FILOSOFISK TIDSKRIFT

NR 4
1985

FILOSOFISK TIDSKRIFT NR 4 • 1985 • ÅRGÅNG 6

Filosofisk tidskrift utges av Stiftelsen Filosofisk tidskrift
och Bokförlaget Thales

Filosofisk tidskrift har som syfte att bidra till en allsidig och fruktbar diskussion av filosofiska problem, samt att på ett lättfattligt sätt informera om aktuell filosofisk forskning. Den vänder sig inte enbart till fackfilosofer, utan vill framför allt nå en bredare läsekrets av filosofiskt intresserade personer.

Tidskriften står öppen för alla filosofiska riktningar, men den vill undvika bidrag som man inte kan tillgodogöra sig utan speciella förkunskaper eller tekniska färdigheter. Utöver längre artiklar på omkring 10—25 sidor, tar tidskriften gärna emot även kortare bidrag och inlägg av notiskaraktär.

Redaktör och ansvarig utgivare: Lars Bergström
Redaktionskommitté: Göran Hermerén (Lund), Göran Lantz (Uppsala), Sven-Eric Liedman (Göteborg), Åke Löfgren (Karlstad), Giuliano Pontara (Stockholm) och Dag Prawitz (Stockholm)
Produktion, prenumerationer och annonser: Bokförlaget Thales, Box 50034, 104 05 Stockholm, redaktionssekreterare Astrid Thorson, tel 08 162618 (arb), 08 7686512 (bost)
Prenumeration: Tidskriften kommer ut med 4 nummer/år, pris SEK 80:—, lösnummer SEK 25:—/ex, postgiro 507991 - 8 Filosofisk tidskrift
Annonspriser: 1/1-sida SEK 250:—, 1/2-sida SEK 150:—, 1/4-sida SEK 100:—
Upplaga: 1 000 ex/nummer
Tryckt hos Holms Gårds tryckeri, Edsbruk 1986

ISSN 0348-7482

Staffan Carlshamre

Världsandens mål

- några sidor om Hegels syn på historien

Vad som följer är en uppsats om Hegels historiefilosofi. Det handlar naturligtvis inte om att ge en systematisk framställning av hela Hegels syn på historien. Min utgångspunkt har helt enkelt varit att försöka "ge god mening" åt två av de drag i denna historiesyn som av moderna läsare oftast uppfattas som märkvärdiga, eller rent av absurda. Det ena av dessa drag är åsikten att de centrala handlande agenterna i historien inte är enskilda människor, utan "större" subjekt sådana som den grekiska eller den germanska "folkanden", eller t o m ett enda subjekt för hela den historiska utvecklingen, "världsanden". Det andra är åsikten att världshistorien som helhet har ett mål, ett sluttillstånd som den "strävar" mot.

Dessa teser är naturligtvis inte oberoende av varandra - att se ett förlopp som målinriktat är ett väsentligt inslag i att se det som producerat av en enda rationell agent - men jag skall behandla dem var för sig. För att göra dessa åsikter begripliga måste jag dock börja med att säga något om vad Hegel kan ha menat med att världshistorien är ett "andligt" förlopp. Härav uppsatsens disposition i tre huvudavsnitt. Några egentliga exegetiska anspråk gör jag inte - det handlar mera om att ge rimliga tolkningsförslag än att argumentera med omfattande textstöd för att min tolkning är "den korrekta".

1. Historien utspelar sig på andlig mark

En lämplig utgångspunkt för en skildring av Hegels historiefilosofi är ett påstående från Förnuftet i historien: "Historien utspelar sig på andlig mark". Vad betyder det?

För den ytlige betraktaren kan det förefalla som om historiker behandlar de mest olikartade företeelser. Några av dem, idéhistorikerna, skriver mycket riktigt om andliga ting, om tankar, men det överväldigande flertalet verkar skildra helt andra sorters saker. Det handlar om ond bråd död på slagfältet, om jordbruksmetoder i forn-

tida kulturer, om utbredningen av bruket av stigbygel från Indien till Europa under den tidiga medeltiden osv.

