peter, Freeman, Howard E. & Lipsey, Mark W. (2004). *Evaluation. A systematic Approach*. 7 uppl. Thousand Oaks: Sage Publications.

Sandahl, Rolf (1999). *Myndigheternas syn på resultatstyrningen*. ESV 1999:20. Stockholm: Ekonomistyrningsverket.

Sandahl, Rolf & Petersson, Gustav Jacob (2016). *Kausalitet. I filosofi, politik och utvärdering*. Lund: Studentlitteratur.

Scriven, Michael (1973) "Goal- free evaluation", i House, Ernest, R. (ed.) *School evaluation. The politics and process*, s. 319-328. Berkeley, CA: McCutchan.

Shadish, William R., Cook, Thomas D. & Leviton, Laura C. (1991). *Foundations of Program Evaluation. Theories and Practice*. Newbury Park: Sage Publications.

Shadish, William R., Cook, Thomas D. & Campbell, Donald T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.

Smith, M. F. (1994) "Evaluation: Review of the Past, Preview of the Future", *Evaluation Practice*, vol. 15, nr 3 1994, s. 215-227. Greenwich, Connecticut: Jai Press Inc.

SOU 1967:11. *Programbudgetering* (Stockholm: Fritzes)

SOU 1934:1. *Arbetslöshetsutredningens betänkande*.

Weiss, Carol H. (1972). *Evaluation Research: Methods of Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.

Weiss, Carol H. (1997). "Theory-based evaluation: Past, present and future", i Rog, Debra J. & Fournier, Deborah (eds.) *Progress and Future Directions in Evaluation; Perspectives on Theory, Practice and Methods*, s. 41-55. New Directions for Evaluation, No 76, Winter 1997. San Francisco: Jossey-Bass Publishers.

Weiss, Carol H. (1998). *Evaluation*, 2 uppl. New Jersey: Prentice Hall, Upper Saddle River.

Weiss, Carol H. (1988). "Evaluations for Decisions: Is Anybody There? Does Anybody Care?", *Evaluation Practice*, Vol 9, No. 1, s. 5-19, Newbury Park: Sage Publication.

Vedung, Evert (1998). *Utvärdering i politik och förvaltning*, 2 uppl. Lund: Studentlitteratur.

Vedung, Evert (2009). *Utvärdering i politik och förvaltning*, 3 uppl. Lund: Studentlitteratur.

Wildavsky, Aaron (1973). "If planning is everything, maybe it's nothing", *Policy Sciences* 4 (1973), s. 127-153.

White, Howard (2010). "A contribution to current debates in impact evaluation", *Evaluation* 16(2), s. 153-164. London: Sage.

von Wright, Georg Henrik (1971). *Explanation and Understanding*. London: Routledge & Kegan Paul.