

Bayesianism – en vetenskapsteoretisk återvändsgränd

Working-paper Lars Pålsson Syll 2011

En av vetenskapens viktigaste uppgifter är att förklara. Även om vetenskapen inte har något ”tydligt utformat kriterium på vad som är en tillfredsställande förklaring” (Halldén 1974 s 55) ska vi i detta avsnitt försöka beskriva hur vi kan få en klarare uppfattning om vad förklaring innebär och vilken roll den spelar i samhälls- och historievetenskaperna. Vi har därför inga pretentioner på att presentera en allmän förklaringsfilosofi eller att kunna ge annat än delsvår på de svåra frågor som vi ställer. Vi ska försöka särskilja olika slags förklaringar och ange varför vi anser vissa vara mer relevanta och adekvata för samhälls- och historievetenskaperna än andra. Vi kommer att hävda att vissa förklaringar är bättre än andra och att några förklaringar egentligen inte är förklaringar och därför ger svar på annat än det som frågan gäller.

All vetenskap handlar i grunden om att förklara. Att förklara en händelse eller ett förlopp är att ange varför något inträffade, att ange ett svar på frågan ”*varför?*” Ofta innebär detta att ange dess orsak.

Det som utmärker orsakssambandet mellan två händelser är ofta att orsaken i tiden föregår verkan. Att Bo fick löneförhöjning igår orsakade hans ökade konsumtion idag. Men det är inte tillräckligt – och strikt sett inte ens nödvändigt – att X föregår Y för att X ska hållas för orsak till Y. Orsaken måste också på något sätt kunna visas ha frambringat, determinerat eller nödvändiggjort sin verkan. Att Bo bar svarta sockor i förrgår är knappast orsaken till att han ökade sin konsumtion idag. Det krävs att man kan identifiera en nödvändiggörande egenskap hos orsaksfaktorn för att den ska kunna räknas som en förklaringsfaktor. Förklaringen måste kunna ge svar på våra frågor om vad det var som ledde till Y eller varför Y nödvändigtvis inträffade och vad det i så fall var som nödvändiggjorde Y.

En händelse eller ett förlopp kan givetvis ha flera olika orsaker. Ofta innebär detta att det finns flera var för sig tillräckliga orsaker bakom en viss verkan. Bo fick kanske inte bara löneförhöjning utan fick kanske också ärva pengar. Var för sig skulle dessa orsaker vara tillräckliga för att få honom att öka sin konsumtion. Vi har vad man brukar kalla *överdeterminering*.

I vetenskapsteorin har man ofta beskrivit fastställandet av orsaker i en förklaring med hjälp av den hypotetisk-deduktiva metodens *induktiva* och *deduktiva* slutledningar. Denna metod bygger i stor utsträckning på att man i en experimentsituation kan kontrollera och manipulera olika möjliga orsaksfaktorer. I samhälls- och historievetenskaperna har man inte den

möjligheten och därför har man i dessa ofta fått tillgripa andra metoder. I samhället är det olika aktörer som handlar och när vi vill förklara en handling försöker vi inhämta information om dessa aktörers föreställningar och uppfattningar. Vi försöker identifiera aktörernas *motiv*. Den av von Wright (1971) beskrivna *praktiska slutledningen* kan beskrivas schematiskt på följande sätt:

- 1 A har för avsikt att öka sitt välbefinnande
 - 2 A anser att han har fått större inkomster att röra sig med
 - 3 A tror att han när han har fått större inkomster att röra sig med inte kan öka sitt välbefinnande om han inte köper en ny bil.
-
- 4 A köper en ny bil

Många av de gängse förklaringarna till aktörers handlande har den här strukturen. Genom att ange motivet bakom en viss handling anser man sig ha förklarat den. Många gånger räcker detta. Men andra gånger – speciellt i vetenskapliga sammanhang – vill man gå längre. Vad ligger bakom motiven? Hur ska de relateras till den *situation* aktören befinner sig i? Vilken roll spelar samhällsstruktur, institutioner etc? Vi får då en förklaringsmodell som ser ut så här:

Bakgrundsvariabler => Attityder, uppfattningar => Handlingar

Vi försöker alltså att gå bakom aktörers motiv och identifiera vilka de relevanta bakgrundsvariablerna som kan förklara motiven är för att på så sätt fastare *grunda* förklaringen.

För att få svar på frågor av typen ovan, och för att få reda på vad som ligger bakom motiv kan ofta den av Charles S Peirce (se t ex Peirce 1990, Mirowski 1987, Kirkeby 1994b, Wible 1994, Niniluoto 2002) utvecklade *abduktionen* vara behjälplig. Denna kan schematiskt beskrivas på följande sätt:

- 1 Alla som får löneförhöjning ökar sin konsumtion
 - 2 A ökar sin konsumtion
-
- 3 A har fått löneförhöjning

Utifrån en *regel* (1) och ett *resultat* (2) kan vi dra *slutsatsen* (3). Resultatet (2) är här något vi inledningsvis inte har en förklaring till men som vi kan ge en möjlig förklaring (3) när vi tolkar den utifrån en regel (1). Men till skillnad från deduktionens nödvändiga och induktionens sannolika slutsatser är abduktionens slutsatser bara möjliga (A kan ju ha ärvt pengarna).

Abduktion är inte som deduktionen logiskt nödvändig eller som induktionen en empirisk generalisering. Abduktionen påminner i själva verket mycket om ett detektivarbete. Olika tolkningsramar prövas tentativt, problemet rekontextualiseras och med hjälp av kreativitet och fantasi upptäcks nya sammanhang och innebörder som gör att man kan lösa gåtan eller förklara en händelse eller ett förlopp. Vi kan inte med säkerhet veta att de nya sammanhangen och innebörderna ger en *sann* kunskap, men det finns en möjlighet att de ger *bättre* och *fördjupad* kunskap. Istället för att bara se ett konsumtionsmönster kan man, genom att inom ramen för ett inkomst- och fördelningssammanhang, se att Bo:s inköp av en ny Mercedes (som ersätter den gamla Folkan) avspeglar högre inkomster och förväntat statusbeteende (rika människor kör Mercedes).

Ibland hävdas det – oriktigt – att en förklaring inom vetenskapen normalt sett bara utgör en mer generell beskrivning av en företeelse. Paul Samuelson (1972 s 762) menar exempelvis att Newton – i sin ”förklaring” av himlakropparnas banor – efter att ha ”beskrivit ’hur’ inte ödslade tid på det ofruktbara ’varför’.” Att ställa frågor om orsaker och söka drivkrafter bakom företeelser är enligt denna uppfattning ”metafysik”. I själva verket är det väl snarare så som vi visat, att en samhällsvetenskap som inte ställer sådana frågor, knappast kan betraktas som en vetenskap.

Varför abduktion är bättre än bayesianism

På senare år har det förekommit en omfattande diskussion bland filosofer och vetenskapsteoretiker om abduktionens förtjänster och brister. Traditionellt har man i huvudsak försökt karakterisera den vetenskapliga metodens basala struktur i termer av induktion och/eller deduktion.

Den vetenskapliga metoden ska helst både vara *ampliativ* – utöka våra kunskaper – och öka vår epistemiska *tillförsikt* (beviskraft) för de resultat den producerar. Den bästa balansen mellan dessa två mål fås med hjälp av den beskrivning av den vetenskapliga metoden som ges av abduktionen, snarare än med hjälp av induktion eller den hypotetisk-deduktiva metoden. Detta gäller också för samhälls- och historievetenskaperna.

Att den vetenskapliga metoden ska utvidga våra kunskaper är egentligen en självklar utgångspunkt för ett kritisk-realistiskt förhållningssätt, där våra kunskaper och förståelsen av dem inte ses som begränsade till enbart våra sinnesintryck. ”Bakom” dessa går det att med hjälpa av vetenskapen uppdaga ny och kanske djupare kunskaper än de våra sinnen omedelbart kan förse oss med. Med de metoder vetenskapen tillhandahåller kan vi få fram något som vi inte visste innan och förhoppningsvis är denna kunskap också beviskraftig.

Det är dock inte alltid så lätt att förena ampliatio och epistemisk beviskraft. Vad är det som ger beviskraft åt en teori/hypotes/förklaring snarare än en annan när vi försöker gå utöver våra sinnesintryck? En rent deduktiv metod skulle försäkra oss om att slutsatsen är lika beviskraftig som de premisser den bygger på. Problemet är bara att deduktionen också är helt icke-ampliativ. Ska vi använda oss av en ampliativ metod måste vi därför acceptera att beviskraften inte kan vara av deduktiv kaliber. Våra data garanterar aldrig att endast *en* teori/hypotes/förklaring är giltig. Innebär detta att *alla* teorier/hypoteser/förklaringar har samma grad av giltighet eller beviskraft? Är alla katter grå? Inte nödvändigtvis! Om detta är en riktig hållning kan inte enbart avgöras på formallogiska grunder utan måste föras tillbaka på hur vår värld ser ut och är strukturerad. Att en metod inte är den bästa i alla möjliga världar utesluter ju inte att den fungerar väl i den värld vi lever i.

Ur ett perspektiv sett gäller ampliatio egentligen också för deduktionen. ”Om x , så y ” ger ju ingen beviskraft åt y om vi inte redan gett densamma åt x . För att kunna göra detta måste vi – åtminstone utanför matematiken och den rena formallogiken – föra ett ampliativt resonemang kring substantiella påståenden om saker och ting.

Det vi med hjälp av en ampliativ metod kommer fram till är alltid mer eller mindre tentativt. Till skillnad från deduktionens antingen /eller kan en ampliativ metods slutledningar alltid att kunna ändras/modifieras/förkastas som resultat av att ny information dyker upp eller en bättre analys genomförs. Ny information och bättre analys kan leda till att den tidigare tillförsikt vi hade i en teori/hypotes/förklaring/ helt försvinner.

Induktionens problem är att den bygger ampliatio från ”några” instanser till ”alla” på generalisering. Denna ”mer av samma” metod utökar våra kunskaper på ett enbart ”horisontellt” sätt. Inga nya entiteter, relationer, strukturer etc dyker upp. I det avseendet innebär induktionen en minimal ampliatio av våra kunskaper, baserad på ett underliggande antagande av världen som styrd av regelbundenheter. Induktionens tillkortakommande är uppenbar. Vad vi vinner i beviskraft förlorar vi i ampliatioens styrka. Induktionen är för restriktiv och ger oss inga hypoteser/teorier/förklaringar till orsakerna bakom de observerade fenomenen.

Den hypotetisk-deduktiva metoden å andra sidan möjliggör en kraftfull ampliatio via bekräftelse av uppställda hypoteser/teorier och öppnar upp för användande av icke-observerbara orsaker bakom fenomenen (vilket den rena induktionens ”mer av samma” inte möjliggör). Som Duhem-Quine-problemet exemplifierar hjälper oss metoden dock inte att identifiera vilket av antagandena eller hypoteserna i en teori som är felaktig när teorin inte bekräftas. Om både hypotes A och hypotes B kan förklara X, ger oss den hypotetisk-deduktiva metoden inga möjligheter att diskriminera mellan dem. Vad vi vinner i ampliatio förlorar vi i stringens. Den hypotetisk-deduktiva metoden tillåter helt enkelt för mycket eftersom den inte kan diskriminera mellan olika hypoteser

som är förenliga med evidensen. En metod som inte kan rangordna hypoteserna ”svensk arbetslöshet är en följd av att svenskar är slö, slappa, loja och likgiltiga” och ”svensk arbetslöshet är en följd av globalisering, teknologisk utveckling och ekonomisk politik” är helt enkelt ingen bra metod.

Abduktionen däremot kan rangordna konkurrerande hypoteser och kan tackla Duhem-Quine problemet betydligt lättare än den hypotetisk-deduktiva metoden. Abduktionen uppmanar oss att se bortom de enskilda hypotesernas egenskaper och logiska implikationer för att också bedöma och rangordna hypotesernas förklaringskraft. På så sätt är abduktionen ett mer kraftfullt verktyg än den hypotetisk-deduktiva metoden. Detta rimmar också väl med uppfattningen att abduktionen inte bara är en *justifikationslogik* utan också en *upptäcktslogik*.

Att det föreligger en trade-off mellan ampliatio och epistemisk tillförsikt beror klart på att det finns en slags risk i all ampliatio och ju mer risk vi är beredda att ta, ju mindre epistemisk säkerhet får vi leva med. Vi vet mer, men är mindre säkra på det vi vet. Vill vi vara mer säkra på den kunskap vi har tvingas vi avstå från ny kunskap och dess medföljande risker.

Nu tror jag att den bästa metoden för att hitta en balans mellan epistemisk säkerhet och ampliatio är användande av abduktion i stället för induktion och deduktion.

