EXPERIMENTBEGREPPETS BETYDELSE INOM SANNOLIKHETS- OCH BESLUTSTEORIN

Jan Odelstad

I

Vad för sorts företeelser är argument för sannolikhetsfunktioner? I framställningar av sannolikhetskalkylen kan man finna två svar på den frågan, nämligen satser och mängder. Inom matematiken är det gängse att formulera sannolikhetskalkylen med mängdläran som bas och alltså låta argumenten för sannolikhetsmåttet vara mängder. Bland filosofer är det däremot inte ovanligt att låta sannolikhetsmåttet tilldela värden till satser.

Denna skillnad i formulering av sannolikhetskalkylen mellan matematiker och filosofer förefaller lätt att förstå. Satslogiken har traditionellt en central roll inom filosofin, vilket den inte på samma sätt har inom matematiken. Och för matematiken i vår tid har mängdläran ofta varit den disciplin som lagts till grund för de flesta av matematikens olika grenar. Nu finns det ju ett uppenbart samband mellan satser och mängder (nämligen sanningsmängder) och mellan satslogik och mängdläran (satslogiken är ju en boolsk algebra och isomorf med en mängdalgebra). Man kan därför fråga sig om det finns någon väsentlig skillnad mellan att låta sannolikhetsmåttet ha satser eller mängder som argument.

Jag misstänker att en matematiker skulle hävda att det finns tekniska skäl att välja mängder. Men det är inte dessa tekniska aspekter som jag här ska uppehålla mig vid. I stället ska jag ta upp en annan skillnad mellan det traditionella sättet att framställa sannolikhetskalkylen i matematiken och hur man ofta gör i filosofin - en skillnad som inte direkt men möjligen indirekt har med skillnaden mellan användandet av mängd och sats att göra. Det gället experimentets roll.

Enligt den gängse matematiska framställningen av sannolikhetskalkylen representerar mängder, som utgör argument för sannolikhetsmåttet, händelser. I stället för att representera en händelse med en mängd kan man ju beskriva den med en sats. Så långt är det inte så stor skillnad mellan de två olika formuleringarna av sannolikhetsteorin. Men det är i den traditionella matematiska framställningen betydelsefullt att händelser är utfall av experiment. (I stället för "experiment" används ofta i detta sammanhang termen "slumpmässigt försök".) Någon motsvarighet till detta är inte vanligt att finna i framställningar av sannolikhetskalkylen baserad på satsbegreppet. Och det är experimentbegreppets ställning som är den väsentliga skillnaden mellan det traditionella matematiska och det vanliga filosofiska sättet att formulera sannolikhetskalkylen.

Självklart kan man ge experimentbegreppet en central ställning även i en framställning av sannolikhetskalkylen baserad på satsbegreppet. Men medan det är mycket vanligt att se en mängd som en delmängd av ett "universe of discourse" är detta inte lika självklart då det gäller satser. Detta kan vara en del av en förklaring till varför inte experimentbegreppet fått samma roll bland filosofer som bland matematiker.

Vilken är då experimentbegreppets roll inom sannolikhetsteorin? En sannolikhetstilldelning gäller alltid utfallen av ett experiment, dvs är alltid definierat för ett utfallsrum. Vad som ska tilldelas sannolikhet bestäms alltså av det ifrågavarande experimentet. Detta karakteriserar också den kunskapssituation mot bakgrund av vilken sannolikhetstilldelningen görs. Det är därför viktigt att observera att ett sannolikhetsmått gäller i ett experiment (relativt en experimentsituation). Experimentet bestämmer alltså situationen, och experimentets utfall är vad som sker i denna. Man måste därför skilja mellan situationen (experimentet) och vad som sker i denna (experimentets utfall), i synnerhet skilja förändring av experimentet från hur experimentet utfaller. Det är här fråga om två "nivåer" och en sammanblandning av dessa kan ge upphov till konstigheter, t ex viss form av självreferens. Jag tror att de s k paradoxer som idag diskuteras livligt inom vissa grenar av "filosofisk" sannolikhets- och beslutsteori till en del har sin rot i en dylik sammanblandning, och jag ska återkomma till detta längre fram. Låt oss emellertid först se på vilken betydelse skillnaden mellan information om experimentet och information om experimentets utfall har för förståelsen av betingad (konditional) sannolikhet.

II

Konditionalisering, dvs övergången till betingad sannolikhet, är avsett att tillämpas då man erhåller (ny) kunskap om utfallet av ett experiment. Däremot är det inte tillämpbart då man erhåller ny kunskap om experimentet, dvs att experimentet är annorlunda än man trodde. Det är nämligen i dylika fall inte ens meningsfullt att konditionalisera. $P(A|B)$ definieras ju som $P(A \cap B)/P(B)$ och B skall uppenbarligen vara en delmängd av utfallsrummet. Det man konditionaliserar till skall alltså vara ett utfall för experimentet, dvs något som sker i situationen och inte något som förändrar denna.

För att belysa skillnaden mellan information om experimentsituationen och om utfall för experimentet tänkte jag anknyta till Gärdenfors (1981). Gärdenfors diskuterar där beskrivandet av ofullständig kunskap med hjälp av sannolikheter och tar upp problemet hur sannolikhetsbedömningen förändras då vi får ny kunskap. Han menar att inom den bayesianska sannolikhetsteorin finns ett traditionellt svar, nämligen genom konditionalisering, men att det finns "situationer där det tycks som om konditionalisering inte beskriver den rationella förändringen av en sannolikhetsbedömning då man får ny kunskap" (sid 6). Gärdenfors ger ett exempel för att illustrera detta och jag ska här återge det.

> Antag att du får reda på att det finns tre negerstammar i Kenya: kikuyu (K), luo (L) och masai (M) och att befolkningen kan delas in i bofasta jordbrukare (J) och nomadiserande boskapsskötare (B). Antag vidare att detta är i stort sett allt du vet om Kenyas befolkning. Du ställs nu inför en vadhållningssituation där det

gäller att avgöra till vilka kategorier en viss kenyan tillhör. Du måste därför göra en bedömning av hur sannolika de olika kombinationerna av egenskaper är. Låt oss för enkelhetens skull anta att du på grund av den bristande informationen finner var och en av de sex kombinationerna lika sannolik. Din bedömning kan representeras på följande sätt:

		K	L	M
(3)	J	1/6	1/6	1/6
	B	1/6	1/6	1/6

Därefter får du ny kunskap. Alla masaier är boskapsskötare. Hur påverkar detta din sannolikhetsbedömning? Det verkar inte irrationellt om du anser att denna information inte påverkar din bedömning av stammarnas inbördes fördelning, så att du fortfarande tror att det är lika stor chans att finna den utpekade individen i en stam som i en annan. Det är också rimligt att informationen inte påverkar din bedömning av fördelningen av jordbrukare och boskapsskötare inom kikuyu och luo. Om detta gäller kan din bedömning efter den nya informationen representeras på följande sätt:

		K	L	M
(4)	J	1/6	1/6	0
	B	1/6	1/6	1/3

Antag nu i stället att du redan från början hade fått reda på att det finns lika många jordbrukare som boskapsskötare i Kenya, men att du fortfarande inte visste något om stammarnas inbördes fördelning. Den sannolikhetsfördelning som representeras av schemat (3) verkar fortfarande rimlig. Men om du sedan får den nya informationen att inga masaier är jordbrukare, så är (4) inte längre en rationell beskrivning av det nya kunskapsläget eftersom fördelningen mellan J och B är ojämn trots att du vet att de är lika många. Snarare kommer den nya fördelningen att se ut ungefär så här:

		K	L	M
(5)	J	1/4	1/4	0
	B	1/8	1/8	1/4

Nu är det inte längre rimligt att hålla fast vid antagandet att stammarna är lika stora, utan det måste gälla för den nya fördelningen P' att P'(J)=P'(B)=1/2.