Denna mångfald är enligt Hegel bara skenbar. I själva verket är all historia en tänkandets historia, och alla historiker idéhistoriker.

Odysseus återvänder till Ithaka och dödar friarna. Varför? Hur förklarar man en (föregiven i det här fallet) historisk händelse? Ett förslag är detta: han slår ihjäl dem för att de är friare.

Men antag att de verkliga friarna, Odysseus ovetande, har gått hem var och en till sitt för att sköta olika privata bestyr, medan de personer som nu finns i salen, och som Odysseus tar livet av, i själva verket är en samling hedervärda unga män som just anlänt för att trösta Penelope och erbjuda Telemachos sin hjälp i hans utsatta situation. I så fall är ju den föreslagna förklaringen inte längre giltig. De personer som ligger och badar i sitt blod är inte friare, och kan följaktligen inte vara ihjälslagna därför att de är friare. Ändå tycker man intuitivt att det är samma faktorer som styr Odysseus handlande i båda de tänkta situationerna, och slutsatsen ligger nära till hands: Odysseus slår inte ihjäl friarna för att de är friare, utan för att han *tror* att de är friare.

Förklaringen är givetvis inte fullständig. Kommer jag hem till min lägenhet efter en utrikes resa och finner den belamrad med friare så blir jag inte glad, men jag slår inte ihjäl dem. Ändå kan vi anta att jag har precis den trosföreställning som i Odysseus fall skulle förklara hans brutala handling.

Skillnaden är naturligtvis - om vi bortser från den rent fysiska förmågan att genomföra massmordet - att även om jag och Odysseus har de relevanta teoretiska föreställningarna om verkligheten gemensamma, så har vi olika normer och värderingar. Vi har inte samma åsikter om hur friare bör behandlas. Hans värderingar är den grekiska krigaradelns, medan mina är produkten av tusen år av kristendom och humanism.

Gör vi detta tillägg, och tillåter oss att ta en mängd andra faktorer för givna (det här är ett exempel, inget argument!) så kan vi säga att vi har fått en förklaring till mordet på friarna som enbart tar hänsyn till "andliga" faktorer, nämligen tankar hos Odysseus. Närmare bestämt har vi urskilt två olika sorters "tankar" som båda måste till för att förklaringen skall bli fullständig: å ena sidan föreställningar om världens faktiska beskaffenhet, vad Odysseus tror om den situation han befin-

ner sig i, och å andra sidan en uppsättning normer och värderingar beträffande vad som bör göras i en sådan situation.

Mer abstrakt kan man formulera Hegels åsikt så här: Historien är inte bara en följd av händelser vilka som helst, som historikern skall uppspåra orsakerna till - i den tidigare historien, i "yttre" omständigheter av olika slag, i det mänskliga psyket o s v - utan en serie mänskliga handlingar, som förklaras genom att vi lyckas rekonstruera de skäl eller motiv som ligger bakom dem.

Men även om man skulle vara villig att ge Hegel hans poäng beträffande förklaringar till historiska händelser, så intresserar sig ju historikern inte bara för förklaringar. Han vill också veta något om händelserna själva, resultaten av de historiska handlingarna. Hur andliga de faktorer än är som styr Odysseus handlande så är blodbadet verkligt - och kan väl inte reduceras till tankar hos de inblandade agenterna?

Vill Hegel bemöta den invändningen så svarar han förmodligen med en motfråga. Historikern intresserar sig inte för allt som har hänt i det förflutna. Alla händelser är inte historiska händelser. Vad gör en händelse historisk? Rimligtvis att den har betydelse för det fortsatta historiska förloppet — d v s att den kommer att ingå i förklaringen av senare historiska händelser. Men om det vi har sagt om historiska förklaringar är riktigt så betyder det att mordet på friarna har historiskt intresse bara om det kommer att ingå i senare historiska agenters världsuppfattning, och på så sätt påverka deras handlingar.

I det perspektivet bortfaller skillnaden mellan ett verkligt och ett imaginärt blodbad som historiskt irrelevant. Det enda som betyder något är att det *betraktas* som verkligt.