Peirce och abduktionens upptäcktslogiska status

Abduktion bygger som vi sett på att vi gör en slutledning utifrån det faktum att om en viss hypotes skulle *förklara* hypotesen så är hypotesen *sann*. Generellt gäller givetvis att flera olika hypoteser skulle kunna förklara evidensen, så man måste innan man kan göra slutledningen kunna utesluta de andra hypoteserna. Schematiskt skulle vi kunna beskriva abduktionen som:

X är en samling data/fakta/observationer
Hypotesen H förklarar X
Ingen annan hypotes förklarar X lika väl som H

H är därför troligen sann

Här framgår också abduktionens karaktär av kontrastförklaring. Olika hypoteser kontrasteras mot varandra och den som givet vår bakgrundkunskap bäst förklarar ”varför X?” väljs som den troliga (eftersom vi inte har en deduktiv argumentation) förklaringen. Att det inte kan handla om annat än sannolikhet/probabilitet/trolighet/rimlighet framgår redan ur Peirces beskrivning av skillnaden mellan de olika slutledningsformerna (CP 2.623):

Deduktion Regel: Alla böner från dessa säckar är vita
 Fall: Dessa böner är från denna säck

Resultat: Dessa bönor är vita

Induktion Fall: Dessa bönor är från denna säck
Resultat: Dessa bönor är vita

Regel: Alla bönor från denna säck är vita

Abduktion Regel: Alla bönor från denna säck är vita
Resultat: Dessa bönor är vita

Fall: Dessa bönor är från denna säck

I deduktionen är det uppenbart att konklusionen är nödvändig, medan i induktionen och abduktionen konklusionerna endast är rimliga/troliga/sannolika. Detta är priset vi får betala för att våga oss utanför den rena logiska formens säkerhet.

Peirce betonar abduktionens pragmatiska karaktär: ”våra slutledningar [har inte] anspråk på att vara mer än på det här viset erfarenhetsmässiga och provisoriska” och ”slutsatsen är giltig så långt vår möjliga erfarenhet sträcker sig, och det är ju också allt vi påstår” (PoK s 162). Detta utesluter dock inte dess karaktär av logisk slutledning (PoK s 237):

Vi måste hålla i minnet att abduktionen visserligen blott i ringa grad belamras av logiska regler, men att den icke desto mindre innebär logisk slutledning, låt vara att den framlägger sina slutsatser endast i form av problem eller gissningar men dock i fullständig och otvetydig logisk form ... Den abduktiva slutledningsformen är följaktligen denna:

- (Z) Ett överraskande faktum, C, observeras;
 men om A vore sant, så skulle C vara en självklarhet;
 följaktligen finns det skäl att anta att A är sant

Redan i sina *Harvard Lectures* (1865) gör Peirce klart att abduktionen handlar om slutledning till en förklaring: ”We find that light gives certain peculiar fringes. Required an explanation of the fact. We reflect that ether waves would give the same fringes. We have therefore only to suppose that light is ether waves and the marvel is explained” (W 1.267) Här framgår också hur slutledningen sker via analogi.

Runt sekelskiftet 1900 karakteriserar Peirce abduktion (retroduktion) som ”the first starting of a hypothesis and the entertaining of it, whether as a simple interrogation or with any degree of confidence” (CP 6.525) Abduktion består i “studying facts and devising a theory to explain them.” (CP 5.145)

Vetenskapsteoretikern N R Hanson föreslog redan 1958 att Peirces idéer om abduktionen sådana de framgår i (Z) skulle kunna användas som en upptäcktslogik. Även om man kan invända att eftersom A redan ingår i premisserna, så förklarar schemat inte i sig hur man kom fram till A, kan man som Hanson (1961) hävda att abduktionen ger goda skäl att föra fram en hypotes som potentiell förklaring. Även om abduktionen inte skulle anses bevisa riktigheten av en hypotes så kan den ge *prima facie* skäl att betrakta den som en möjlig förklaring och fortsätta undersöka den. Detta är en *svag* variant av abduktion, där abduktionen endast pretenderar på att vara ett sätt att upptäcka hypoteser som är värda att undersöka utan att för den skull i sig ge ett berättigande eller stöd för hypotesen. För det senare krävs fler experiment och tester eller kraftfullare analys innan vi kan avgöra hur pass hållbar hypotesen är.

Paavola (2004) har argumenterat för att abduktionen likväl kan ses som en upptäcktslogik om man skiljer på definitions- och strategiska regler. När vi spelar schack räcker det inte att känna till vilka drag man får göra, för att spela bra. Likaså i slutlednings- och förklaringspelet. Genom att anamma en speciell strategi för abduktionen – att söka förklaringar som är förenliga med våra bakgrundskunskaper – kan vi inskränka vårt sökande efter bra förklaringar till en mindre mängd alternativ än om vi började från ”scratch”. Givet en sådan strategi behöver jag inte ödsla tid på hypotesen att jag skulle skriva den här essän för att jag får 1 miljard kronor när den publiceras. En sådan hypotes passerar helt enkelt inte mitt ”strategifilter”. I detta avseende gör abduktionsfiltret det lättare att upptäcka sanna/sannolika/bra hypoteser eftersom jag inte behöver lägga tid och energi på vilda och långsökta hypoteser.

Det kan ju också vara så att även om A finns med i premisserna så har vi inte förrän vid abduktionen blivit varse hur A hänger samman med C. Liksom analogi kan abduktionen möjliggöra “to discover things before hidden” (P 1889:15). Peirce (CP 5.181) skriver: ”It is true that the different elements of the hypothesis were in our minds before; but it is the idea of putting together what we had never before dreamed of putting together which flashes the new suggestion before our contemplation.” Malthus befolkningsteori var välkänd redan sedan slutet på 1700-talet men det var först när Darwin såg hur den kunde förklara det naturliga urvalet som den kom att spela en avgörande roll i vår uppfattning om evolutionen. Vem skulle vilja påstå att Darwin inte ”upptäckte” den moderna evolutionsteorin bara för att Malthus teori redan var känd? En upptäckt kan ju ibland vara att se ett mönster, hur delar passar i ett sort pussel – även om pusselbitarna var för sig redan finns där.

Urvalskriterier

Hur ser då en bra abduktion ut? Ett krav som är naturligt för en kritisk-realist att ställa är att den upprättar en kausal relation mellan explanans och explanandum. Att säga att H är den bästa förklaringen till X är samtidigt att säga att - av de hypoteser vi jämför - den kausala historia H målar upp stämmer bäst överens

med vår grundkunskap. Återigen uppmärksammas vi på förklaringens kontrastiva karaktär. Det går inte att avgöra vilken av flera potentiella förklaringar som är bäst utan att ta hänsyn till relevant grundkunskap.

Självklart finns flera andra karakteristika som brukar nämnas när man försöker beskriva bra förklaringar: konsiliens, djup, enkelhet, sparsamhet, precision. Men även om dessa många gånger är önskvärda är de inte som jag ser det självklara eller ens avgörande för vår bedömning av de potentiella förklaringarna. I stor utsträckning är de pragmatiska och domänspecifika till sin karaktär.

Om förklaringskraft (enkelhet, unification, koherens) har med sanning att göra måste argumenteras. Enkelhet, unification m m *kan* utgöra kriterium för teorival, men *behöver* inte göra det. Dessa kriterier uttrycker främst enskilda forskares mer eller mindre idiosynkratiska preferenser för den ena eller andra egenskapen hos en teori/hypotes. *Ceteris paribus* är det över lag klart att föredra den enklare (unifierade etc) teorin/hypotesen A framför den mer komplicerade, men i övrigt helt ekvivalenta, B. Detta går att försvara utifrån rent tanke- och kognitionsekonomiska eller estetiska överväganden. Självklart vill vi ha så mycket ”bang for the buck” som möjligt. Men det går inte *a priori* att hävda att enklare teorier/hypoteser skulle vara sannare/sannolikare/bättre än sina rivaler. Detta till trots ligger ett sådant antagande till grund för samhällsvetenskapens och historiska studiers ekonometriska kurvpasande, där man ju faktiskt utgår från att den enklaste kurvan ger den bästa prediktionen eller bilden av det verkliga förhållandet. På grund av mät- och observationsfel är väl den enda rimliga ansatsen att på förhand bestämma den eftersökta enkelhetsnivån (för att undvika ”overfit”) och sedan bland teorier som uppfyller kravet söka den som är närmst observationsdatan.

En bra förklaring ska uppfylla vissa pragmatiska kriterier.

Den ska vara *relevant* (att jorden snurrar runt solen är ingen förklaring till varför USA skickade trupper till Irak).

Om fenomenet vi vill förklara ska kunna gå att förändra är det nödvändigt att förklaringen på något sätt lyfter fram en faktor vi kan manipulera eller *kontrollera* (om vi vill minska antalet dödsolyckor i trafiken är hastigheten antagligen är bättre faktor att peka på än att människor är av kött och blod, eftersom den förra faktorn går att ändra på till skillnad från den andra). Detta är också ett av skälen till varför kausala förklaringar är att föredra. Kausala faktorer går att påverka genom intervention och givet att förklaringar söks för att förändra saker och ting är kausala förklaringar att föredra.

En förklaring som gäller inte bara under mycket specifika omständigheter utan också har en egenskap som man i spelteorin kallar ”trembling hand”- stabilitet är i regel önskvärt. *Stabilitet* innebär att en valid förklaring till ett fenomen som gäller även i grannskapet av de undersökta värdena i regel är att föredra framför en förklaring som inte tillåter att vi darrar på handen.

Inferens till den bästa förklaringen – en fråga om att vara bäst eller sannolik?

Är verkligen den bästa förklaringen alltid sann? Eller är det snarare så att den sanna (mest sannolika) förklaringen också är den bästa? Varför skulle den bästa förklaringen också vara sann/trolig/sannolik? För att kunna besvara dessa frågor måste vi titta närmre på vad för slags slutledning abduktion egentligen är.

Peter Lipton skiljer i sin *Inference to the Best Explanation* (2004) mellan två olika typer av abduktiv inferens. Den första typen – *inference to the likeliest potential explanation* – innebär att vi sluter oss till en förklarings sanningshalt utifrån det faktum att förklaringen är den mest sannolika förklaringen. Det avgörande är här hur sannolik hypotesen H är. Om den är mycket sannolik drar vi slutsatsen att den är sann/riktig. Om den däremot endast har en låg grad av sannolikhet drar vi slutsatsen att den inte är sann/riktig. Den andra typen – *inference to the loveliest potential explanation* – innebär att den bästa förklaringen sannolikt också är sann/riktig. Det avgörande är här hur pass bra hypotesen H är, hur mycket och djupt den ökar vår förståelse av det vi vill ha förklarat. Om den är väldigt bra drar vi slutsatsen att den också är mycket sannolik.

Den första typen reducerar abduktionen till en fråga om sannolikhet och gör abduktionen snarlik *bayesianism*. Om vi vill ha en slutledningsmodell som ska beskriva hur vi kommer fram till att en förklaring är mer sannolik än de andra hjälper det oss föga att vi väljer den mest sannolika förklaringen. Förklarings epistemiska beviskraft kan då bara avgöras av *externa* överväganden (hypotesens sannolikhet). Detta är dock inte den väg vi ska välja.

Ett av problemen med bayesianismen är att den inte ger någon vägledning till hur vi upptäcker de hypoteser vi arbetar med. Deskriptivt ger bayesianismen inte heller någon psykologiskt trovärdig bild av hur människor hanterar sannolikheter. Bayes regel formulerar endast ett samband mellan sannolikheter *ex ante* ($P(H)$) och sannolikheter *ex post* ($P(H/e)$) och säger inget om hur vi etablerat hypotesen H eller dess sannolikhet *ex ante* (även om Bayes regel ger oss möjligheter att ur dem beräkna de konditionala sannolikheterna $P(H/e)$), eller *vilket* evidens e vi ska titta efter för att kunna börja kalkylera revisionerna av sannolikheterna. Just här är abduktionen överlägsen, eftersom den pekar ut vilket evidens som skulle vara epistemiskt relevant för H utifrån överväganden om huruvida evidensen skulle *förklara* H! I sökandet efter en förklaring av evidensen hjälper abduktionen oss att konstruera alternativa förklaringshypoteser.

Ett annat problem för bayesianismen är att den måste uppfatta konjunktionen av flera observationer som mindre sannolik än summan av de olika observationernas sannolikheter var för sig. En observation måste minska sannolikheten av den andra observationen. Detta är inte helt lätt att köpa. Varför skulle observerandet av att grannen köpt en svart bil minska sannolikheten för att jag ska observera att min fru skaffat sig en vit blus? Om jag konstaterar att grannens bil är svart, konfirmerar detta hypotesen att alla bilar är svarta, och

man kan kanske tänka sig att det också konfirmerar antagandet att nästa icke-svarta ting jag kommer att observera inte är en bil. Men utanför kontrollerade experiment är det svårt att se det rimliga i ett sådant antagande.