De båda fördelningarna (4) och (5) utgör således rimliga förändringar av den sannolikhetsfördelning som beskrivs av (3), då samma nya information (att inga M är J) läggs till. Detta exempel visar att ett sannolikhetsmått inte kan innehålla all information om den ofullständiga kunskap som en person har vid ett visst tillfälle, eftersom en förändring av en sannolikhetsfördelning i så fall skulle vara entydigt bestämd av den nya informationen.

> Vi kan också jämföra med vad som hade inträffat om vi hade tillämpat konditionalisering på den fördelning som ges av (3). Den nya informationen att inga M är J skulle då ge fördelningen
>
> | | | K | L | M |
> |---|---|---|---|---|
> | (6) | J | 1/5 | 1/5 | 0 |
> | | B | 1/5 | 1/5 | 1/5 |
>
> Nu vill jag betona att de exakta siffrorna inte är avgörande för ett exempel av den här typen. ... Avsikten med exemplet är att visa att det finns situationer där sannolikhetsförändringen inte är entydigt bestämd av det givna sannolikhetsmåttet och den nya informationen och att speciellt konditionalisering inte självklart leder till den enda "förnuftiga" förändringen av en sannolikhetsbedömning. Denna poäng har jag uppnått så snart man kan peka på en sannolikhetsfördelning som inte är exakt den som förestavas av konditionalisering.
>
> (Gärdenfors (1981) sid 6-7)

Du presenteras alltså för en kenyan och frågan är vilken kategori han kommer att avslöja sig tillhöra. Det finns sex möjliga "atomära" utfall (elementarhändelser), nämligen JK (jordbrukande kikuyu), JL (jordbrukande luo), JM (jordbrukande masai), BK (boskapsskötande kikuyu), BL (boskapsskötande luo) och BM (boskapsskötande masai). För mina syften hade det varit bättre om Gärdenfors hade beskrivit situationen så, att Du slumpmässigt väljer en kenyan och frågar Dig vilken grupp han tillhör; Gärdenfors resonemang kan lika bra genomföras för det fallet. Det är mer i enlighet med vanligt språkbruk att kalla detta för ett experiment. Men "experiment" skall inom sannolikhetsteorin tolkas vitt; observationer kan t ex ha karaktären av experiment. I stället för experiment talar man ju ofta om (slumpmässigt) försök.

På basis av den information Du har om experimentet gör Du en sannolikhetstilldelning till elementarhändelserna. Gärdenfors tänker sig att inledningsvis lika stor sannolikhet tilldelas varje elementarhändelse, nämligen 1/6.

Gärdenfors intresserar sig nu för hur Du bör ändra sannolikhetstilldelningen när Du får ny information, nämligen att alla masaier är boskapsskötare. Jag vill börja med att se på vad som enligt honom skulle gälla om konditionalisering tillämpats på den ursprungliga (jämna) sannolikhetsfördelningen.

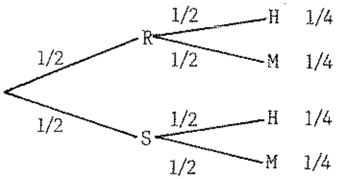
Gärdenfors hävdar följande:

> Den nya informationen att inga M är J skulle då ge fördelningen
>
> | | K | L | M |
> |---|---|---|---|
> | J | 1/5 | 1/5 | 0 |
> | B | 1/5 | 1/5 | 1/5 |
>
> (sid 7).

Detta är såvitt jag förstår fel. Konditionalisering säger nämligen i detta fall ingenting. Konditionalisering skall användas då man får ny kunskap angående <u>utfallet av experimentet</u>. Men informationen att alla masaier är boskapsskötare är ingen information om hur experimentet har utfallit, dvs att det inte finns jordbrukande masaier är ingen händelse inom experimentets ram. Det är i stället information om själva experimentet och ska ligga till grund för sannolikhetstilldelningen för detta. Gärdenfors blandar ihop "alla masaier är boskapsskötare" (eller "inga masaier är jordbrukare") med "ifrågavarande kenyan är inte masai och jordbrukare" - underförstått att han skulle kunna ha varit det <u>givet</u> experimentsituationen. Det senare är information om experimentets utfall och i dylika fall menar bayesianer att konditionalisering är tillämpbart och den ger för övrigt det resultat Gärdenfors anger, vilket är lätt att inse.

Det är alltså viktigt att skilja på information om experimentet (om hurudant det egentligen är) och information om experimentets utfall. Jag ska försöka ytterligare belysa detta genom att se närmare på två exempel.

<u>Pennexperimentet</u>: Antag att man har fyra pennor, två röda och två svarta. En av de röda pennorna är hård och en är mjuk, och likadant är det med de svarta. Antag att vi slumpmässigt väljer en penna. Detta experiment har fyra elementarhändelser, nämligen RH (röd och hård), RM (röd och mjuk), SH (svart och hård) och SM (svart och mjuk). Den naturliga sannolikhetstilldelningen för detta experiment är att alla elementarhändelser tilldelas sannolikheten 1/4. Experimentet kan lämpligen beskrivas med ett "träddiagram".



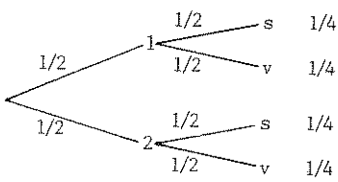
(Observera att "nivåerna" i trädet egentligen inte avspeglar en tidsordning; färg utfaller inte före hårdhet. Möjligen kan man tänka sig att man observerar färg före hårdhet.)

Antag nu att man får veta att man vid det slumpmässiga valet av en penna inte har valt en svart och mjuk. Vilken blir då sannolikhetstilldelningen till de övriga elementarhändelserna? Jo, naturligtvis 1/3. Ty

$$P(RH|\sim SM) = \frac{P(RH\cap\sim SM)}{P(\sim SM)} = \frac{P(RH\cap(RH\cup RM\cup SH))}{P(RH\cup RM\cup SH)} = \frac{P(RH)}{P(RH\cup RM\cup SH)} = \frac{1}{3}.$$

Och analogt visas $P(RM|\sim SM) = P(SH|\sim SM) = 1/3$. (Här och i fortsättningen utelämnar jag mängdklamrarna och skriver t ex RH,~SM och RH∩~SM i stället för {RH},~{SM} och {RH}∩~{SM} respektive.)