Ett annat, och för Hegel viktigare, problem är att resultatet av en handling inte alltid blir det av agenten avsedda. Detta kan t ex bero på den materiella världens ovillighet att foga sig efter våra önskningar - vad hade hänt om Odysseus båge gått och blivit skör under hans långa frånvaro?

Hegel vill säkerligen inte helt förneka betydelsen av sådana faktorer, utan bara hävda att de saknar vikt för det historiska förloppet i stort. Åtminstone i den mänskliga historiens tidsperspektiv kan naturen ses som en konstant - skiftar de naturliga förutsättningarna från en tidpunkt till en annan är det som resultat av mänskliga ingripanden.

(Enligt Hegel finns det inte någon utveckling i egentlig mening i naturen, alla naturförlopp är cykliska.)

Men att våra handlingar inte alltid ger de resultat vi tänkt oss kan också bero på att många människors handlingar samverkar, och tillsammans ger ett resultat som ingen av de handlande avsett. Tänk t ex på sådana fenomen som konjunkturcykler och börskrascher – eller på när hundra personer samtidigt hoppar ner i en livbåt, med påföljd att båten sjunker.

Sådana komplikationer behandlar Hegel under rubriken "förnuftets list", men jag lämnar dem för tillfället åt sidan.

2. Vem är det som handlar i historien?

Som jag har tolkat Hegels påstående att världshistorien utspelar sig på andlig mark betyder det alltså att historikerns ämne är människors handlingar, förklarade i termer av deras föreställningar och motiv. Även om vi bortser från hur pass rimlig en sådan uppfattning är i sak, så finns det en svårighet med att tillskriva just Hegel den, som jag hittills inte har berört.

Enligt Hegel är ju historien inte alls enskilda människors verk. Den som handlar i historien, och vars bevekelsegrunder historikern försöker förstå, är i själva verket en enda överindividuell agent, "världsanden", som under tidernas lopp i tur och ordning antar gestalten av ett antal olika "folkandar". Visserligen kan världsanden inte förverkliga sina mål annat än via enstaka människors ansträngningar, men det är ändå *världsandens* mål som förklarar historiens gång, inte deras.

Speciellt inom den huvudtradition i brittisk filosofi som är empiristisk i kunskapsteorin, utilitaristisk i moralfilosofin och liberal i politiska sammanhang, har det ofta ansetts vara en elementär filosofisk insikt att det egentligen bara är enskilda människor som gör historia. "Nationer" och "klasser" är bara bekväma fiktioner som man tar till för att åstadkomma en hanterbar beskrivning, när de verkliga förhållandena blir så komplicerade att de inte på ett rimligt sätt låter sig omtalas direkt. Ur det perspektivet blir Hegel och Marx paradexempel på tänkare som begått misstaget att "reifiera" dessa fiktiva entiteter.

Denna individualistiska tradition är en utmärkt utgångspunkt om man vill hitta det korn av sanning som ligger i Hegels motsatta upp-

fattning. Läser man på andra ställen i t ex Humes eller Russells skrifter så finner man ju att de betraktar också enskilda personer som fiktioner. En individ är egentligen ingenting annat än "a bundle of perceptions", *en samling tankar* i vidaste mening.

Varför löper man då inte linan ända ut och säger, att det är de enstaka tankarna som är de historiska agenterna? Varför inte säga att det var Odysseus vrede som slog ihjäl friarna, snarare än "hela" Odysseus? (Det har ofta hävdats att det snarast är så Homeros själv ser på saken. Se t ex Dobbs, kap 1, eller Feyerabend, s 205 ff.)

Svaret på den frågan är naturligtvis att de tankar som tillsammans utgör en person i praktiken inte är så oberoende och självständiga i förhållande till varandra som Hume ibland tycks tänka sig. Jag skall beröra två sätt på vilka en människas tankar, och därmed de av dessa tankar motiverade handlingarna, i allmänhet förefaller höra samman med varandra. (Jag menar naturligtvis inte att just dessa två är de enda, eller ens de viktigaste, faktorerna i sammanhanget. I själva verket kan man ta nästan vilken som helst av de faktorer som förespråkare för olika typer av "psykologisk kontinuitet" som svar på frågan vari "personlig identitet" består, alltsedan Locke och framåt, har dragit fram som väsentliga för personers enhet över tiden och generalisera dem till hela samhällen. Minnets roll, t ex, för att konstituera en person motsvaras av den historiska medvetenhet som bl a historikern står för.)