Bayes teorem sätter gränser för vad som kan betraktas som rationella försanthållandet. Även om de ibland kan vara för vaga eller starka är Bayes teorem i normalfallet inte *per se* oförenligt med abduktion. Som jag ser det är dock abduktionen det övergripande förhållningssättet och förklarar hur vi kommer fram till de försanthållanden vi har.

Genom att i stället uppfatta abduktionen i termer av den andra typen blir det *interna* överväganden om sannhet/riktighet som faller avgörandet. Vi vill så att säga visa på vilka drag i en förklaring som gör att vi också tro att det är den mest sannolika. Sannolikheten bedöms utifrån förklaringsövervägandena, inte tvärsom. Förklaringsövervägandena leder oss till en uppfattning om vad som är sant. Har vi olika men förenliga förklaringar kan vi ha flera förklaringar som är bäst. Är förklaringarna i alternativmängden däremot konkurrerande kan vi bara ha ett alternativ som det bästa. Och är den bästa inte bra nog får vi givetvis förbli obestämde.

Bayesiansimen är ytterligare ett försök att förvandla induktion till deduktion. Paradigmet för många filosofer som har arbetat med härledningslogik har varit studiet av deduktiva regler. Detta har medfört ett slags jakt på "den vises sten" bestående av syntaktiskt och ämnesneutrala (abstrakta) redogörelser för alla *prima facie* rationella, transparenta och algoritmiska härledningsregler, vars efterföljande bara skulle vara en enkel fråga om att förstå deras logiska form. Kruxet är bara att det inte existerar någon sådan "den vises sten". Bayesianismen är i grunden ytterligare ett försök i denna riktning, som förutom deduktiv koherens också kräver probabilistisk koherens i individens grad av tilltro. Men den blir därför också lika tom på innehåll som formallogiken eftersom den – till skillnad från abduktion och andra varianter av ampliativ och kontextberoende induktion – inte har något substantiellt att säga om *vad* individen ska härleda eller försanthålla. Att som utgångspunkt antar att vetenskapligt tänkande och slutledningar i grunden bara handlar om att uppdatera försanthållanden (H) i ljuset av ny information (e) för ingen vart. Vetenskap handlar självklart också – och mer fundamentalt – om att vara med om att skapa och vidmakthålla dessa försanthållanden som vi kallar vetenskap.

Bayesianismens kontextlöshet är problematisk, eftersom *de facto* kontexten visar hur vi låter förklaringsöverväganden inverka på våra slutledningar och kan avgöra vad som är relevanta förklaringar och hur alternativa förklaringar ska rangordnas. Vad som är bra förklaringar och/eller sannolika förklaringar hänger intimt samman med den specifika och relevanta information som finns tillgänglig i varje specifikt fall. Det är utifrån sådana specifika, konkreta och kontextuella fall som vi eventuellt kan abstrahera fram gemensamma slutledningsmönster. Men vi kan inte börja *ab ovo* från allmänna, axiomatiska överväganden av deduktiv logisk karaktär. Som abduktionen visar är det först

inom ramen för den kontextuellt situerade problematiken som vi också kan få svar på frågan hur sannolik den bästa förklaringen är. Att som bayesianismen – utifrån deduktiv-logiska sannolikhetsöverväganden – vända upp och ner på relationen mellan goda och sannolika förklaringar, är inte fruktbart.

Utifrån dessa överväganden kan man klart ställa sig frågan om man skulle kunna tänka sig att kombinera abduktion och bayesianism? Även om några forskare ställt sig (svagt) positiva till detta (varav Lipton kanske är den mest namnkunnige) är jag tveksam.

Bayesianismen i sin strikta form ger inte utrymme för att lyfta ut sannolikheten av slutsatsen av en probabilistisk argumentation och exempelvis säga att vi på grundval av ett fört resonemang kan acceptera en hypotes som riktig på basis av vår evidens. Bayesianismen har egentligen ingen *acceptansregel*, utan säger bara att en sannolikhet *ex post* har ett visst värde och att detta värde utifrån ny evidens kan uppdateras med hjälp av ytterligare konditionalisering via Bayes teorem. Abduktionen däremot har en acceptansregel – den auktoriserar acceptande av en hypotes om denna är den bästa förklaringen av evidensen. Abduktion är inte *bara* en fråga om *teorikonfirmation* – en jämförelse mellan $p(H)$ och $p(H|e)$ – utan också ett normativt ställningstagande som innebär att vi tar ställning till vad vi utifrån evidens och förnuft är berättigade att hålla för sant. Och detta sträcker sig ofta långt utöver vad en jämförelse mellan $p(H)$ och $p(H|e)$ kan ge. Konfirmation är varken nödvändigt eller tillräckligt för vad vi ska uppfatta som förnuftsensliga ställningstaganden. Att sannolikheten för H har ökat medför *per se* inte att det skulle vara förnuftsensligt att tro på H (som våra demonexempel visat). Och vetenskapshistorien har en uppsjö av exempel som visar att även om H inte kunnat konfirmeras av ett evidens, kan det likväl vara förnuftsensligt att hålla fast vid H i brist på hållbara alternativ.

Bayesianismen är inte heller ampliativ i någon substantiell mening. Det enda den säger är att vi ska uppdatera våra grader av tilltro på ett speciellt koherent sätt. Abduktionen å sin sida levererar nya hypoteser och teorier som informations- och innehållsmässigt går utöver de evidens som gav upphov till dem, och som utökar vår kunskap och förståelse.

Bayesianismen är som vi sett väldigt flexibel, åtminstone i den meningen att den inom ramen för kravet på koherens och specifik konditionalisering i princip accepterar vilka som helst sannolikhetskattningar. I mitt tycke är bayesianismen *för* flexibel och att inbegripa abduktion i bayesianismen genom att – som några föreslagit – låta förklaringsöverväganden påverka sannolikheter *ex ante* är till intet förpliktigande. Däremot skulle man kanske kunna tänka sig att låta förklaringsöverväganden spela en normativ roll vid fastställandet av sannolikheter *ex ante*, men det är tveksamt om bayesianer skulle acceptera detta, eftersom det skulle innebära att det inte längre bara är sannolikheter som bestämmer vad som utgör förnuftsensliga begränsningar på våra försanthållanden.

En annan möjlighet som övervägts är att låta förklaringsöverväganden inverka på hypotesens förklaringsvärde i förhållande till evidensen – likelihoodvärdet $p(e|H)$. Men ur deskriptiv synpunkt är det tveksamt – som vårt exempel med ”base-rate” felslutet visat – om det verkligen går att hävda att den bästa förklaringen verkligen är den som har störst likelihood. Åter igen är det tveksamt om detta skulle gå att ackommodera inom den bayesianska ramen. Ur bayesiansk synpunkt är det bara fråga om hur de redan existerande sannolikheterna ändras givet den nya evidensen. Abduktionens inkorporerande av överväganden kring relevans och fruktbarhet går inte att få med i enkla sannolikhetskalkylens endimensionella överväganden.

Så min slutsats måste bli: om abduktion och bayesianism ska kunna förenas måste det innebära att en i hög grad modifierad variant av bayesianism om möjligt bistår abduktionen.

Abduktion och bakgrundskunskaper

Tillvägagångssättet är då att utifrån Garfinkels idéer kring förklaringar som i grunden kontrastiva och ställa varför-frågan i termer av ”varför x, snarare än y, z ...”. Våra substantiella bakgrundskunskaper – och detta är helt avgörande för abduktionens framgång – gör att vi på förhand kan utesluta en hel rad hypoteser som i det aktuella fallet irrelevanta. Detta gör att vi får en hanterlig mängd *potentiella* förklaringar att utgå från när vi sedan försöker välja den bästa av dessa potentiella förklaringar som den *verkliga* förklaringen. Självklart kommer vi inte undan krav på att verkligen överväga om den valda hypotesen är *tillräckligt bra* (om inte får vi skjuta upp vår bedömning tills vi har mer fakta i målet) och att vi verkligen tagit alla relevanta alternativhypoteser under övervägande. Men har vi gjort detta föreligger *prima facie* skäl att acceptera den valda hypotesen H. Även om vi aldrig kan vara absolut säkra på att hypotesen är sann, utgör likväl det faktum att en hypotes är den bästa förklaringen till evidensen ett giltigt skäl att acceptera den.

Här finns också en viktig *kumulativ* aspekt. När vi väl accepterat H som den bästa förklaringen kan vi inkorporera denna i vår bakgrundskunskap och försanthållanden, vilket möjliggör för oss att i framtiden kunna förbättra vår förståelse, göra bättre förutsägelser, hantera ny information m m. På så vis ökar också koherensen i våra bakgrundskunskaper när en ny pusselbit läggs till helheten. Att sluta sig till den bästa förklaringen H av vår evidens är ju liktydigt med att utifrån vår bakgrundskunskap identifiera tänkbara hypoteser och välja den som passar ihop med denna och därigenom ökar dess koherens ur förklaringsynpunkt.

Förklaring, underdeterminering och förståelse

Enligt den form av abduktion som jag förespråkar här utgör våra förklaringsöverväganden en vägvisare för våra slutledningar. Givet de evidens vi

har framför oss försöker vi sluta oss till vad som – givet att det är sant – bäst skulle förklara evidensen.

När jag just hemkommen från arbetet kommer in i köket och ser köksbordet nerspillt med sylt försöker jag förstå vad som har hänt. Jag vet att mina barn brukar komma hem tidigare än jag och att de till mellanmål föredrar smörgåsar med sylt. Jag sluter mig till att det är något av barnen som spillt sylten på bordet. Självklart skulle det kunna finnas andra förklaringar – kanske har syltmonstret varit framme, eller kanske är dolda kameran på besök – men jag drar slutsatsen om mina barn därför att - givet min bakgrundskunskap - detta är den bästa förklaringen. Det är självklart inte den enda tänkbara förklaringen. Även om premisserna - min beskrivning av situationen och min bakgrundskunskap – är sanna kan jag likväl göra en felaktig konklusion. Det handlar alltså inte om en deduktiv slutledning utan om en induktiv slutledning i bemärkelsen att vi rör oss med sannolikheter och inte om logisk sanning. Utfallet är alltid underdeterminerat. Informationen vi har om syltpreferenser och tider för barnens hemkomst efter skolan garanterar inte att vi bara har *en* förklaring. Kanske var det jag själv som spillde sylten vid frukosten och glömde torka upp.

Underdetermineringen innebär att för att ta ställning till vilken förklaring som är riktig/bäst räcker det inte med deduktion. Jag måste argumentera för att min slutledning är berättigad. Förhoppningsvis är den också sann. När jag drar slutsatsen om barnen gör jag det för att den i mitt tycke utgör den bästa förklaringen och för att jag också tror att det verkligen var så att något av barnen kom hem före mig och spillde sylt vid mellanmålet. Förсанthållandet följer av förklaringen. Jag kan inte både hävda att x är den bästa förklaringen och att icke-x är sann. Underdetermineringen tvingar oss att alltid *berättiga* vårt beslut om vad vi hävdar verkligen har ägt rum, eftersom det finns alternativa hypoteser som vi därigenom förkastar.

Syftet bakom förklaring är att förstå. När vi tycker oss ha fått en bra förklaring till den spillda sylten känns det som om vi förstår vad som hänt. ”Aha, var det därför köksbordet var nersmetat med sylt. Nu förstår jag!” Det bästa sättet att uppnå denna förståelse är i de flesta fall (det finns undantag) att använda sig av en kausal förklaringsmodell. Att förklara ett fenomen är att ge information om dess kausala historia och beskriva ”mekanismen” som länkar orsak till verkan. Att *förstå* något är enligt detta synsätt att ha kunskap om *orsaker*. För att förstå varför bilen stannade på motorvägen räcker det att veta att motorn skar. Att jag sen inte besitter bilmekanikerkunskaper och vet *varför* motorn skar gör inget. Att förstå kräver inte att du vet orsaken bakom orsaken. Det avgörande är att kausala förklaringar förklarar därför att de är kausala.

Intresse och alternativmängd

Nu är det ju dock inte all information om kausal historia som verkligen förklarar. Detta beror på att vi har olika *intressen* när vi ställer varför-frågor. En följd av att vi betonar kontraster är att vi aldrig förklarar en händelse *per se* utan

bara en *aspekt* av händelsen. Denna aspekt ges av det intenderade alternativmängden. När jag vill veta hur köksbordet kunnat bli nersmetat med sylt har jag ett annat alternativrum (sönerna, döttrarna, frun) än kemisten (adhesionskraft, viskositet, friktion). Våra intressen är olika och därför tar vi fasta på olika aspekter av händelsen.