<u>Kulexperimentet</u>: Antag att man har två urnor, 1 och 2. Var och en av urnorna innehåller en svart och en vit kula. Man väljer en urna slumpmässigt och ur denna tar man sedan på måfå en kula. Detta experiment har följande elementarhändelser: 1s (svart kula ur urna 1), 1v (vit kula ur urna 1), 2s (svart kula ur urna 2) och 2v (vit kula ur urna 2). Den naturliga sannolikhetstilldelningen för detta experiment är att alla elementarhändelser tilldelas sannolikheten 1/4. Experimentet kan lämpligen illustreras med följande träddiagram:



Antag att man får veta att man vid det slumpmässiga valet av kula inte valt en vit kula ur urna 2. Därmed ändras sannolikhetstilldelningen till de övriga elementarhändelserna så att de alla får sannolikheten 1/3. Ty

$$P(1s|\sim 2v) = \frac{P(1s \cap \sim 2v)}{P(\sim 2v)} = \frac{1}{3}$$

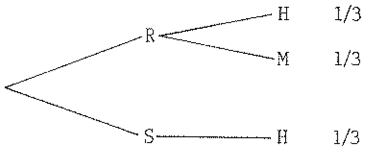
och analogt $\quad P(1v|\sim 2v) = P(2s|\sim 2v) = 1/3.$

I båda experimenten har vi använt oss av "konditionalisering" (Bayes' regel) för att bestämma den nya sannolikhetstilldelningen när vi fått ny information (i pennexperimentet att vi inte valt en svart och mjuk penna, och i kulexperimentet att vi inte valt en vit kula ur urna 2). Och det är knappast något som helst kontroversiellt med detta. Låt oss nu tänka oss att just när Du ska till att utföra penn- respektive kulexperimentet så får Du veta att den svarta och mjuka pennan är bortplockad respektive att den vita kulan ur urna 2 är borttagen. Hur ändras Din sannolikhetstilldelning då?

Antag att man skulle vilja bestämma detta genom konditionalisering, dvs bestämma sannolikhetstilldelningen till det nya experimentet ur sannolikhetstilldelningen till det gamla genom konditionalisering. Låt oss se på pennexperimentet först. Den nya informationen är ju att den svarta och mjuka pennan är bortplockad vilket inom ramen för använda beteckningar man kan känna sig frestad att representera som ~SM; att den svarta och mjuka är bortplockad innebär ju att man inte valt den. Som tidigare påpekats gäller:

$$P(RH|\sim SM) = P(RM|\sim SM) = P(SH|\sim SM) = 1/3.$$

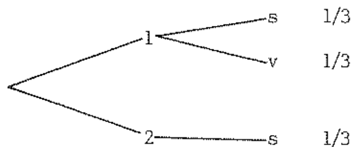
Sannolikhetstilldelningen till det nya pennexperimentet ges alltså av följande träddiagram.



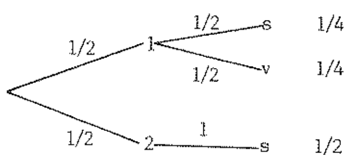
Låt oss så tillämpa samma resonemang för kulexperimentet. Vi vill bestämma det nya experimentets sannolikhetstilldelning ur det gamlas genom konditionalisering. Den nya informationen är ju att den vita kulan i urna 2 är bortplockad. Inom ramen för använda beteckningar är det måhända frestande att representera det med ~2v; att den vita kulan ur urna 2 är bortplockad innebär ju att den inte valts. Som tidigare framhållits gäller:

$$P(1s|\sim 2v) = P(1v|\sim 2v) = P(2s|\sim 2v) = 1/3$$

och sannolikhetstilldelningen till det nya kulexperimentet ges av följande träddiagram:



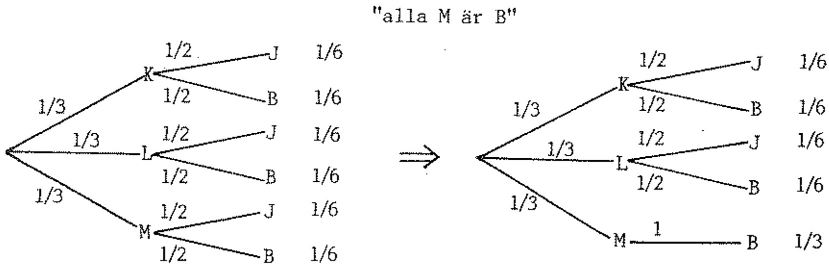
Men det är uppenbart att det har gått snett här. Sannolikheten för att välja urna 2 påverkas ju inte av att den vita kulan i den urnan är bortplockad. Sannolikhetstilldelningen till det nya kulexperimentet är naturligtvis:



Skillnaden mot pennexperimentet är att sannolikheten för att välja en svart penna minskar då man plockar bort den mjuka svarta pennan, medan sannolikheten för att välja urna 2 inte påverkas av att man plockar bort den vita kulan.

Felet man gör vid konditionalisering i kulexperimentet är förstås att man övergår från händelsen att den vita kulan är bortplockad ur urna 2 till ~2v som är händelsen att man inte valt den vita kulan ur urna 2. Och det förra är ingen möjlig händelse inom experimentets ram. På samma sätt är ju händelsen att den svarta och mjuka pennan är bortplockad en annan händelse än ~SM som ju är händelsen att man inte valt den svarta och mjuka (men kunde ha gjort det). Det är alltså lika felaktigt att använda konditionalisering i penn-experimentet som i kulexperimentet, men i det förra fallet råkar resultatet ändå bli riktigt.

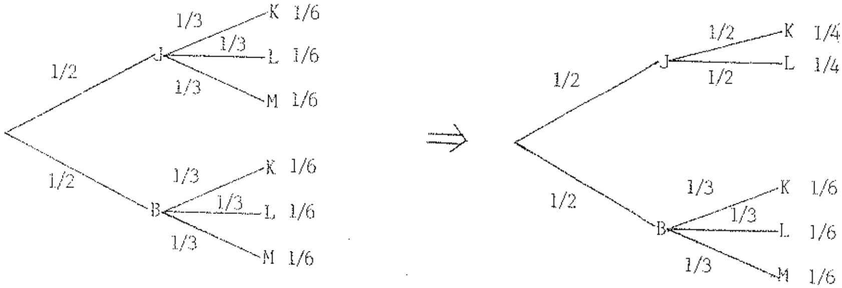
Observera att Gärdenfors' diskussion av stamexemplet i viss utsträckning kan ses som analogier med penn- och kul-experimentet. Om vi först ser stammarna som urnor och sysselsättning som kulfärg så innebär att vi får veta att alla masaier är boskapsskötare inte att sannolikheten för stamtillhörighet ändras. Följande figur får illustrera detta:



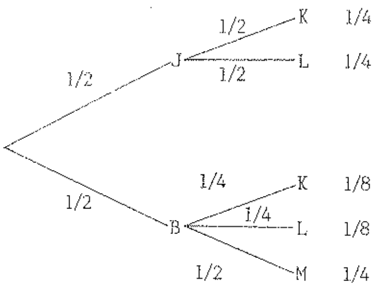
Och resultatet är alltså det som ges i schema (4).