Den första punkten är den enklaste. När det gäller människor jag känner behöver jag ofta inte alls känna till några detaljer om vad som i en given situation rör sig i deras själsliv, för att kunna förutsäga hur de kommer att handla. En människas tankar och handlingar förenas av att hon över ganska långa tidsrymder behåller en viss "karaktär" – hon känner och gör liknande saker i situationer som liknar varandra tillräckligt mycket. Men även mellan olika typer av situationer bevarar hon en viss enhetlighet i uppträdande och reaktionsmönster, en "stil". Detta resonemang kan uppenbarligen generaliseras till större sammanhang än enstaka personer. För att förutsäga att jag och Odysseus kommer att bete oss olika i situationen med friarna behöver man inte ha tillgång till några detaljerade informationer om någon av oss. De räcker ganska långt att veta att han är grekisk kung vid tiden för trojanska kriget, medan jag är svensk akademiker vid slutet av 1900-

talet. Man talar om "nationalkaraktär" - Hegel talar i stället om "folkandar". Vi är formade av vårt historiska ursprung i en viss tid och en viss social omgivning i så hög grad, att våra personliga idiosynkrasier ofta spelar liten roll för att bestämma hur vi handlar. (I Eyvind Johnsons "Strändernas svall" får vi se Odysseus närmast mot sin egen vilja utföra den handling som krävs av honom.)

Den speciella enhetlighet hos en människas tankar och handlingar som vi kallar karaktär är alltså, tror jag, en av de faktorer som får oss att betrakta "hela" människor, snarare än enstaka sinnestillstånd, som upphov till handlingarna ifråga. Ett analogt resonemang får Hegel att postulera ännu större handlande agenter.

Den andra punkt jag vill anföra i sammanhanget har att göra med vad Collingwood kallar "presuppositioner".

När Hume hävdar att en person inte är något annat än "en samling tankar", så bygger han på åsikten att var och en av de "tankar", som tillsammans utgör hela personen ifråga, egentligen är *självständig*. Det gäller t ex för vilken som helst av mina tankar att det kunde finnas ett själsliv som bara innehöll denna enda tanke. Den typ av "tankar" som Hume själv haft i tankarna är förmodligen saker i stil med färgupplevelser. Tänk dig ditt synfält fullkomlig blått - vore det inte möjligt att du under hela ditt liv bara hade denna enda upplevelse? Andra upplevelser skiljer sig enligt Hume från denna blåvision endast med avseende på styrka och komplexitet.

För att väcka litet misstänksamhet mot denna doktrin om tankars radikala självständighet räcker det att ett ögonblick kontemplera ett mera normalt exempel på en tanke. Ta den här t ex: "Jag undrar om jag hinner in på banken före klockan tre, eller om jag måste försöka låna en hundralapp av någon tills i morgon?". Är det verkligen möjligt att tänka sig ett själsliv som under hela sin existens bara innehåller denna enda tanke?

Poängen med exemplet är naturligtvis att vissa tankar förutsätter andra tankar. Den som inte haft andra tankar - förknippade med pengar, andra människor, tidmätning, sociala institutioner av olika slag o s v - kan inte förstå frågan i exemplet. Mitt själsliv är inte bara en följd av tankar i godtycklig ordning. Senare tankar "presupponerar" tidigare tankar — de vore "otänkbara" utan dem.

Mig förefaller det uppenbart att dessa presuppositionsförhållanden

är en annan del av bakgrunden till att vi räknar med att tankars och handlingars subjekt är i tiden utsträckta personer. Att det är på det viset är i sin tur förbundet med en annan sak som är väsentlig för Hegel — det faktum att människor är "bildbara". Att vara "pedagogisk" är till stor del helt enkelt att presentera tankar i rätt ordning.

