

PER LINDSTRÖM

Utilitarismen och det kontrafaktiska

Enligt *utilitarismen* är en handling (moraliskt) *rätt* om och endast om den uppfyller det *utilitaristiska villkoret* att det, i den givna situationen, inte fanns någon annan handling H, som agenten skulle kunnat utföra, och sådan att om han (hon) utfört H, skulle världen blivit en "bättre" värld. (Det är i termer av hur detta "bättre" definieras som man skiljer mellan de olika formerna av utilitarism.) Om en given handling är rätt eller inte bestäms alltså av sanningsvärdet hos ett antal satser, låt oss kalla dem *H-satser*, av formen:

(*) Om P utfört (handlingen) H, skulle S varit (blivit) fallet.

En sats av denna form och mer allmänt av formen:

(+) Om A varit fallet (inträffat, etc), skulle S varit fallet (inträffat, etc)

sågs vara *kontrafaktisk* (en *kontrafaktisk villkorssats*).

Man kan tänka sig (minst) två olika tolkningar av "skulle" i (+). I den *starka* tolkningen utläses "skulle" som "skulle med nödvändighet", i den *svaga* som "skulle faktiskt".

Den *starka* tolkningen är kanske den mest naturliga men utilitarismen förutsätter den *svaga*. Ty antag att världen inte är deterministisk; det tycks ju för övrigt vara fallet. (Om motsatsen gäller är det ingen relevant skillnad mellan de två tolkningarna; dessutom blir utilitarismen ointressant: alla handlingar är rätt eftersom agentens beteende är determinerat och det alltså inte finns några möjliga alternativa handlingar.) Normalt kommer då det starkaste påstående S för vilket (*) (med den *starka* tolkningen) är sann att vara en disjunktion (S_1 eller ... eller S_n) av (fullständiga) beskrivningar av universums historia. Låt S_0 vara en beskrivning av universums faktiska historia. Det finns då en,

långt ifrån försumbar, risk att det t ex är "bättre" att S_1 än att S_0 som är "bättre" än att S_2 . Men det är då omöjligt att uttala sig om vilket som är bäst, att S_0 eller att S ; dessa båda är ojämförbara. Det går alltså inte att jämföra vad som (med nödvändighet) skulle hänt om P valt H , dvs att S är sann, med vad som faktiskt inträffade. Och därmed skulle det vara omöjligt att tillämpa utilitarismen. (Att jag här och i fortsättningen använder begrepp, som jag i andra sammanhang förklarat vara icke-deskriptiva, som om de vore deskriptiva kan förefalla inkonsekvent, men är det inte: vi befinner oss nu på "ett annat plan".)

För att utilitarismen skall kunna tillämpas på grundval av sanningsvärdet hos H -satser fordras alltså, tycks det, om (eftersom) världen inte är deterministisk, att dessa satser ges den svaga tolkningen. (Om ingenting annat sägs, är det i det följande *den svaga tolkningen* som gäller.)

Innan vi tittar närmare på H -satser skall vi betrakta kontrafaktiska satser av följande typ. Låt R vara en (sann, inte pseudo-) slumpmekanism och låt u vara ett möjligt utfall av R , för vilket sannolikheten är $\neq 0, 1$. Antag att R inte aktiveras vid t och betrakta satsen:

(++) Om R hade aktiverats vid t , skulle R givit utfallet u .

Sådana satser kallar vi *R-satser*. Givet den starka tolkningen är förstås alla R -satser falska. Men givet den svaga?

Vid första påseende kan (++) ge intryck av att vara en ganska ordinär påståendesats som är sann om världen uppvisar vissa egenskaper, falsk annars, men vid närmare betraktande försvinner detta intryck. Om man tänker efter visar det sig, skulle jag vilja påstå, att det inte bara är svårt att tillskriva (++) något sanningsvärde, det är omöjligt. Det är inte bara omöjligt att veta om (++) är sann eller falsk, det finns inget att veta. Det finns ingen egenskap (beskaffenhet) hos världen som svarar mot att (++) är sann. Och att hänvisa till s möjliga världar, i vilka R aktiveras, hjälper inte: i somliga av dessa ger R utfallet u , i andra inte. "Den värld" i "Den värld som skulle blivit verklig om R aktiverats vid t " refererar alltså inte till en bestämd (möjlig) värld; uttrycket saknar referens. Och detta beror inte på att A i (++) , dvs "R aktiveras vid t " skulle vara, i någon mening, ofullständig: denna sats kan ersättas med en fullständig beskrivning av världen vid t , resultatet kommer fortfarande att sakna sanningsvärde. Det är utmärkande för en (sann) slumpmekanism att det saknar mening

att uttala sig om vilket utfall den skulle givit om den aktiverats.