Speciellt i öppna historiska sammanhang där den kausala kedjan är näst intill oändlig räcker fakta själv inte som specifikation av vad vi vill ha svar på. Här behöver vi också specificera en tydlig alternativmängd för att i någon rimlig mening kunna få ett relevant svar på frågan vi ställer.

Alternativmängden *fokuserar* vårt *intresse* runt en viss *aspekt* av det som vi vill förklara. Att förklara ”varför är bordet nersmetat med sylt?” kräver något annat än att förklara ”vem var det som smetade ner bordet med sylt?” För mig är det lätt att förklara varför jag föredrar Mozart framför Haydn, men desto svårare att försöka förklara varför jag över huvud tycker om Mozarts musik. I det förra fallet förklara jag varför jag föredrar X givet ”X eller Y” vilket underlättar förklarandet. I det senare fallet ska jag rätt av förklara varför jag föredrar X utan den specificerade alternativmängden, vilket försvårar förklarandet.

Genom att ställa kontrasterande frågor (”varför X snarare än Y?”) *specificeras* frågeställningen naturligt i riktning mot att ange en orsak som på något vis utgör en *skillnad* mellan den valda förklaringen och dess alternativ. Om jag som svar på frågan vem av barnen (David, Pelle, Erik) som spillde sylt anger förklaringen att det måste ha varit David eftersom han kommit hem tidigare från skolan, håller detta bar som en rimlig förklaring om Pelle och Erik inte också kommit hem tidigare. I regel bygger en förklaring av ”X snarare än Y” på att vi anger en orsak till X och en frånvaro av en motsvarande orsak till Y. Vi söker en differens, en orsak till X som Y saknar.

Potentiella och verkliga förklaringar

Vi kan självklart inte dra några slutsatser utifrån det faktum att en viss förklaring är en möjlig förklaring. Den måste också vara den bästa förklaringen i alternativmängden. Den bärande idén i vår abduktionsmodell är ju att förklaringsöverbäganden fungerar som en ledtråd för vår slutledning. Om den bästa förklaringen till att köksbordet är nersmetat med sylt, givet vår bakgrundskunskap och alternativa hypoteser, är att David spillt ut den, är vår slutsats att David spillt ut sylten. För att en *potentiell* förklaring också ska vara en *verklig* förklaring måste den vara sann eller i hög grad sannolik. Detta kan vi bara avgöra efter det att vi dragit vår slutsats. Vi sluter oss till att den bästa av de tillgängliga potentiella förklaringarna i alternativmängden är en verklig förklaring.

Nu kan detta låta som en herkulisk uppgift. Hur stor är alternativmängden? I normalfallet kanske ändå inte så stor eftersom man kan argumentera för att det bara är *troliga* alternativ som kommer i fråga. Allt som vi mot vår bakgrundskunskap bedömer som irrelevant kan läggas oss sidan (syltmonster

existerar inte, grannens barn kan inte ta sig in i vårt kök). Vi har alltså i praktiken ett tvåstegsförfarande: först sällar vi ut de irrelevanta alternativen, sedan väljer vi ur den kvarvarande mängden den bästa förklaringen.

Abduktion, sanning och realism

Abduktionen är i sig också ett försvar för vetenskaplig realism. Vi är realister därför att detta ger den bästa förklaringen till de framsteg vi ser inom vetenskapen. De andra förhållningssätten kan inte lika bra *förklara* vetenskapens utveckling, därför hävdar vi att realismens beskrivning av verkligheten också är den *sanna/riktiga*.

Att något är *sant* räcker uppenbarligen inte till för att *förklara* något. En hypotes måste uppfylla ytterligare krav för att kunna räknas som en förklaring. Dessa krav kan uppfyllas mer eller mindre och måste på något vis vägas samman när man ska komma fram till vilken förklaring som är den bästa. Som jag tidigare argumenterat är det dock bara sanna hypoteser som kan ge den utökade förståelse som är målet för förklaringar. Avgörande för uppnående av denna förståelse är relevanta kausala historier. Och som vi också konstaterat beror relevans på bl a intressen, aspekter och historisk kontingens. Det som är en förklaring för en vit aristokrat på 1500-talet utgör inte nödvändigtvis en förklaring för en afro-amerikansk professor på 2000-talet.

När vi ska ta ställning till vilken av rivaliserande hypoteser som är riktig brukar vi söka mer data för att kunna ta ställning. Ofta är det så att även om de nya datan är förenliga med flera hypoteser, är det kanske bara en hypotes som kan förklara dem. Här hjälper oss abduktionen att dra rätt slutsats utifrån förklaringsöverbäganden som möjliggör ett eliminerande av ett antal hypoteser.

När vi utgår från kontrastiva hypoteser är det klart att en faktor som är gemensam för flera av hypoteserna mycket väl kan *orsaka* företeelsen vi studerar, men den hjälper oss inte att *förklara*. Om både Pelle och David är hemma från skolan är detta ingen hjälp i förklaringen att det var David och inte Pelle som spillde sylten. Här ser vi hur fokuserande på att förklara kan hjälpa oss att hitta relevanta kausala orsaker. Att både David och Pelle är rödhåriga är inte – även om det är sant – relevant för att förklara syltspillandet. Genom att först söka en förklaring finner vi också en relevant orsak (Davids släpphänthet).

Att David är släpphänt kan *förklara* varför det var han snarare än Pelle som spillde sylt, men var det verkligen *orsaken* till att han spillde? Orsaken kanske var att Pelle knuffade till honom just när han skulle sylta sin macka. Här har vi en situation där det föreligger en differens mellan en potentiell förklaring och den verkliga orsaken och där det är lättare (eftersom vi inte direkt haft möjlighet att observera barnen) att tänka i förklarings- snarare än orsakstermer.

Abduktion som heuristik

I modern vetenskapsteori har vi ofta att välja mellan Skylla (deduktivism, positivism) och Karybdis (relativism, social konstruktivism). Som jag ser det är

abduktionen ett sätt att styra mellan dessa grund. Abduktionen fokuserar på den process varigenom vi producerar förklaringar och drar våra slutsatser, i stället för att bara se på de färdiga produkterna (vad *är* en förklaring, vad *är* en slutsats etc). Abduktionen är med andra ord mer av en *heuristisk* karaktär än de andra vetenskapsmetoderna, som snarare har en *epistemisk* karaktär. Ser vi abduktionen som en upptäcktslogik är detta föga förvånande. Ska vi upptäcka ett grundläggande mönster, en relation, eller en struktur, är vi mer hjälpta av antydningar och ledtrådar som pekar ut en viss riktning för vårt sökande, snarare än att få en lista på kriterier som mönstret – när det väl är upptäckt – måste uppfylla för att räknas som säker eller vetenskaplig kunskap.

I historievetenskapen ställs vi ofta inför ett liknande problem. Flera olika hypoteser ställs upp som förklaringar till den svenska industrialiseringen, den amerikanska revolutionen eller första världskrigets utbrott. En del förklaringar är förenliga med varandra – både skottet i Sarajevo och tysk militarism kan vara förklaringar till första världskrigets utbrott. Det räcker inte med att förklaringar är olika för att de ska utesluta varandra.

Framför allt Kim (1988, 1989) har försökt analysera de situationer då olika förklaringar utesluter varandra. När vi har kontrastförklaringar har vi en sådan situation. Antingen var det Pelle, Erik eller David som spillde sylt. Om vi kommer fram till att förklaringen ”David spillde sylten” är korrekt, utesluter det att ”Erik spillde sylten” eller ”Pelle spillde sylten” är korrekta.

En annan vanlig typ av uteslutningssituation är när vi bedömer en förklaring som *icke-relevant* vad avser förklaringsobjektet. Huruvida en förklaring är relevant eller inte kan i normalfallet ofta indikeras av den alternativmängd man laborerar med i det specifika fallet. Kontrasterna (”fact and foil”) säger oss något om vad för slags svar vi förväntar oss på den erotetiska varför-frågan. Om det föreslagna alternativet avviker från ”mönstret” bedöms det i regel som irrelevant.

Den deduktiva modellen ger varken nödvändiga eller tillräckliga villkor för en förklaring. Speciellt i historievetenskaperna har vi sällan förklaringar som lever upp till de krav som den deduktiva modellen (Hempel-Oppenheimer) ställer upp.

Även om vi utifrån min skugga kan räkna ut hur lång jag är, är det knappast rimligt att säga att skuggan förklarar varför jag har den längd jag har. *Kausal relevans* är uppenbarligen ett avgörande kriterium för att vi ska kunna tala om förklaring. Problemet för den deduktiva metoden är bara att det är utsiktslöst att försöka fånga den dimensionen enbart med hjälp av logik.

Kontrastförklaringar igen

Våra liv är fulla av förklaringar. Både till vardags och när vi ägnar oss åt vetenskap. Barnets fråga ”varför måste jag gå i skolan idag” och professors fråga ”varför bestämde sig USA:s regering för att invadera Irak” föranleder oss att ge förklaringar. För det mesta är dessa okomplicerade och föranleder inte

speciellt mycket reflektion. Men ibland, speciellt när flera olika – och kanske oförenliga – förklaringar presenteras, ställer förklaringar till problem. Även om vi – vilket långt ifrån är säkert – är överens om vad en förklaring *är*, kan det vara komplicerat att välja. Vilka förklaringar är mest trovärdiga? Vilka förklaringar är acceptabla och vilka kan förkastas? Vilka förklaringar ger verkligen svar på det vi frågar om och vilka brister i något avseende? Vilken är bäst?

Vilka förklaringar vi får beror många gånger på hur vi ställer frågan. Om frågan ställs som ”Varför slog Pelle Lisa?”, ”Varför slog *Pelle* Lisa?” ”Varför slog *Pelle Lisa*?” eller ”Varför slog Pelle Lisa?” ger upphov till olika förklaringar. Detta beror på att betoningen (eller frånvaron av betoning) indikerar att frågeställaren tänker sig olika möjliga alternativ som svar. På den första frågan är *alternativmängden* kanske (slå, kyssa, skälla ut), på den andra (Bo, Anders, Pelle), på den tredje (Marie, Kristina, Lisa) och på den fjärde (Pelle är kolerisk, Lisa är provokativ, det senmoderna samhället försvagar jagkontrollen). Självklart ger dessa frågor i normalfallet upphov till olika förklaringar. Många gånger förkastar vi en förklaring därför att den helt enkelt anses tillhöra fel alternativmängd. Att vi lever i en senmodern värld är inget relevant svar på frågan ”varför slog *Pelle* Lisa?” Ofta är det nödvändigt att en viss kontrast lyfts fram i frågeställningen för att man ska tycka att förklaringen är lyckad. När vi undrar varför kvinnor har så låga löner är den underförstådda kontrasten männens löner, och att som förklaring åberopa det faktum att kvinnor ofta är kortare än män duger inte långt eftersom denna ”förklaring” inte ger en förklaring till den avsedda kontrasten.

Att fråga ”varför?” är som vi tidigare sett egentligen alltid att fråga ”varför X snarare än Y?” En förklaring kan bara ges mot bakgrund av en speciell alternativmängd. Om vi således har två olika förklaringar kan vi jämföra dem genom att se på deras alternativmängd. Vad som *prima facie* förefaller vara helt oförenliga förklaringar är i själva verket olika alternativmängder. Detta understryker också vikten av vad vi skulle kunna kalla *förklaringsrelativitet* och att förklaringar oftast inte är förklaringar av X eller Y utan snarare ”varför X snarare än Z, P, V?” eller ”varför Y snarare än M, N, Q?” Denna förklaringsrelativitet är också en av anledningarna till att abduktionen fungerar. Genom att förklaringar förutsätter alternativrum reduceras realiter mängden av kontrasterande hypoteser som ska jämföras och det blir lättare att vaska fram den bästa förklaringen.

Kontraster som vägledning för slutledning

Om vi vill förklara varför jag var på middag hos Bergströms i går kväll kan kännedom om att jag fått en inbjudan dit vara tillräcklig. Men om jag även blivit inbjuden till Nilssons är så inte fallet. Lika lite utgör information om att jag inte blivit inbjuden av Perssons en förklaring till att jag gick till Bergströms snarare än Nilssons.

