PER LINDSTRÖM

David Lewis' teori för kontrafaktiska satser

En sats sägs vara *kontrafaktisk* (en *kontrafaktisk villkorssats*) om den är av formen:

(*) Om A varit fallet (inträffat, etc), skulle S varit fallet (inträffat, etc).

(Här är A antecedenten och S succedenten i (*).) Sedan Nelson Goodmans klassiska uppsats "The problem of counterfactual conditionals" från 1947 har det stått klart att det inte alls är lätt att redogöra för exakt vad (*) betyder, dvs under vilka omständigheter (*) är sann. Goodman föreslår ingen lösning av problemet. David Lewis, å andra sidan, har lagt fram en teori, dvs en sanningsdefinition, för kontrafaktiska satser i boken *Counterfactuals* (1973). I det följande skall jag försöka visa att Lewis' teori inte är godtagbar. Lewis' teori är i korthet följande. Med varje möjlig värld a, inklusive den faktiska världen, förknippar vi en mängd V_a av möjliga världar, de som (i någon lämplig mening) är möjliga relativt a – a finns alltså med i V_a – och en relation $\leq_a$ på V_a. Formuleringen $b \leq_a c$ betyder att b är minst lika lik a som c. Relationen $\leq_a$ är (antas vara) reflexiv, transitiv, konnex och sådan att $a <_a b$ för alla $b \neq a$. Lewis' teori säger nu att:

(L) (*) är *sann i* a om och endast om antingen (i) A är falsk i alla b i V_a eller (ii) det finns ett b i V_a sådant att A är sann i b och S är sann i alla c i V_a sådana att A är sann i c och $c \leq_a b$.

Här kan A och/eller S (möjligen) själva vara kontrafaktiska satser – (L) skall då uppfattas som en klausul i en induktiv definition av sanning för sådana satser – men jag skall begränsa mig till fallet att A och S är ordinära icke–kontrafaktiska satser.

Tanken är alltså, ungefärligen uttryckt, att (*) är sann i a om antingen (i) detta gäller trivialt, eftersom det i a är omöjligt att A

Filosofisk tidskrift 1997 nr 4, 23-26

skulle vara sann, eller (ii) S är sann i alla världar i V_a i vilka A är sann och som i övrigt är tillräckligt lika a. Om A är sann i a, är (*) sann i a om och endast om S är sann i a.

Om det verkligen finns några s k möjliga världar och vad för slags entiteter dessa i så fall är, dvs hur tal om möjliga världar skall förstås, och om de övriga antaganden som Lewis inför är rimliga, skall jag inte diskutera; frågorna är inte relevanta för mina nuvarande syften. Strategin är snarare att "ge" Lewis så mycket han kan begära, att visa att hans teori inte är godtagbar även om man accepterar hans (mer eller mindre dubiösa) bakgrundsantaganden.

Man kan naturligtvis betvivla att V_a och $\leq_a$ är välbestämda. Men detta är inte nödvändigtvis någon allvarlig svårighet: kontrafaktiska satser har kanske inte alltid en välbestämd betydelse; det enda vi kan kräva är att denna betydelse skall vara entydigt bestämd givet V_a och $\leq_a$, ett villkor som Lewis' teori uppfyller.

Lewis diskuterar i sin bok, som ett exempel på (vad han anser vara) en felaktig slutledning, följande:

(1) (Det är nödvändigt sant att) om P givit sig av klockan fem i morse, så hade han givit sig av före sex.
(2) Om P givit sig av före sex, hade han kommit fram före tolv.
(3) Om P givit sig av klockan fem, hade han kommit fram före tolv.

(1) är förstås sann. (2) skulle, föreslår Lewis, vara sann under följande omständigheter: P gav sig av precis klockan sex, prövade en genväg, som visade sig avkorta den vanliga restiden, åtta timmar, med två timmar, och kom fram precis klockan tolv. Av detta följer, kan vi kanske tänka oss, att (2) är sann. Slutligen kan man tänka sig att (3) är falsk av följande skäl: om P givit sig av klockan fem, hade han varit så sömnig att han inte kommit på tanken att pröva genvägen och resan skulle därför som vanligt tagit åtta timmar och han skulle inte kommit fram förrän klockan ett. (Man kan tycka att detta innebär att den första förekomsten av "före" i (2), men inte de övriga förekomsterna i (1), (2), (3), betyder "före, men inte alltför långt före"; och då är det ju inte så konstigt att (3) inte följer av (1) och (2).)

För att visa att detta harmonierar med (L) tänker sig Lewis att det för varje t sådant att $t > 0$ (och t inte alltför stort) finns en möjlig värld b_t i V_a, där a är den (tänkta) verkliga världen, sådan att P ger sig av klockan t minuter i sex i b_t. I vissa av dessa, bl a de för vilka t ex $t \leq$

15, antas resan ta (ungefär) lika lång tid som i a och P kommer alltså fram klockan t minuter i tolv i b_t. Om $t < t' \leq 15$, antar vi att $b_t <_a b_{t'}$; m a o ju kortare tid t före sex P ger sig av i b_t desto mer lik a är b_t. Vi antar också att om $t > 15$, så $b_{15} \leq_a b_t$ och dessutom att resan tar åtta timmar, och P alltså kommer fram klockan ett, i b_{60}.