Om vi ser sysselsättningen som urna och stammarna som kulfärg så innebär att alla masaier är boskapsskötare inte att sannolikheten för sysselsättning ändras. En möjlig förändring illustreras av följande figur:

"alla M är B"



Nu är inte det vad Gärdenfors redovisar som schema (5). Träddiagrammet till höger blir hos Gärdenfors i stället:



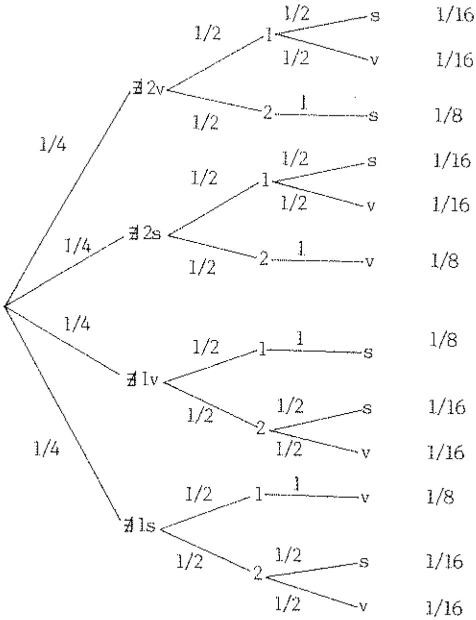
Det finns naturligtvis många olika möjligheter och Gärdenfors motiverar inte närmare sitt val.

Schema (6) slutligen är alltså det vi skulle få om vi konditionaliserade till att ifrågavarande kenyan inte är masai och jordbrukare och om situationen ses analog med penn-experimentet så är schema (6) också resultatet av att vi fått veta att alla masaier är boskapsskötare.

Visar inte detta att Gärdenfors har rätt? Konditionalisering är bedräglig! Mja, det är inte riktigt så enkelt. Ingen än så hängiven bayesian skulle drömma om att tillämpa konditionalisering på det sätt jag skisserat för att bestämma sannolikhetstilldelningen i det nya kulexperimentet. Konditionalisering är ju inte avsett att fungera i detta sammanhang utan ska användas då man fått information om utfallet av det experiment man betraktar. Konditionalisering är m a o alltid konditionalisering relativt möjliga utfall för experimentet. Gärdenfors argumenterar mot konditionalisering genom att visa att det i sammanhang där det aldrig varit avsett att tillämpas ger konstiga resultat. Ingen bayesian behöver sova dåligt om natten på grund av detta. Däremot understryker Gärdenfors' exempel vikten av att använda konditionalisering i det sammanhang för vilket det är avsett.

Penn- och kulexperimenten understryker det Gärdenfors säger om att allt vad man tror i en situation inte kan ges enbart av sannolikhetstilldelningen. Penn- och kulexperimenten har ju analoga sannolikhetstilldelningar (se trädrepresentationerna), och den information man får om den nya varianten av de båda exeprimenten är analog (åtminstone i viss mening), men ändå blir sannolikhetstilldelningen i de nya experimenten inte analoga (se trädrepresentationerna). Men ingen kan rimligen ha hävdat att allt det vi tror om ett experiment fångas i sannolikhetstilldelningen - möjligen det vi tror om dess utfall. Det är bl a därför viktigt att begränsa konditionalisering till möjliga utfall av experimentet.

Informationen om att ett experiment har ändrats kan naturligtvis självt vara utfallet för ett experiment (fast i så fall ett annat än det första) och i detta kan man konditionalisera till denna information. Låt oss se på kulexperimentet igen. Antag att situationen är sådan att vi vet att precis en av kulorna i en av urnorna är bortplockad, dvs en av urnorna innehåller en svart och en vit kula medan den andra bara innehåller en kula som antingen är svart eller vit, och att vilket som är fallet är helt slumpmässigt. Följande träddiagram, där Ø2v är händelsen att urna 2 inte har någon vit utan bara en svart kula (analogt med Ø2s, Ø1v, Ø1s), illustrerar situationen.

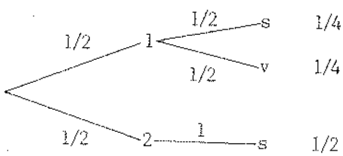


Observera att

P(1s) = 1/16 + 1/16 + 1/8 = 1/4 , P(1v) = 1/16 + 1/16 + 1/8 = 1/4

P(2s) = 1/8 + 1/16 + 1/16 = 1/4 , P(2v) = 1/8 + 1/16 + 1/16 = 1/4

precis som i det ursprungliga kulexperimentet. Om vi nu får veta att det är den vita kulan i urna 2 som är bortplockad så har alltså händelsen ∄2v inträffat och den situationen vi då har illustreras av följande träddiagram



som alltså är erhållet genom konditionalisering till ∄2s.

I detta utvidgade kulexperiment kan vi alltså konditionalisera till händelsen att den svarta kulan i urna 2 är bortplockad och resultatet överensstämmer helt med vad vi väntar oss.

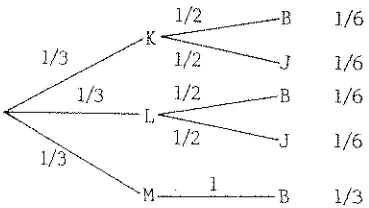
I stället för att säga att vi utvidgar experimentet kan vi säga att vi förlänger det - vi förlänger ju träddiagrammet. Uppenbarligen kan det ursprungliga kulexperimentet förlängas på andra sätt, så att P(1s) = P(1v) = P(2s) = P(2v) och så att konditionalisering till att den vita kulan ur urna 2 är bortplockad ger önskvärt resultat.

En utvidgnng/förlängning analog med den som nyss skisserats för urnexemplet kan också tillämpas på Gärdenfors' exempel. Vi antar att vi får veta att en av kombinationerna av stamtillhörighet och sysselsättning inte förekommer, men inte vilken som är utesluten. $K \subseteq B$ är händelsen att alla kikuyu är boskapsskötare och analogt med $K \subseteq J$, $L \subseteq B$ osv. Träddiagrammen i fig 1 och 2 på de följande två sidorna beskriver två möjliga sannolikhetstilldelningar. I fig 1 uppträder stamtillhörigheten som urna och i fig 2 har sysselsättningen den rollen. Observera att för båda träddiagrammen gäller att

$$P(JK) = P(JL) = P(JM) = P(BK) = P(BL) = P(BM) = 1/6 ,$$

dvs vi har sannolikhetstilldelningar som överensstämmer med schema (3) hos Gärdenfors.

Träddiagrammet i fig 1 svarar i viss mening mot Gärdenfors' schema (4). Om $M \subseteq B$ inträffar, dvs om alla masaier är boskapsskötare, så erhåller vi



som just är schema (4) och detta <u>genom konditionalisering</u> till $M \subseteq B$.