De kontraster som vi väljer att utgå från är sällan tillfälliga, utan beror i stor utsträckning på vår vilja att avgöra vilken av de konkurrerande hypoteserna som är riktig. En felaktigt vald alternativmängd kan omöjliggöra detta. ”Differensmetoden” förutsätter om den ska fungera väl att de alternativ man laborerar med har en till stora delar gemensam kausal historia. X och Y har en gemensam historia a, b, c men skiljer sig åt vad gäller d, vilket förklarar varför X är fallet men inte Y. Både Pelle och David kom hem till mellanmålet och åt sylt, men till skillnad från Pelle är David en slarver, vilket förklarar varför det var han som spillde sylt och inte Pelle.

Resonemanget ovan är ett exempel på *kausal triangulering*, som vi använder oss av i abduktionen. Alternativmängden hjälper oss att välja ut den del av den kausala historien som kan ge en bra/relevant/adekvat förklaring. Det kriterium vi använder för att värdera en förklarings kvalitet är samma som vi använder när vi drar slutsatser om vad som är orsaken till det vi söker en förklaring av. Davids slarvighet är *orsaken* till att han spillde sylt, och det är också den förklaringsmässiga differens som ger den bästa *förklaringen* till att köksbordet var nersmetat med sylt. Genom att fokusera på de kontrasterande förklaringshypoteserna i alternativmängden kan vi också få ledning för att dra kausala slutsatser.

I stället för att börja med en nästintill oändlig mängd hypoteser om orsaken till ett fenomen, reducerar kontrastiva förklaringsöverbärganden mängden till mer hanterliga proportioner. De hjälper också forskaren att välja ut lämpliga experiment och observationer att göra i syfte att fastställa kausala förhållanden. Jag observerar mina barns fingerfärdighet och kommer fram till att den *troligaste* förklaringen till att köksbordet är nersmetat med sylt är att David (och inte Pelle eller Erik) spillt sylt, eftersom detta är den *bästa* förklaringen av evidensen.

Även om differensmetoden är den typiska för de flesta förklarings- och slutledningsförsök vi gör är det inte den enda metoden. Ibland måste vi använda *överensstämmelsemetoden*, där vi försöker vi förklara en likhet snarare än en differens. Vi söker en orsak som producerar samma verkan i olika fall. Risker är förstås att vi kan ha flera orsaker som ger samma verkan. Om både A och B blir arbetslösa och det enda i övrigt de har gemensamt är att de är rödhåriga, kan metoden peka ut detta som orsaken till att de blev arbetslösa. I normalfallet leder oss våra bakgrundskunskaper till att dock bortse från denna irrelevanta likhet.

Varför bayesianism är en återvändsgränd

I detta avsnitt ska jag försöka fördjupa och klargöra varför jag inte tror att den bayesianska vetenskapsanalysen är rätt väg att gå för samhälls- och historievetenskaperna.

Oberoende av hur vi uppfattar sannolikheter – och teorierna är många – så finns det en gemensam förankring i sannolikhetskalkylen. Oberoende av hur vi tolkar sannolikheter kan sannolikhetskalkylen appliceras på nästan alla kända teorier. För att underlätta vårt ställningstagande till bayesianismen och kritiken av denna kan det vara lämpligt att ge några grunder i denna kalkyl.

I grunden handlar sannolikhetskalkylen om hur man utifrån kända sannolikheter kan beräkna ytterligare sannolikheter. Det finns några olika axiomatiseringar, men den jag ska ge här är den vanligaste och går tillbaka på Kolmogorov.

Den grundläggande formen är

$$p(A) = a$$

där A representerar ett påstående och a är ett tal. $p(A) = a$ utläses ”sannolikheten för A är a.”

I de sammanhang som vi diskuterar här gäller frågan oftast vilka sannolikheter vi kan tillskriva teorier och hypoteser (H) utifrån olika evidens (e). Därför har jag här valt att exemplifiera påståenden med symbolerna H och e.

Konditionala sannolikheter skrivs som

$$p(e | H) = a, \text{ vilket utläses ”sannolikheten av e givet H är a.”}$$

Definition 1: e och H är oberoende $\leftrightarrow p(e) = p(e | H)$.

Definition 2: e och H är ömsesidigt uteslutande om det är omöjligt för båda att vara sanna.

Axiom 1: $0 \leq p(e) \leq 1$.

$$0 \leq p(e | H) \leq 1.$$

Axiom 2: Om e är säker, är $p(e) = 1$.

Axiom 3: Om e och H är ömsesidigt uteslutande är

$$p(e \wedge H) = p(e) + p(H).$$

Med hjälp av dessa tre axiom kan vi härleda

Theorem 1: $p(e \wedge \neg e) = p(e) + p(\neg e)$.

Theorem 2: Om e och H är ekvivalenta, så är $p(e) = p(H)$.

Theorem 3: $p(e \vee H) = p(e) + p(H) - p(e \wedge H)$.

Följande axiom visar hur man beräknar konjunktioner av sannolikheter:

Axiom 4: $p(e \wedge H) = p(e)p(H | e)$

Om $p(e) \neq 0$ får vi

Theorem 4: $p(H | e) = p(e \wedge H)/p(e)$.

Från beskrivningen av oberoende i definition 1 får vi

Theorem 5: $p(e \wedge H) = p(e)p(H)$.

Om både $p(e)$ och $p(H)$ är skilda från noll, följer också från definition 1

Theorem 6: $p(e) = p(e | H) \leftrightarrow p(H) = p(H | e)$.

Nu kan vi – givet att $p(e) \neq 0$ – härleda

Theorem 7 (inversa sannolikhetslagen):

$$p(e | H) = [p(e)p(H | e)] / p(H), \text{ alternativt } (p(H) \neq 0)$$

$$p(H|e) = [p(H)p(e|H)] / p(e).$$

Teorem 8: (Bayes teorem):

$$p(e|H) = \frac{p(H|e)p(e)}{p(H|e)p(e) + p(H|\neg e)p(\neg e)} \text{ alternativt } (p(H) \neq 0)$$

$$p(H|e) = \frac{p(e|H)p(H)}{p(e|H)p(H) + p(e|\neg H)p(\neg H)}.$$

$p(H)$ är den sannolikhet *ex ante* som tillskrivs en viss hypotes H innan evidensen e tas med i beräkningen. $p(H|e)$ är sannolikheten *ex post*, sannolikheten efter det att evidensen e tagits med i övervägandena. Bayes teorem visar att sannolikheten *ex ante* ska multipliceras med skalfaktorn $p(e|H)/p(e)$ för att sannolikheten *ex post* ska erhållas. $p(e|H)$ är ett mått på hur sannolik evidensen är givet H . Om e följer deduktivt ur H antar den värdet 1, och om $\neg e$ följer ur H antar den värdet 0. Evidensen sägs *stödja* hypotesen i samma utsträckning som hypotesen *förutsäger* evidensen. Detta innebär att om $p(e)$ har ett mycket högt värde så får hypotesen inget större stöd av att bekräftas av evidensen. Om däremot hypotesen skulle förutsäga något ingen trodde var möjligt, skulle evidens för en sådan förutsägelse ge ett starkt stöd åt hypotesen.

Alla de sannolikheter som bayesianismen laborerar med utgår från en fond av det vi – efter Popper – kallat *bakgrundkunskaper*. Även vi ska följa denna konvention. De resonemang om enkla och konditionala sannolikheter som vi försker mot en given bakgrundskunskap.

Den i dag helt ledande skolbildningen inom bayesianismen är den subjektivistiska skolan, enligt vilken de sannolikheter som figurerar i sannolikhetskalkylen ska tolkas som subjektiva *grader av tilltro*. Vilken grad av tilltro människor har tänker man sig att man kan få kännedom om genom att ställa de inför olika slags vad. Metoden – som i grunden bygger på en spelanalogi mellan kunskapsformer och spel utvecklades av Ramsey, de Finetti och Savage – sägs visa att människors sannolikhetsbedömningar måste vara konsistenta i enlighet med sannolikhetskalkylen. Om inte, kan man upprätta så kallade *Dutch Books* mot de vars bedömningar är inkonsistenta, vilket innebär att de oavsett utfallet av vadet alltid kommer att förlora. Detta innebär också – eftersom vi ovan visat att Bayes teorem är deducerbart ur sannolikhetskalkylens axiom – att konditionalisering av sannolikheter *per se* är okontroversiellt (åtminstone i icke-deskriptiva mening).

Fortsättningsvis kommer vi – om inte annat anges – när det talas om bayesianism att avse den subjektivistiska versionen av bayesianism. Enligt denna har individer (inklusive forskare och filosofer) alltid sannolikhetsuppfattningar om företeelser och förhållanden runt omkring oss. Dessa sannolikhetsuppfattningar – grader av tilltro – utgör de sannolikheter *ex ante* som bayesianer sedan använder rätt upp och ner för att göra sina beräkningar av konditionala sannolikheter.

Precis som system och modeller baserad på den deduktiva logiken behöver enligt bayesianerna inte dessa sannolikheter *ex ante* berättigas. De är det bayesianska systemets axiom på vilka sedan berättigade konditionala sannolikheter kan beräknas och användas i konfirmations- eller annat syfte inom vetenskap och vardagsliv. Och det gör inte så mycket att dessa sannolikhetskattningar är så subjektiva och kanske därför väldigt spretiga. Genom konditionalisering mot bakgrund av ny evidens kommer de ändå att med tiden *konvergera* ("wash out") hävdar bayesianerna.

Mot denna bakgrund menar sig också bayesianer kunna lösa exempelvis Duhem-Quine-problemet genom att helt enkelt applicera Bayes formel. De antaganden vars sannolikheter *ex post* blir låga bör helt enkelt överges.

Bayes teorem

Bayes teorem är i sig inte kontroversiellt. Vad som däremot är kontroversiellt är det sätt på vilket bayesianer använder teoremet när man hävdar att det rätta sättet att *värdera* hypoteser är att använda de konditionala sannolikheter som evidensen ger upphov till i enlighet med Bayes teorem som våra nya sannolikheter *ex post*. Det stöd eller konfirmering som evidens ger brukar anges som (se Howson & Urbach 1993 s 117):

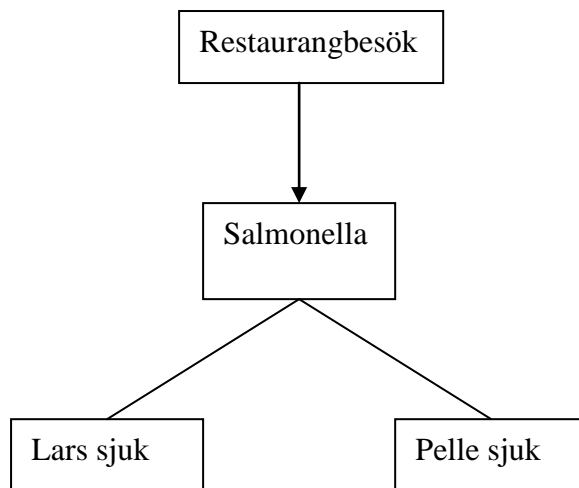
$$e \text{ stödjer } H \text{ då } p(H|e) - p(H) > 0$$

eller alternativt

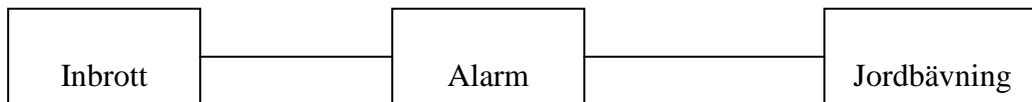
$$e \text{ stödjer } H \text{ då } p(e|H) / p(e|\neg H) > 1.$$

Eftersom en *likelihood* är sannolikheten för evidensen givet en hypotes, och hypoteser i vetenskapliga sammanhang ofta beskriver hur kausala relationer ser ut, blir $p(e|H)$ en naturlig beräkning. *Likelihoodkvoten* $p(e|H) / p(e|\neg H)$ anger vilken *inverkan* evidensen har på sannolikheten *ex post*.

Bayesianer använder sig också av begreppet *probabilistisk relevans*. Enligt detta är e och H probabilistiskt relevanta gentemot varandra endast om (jfr Teorem 6) $p(e) \neq p(e|H)$ och $p(H) \neq p(H|e)$. Probabilistisk relevans kan självklart påverkas av externa faktorer. Om C (Lars och Pelle åt mat på samma restaurang) föregår A (Lars blev magsjuk) och B (Pelle blev magsjuk) ändras $p(B|C)$ av vår kännedom av $p(A|C)$. Om däremot den gemensamma kausala faktorn D (salmonellabakterier i kycklingfilén) blir oberoende känd, blir A och C också oberoende av varandra. D fungerar då (se Reichenbach 1956 s 189) som *avskärmande faktor* ("screening off").