Om (eftersom) världen inte är deterministisk (det finns slumpmekanismer) finns alltså, förefaller det, satser (+), t ex R-satser, som saknar sanningsvärde (trots att det inte är något fel på vare sig A eller S).

Att det finns H-satser som saknar sanningsvärde följer nu (trivialt). Låt t ex H vara handlingen att aktivera R och låt S vara "R ger utfallet u". Det följer då att (*) varken är sann eller falsk. Och alltså kan det, om H var en för P tillgänglig handling, mycket väl vara omöjligt att bedöma den handling P faktiskt utförde med hjälp av det utilitaristiska villkoret.

Men detta resonemang och denna slutsats är kanske alltför "teoretiska" för att göra något större intryck på utilitaristen: det kanske inte spelar någon större roll i praktiken (hur man nu kan veta det) om detta är riktigt eller inte. Det kan därför vara relevant att betrakta satser av formen:

(**) Om A varit fallet, skulle Q utfört H.

där A dessutom kan vara av formen "P utför H". Vi antar då att Q är en agent som, även om A varit fallet (P utfört H'), fritt kan välja mellan att utföra H och att avstå (utföra någon annan handling). Satser av denna typ (med den svaga tolkningen), låt oss kalla dem *H'-satser*, kan inte rimligtvis ignoreras av utilitaristen. (Enligt det utilitaristiska villkoret är alltså huruvida en handling utförd av P vid t är rätt eller inte egendomligt nog avhängigt av hur andra handlar, respektive skulle handlat om ..., vid varje senare tidpunkt.)

Om vi accepterat att R-satser saknar sanningsvärde förefaller det svårt att undgå slutsatsen att detsamma gäller om H'-satser. Därmed inte sagt att fria agenter är (att likna vid) slumpmekanismer. Men hur en fri agent skulle handlat om ... är *obestämt* i samma fundamentala mening som vilket utslag en slumpmekanism skulle givit om den aktiverats.

Jag skall försöka göra denna slutsats lite mer övertygande. Antag först att om A varit fallet, skulle Q haft två handlingar att välja mellan, H_1 och H_2 . Antag vidare att det är Q fullständigt egalt vilken av dessa han väljer, (det omedelbara) resultatet av att välja den ena är, såvitt Q kan förstå, alldeles likvärdigt med (det omedelbara) resultatet att välja den andra och handlingarna själva är precis lika besvärliga (eller lika tilltalande) att utföra. Det finns alltså inget skäl, och inte heller, som

vi normalt tänker oss saken, någon orsak, till att han skulle valt t ex H_1 i stället för H_2 . Men det är då svårt (omöjligt?) att förstå vad det skulle innebära att t ex satsen "Om A varit fallet, så skulle Q utfört H_1 " är sann (eller falsk).

Ett mer konkret exempel av lite annan karaktär är följande:

(#) Om Q vetat att det skulle bli snöstorm, hade han inte tagit bilen.

Q kände alltså inte till väderprognosen, tog bilen, hamnade mitt i snöstormen, tappade kontrollen över bilen, och gav upphov till en trafikolycka. För att något tona ner sitt ansvar för olyckan, påpekar Q att han inte vetat, inte ens haft skäl att misstänka, att det skulle bli svårt väder. Om han vetat det, skulle han naturligtvis låtit bilen stå. Q är alltså själv benägen att tro på (#) ... till att börja med.

För dem som känner Q är saken mindre klar: för dem verkar det lika sannolikt att Q, även om han lyssnat på nyheterna, skulle tittat ut, funnit att vädret fortfarande verkade hyggligt och bestämt sig för att, om det skulle behövas, köra lite, eller mycket, fortare än vanligt för att komma till jobbet innan ovädret bröt ut.

Vem har då rätt, Q eller den som tvivlar? Ingendera, skulle jag vilja påstå. Den situation som Q skulle befunnit sig i, om han känt till att det skulle bli oväder, skulle inte vara tillräckligt utpräglad för att bestämna vilka hänsyn Q, som fri agent, skulle valt att låta väga tyngst.