FIG. 1

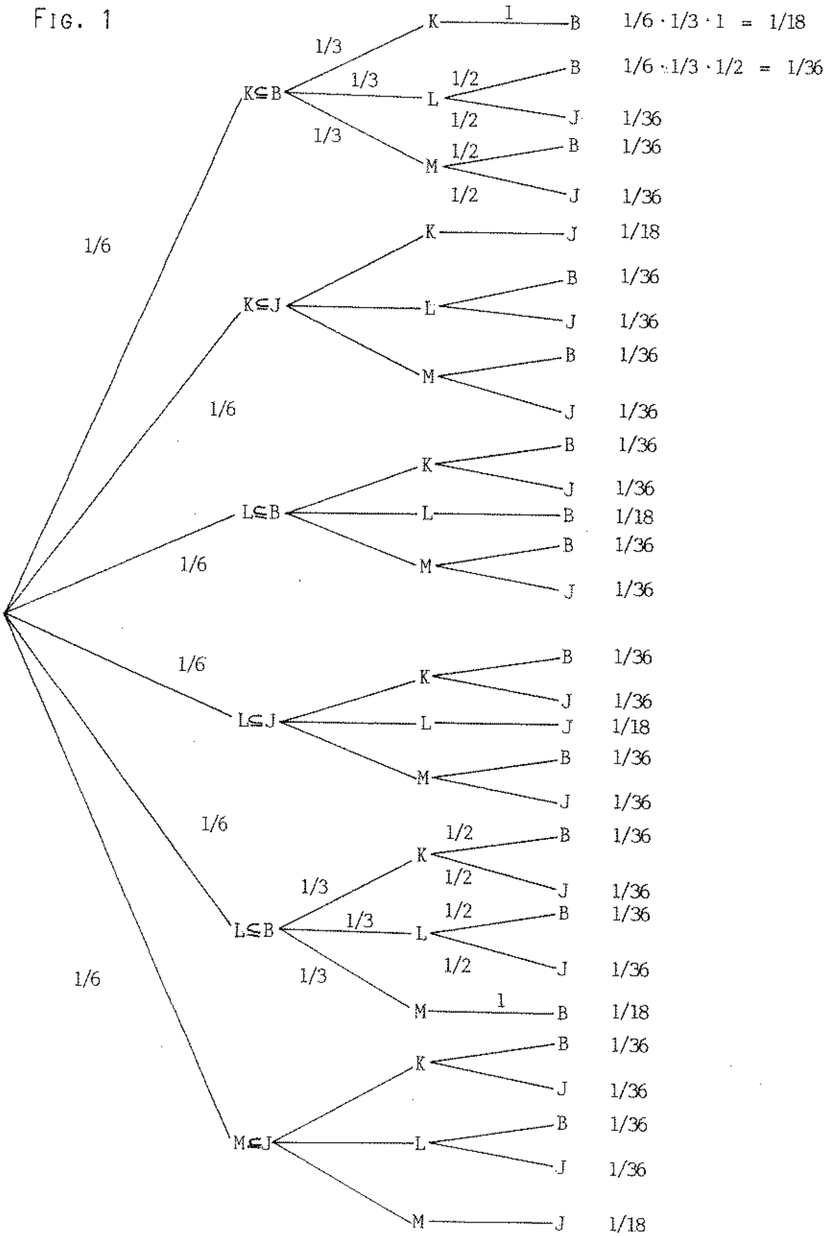
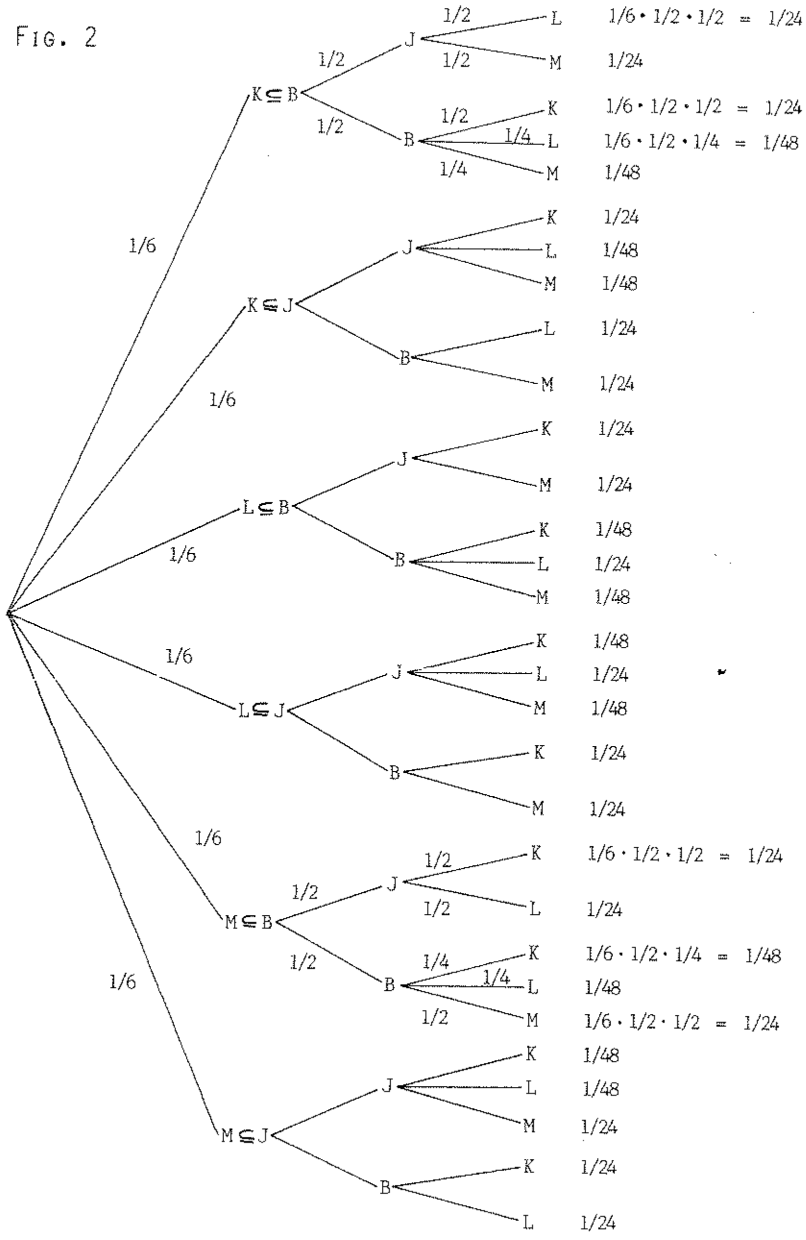
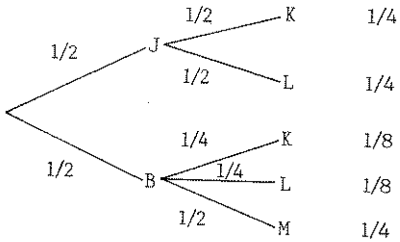


FIG. 2


1/2
1/2
1/2
1/2
1/2
1/4
1/4
K⊆B
J
B
L
M
K
L
M
1/6 · 1/2 · 1/2 = 1/24
1/24
1/6 · 1/2 · 1/2 = 1/24
1/6 · 1/2 · 1/4 = 1/48
1/48
1/6
K⊆J
1/24
1/48
1/48
1/24
1/24
1/6
L⊆B
1/24
1/24
1/48
1/24
1/48
1/6
L⊆J
1/48
1/24
1/48
1/24
1/24
1/6
M⊆B
1/2
1/2
1/2
1/2
1/4
1/4
1/2
1/6 · 1/2 · 1/2 = 1/24
1/24
1/6 · 1/2 · 1/4 = 1/48
1/48
1/6 · 1/2 · 1/2 = 1/24
1/6
M⊆J
1/48
1/48
1/24
1/24
1/24

Träddiagrammet i fig 2 svarar däremot i viss mening mot Gärdenfors' schema (5). <u>Genom konditionalisering</u> till $M \subseteq B$ erhåller vi

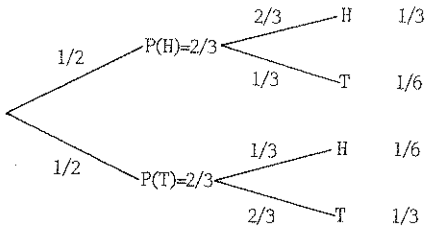


vilket är Gärdenfors' schema (5).