På motsatt vis kan två oberoende faktorer bli beroende via en tredje faktor. Även om det inte föreligger något uppenbart beroende mellan inbrott (A) och jordbävningar (B) kan de bli (probabilistiskt) beroende om ett vaktbolag exempelvis väljer att inte vidta åtgärder när de får ett inbrottslarm (C) eftersom det anser att sannolikheten för ett inbrott givet att en jordbävning ägt rum är liten.



När det gäller själva konditionaliseringen är det givet att den inte i någon strikt logisk mening är en valid härledning. Även om den är försvarbar i många sammanhang, är den inte universellt odiskutabel. Den förutsätter ju tidsdimensionellt inte bara att de skattade sannolikhetsvärdena är konsistenta, utan också att vår sannolikhetsfunktion $p(\bullet)$ är *tidskonsistent*, vilket ju långt ifrån är säkert. Jag kan mycket väl innan e instantierats ha ett värde på $p(H|e)$ och efter det att e inträffat ha ett annat värde på $p(H|e)$ (som då sammanfaller med $p(H)$ eftersom $e=1$) utan att detta verkar orimligt. I *strikt* bayesiansk tappning måste dock de två värdena på $p(H|e)$ sammanfalla (Howson 2000 är exempel på bayesian kritisk till denna rigida hållning).

Anledningen till att bayesianer håller fast vid sina sannolikheter *ex post* beror på att de menar att det är svårt att i vetenskapliga eller vardagliga sammanhang tänka sig att $\Delta e \rightarrow \Delta p(e|H)$. Om vi har en normal kortlek (H) är sannolikheten för att dra ett kort vilket som (e) – $p(e|H)$ – lika med 1/52. Denna sannolikhet

ändras inte av att vi t ex drar just hjärter ess ur kortleken. Där av slutsatsen att de två värdena ovan på $p(H|e)$ sammanfaller.

Bayesianismens påstådda fördelar

Varifrån kommer sannolikheterna *ex ante* ifrån? Om detta har bayesianer många olika idéer. Objektivistiska bayesianer som Carnap (1950) menar att de kommer ur språkets struktur, medan subjektivisterna som de Finetti (1937) griper dem ur luften. De flesta bayesianer menar dock att de inte behöver berättiga dem. De är vad de är. Punkt slut.

Bayesianismen är i mångt och mycket analog med den deduktiva logiken. Vad är det då som får bayesianer att tro att deras teori skulle klara sig bättre än den deduktiva logiken i vetenskapens värld?

Korb (2004) argumenterar för att bayesianismen är kvalitativ och semi-formell och därför applicerbar på informella logiska system (som vetenskap, icke-semantisk argumentation m m) på ett helt annat sätt än den deduktiva logiken.

Korb menar att likelihoodkvoten t ex kan ge en förklaring till varför man – som Popper brukade understryka – alltid ska utsätta hypoteser för *stränga test*. Intuitivt kan man beskriva ett strängt test som ett test där en hypotes som är falsk har låg sannolikhet att överleva. Antag att hypotesen förutsäger e (vilket gör $p(e|H)$ hög), medan konkurrenterna inte gör det (vilket gör $p(e|\neg H)$ låg). Likelihoodkvoten $p(e|H)/p(e|\neg H)$ blir då väldigt hög. Därför blir ett strängt test i hög grad konfirmerande om hypotesen passerar det och i hög grad falsifierande om det inte passerar det. Mot detta kan man väl invända att bayesianismen inte tar hänsyn till det faktum att man efter ny evidens kan få $\Delta p(e|H)/p(e|\neg H)$ i stället för $\Delta p(H)$.

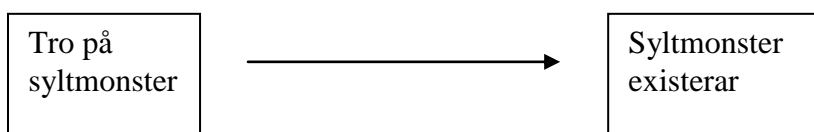
Korb anför också ett exempel som han menar visar på bayesianismens fördelar framför abduktion. Antag att

$$\begin{array}{l} p(e|H) \text{ är hög} \\ e \text{ observeras} \\ \hline p(H|e) \end{array}$$

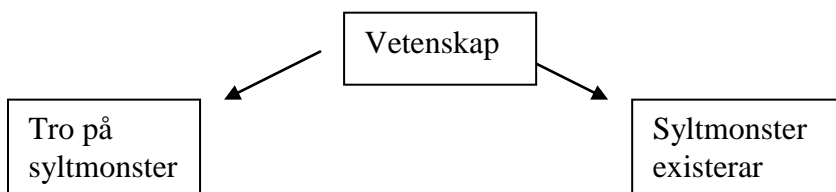
Detta är ett exempel på en abduktion som baseras på ”affirming the consequent” som Korb dock menar är otillfredsställande eftersom det kan finnas en uppsjö med alternativa hypoteser $\neg H$ som lika väl kan förklara e (och att abduktionen bara är rimlig om man antar att $p(e|\neg H)$ är låg). Enligt Korb undviker vi detta felsslut om vi i stället använt likelihoodkvoten $p(e|H)/p(e|\neg H)$ som kriterium på konfirmation. Enligt Korb är därför bayesianismen ett exempel på att ”affirming the consequent” inte är ett felsslut om man låter affirmationen bero på likelihoodkvoten $p(e|H)/p(e|\neg H)$ och inte som hos abduktionen enkom

på $p(e|H)$. ”It is only in Bayesian considerations of likelihood and prior that a principled distinction between good and bad applications of this form of reasoning has ever been made” (Korb a a s 13)

Korb menar att bayesianismen också kan tackla problemet med ”demonförklaringar”. Om vi antar att syltmonstret existerar (H) kan vi förklara sylten på bordet (e) – $p(e|H)$ blir hög. Oberoende av hur mycket ny sylt vi ser utspild på bordet så stärker detta bara förkättrade syltmonsteranhängare i riktigheten av deras monstertro och $p(H)$ ökar. Självklart är detta problematiskt för icke-monsteranhängare bland bayesianer. Korb menar att vetenskapligt mer hållbara hypoteser än syltmonster kan avskärma H, så att både e och H förklaras av denna vetenskapliga hypotes. I stället för



skulle vi få



där alltså vetenskapen både kan förklara varför vissa människor tro på syltmonster och varför det är orealistiskt att syltmonster skulle existera.

Om detta verkligen är en framkomlig strategi avhänger hur vi ställer oss till avskärningsmöjligheten. Men även om vi skulle acceptera den är det inte klart hur detta ska kunna beskrivas som en specifik bayesiansk strategi. Appellerandet till framtida vetenskapliga möjligheter går ju väsentligt utanför rena sannolikhetsöverväganden. För en icke-strikt bayesian som Korb är detta dock inte något problem eftersom – även om den bayesianska analysen skulle kunna anses bidra med mycket av vikt – ”the point of this bayesian analysis completed will be simply to provide a useful *framework* for identifying and dealing with inferential relations between statements, one which supplements and must be supplemented by considerable additional analytical resources” (Korb a a s 16).

Från korrelationer till kausalitet

De flesta vetenskapsteoretiker är överens om att korrelationer inte *obetingat* implicerar kausalitet. Frågan är då hur omfattande betingningen måste vara för

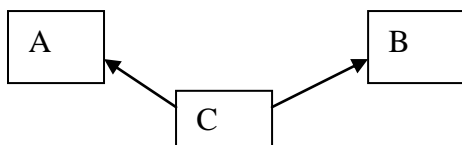
att vi ska kunna gå från korrelationer till kausala relationer. Utgångspunkten brukar vara Reichenbachs (1956 s 157) princip om *gemensam orsak*: ”if an improbable coincidence has occurred, ther must exist a common cause.” Enkelt uttryckt innebär denna att om det föreligger en korrelation mellan A och B så måste detta bero på antingen (1) $A \rightarrow B$ (2) $B \rightarrow A$ (3) $C \rightarrow A \wedge B$ (där C antingen är den enda gemensamma orsaken eller disjunktionen av alla möjliga gemensamma orsaker. Reichenbach borde ha tagit med även en fjärde möjlighet, nämligen att korrelationen mellan A och B också kan bero på en simultan kombination av (1) – (3)). Och om vi har ett system med tre variabler kan bara följande kausala relationer gälla (under förutsättning att paren (A,B) och (B,C) är direkt relaterade): 1) $A \rightarrow B \rightarrow C$ 2) $A \leftarrow B \leftarrow C$ 3) $A \leftarrow B \rightarrow C$ 4) $A \rightarrow B \leftarrow C$.

Reichenbach (s 159) kallar relation 3) *fork open toward the future* och 4) för *fork open toward the past* och menar att de stödjer olika konditionalt oberoende relationer. I fall 3) har vi $p(A|C) \neq p(A)$ men $p(A|C \wedge B) = p(A|B)$, och i fall 4) tvärtom $p(A|C) = p(A)$ och $p(A|C \wedge B) \neq p(A|B)$. För Korb och andra bayesianer är ”the recursive application of the principle over models with many variables ... very powerful” (Korb s 20). Men är den verkligen det?

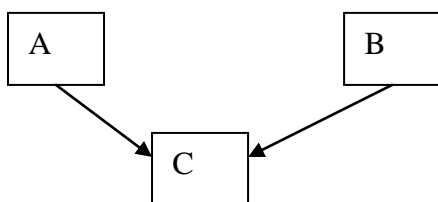
Det beror på hur vi specificerar ”models with many variables”. Om man som bayesianska nätverksteoretiker gör antaganden om ”faithfulness” (vilket enkelt uttryckt innebär att bayesianska kausala modeller antas vara trogna verkligheten och att kausala relationer kommer till uttryck i korrelationer), ”stability”, ”independencies” må så vara hänt. Med bara rätt antaganden om det kausala system vi studerar kan vi läsa av information om kausalitet från de bayesianska näten. Mindre är inte att vänta av axiomatiska, slutna modeller av artificiella slutna system. Men det finns ingen *generell* formel för hur man går från korrelationer till kausalitet!

En probabilistisk relation (korrelation) mellan A och B kan bero på:

(1) De har en gemensam orsak



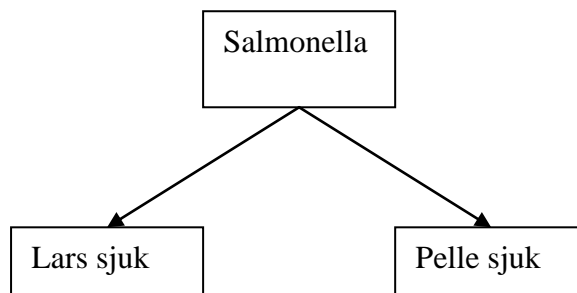
(2) De kan tillsammans ge upphov till en verkan



(3) När A och B tillhör olika system (populationer) som interagerar (mixas) kan det emergenta systemet (populationen) uppvisa egenskaper som inte uppfyller exempelvis Markovkravet (som Simpsons paradox).

Om vi har ett "integrerat" system är ofta delsystem inom dess ram koordinerade med varandra. Om vi (se Reynolds 1987) har en flock med fåglar, koordineras deras flygturer kanske på så vis att de enskilda fåglarna följer en slags beteenderegulering som säger att de ska upprätthålla ett visst avstånd till fåglar i dess närhet. När en fågel ändrad flygriktning följer andra efter och vi får en koordinerad flygning. Fåglarnas rörelser är koordinerade, men det existerar ingen "lokal" gemensam orsak, utan vi har ett slags "jämviktskorrelation" som bibehålls trots externa rubbningar. Även om vi kan ange föregående händelser i till koordinerade händelser i tidrummet är det antagligen mycket svårt att identifiera händelser som skärmar av fåglars tidsrumsliga lokalisering. Rent generellt kräver vi inte att korrelationer mellan delar av ett system kräver en förklaring i termer av en gemensam orsak bara för att delarna i jämvikt uppvisar hög grad av korrelation .

Om det däremot föreligger en rent spatial separation mellan mikrostatiska faktorer i olika makrostatiska enheter kan man möjligen hävda att vi har oberoende mellan de mikrostatiska faktorerna och att Reichenbachs principer håller. Ta vårt tidigare exempel med restaurangbesök:



Här är det ju uppenbart att den salmonella som smittar rent fysiskt delas upp i två separata processer (som förblir separata eftersom den salmonella som går in i Lars mage inte är samma salmonella som går in i Pelles mage) och att vi därigenom (givet den karakterisering vi har gett av de kausala processerna) har oberoende och att principen om gemensam orsak och den kausala Markovegenskapen håller. Arntzenius (Stanford Enc of Phil s 13) skriver:

The core truth of common cause principles thus in part relies on our choice as how to partition the world into unified and independent objects and quantities, and in part on the objective, temporally asymmetric, principles that lie at the foundation of the statistical mechanics.