Låt oss nu, i likhet med Lewis och för exemplets skull, tänka oss att världarna b_t är (i allt väsentligt) de enda vi behöver ta hänsyn till. Det följer då av (L) att (2) är sann (i a): "P ger sig av före sex" är sann i b_{10} och "P kommer fram före tolv" är sann i alla b sådana att $b \leq_a b_{10}$ i vilka "P ger sig av före sex" är sann, dvs, enligt antagande, de b_t för vilka $0 < t \leq 10$. (3), å andra sidan, är falsk, eftersom "P kommer fram före tolv" är falsk i b_{60}.

(1) och (2) är alltså sanna och (3) är falsk; (3) följer därmed inte av (1) och (2).

Så långt kan det hela kanske verka ganska övertygande. Men betrakta nu satsen:

(4) Om P givit sig av före sex, hade han kommit fram tidigast fem i tolv.

Enligt (L) är detta, under de givna omständigheterna, en sann sats. Ty "P ger sig av före sex" är sann i t ex b_5. Dessutom kommer P fram tidigast fem i tolv i alla b_t sådana att $b_t \leq_a b_5$. Om vi som tidigare tänker oss att världarna b_t är de enda vi behöver ta hänsyn till, följer alltså att (4) är sann.

Men denna slutsats är knappast acceptabel: varför skulle P, om han givit sig av före sex, inte kunnat göra detta, kanske inte klockan fem, men t ex tio i sex med resultatet att han kommit fram tio i tolv? "P gav sig av före sex" skulle då vara sann och "P kom fram tidigast fem i tolv" falsk. Av detta följder att (4) är falsk. En teori som medför att (4), under de givna omständigheterna, är sann kan alltså inte vara annat än felaktig.

Motexemplet (4) kan förenklas. Betrakta satsen:

(5) Om P givit sig av före sex, hade han givit sig av tidigast fem i sex.

Här är antecedenten sann i b_5 och succedenten sann i alla b_t sådana att $b_t \leq_a b_5$. Enligt (L) är alltså (5) sann, under samma förutsättningar som tidigare, något som förstås inte heller det är godtagbart. (Allt vi nu behöver anta om $\leq_a$ är att om $c \leq_a b_5$ så ger sig P av tidigast fem i sex i c.)

Skälet till att Lewis' teori felaktigt klassificerar (4) och (5) som sanna är lätt att förstå. Antecedenten i (2), (4), (5), "P ger sig av före sex", kan realiseras på många (mer än ett) olika sätt: P ger sig av t minuter i sex, t = 1, 2,.... För succedenten "P kommer fram före tolv" i (2) spelar det ingen roll hur antecedenten realiseras, givet att vi begränsar oss till vad vi kan kalla *tillåtna realiseringar* och alltså utesluter bl a t = 60. Men för satserna "P kommer fram tidigast fem i tolv" och "P ger sig av tidigast fem minuter i sex" i (4) respektive (5) är detta av betydelse: i vissa av dessa tillåtna realiseringar är satserna sanna, men i andra, t ex b_{10}, är de falska. Och det är detta senare som medför att (4) och (5) är falska.

Att i (2) inte betrakta b_{60} som en tillåten realisering av "P ger sig av före sex" är kanske inte onaturligt. Men enda möjligheten att, i Lewis' teori, klassificera (4) och (5) som falska är att inte betrakta något b_t för $t \leq 5$ som en tillåten realisering av "P ger sig av före sex", och det verkar omöjligt att motivera. Dessutom kommer i så fall t ex den falska satsen "Om P givit sig av före sex, hade han givit sig av senast fem i sex" att klassificeras som sann. Vi har diskuterat (4) och (5) under vissa (av Lewis introducerade) förenklande antaganden rörande världarna b_t och relationen $\leq_a$. Men det är nu enkelt att åstadkomma ett regelrätt bevis för att Lewis' teori inte kan accepteras. Låt A vara en sats till vilken det finns b, c i V_a sådana att $a <_a b <_a c$ och A är sann i b och c (b och c är tillåtna realiseringar av A) (men falsk i a), t ex satsen "Världen är $\leq_a c$ (men $>_a a$)". Låt sedan S vara en sats som är sann i en värld d om och endast om $d \leq_a b$, t ex "Världen är $\leq_a b$". Enligt (L) är då (*) sann i a. Men detta är inte riktigt eftersom A är sann i c (c är en tillåten realisering av A) medan S är falsk i c.

Som framgår av detta bevis, kan (5) ytterligare förenklas:

(5') Om P givit sig av klockan fem eller (strax före) sex, så hade han givit sig av (strax före) sex.

Denna sats är sann enligt Lewis' teori, men i själva verket falsk.

Lewis' teori fungerar alltså (möjligen) för (*) i de fall där A kan realiseras på, vad S beträffar, i allt väsentligt (högst) ett sätt, men inte annars. Frågan om Lewis' teori kan modifieras till en godtagbar teori för kontrafaktiska satser, eller om det alls finns någon sådan teori, låter jag vara öppen.