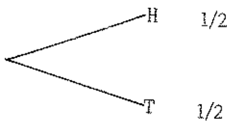
Vi kan alltså just genom konditionalisering till alla masaier är boskapsskötare erhålla Gärdenfors' schema (4) och (5) som är avsedda att visa konditionaliseringens brister. Vad vi har gjort är alltså att vi konstruerat två olika situationer, som båda har en jämn sannolikhetsfördelning till de sex möjliga kombinationerna av stamtillhörighet och sysselsättning och sådana att vid konditionalisering till att alla masaier är boskapsskötare det ena ger schema (4) och det andra schema (5).

De förlängningar av såväl kulexperimentet som Gärdenfors' experiment som jag föreslagit är ju lite ad hoc och kan raffineras avsevärt. Särskilt i Gärdenfors' exempel är det naturligt att tillämpa en idé som går tillbaka åtminstone till Bruno de Finetti, och som Gärdenfors har varit inne på i andra sammanhang (se t ex Gärdenfors och Sahlin (1982)). Den går ut på att då man inte vet sannolikhetstilldelningen i ett experiment innebär det att man har en sannolikhetsdistribution över olika sannolikhetstilldelningar. Man utvidgar alltså experimentet så att olika sannolikhetstilldelningar ingår som (del)utfall för experimentet. Ett exempel kan här vara på sin plats och för att det ska kunna illustreras lätt väljer jag ett något orealistiskt.

Antag att Du får veta att ett visst mynt är osymmetriskt så att sannolikheten för att få head är 2/3 <u>eller</u> sannolikheten för att få tail är 2/3. Däremot får Du inte veta vilket av de två alternativen som är det riktiga. Du får alltså inte veta om myntet är "biased" mot head (H) eller tail (T). Hur stor är sannolikheten att få H vid ett kast med myntet? Jo, i de Finettis anda är svaret 2/3 med sannolikheten 1/2, och 1/3 med sannolikheten 1/2, dvs 1/2 (ty $2/3 \cdot 1/2 + 1/3 \cdot 1/2 = 1/2$). Situationen kan illustreras med följande träddiagram:



Situationen ska alltså <u>inte</u> ses som



utan detta är bara en sammanfattning (förkortning) av beskrivningen ovan.

Att $P(H)=2/3$ och $P(T)=1/3$ är alltså ett utfall för en del (ett steg) i experimentet, och det är förstås också $P(H)=1/3$ och $P(T)=2/3$. Fördelen med den förlängda beskrivningen blir tydlig om man ser på flera kast. Hur stor är sannolikheten att få H i fjärde kastet om (a) de tre första har resulterat i H och (b) de tre första i T? Å ena sidan kan man tycka att eftersom kasten är oberoende av varandra så är sannolikheten för H i fjärde kastet i båda fallen lika. Å andra sidan är det rimligt att sannolikheten för H i fjärde kastet i fallet (a) är större än i fallet (b); har vi fått tre H i rad så ökar det sannolikheten för att $P(H)=2/3$. Sannolikheten för $P(H)=2/3$ givet tre H i rad är närmare bestämt:

$$\frac{1/2 \cdot 2/3 \cdot 2/3 \cdot 2/3}{1/2 \cdot 2/3 \cdot 2/3 \cdot 2/3 + 1/2 \cdot 1/3 \cdot 1/3 \cdot 1/3} = 8/9$$

Sannolikheten för att få H i fjärde kastet då de tre första resulterat i H är därför 2/3 med sannolikheten 8/9 och 1/3 med sannolikheten 1/9, dvs

$$8/9 \cdot 2/3 + 1/9 \cdot 1/3 = 17/27$$

Sannolikheten för P(H)=2/3 givet tre T i rad är

$$\frac{1/2 \cdot 1/3 \cdot 1/3 \cdot 1/3}{1/2 \cdot 1/3 \cdot 1/3 \cdot 1/3 + 1/2 \cdot 2/3 \cdot 2/3 \cdot 2/3} = 1/9$$

och alltså sannolikheten för H i fjärde kastet då de tre tidigare resulterat i T

$$1/9 \cdot 2/3 + 8/9 \cdot 1/3 = 10/27$$

Detta motsäger alltså att sannolikheten för H i fallen (a) och (b) ska vara lika. Men observera att det på intet sätt motsäger att kasten är oberoende. Att kasten är oberoende innebär nämligen att sannolikheten för att få H i ett kast givet P(H)=2/3 (resp P(T)=2/3) är oberoende av resultaten i tidigare kast; däremot beror alltså sannolikheten för att P(H)=2/3 på tidigare kast.

I mer realistiska situationer har man inte att göra med en sannolikhetsdistribution över enbart två sannolikhetstilldelningar utan över många, kanske alla. Det innebär naturligtvis matematiska komplikationer men grundtanken är densamma. I Gärdenfors' exempel har man i en version en jämn distribution till alla sannolikhetstilldelningar, i en annan en (ev jämn) distribution till alla sannolikhetstilldelningar som gör alla stammar lika sannolika, i en tredje version en distribution till alla sannolikhetstilldelningar som gör sysselsättningarna lika sannolika osv. En hel del av vad vi vet om Kenyas befolkning kan fångas i olika sannolikhetsdistributioner till sannolikhetstilldelningar. Kanske är det rent av så att den kunskapssituation man har kan karakteriseras godtyckligt noggrant med hjälp av experimentbegreppet - det är bara en fråga om att analysera experimentet tillräckligt. Det som i <u>en</u> analys av situationen blir information om att experimentet har ändrats kan i <u>en annan</u>

analys bli information om utfallet av experimentet (som då alltså är ett annat än det förra) och konditionalisering kan då bli meningsfull.

III

Konditionalisering spelar en inte oviktig roll inom beslutsteori och även där används den ibland ovarsamt och felaktigt. Enligt Jeffreys variant av maximen om maximering av förväntad nytta (jag följer här framställningen i Rabinowicz (1985)) ska man välja så att man maximerar den förväntade "desirability", där förväntad "desirability" hos en handling A ges av formeln

$$\sum_{i=1}^{n} P(S_i|A) \cdot D(S_i\&A)$$

där $S_1,\ldots,S_n$ är en partition vars element lämpligen kan kallas tillstånd medan P och D är agentens subjektiva sannolikhets- repektive "desirability"-funktion. P och D opererar på propositioner varför 'A' och 'S_i' för att vara exakt representerar propositionerna att en given handling utförs eller att ett givet tillstånd inträffar.