(4) A och B kan råka ha samma temporala utveckling (jfr tidsserier och problemet med trendrensning), vilket gör att de kan uppvisa probabilistiskt beroende över tiden, utan att det egentligen föreligger någon kausal relation. Elliott Sober (1988) anför som exempel att priset på bröd i London och vattennivån i Venedigs kanaler båda har stigit samtidigt under en längre tid. Här är det svårt att tänka sig att det finns någon variabel som skulle skärma av dem från varandra, förutom tid. Om vi för in en tidsvariabel har vi en gemensam orsak och konditionalt oberoende ($p(A|B \wedge t) = p(A|t)$) uppnås. Ur kausal synpunkt är detta dock knappast tillfredsställande för en bayesiansk nätverkare eftersom tid är en variabel som vi inte ens i princip kan intervensera på. Och då kan det inte heller vara en variabel i en bayesiansk kausalmodell.

(5) A och B kan uppstå till följd av en probabilistisk orsak

I alla fem fallen kan vi felaktigt tro att det bakom en probabilistisk relation mellan A och B också föreligger en kausal relation mellan A och B, vilket som vi sett inte behöver vara fallet.

Hur ska bayesianer reagera på allt detta? Korb et al (2005, 11-12) tycker dels att de anförda exemplen ger uttryck för ett slags pedanteri och att lösningen på

the constraint-driven anomalies to [the Reichenbach principles] is to rewriting [the Reichenbach principles] so as to exclude them". Man får helt enkelt acceptera att bayesianska nätverk är ofullständiga: "So we agree that [the Reichenbach principles] has limits – namely the limits of our modeling and learning methods that determine what variables and arcs can be introduced into causal models.

Vår slutsats blir en annan. Probabilistiskt beroende mellan faktorer kan bero på andra saker än kausalitet. Bayesianska metoder fokuserar framför allt på fallet med gemensam orsak och försöker redogöra för hur vi ska hantera detta problem under antagande av att vi har ett underliggande *deterministiskt* system. De andra fallen behandlas sparsamt eller inte alls.

Problemet är i grunden att oberoende av vilka olika bayesianska ansatser som används så innebär modellspecifikationerna genomgående att vi måste *antas veta* att *alla* relevanta faktorer är representerade i modellen och att alla resttermer är genuina slumpstermer (som alltså inte bara beror på att vi för tillfället saknar information om dem). Och är det något korrelations- och ekonometriska studier visat så är det att vi sällan har den kunskapen eller uppfyller de villkoren! Våra slutna modeller är inte fullständiga representationer av vår öppna värld och därför kan vi i verkligheten sällan sluta oss till kausalitet från korrelationer. Orsaker kan öka sannolikheten av sina effekter, men *behöver* inte göra det. Cartwright (Monist 2001) formulerar den nödvändiga slutsatsen:

If you see a probabilistic dependence and are inclined to infer a causal connection from it, think hard. Consider the other possible reasons that that dependence might occur and eliminate them one by one. And when you are all done, remember – your conclusion is no more certain than your confidence that you have eliminated all the possible alternatives.

Det är ju också så att *kontextselektion* spelar stor roll. Ta följande exempel (jfr Collins 2000): Ada slår en frispark och bollen är på väg mot målet. Beda ställer sig i bollens väg och sedan ställer sig Cilla mellan Beda och målet. Beda får stopp på bollen, men kan man verkligen säga att hon förhindrade att det blev mål? Antagligen ja, men inte på samma sätt som om vi ersätter Cilla med en 50 m bred och 20 m hög vägg. Här blir svaret ett annat eftersom kontexten i vilken frågan ställs är annorlunda. Trots att varken Cilla eller väggen de facto ingriper förändra sannolikheten drastiskt. De svar du får på frågan vad som orsakade x beror på vilken kontext x relateras till, vilket är en av de centrala momenten i abduktionens kontrastförklaringar.

Från kausalitet till korrelationer

Ovan har vi sett att det inte är oproblematiskt att utifrån korrelationer dra slutsatser om kausalitet. Men även det omvända problemet föreligger.

Även om orsaker producerar eller ger upphov till sina verkningar så kan vi bara i *speciella fall* utgå från att sannolikheten av verkan V ökar när dess orsak O föreligger ($p(V|O) > p(V|\neg O)$). Problemet med bayesianismen är att den glömmer bort kvalifikationen ”i speciella fall”. Låt oss ta exemplet med Simpsons paradox för att åskådliggöra problemet:

(1) Antag att O = ta ett läkemedel, V = tillfrisknande, K = egenskapen att vara kvinna – då kan vi mycket väl ha följande sannolikheter

- (1) $p(V|O) > p(V|\neg O)$
- (2) $p(V|O \wedge K) < p(V|\neg O \wedge K)$
- (3) $p(V|O \wedge \neg K) < p(V|\neg O \wedge \neg K)$

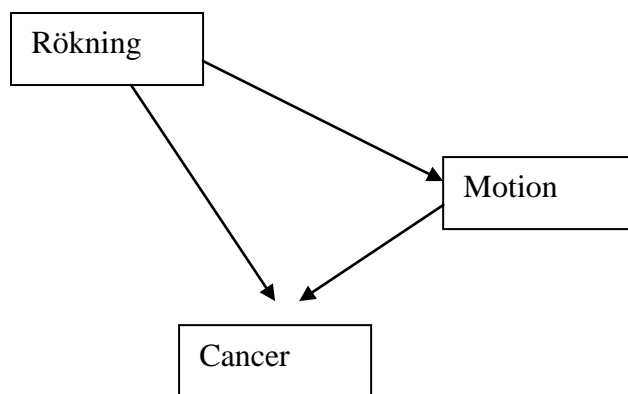
Självklart är det paradoxalt att olikheten i (1) skiljer sig från de i (2) och (3) givet den naturliga *kausala* tolkningen av ekvationerna. I normalfallet brukar man här uppfatta K som en confounder och därför välja att konditionalisera sannolikheterna för O och V på denna faktor. I så fall, om K bedöms vara den *enda* confoundern, motsvarar (2) och (3) hur effektivt läkemedlet är i respektive populationer, (1) motsvarar läkemedlets evidentiella vikt när vi inte har någon information om patientens kön, och paradoxen upplöses.

För en bayesian blir frågan: om en ny patient e dyker upp på mottagningen, ska vi använda läkemedlet eller inte på e ? Om vi känner till könet på e ska vi

alltså enligt (1) använda läkemedlet, men om vi inte känner till könet på e ska vi i enlighet med (2) och (3) inte använda det. Detta känns *prima facie* inte rimligt. Först om vi kan se att det finns en kausal faktor kopplad till K – som t ex att blodtrycket B ser annorlunda ut för K och $\neg K$ – blir det rimligt. Läkemedlet påverkar inte vårt kön, men däremot blodtrycket, som ser annorlunda ut hos könen.

(2) En analog situation har vi om vi tänker på fallet med hälsoeffekter av rökning.

Antag att O = att röka, V = cancer, K = motionerande. Situationen skulle då kunna beskrivas som:

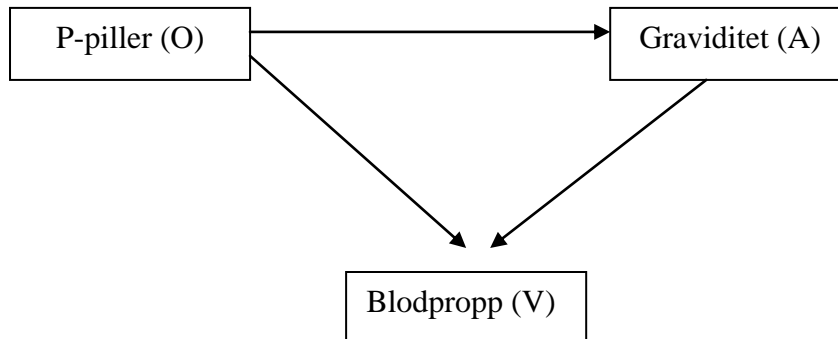


Här är det inte självklart att $p(V|O) > p(O)$ och vi kan mycket väl ha $p(V|O \wedge K) > p(V)$ och $p(V|O \wedge \neg K) > p(V)$. Alltså kan vi ha:

- (4) $p(V|O) < p(V|\neg O)$
- (5) $p(V|O \wedge K) > p(V|\neg O \wedge K)$
- (6) $p(V|O \wedge \neg K) > p(V|\neg O \wedge \neg K)$

Här kan man väl intuitivt tycka att den delning (partition) som görs utifrån motion verkar rimlig och att vi till skillnad från läkemedelsexemplet inte hamnar i en situation där en irrelevant delning (kön) gör så att en genuin orsak (läkemedel) förefaller irrelevant, eller där en irrelevant faktor (kön) ser ut som en orsak. Alla korrelationer mellan två faktorer (O och V) kan vändas om genom att vi tar med en tredje faktor (K) som korreleras till de två ”på rätt sätt”. När K är en *kausalt relevant* faktor är det rimligt att använda (2) och (3) respektive (5) och (6) för att värdera den kausala relationen mellan O och V. Om K däremot är *kausalt irrelevant* gentemot V finns det ingen anledning att använda K som kontrollvariabel, utan (1) och (4) bör användas.

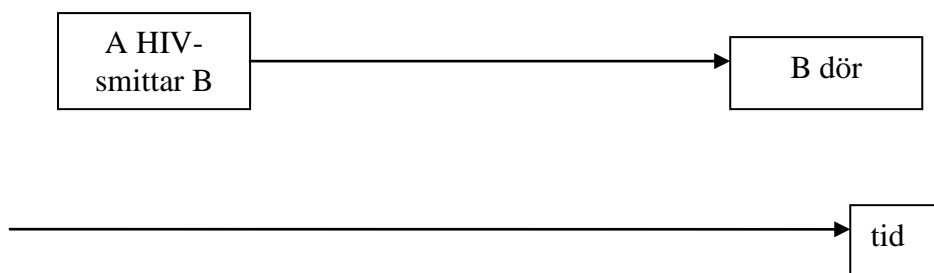
(3) Enligt bayesianer bör $O \rightarrow V$ innebära att $p(V|O) > p(V|\neg O)$. Detta är Patrick Suppes (s 41) definition av *prima facie orsak*. Som Germund Hesslow (1976) visat behöver detta dock inte stämma. Låt oss föreställa oss följande situation:



Att ta p-piller ökar sannolikheten för att drabbas av blodpropp. Men samtidigt minskar det risken att drabbas av graviditet, som vi vet i sig ökar risken för blodpropp. Sammantaget innebär detta att även om $O \rightarrow V$, så kan $p(V|O) \leq p(O)$. Vi kan alltså ha en situation där en orsak minskar sannolikheten av sin effekt eller där nettoeffekterna blir noll

Hur ska en anhängare av bayesianska nätverk förhålla sig till detta? Han kan bara använda sig av standardproceduren i form av konditionalisering över confoundern graviditet (A) om han gör en kvalificerad gissning att det är A som ”ställer till det”. Och även om han i varje enskilt fall skulle kunna beskriva en existerande temporalt kontinuerlig process med en kausal graf mellan orsak och verkan så garanterar inte detta att det alltid finns en nod mellan två andra noder i grafen (se kap5 och 7 i Cartwright 1999). Låt oss ta ett exempel:

(4) Antag att individ A HIV-smittar individ B (O) och att B senare avlider (V):



Vad som kan ställa till problem här är:

(A) Vi har en ”typ-händelse” utan direkt närhet mellan O och V.

HIV-infektion leder ofta till döden, men i det här fallet går det kanske lång tid mellan O och V.

(B) Det existerar ingen unik kausal länk mellan O och V.

HIV-infektionen leder ofta till AIDS, som i sin tur ofta leder till dödliga diaréer, lunginflammationer m m.

(C) Det föreligger ingen unik transmission av kausal information.