Och Q märker själv, kan vi tänka oss, när han funderar närmare på saken, att riktigt säker kan han inte vara: han skulle möjligen tagit en chans, men skulle han ändå inte tagit det säkra före det osäkra och låtit bilen stå (och kommit för sent till ett för honom själv högst betydelsefullt sammanträde)? Ju mer han grubblar, desto mer osäker blir han: frågan verkar egendomligt undanglidande, inte ens med den största ansträngning kan han förstå hur den skall kunna avgöras, eller ens vad den innebär. Och skälet till detta är alltså inte att det är svårt (omöjligt) att veta om (#) är sann eller falsk, skälet är, skulle jag vilja påstå, att (#) i själva verket är varken är sann eller falsk.

(Det kan dessutom meddelas att Q's handlande från utilitaristisk synpunkt (kanske) inte var så dåligt som man skulle kunna tro. I den andra i krocken inblandade bilen fann polisen nämligen en tidsinställd bomb. Avsikten var att denna skulle placeras på något lämpligt ställe på Stockholms Central där den skulle detonerat tio minuter senare något som skulle orsakat ojämförligt mycket mer lidande än den re-

lativt obetydliga krocken ... om inget annat kommit emellan, vill säga.)

Meningen är att denna slutsats skall kunna generaliseras till godtyckliga H'-sats, alltså inte enbart de i vilka Q tänks vara vankelmodig: ingen (tänkt) situation är tillräckligt utpräglad för att *bestämma* vilka hänsyn Q skulle valt att låta väga tyngst eller om han alls brytt sig om att väga olika hänsyn (i vanlig mening) mot varandra. Så länge Q är en fri agent (detta förutsätts ju vara sant i en H'-sats), som alltså kan handla som det faller honom in, är det obestämt om (**) är sann eller inte, m a o (**) saknar sanningsvärde.

Påståendet att *alla* H'-sats saknar sanningsvärde kan förefalla uppenbart oriktigt. Skälet är att man normalt förutsätter att agenten i den hypotetiska situationen skulle handlat som han brukar handla i liknande (faktiska) situationer eller att man på grundval av vad man vet (anser sig veta) om agentens karaktär kan sluta sig till hur han skulle betett sig. (Alternativt kan vi tänka oss att (**) är en förkortning av "Om A varit fallet och Q betett sig som vanligt, skulle Q utfört H", en sats som kan vara sann eller falsk.) Men vad har man egentligen för skäl att förutsätta det? Det händer att "folk" beter sig på fullständigt oväntade sätt. Och inte bara det: en agent är, i alla de situationer i vilka han är (i verklig mening) fri, fri att bete sig på ett fullständigt oväntat sätt – även om de flesta av oss praktiskt taget aldrig använder (ser något skäl eller ens kommer på tanken att använda) oss av denna frihet. Skulle då agenten i den hypotetiska situationen betett sig ungefär som vanligt eller på ett oväntat sätt? Såvitt jag kan se har den frågan (i många fall) inget svar och alltså har påståendet att han skulle betett sig på ett speciellt sätt inget sanningsvärde.

Om man, trots att (**) saknar sanningsvärde, meningsfullt, dvs sant eller falskt, kan påstå att Q sannolikt (med en viss sannolikhet) skulle valt H blir beroende av exakt hur påståenden av denna typ skall förstås, något som jag inte kan gå in på.

Vad utilitarismen beträffar innebär allt detta följande: Om det inte finns några fria agenter, kan utilitarismen aldrig tillämpas, annat än (möjligen) på ett alldeles trivialt sätt: alla handlingar är rätt. Antag alltså att det finns fria agenter. För att utilitarismen skall kunna tillämpas fordras då, för det första, att H'-sats (och H-sats) sats ges den svaga tolkningen och, för det andra, att dessa sats har ett bestämt sanningsvärde. Men dessa båda villkor är oförenliga (med ett litet fåtal

undantag som knappt är värda att nämnas). Eller, mer försiktigt, utilitaristen kan inte utan vidare förutsätta att de är förenliga; han (hon?, finns det kvinnliga utilitarister?) måste visa (göra troligt) att så är fallet. Och om de är oförenliga, följer att utilitarismen, oavsett hur världen är beskaffad (och även om man bortser från denna läras praktiska, övriga teoretiska och, inte minst, moraliska defekter) inte är någon särskilt lyckad "teori".