Som synes konditionaliserar Jeffrey till handlingen A. Utgående från den tes jag här driver, nämligen att konditionalisering endast kan komma i fråga då det man konditionaliserar till är ett utfall för det experiment man betraktar, verkar detta högst uppseendeväckande. I så fall skulle det nämligen bli ett utfall för det aktuella experimentet att A utförs. Detta är tror jag ett "kategorimisstag". I en valsituation så är de handlingar vi väljer mellan <u>inte</u> utfall för ett experiment. Agentens val av handlingsalternativ är <u>ur agentens synvinkel</u> (såsom handlande varelse) inte ett utfall för ett experiment; vore det så vore det inte fråga om ett val. "Inifrån" sett är valet alltså inget experiment. Vilket inte hindrar att någon (ev agenten själv) kan betrakta situationen "utifrån", och därmed se det som ett slumpmässigt försök hur agenten väljer; om agenten i den givna situationen väljer A eller inte är ju något man kan tillordna sannolikheter. Det är emellertid viktigt att observera att ser man agentens val som ett experiment (slumpmässigt försök) så har man bytt perspektiv och övergått från att se situationen "inifrån" till att se den "utifrån". Och agenten måste i valsituationen se situationen inifrån.

Handlingsalternativen ska alltså inte betraktas som utfall av ett experiment utan i stället som experiment. Då man väljer hur man ska handla väljer man mellan att utföra olika experiment. I beslutssituationen är ens intresse för experimenten av lite speciellt slag: man är intresserad av såväl "nyttan" av att utföra experimentet som "nyttan" av dess utfall. Det kan därför vara lämpligt att i stället för experiment tala om spel (gambles). Att välja mellan olika handlingsalternativ är att välja mellan olika spel. Principen om maximering av förväntad nytta innebär att man väljer det spel som har störst förväntad nytta. Den förväntade nyttan av spelet A som kan resultera i utfallen $S_1,\ldots,S_n$ är

$$\sum_{i=1}^{n} P_A(S_i)\cdot D(S_i\&A)$$

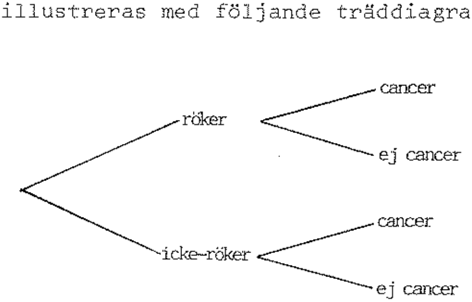
där $P_A(S_i)$ är sannolikheten för utfallet S_i i spelet A och $D(S_i\&A)$ är nyttan av att få spela A och få utfallet S_i.

Den skillnad i förhållande till Jeffrey jag här vill betona är alltså att $P(S_i|A)$ ersätts med $P_A(S_i)$. P_A är alltså sannolikhetstilldelningen till spelet A medan $P(\ |A)$ är sannolikheten betingad till A. (Observera att både $P(S_i|A)$ och $P_A(S_i)$ kan utläsas "sannolikheten för S_i givet A"; i förra fallet bör man kanske egentligen säga "sannolikheten för S_i givet att A inträffat" och i senare fallet "sannolikheten för S_i i experimentet/spelet A".)

Jeffreys variant av teorin om maximering av förväntad nytta har diskuterats en hel del i litteraturen. Bl a har man formulerat exempel där Jeffreys teori har givit ointuitiva resultat. Jag ska här ta upp ett sådant, nämligen det s k Fishers problem.

Antag att Du tycker om att röka men inte att få cancer, och att Du är övertygad om att om någon röker så är sannolikheten att denna ska få cancer större än om han inte röker. Under rimliga antaganden om värdena P(cancer|röker), P(cancer icke-röker), D(cancer och röker) och D(cancer och icke-röker) så gäller att den förväntade "desirability"(ED) för icke-röker är större än för röker. Så långt är allt gott och väl. Antag nu att vetenskapen gör en ny landvinning. Det är inte så att rökning orsakar cancer. Däremot har rökbegäret och cancer en gemensam orsak: har man en viss gen så tenderar man att tycka om rökning och att få cancer. Denna nya insikt om rökningens och cancerna orsaker innebär inte att värdena för P- och D-funktionerna behöver ändras. Den förväntade "desirability" blir därför som tidigare större för röker än icke-röker. Men detta är ointuitivt. Ty under förutsättning att Du föredrar att röka framför att inte röka och rökning inte på något sätt orsakar cancer, så är det rimligt att Du röker. Varför skulle Du inte tillåta Dig nöjet att röka om det i alla fall inte påverkar om Du får cancer eller inte?

Fishers problem tycks alltså innebära en svårighet för Jeffreys teori. Svårigheten ligger i konditionaliseringen till den handling man väljer. Men om man tänker på att konditionalisering bara ska tillämpas då det man konditionaliserar till är utfall för det experiment man betraktar försvinner svårigheten. Det Du har att välja mellan är att delta i röka-spelet eller i icke-röka-spelet. Sannolikheten för att få cancer givet att man röker hör hemma i sannolikhetstilldelningen till ett helt annat experiment där att man får cancer och att man röker båda är utfall. Det experimentet kan t ex ha följande form: Välj en person godtyckligt och observera om han röker och får cancer. Ett dylikt experiment kan illustreras med följande träddiagram:



Empiriska undersökningar ger starkt stöd för hypotesen att P(cancer|röker) är större än P(cancer|icke-röker). Men detta ger inte självklar vägledning för bestämmande av sannolikhetstilldelningen till de två spel Du har att välja mellan. Både röka-spelet och icke-röka-spelet tänks i detta exempel endast ha två utfall; cancer respektive ej cancer. Och om vi accepterar genhypotesen så är sannolikhetstilldelningen densamma i båda spelen. Således gäller

$P_{\text{röka}}(\text{cancer}) = P_{\text{icke-röka}}(\text{cancer})$ och

$P_{\text{röka}}(\text{ej cancer}) = P_{\text{icke-röka}}(\text{ej cancer})$ medan

$D(\text{cancer \& röka}) > D(\text{cancer \& icke-röka})$ och

$D(\text{ej cancer \& röka}) > D(\text{ej cancer \& icke-röka})$.

Alltså ger teorin att den förväntade "desirability" för röka-spelet är större än för icke-röka-spelet, och Du bör välja att röka, dvs teorin ger det resultat som förefaller naturligt under den (onaturliga) genhypotesen.

Det jag här vill betona är alltså skillnaden mellan $P(S_i|A)$ och $P_A(S_i)$ och att det inte finns någon allmän regel att övergå från $P(S_i|A)$ till $P_A(S_i)$, eftersom det rör sig om sannolikhetstilldelningar i olika experiment och man inte kan säga något generellt om hur de hänger ihop som gäller för godtyckliga situationer. Jag förmodar att detta räcker till för att lösa också Newcombs problem som ett beslutsteoretiskt problem - däremot inte som ett demonteoretiskt. Låt oss se närmare på detta.

IV

Robert Nozick presenterar Newcombs problem på följande sätt (se Nozick (1969)).

> Suppose a being in whose power to predict your choices you have enormous confidence. (One might tell a science-fiction story about a being from another planet, with an advanced technology and science, who you know to be friendly, etc.) You know that this being has often correctly predicted your choices in the past (and has never, so far as you know, made an incorrect prediction about your choices), and furthermore you know that this being has often correctly predicted the choices of other people, many of whom are similar to you, in the

particular situation to be described below. One might tell a longer story, but all this leads you to believe that almost certainly this being's prediction about your choice in the situation to be discussed will be correct.