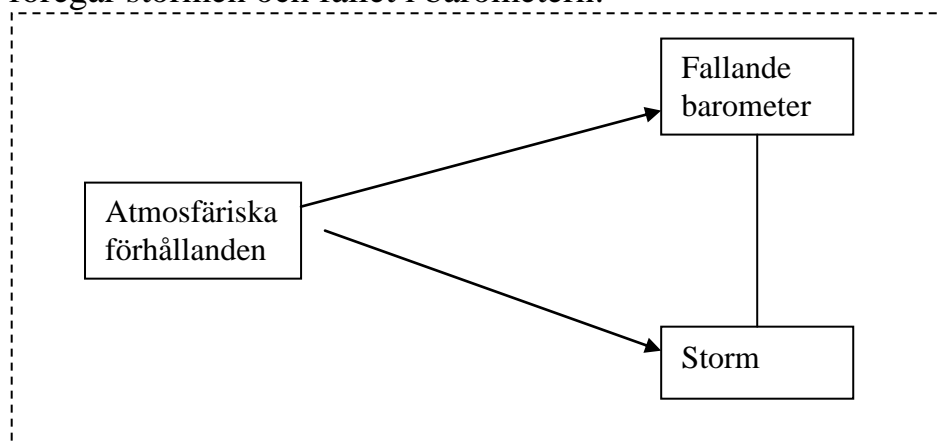
Den ursprungliga smittan kan överföras på flera olika sätt – blodtransfusion, sexuellt umgänge m m.

(D) I bayesianska nät representeras kausalitet alltid i diskret form (för varje nod finns alltid en föregångare och efterföljare), men kausalitet behöver antagligen inte se ut så i verkligheten.

Detta innebär att bayesianernas modeller över kopplingen mellan kausalitet och korrelation bara gäller med vissa förbehåll. Kan bayesianer då rättfärdiga att man oftast bortser från dessa förbehåll? Knappast! De förbehåll man gör i temer av kausala Markovkedjor, stabilitet, ”modularity”, ”faithfulness” o d räcker inte. De innebär oftast att vi bara kan se en strikt koppling mellan korrelationer och kausalitet i system vi har skapat själva och kan manipulera efter eget godtycke. Det är väl bra i sig, men världen i stort ser inte ut så och därför räcker inte dessa bemödanden om syftet med våra undersökningar är kunskap om en öppen värld och inte slutna modeller över den. Som Cartwright skriver (Mäki 2002, s 147):

Nevertheless we want our treatments to be rigorous and our conclusions to follow deductively. And the way to get deductivity when it cannot come out of the relations of the concepts is to put enough of the right kind of structure into the model. That is the trick of building one of these ... models: you have to figure out some circumstances that are constrained in just the right way that results of interest can be derived deductively.

(5) Ett annat klassiskt exempel är fallet med den fallande barometern som är en *prima facie orsak* till den påföljande stormen, men endast är en *skenbar orsak* eftersom den skärmas av från stormen av de atmosfäriska tryckförhållanden som föregår stormen och fallet i barometern.



(6) Ett liknande exempel kan anföras mot *bayesianska nät*, där man ofta utgår från Reichenbachs (1956 s 188) begrepp ”causal betweenness” för att definiera *avskärmning*. B är ”kausalt mellan” A och C om

$$p(C|B) > p(C|A) > p(C)$$

$$p(A|B) > p(A|X) > p(A)$$

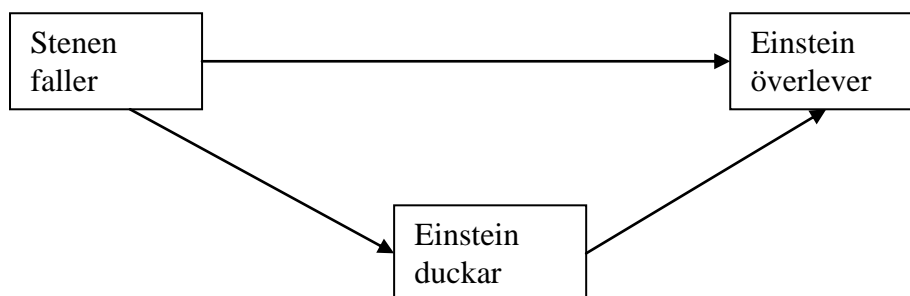
$$p(C|A \wedge B) = p(C|B).$$

B avskärmar A från C (gör A och C statistiskt irrelevanta gentemot varandra) om $p(C|A \wedge B) = p(C|B)$. En kedja $A \rightarrow B \rightarrow C$ där detta gäller uppfyller den så kallade *Markovegenskapen* (vilken, lite mer formellt, innebär att alla *konditionala oberoenden* – en generalisering av probabilistiskt oberoende mellan två variabler till att även gälla flera, exempelvis att $p(A|B \wedge C) = p(A|C)$ – som impliceras av ett bayesianskt nät är sann för det verkliga systemet).

Deborah Rosen (1978) har anfört följande motexempel: Antag att golfaren Jones slår ut snett (A) och att bollen råkar träffa en trädgren (B) och går i hål (C). Här har vi realiter en situation som bryter mot Reichenbachs antagande $p(C|B) > p(C|A) > p(C)$ eftersom $p(C|B) < p(C|A)$. För att rädda antagandet skulle vi vara tvungna att förfina referensklassen (t ex från ”Jones slår ut snett” till ”Jones slår ut bollen i 34,3 graders vinkel med en hastighet av ...”), vilket dock förefaller något långsökt. Det hjälper inte heller i analoga fall inom t ex kvantmekaniken där vi inte har tillgång till dylika förfiningar och vi rör oss med genuint osäkra förlopp (se Salmon i Sosa/Tooley s 149-150).

Reichenbachs princip om en gemensam orsak hävdar egentligen att för varje system som uppvisar probabilistiskt beroende existerar det en kausal modell som besitter Markovegenskapen. Där Reichenbachs princip säger att vi kan gå från korrelation till kausalitet, säger Markovegenskapen motsatsen, att vi kan gå från kausalitet till korrelation. Uppenbarligen stämmer de inte generellt. Bayesianer som Korb menar att detta bör leda till att vi överger tron på Reichenbachs principer som en slags *metafysisk* princip som beskriver universums natur, (Korb et al 2005, 10) ”while retaining it as a fruitful *methodological* principle of inductive method.”

(7) Hitchcock (2001) har anfört ett exempel som visar på ett icke-uppfyllande av faithfulness som beror på *intransitivitet*. En sten lossnar från en klippkant och faller ner (A) mot Einstein som duckar (B) och överlever (C). Om stenen aldrig faller överlever Einstein också, så klart:



Här vore det ju rent löjligt att säga att stenen som faller får Einstein att överleva, för även om vi kausalt har $A \rightarrow B$, $B \rightarrow C$, så gäller inte att $A \rightarrow C$. Vad som ”ställer till det” här är naturligtvis att det kausala beroendet inte kommer till uttryck i ett probabilistisk beroende.

Hesslows tidigare anförda exempel belyser också ett icke-uppfyllande av faithfulness, men förorsakat av ”multipla stigar” som gör att vi inte får någon korrelation mellan balanserande motverkande kausala krafter.

Slutsatsen blir att *generellt* säger sannolikheter i sig inte något om kausalitet och för att lösa paradoxen måste vi inse att premissen om att kausala relationer är direkt översättbara i sannolikhetskalkylens termer inte stämmer. Att det probabilistiska beroendet kan reverseras när vi går från population till stickprov visar på problemet med idén att kausalitet direkt går att översätta i korrelationer (försåvitt man inte väljer att som t ex Grangerkausalitet inom ekonometrin bara studera stickprov och anta *ceteris paribus* fram till tiden för den studerade orsaksrelationen mellan O och V.) Uppenbarligen går detta bara när vi gör vissa antaganden om modellen/populationen/stickprovet. Och då kan man inte säga att det gäller generellt.

Bayes på irrfärd i historievetenskapen

I essän *Har det verkligen hänt? Källkritik och historisk evidens* (Carlshamre 1995), diskuterar filosofen Staffan Carlshamre bland annat bevisvärde hos vittnesmål och visar hur man med hjälp av Bayes teorem kan finna sannolikheten för att en hypotes H är riktig givet att ett visst vittnesmål e säger att den är det.

Carlshamre tar som exempel följande situation:

$$p(H) = 0,2$$

$$p(\neg H) = 0,8$$

$$p(e | H) = 0,9$$

$$p(e | \neg H) = 0,05$$

Med hjälp av Bayes teorem får vi då att $p(H | e) \approx 0,82$. Att sannolikheten för H här stiger så kraftigt beror till stor del givetvis på att ”likelihoodkvoten ” $p(e | H) / p(e | \neg H) = 18/1$. *Ceteris paribus* följer också att om $p(H)$ ökar, så ökar också $p(H | e)$.

Enligt Carlshamre ger den bayesianska ”modellen” en ganska träffande beskrivning av de kontroverser som faktiskt uppkommer kring värderingen av källmaterial. I den vidare diskussionen om värdet av Bayes teorem anför Carlshamre som exempel oenigheten kring orsakerna till Karl XII:s död och framför allt frågan om han mördades av någon på den svenska sidan eller om han sköts ihjäl av fiendesidan. Carlshamre (1995 s 33) konkluderar dock att ”en viktig del av argumentationen rör [sannolikheten *ex ante*] för själva hypotesen

att kungen blev mördad: för att den skall bli tillräcklig för att kompensera den ganska svaga direkta evidensen måste man åstadkomma ett övertygande motiv ... och en rimlig gärningsman.” Så är det säkert, men desto värre för bayesianen, eftersom den typen av överväganden inte kan ha någon som helst plats i dennes överväganden. Man kan inte bara rakt av acceptera uppgivna sannolikheter, utan måste ge sig in på substantiella historiska resonemang om trovärdiga motiv och kausalkedjor. Men varför då ta omvägen om subjektiva sannolikheter?

Carlshamre (1995 s 35) medger att modellen inte är deskriptivt adekvat, men ändå besitter explikatorisk fruktbarhet trots att det föreligger ”en risk att modellen skänker en sorts pseudoprecision.”

Några argument mot bayesianism att utveckla

I själva verket tror jag att man kan rikta en rad allvarliga invändningar mot bayesianism både inom och utom historievetenskapen. Nedan ska jag ange några av de i mitt tycke tyngst vägande invändningarna.

$p(H)$ är arbiträr så till vida att bayesianism bara kräver att $p(H) + p(-H) = 1$. Men varifrån kommer dessa sannolikheter och varför skulle enskilda individers sannolikhetsuppfattningar vara intressant i ett vetenskapligt sammanhang? ,

Bayesianismen belastas av vad som brukar kallas **problemet med gammal evidens**. Bayesianismen kan inte göra reda för hur vi ändrar uppfattning om redan kända evidens, utan bara hur ren okunskap övergår i kunskap, vilket utgör ett problem eftersom vetenskapliga teorier ofta konfirmeras på basis av redan kända fakta.

Bayesianism utesluter inte uppenbara absurditeter. Låt oss t ex anta följande:

e = sylt på bordet

H = syltmonstret har varit hemma

-H = David har varit hemma

$p(H) = 0,2$

$p(-H) = 0,8$

$p(e) = 0,1$

$p(e|H) = 0,9$

$p(e|-H) = 0,2$

vilket om vi applicerar Bayes formel ger oss

$$p(H|e) = \frac{0,9 \times 0,2}{0,9 \times 0,2 + 0,2 \times 0,8} \approx 0,5 > p(H)$$

Här är det väl inte annat än rimligt att konstatera att den bästa förklaringen till e knappast är den förklaring som har högst sannolikhet!

För bayesianism gäller att H konfirmeras om $p(H|e) > p(H)$, men detta visar inte i någon relevant mening att $p(H)$ var en riktig skattning. Omvänt kan en

sannolikhet *ex ante* aldrig vara fel enligt bayesianismen, vilket dock rimmar tämligen illa med vetenskaplig praxis.

Om $p \rightarrow q$ (p kan vara Lund och q Sverige t ex) bör enligt Popper $p(p) < p(q)$, men ej enligt bayesianismen, eftersom sannolikheter *ex ante* så länge de uppfyller konsistenskraven kan vara vilka som helst. Men hur kan man *berättiga* att $p(p) > p(q)$? Eller ska man dra slutsatsen att enkelhet inte har någon epistemisk relevans?

De flesta forskare uppfattar nog vetenskapens uppgift som att bland annat skilja evidens baserat på experiment, teoretiserande och observationer, från evidens baserat enkom på subjektivt tyckande. Detta går dock inte alls ihop med bayesianismen.

Bayesianism innebär en reduktion av osäkerhet till risk, något som Keynes/Knight/Myrdahl visat vara ohållbart ur deskriptiv synvinkel.

De konditionala sannolikheter bayesianer arbetar med är symmetriska (till skillnad från kausala förklaringa). Detta är ibland problematiskt. Antag t ex $e = \text{jag är magsjuk}$, $H = \text{jag äter salmonellasmittad kyckling}$. Symmetrin kräver att $p(e|H) > p(e) \leftrightarrow p(H|e) > p(H)$, vilket i just detta exempel blir smått absurt. Symmetriabsurditeter av detta slag skulle klart kunna delvis undvikas om man tidsdimensionerar sannolikheterna, men då är vi redan lång väg från den rena bayesianismen.