There are two boxes, (B1) and (B2). (B1) contains \$1000. (B2) contains either \$1000000 (\$M), or nothing. What the content of (B2) depends upon will be described in a moment. ...

You have a choice between two actions:

(1) taking what is in both boxes

(2) taking only what is in the second box.

Furthermore, and you know this, the being knows that you know this, and so on:

(I) If the being predicts you will take what is in both boxes, he does not put the \$M in the second box.

(II) If the being predicts you will take only what is in the second box, he does put the \$M in the second box.

The situation is as follows. First the being makes its prediction. Then it puts the \$M in the second box, or does not, depending upon what it has predicted. Then you make your choice. What do you do? (Sid. 114-115.)

Situationen kan beskrivas med följande beslutsmatris:

	S_1	S_2
A_1	\$ M	\$ 0
A_2	\$ 1000+M	\$ 1000

A_1 är handlingen att enbart ta box (B2) och A_2 är handlingen att ta båda boxarna. S_1 innebär att "the being" - i fortsättningen kallad demonen - placerat 1 miljon \$ i box 2 medan S_2 innebär att den inte gjort det.

Förväntat "desirability", förkortat ED, hos handlingarna A_1 och A_2 ges (enligt Jeffreys teori) av följande:

$$ED(A_1) = P(S_1|A_1)\cdot D(S_1\&A_1) + P(S_2|A_1)\cdot D(S_2\&A_1)$$

$$ED(A_2) = P(S_1|A_2)\cdot D(S_1\&A_2) + P(S_2|A_2)\cdot D(S_2\&A_2)$$

Låt oss för enkelhets skull anta att

$$D(S_1\&A_1) = M \qquad D(S_2\&A_1) = 0$$

$$D(S_1\&A_2) = 1000 + M \qquad D(S_2\&A_2) = 1000$$

(Vi har alltså antagit att"desirability" är "linear in money"; se Levi (1975) sid. 370.)

Således gäller att

$$ED(A_1) = P(S_1|A_1)\cdot M$$

$$ED(A_2) = (1-P(S_2|A_2))\cdot(1000+M) + P(S_2|A_2)\cdot 1000 =$$
$$= 1000 + M - P(S_2|A_2)\cdot M$$

Demonen lägger alltså \$M i (B2) omm den förutspår att Du väljer enbart (B2), dvs utför A_1. Och den lägger ingenting i (B2) omm den förutspår att Du tar både (B1) och (B2), dvs utför A_2. Eftersom demonen tycks vara en särdeles begåvning vad beträffar förutsägelser av detta slag är både $P(S_1|A_1)$ och $P(S_2|A_2)$ nära 1 och alltså är $ED(A_1)$ nära M medan $ED(A_2)$ är nära 1000. Du bör alltså välja A_1.

Nozick framhåller att för många är denna slutsats högst ointuitiv och därmed (åtminstone för dessa) inställer sig här ett problem för den jeffreyska beslutsteorin. Demonen har ju redan innan Du gör Ditt val antingen placerat eller inte placerat \$M i (B2). Hur Du väljer kan därför inte påverka vad demonen gör. Varför ska Du då inte ta båda boxarna? Handlingen A_2 dominerar ju f ö A_1 i den meningen att

$$D(S_1\&A_2) > D(S_1\&A_1) \qquad D(S_2\&A_2) > D(S_2\&A_1)$$

dvs vad som än inträffar så är A_2 bättre än A_1. Talar inte det också för A_2?

Diskussionen har varit livlig kring Newcombs problem och vilka slutsatser man kan dra av det (och andra liknande problem). Jag ska inte gå in på den utan enbart påpeka att om man ersätter $P(\ |A)$ med P_A i formeln för ED(A) så upphör Newcombs problem att vara ett beslutsteoretiskt problem. Att demonen är så skicklig på förutsägelser innebär att $P(S_1|A_1)$ och $P(S_2|A_2)$ är nära 1 men säger inget omedelbart om sannolikheterna $P_{A_1}(S_1)$ och $P_{A_2}(S_2)$. Dessa sannolikheter rör ju utfall i helt andra experiment än det där A_1 och A_2 är utfall och som $P(S_1|A_1)$ och $P(S_2|A_2)$ gäller. Vad $P(S_i|A_i)$ innebär för $P_{A_i}(S_i)$ (i=1,2) beror på hur man tror demonen fungerar, dvs vilken demonteori man har. Om demonen utnyttjar "backward causation", "extra sensory perception", stor psykologisk insiktsfullhet eller något helt annat kan ha betydelse för värdet på $P_{A_i}(S_i)$. Beslutsteorin, som tillämpad matematisk teori, kan därför inte ha anspråk på att ge ett svar på hur man ska handla, utan vad som är avgörande för detta är vad man tror om demonen. Newcombs problem är därför, såvitt jag förstår, ett demonteoretiskt och inget beslutsteoretiskt problem.

V

Det som är gemensamt för Fishers och Newcombs problem, och som gör dem till stötestenar för Jeffreys version av teorin om förväntad nytta, är att hur vi handlar inte påverkar vilket tillstånd som inträffar men är däremot ett tecken på vad som kommer att inträffa. $P(S|A)$ har alltså ett helt annat värde än $P(S)$, men det beror inte på att A påverkar (orsakar) att S inträffar eller inte inträffar utan beror på att A är ett tecken på att S kommer eller inte kommer att inträffa. Enligt Jeffreys teori (i den ursprungliga formuleringen som jag här begränsar mig till) blir förväntad "desirability" detsamma om en handling orsakar ett tillstånd eller enbart är ett tecken på det. Detta verkar ofta ointuitivt, vilket Fishers och Newcombs problem visar. Den s k kausala beslutsteorin kan, förmodar jag, ses som framväxt kring denna svårighet. Som torde ha framgått tror jag att egentligen är problemet bara skenbart och beror på en sammanblandning av sannolikhetsmått som hör till olika experiment, närmare bestämt på en sammanblandning av $P(\ |A)$ och P_A.

* * *

Jag vill tacka Sven Danielsson, Peter Gärdenfors, Wlodzimierz Rabinowicz och Sören Stenlund för intressanta diskussioner och värdefulla synpunkter.

Litteraturförteckning

GÄRDENFORS, P.(1981) Hur beskriver man ofullständig kunskap? Filosofisk tidskrift, Årg 2, Nr 1, 1-11.

GÄRDENFORS, P. & SAHLIN, N-E.(1982) Unreliable probabilities, risk taking, and decision making. Synthese, 53, 361-386.

LEVI, I.(1975) Newcomb's many problems, i Hooker, Leach & McClennen (red), Foundations and applications of decision theory, Vol I, 369-383. Reidel, Dordrecht.

NOZICK, R.(1969) Newcomb's problem and two principles of choice, i Rescher et al (red), Essays in Honor of Carl G. Hempel, 114-146. Reidel, Dordrecht.

RABINOWICZ, W.(1985) Ratificationism without ratification: Jeffrey meets Savage. Theory and Decision, 19, 171-200.