Den poäng jag är ute efter för Hegels del bör vara uppenbar. På samma (nåja) sätt som mina nuvarande tankar vore otänkbara utan mina tidigare tankar, så är de i stor utsträckning otänkbara utan de tidigare tankar som ingår i den tradition till vilken jag hör. Det är på något sätt mera naturligt att filosofihistorien börjar med Thales och slutar med Kant och Hegel än tvärtom. Man kan säga att detta är den hegelianska filosofins fundamentala antagande: den ordning i vilken tankar faktiskt tänks är inte godtycklig, utan i någon mening "logisk" Dialektiken skall vara teorin om denna ordning.

Godtar man det här resonemanget så godtar man också att de tankar som i en mening självfallet är mina tankar, i en annan mening inte är det. Det är, om man så vill, traditionen som tänker i mig - och den traditionen, utsträckt så långt det går åt båda hållen i tiden, är det som Hegel kallar världsanden.

Utrustade med denna förståelse av vem världsanden är kan vi kasta litet ytterligare ljus över folkandarna, de historiska agenter som befinner sig på nivån mellan världsanden och oss själva.

Den moderna vetenskapsfilosofin, t ex hos Thomas Kuhn, har på nytt gjort oss förtrogna med tanken att kunskapens utveckling inte bara är ett kontinuerligt förlopp, där upptäckt läggs till upptäckt, utan att den innehåller omvälvningar och diskontinuiteter, då ett helt tankesystem ersätts med ett annat, så att själva ramarna för kunskapssökandet ändras. Det är i ljuset av en sådan uppfattning som Hegels bild av historien som en följd av folkandar skall ses. Även världshistorien är ju för honom en tankehistoria. Det är bara det att det där inte i första hand handlar om teoretiska tankesystem, utan om de praktiska - moraliska - föreställningarnas utveckling. Folkandarna är, om man så vill, det moraliska tänkandets paradigmer.

3. Historiens mål

Att se en serie händelser som uttryck för en rationell agents vilja är bl a att se dem som led i en strävan att förverkliga ett visst mål. Eftersom

världshistorien för Hegel är produkten av en förnuftig agents strävanden, så är det naturligt för honom att tänka sig att den har ett sådant mål — ett tillstånd som kommer att realiseras i slutet av den historiska processen om inga yttre omständigheter lägger hinder i vägen.

Hegel karaktäriserar detta mål på olika sätt i olika sammanhang. Jag skall intressera mig för två sådana karaktäriseringar.

Den ena är konkret och ostensiv. Hegel pekar ut ett visst samhälle vid en viss tidpunkt som det mål mot vilket historien strävar - nämligen det samhälle han själv lever i. För korthetens skull kommer jag att beteckna det måltillståndet som "Preussen 1830".

Den andra bestämningen är mer abstrakt - Hegel beskriver det samhälle som avslutar historien, utan att peka ut det. Historien slutar med världsandens insikt om sin frihet, med insikten att "människan som människa är fri, att andens frihet utgör hennes innersta natur".

För Hegel sammanfaller de båda beskrivningarna. Preussen 1830 är den plats där världsanden först inser att hon är fri. Den åsikten har inte många varit beredda att dela med honom. (Ett undantag är Alexandre Kojève, fast han förlägger historiens slut redan till 1807 — det ögonblick då Hegel avslutar Phänomenologie des Geistes samtidigt som han hör Napoleons kanoner inleda slaget vid Jena.) Det är lätt att se att valet mellan de båda alternativen kan bilda utgångspunkt för två olika politiska tolkningar av Hegel - en konservativ och en radikal.

Jag skall inte försöka försvara Hegels höga uppskattning av det samhälle i vilket han själv levde. Däremot skall jag försöka visa att båda karaktäriseringarna av historiens mål var för sig har en del som talar för sig.

Nyckeln till varför Hegel ansåg att historien slutar i Preussen 1830 är naturligtvis att Hegel omfattade denna åsikt just i Preussen 1830. Vad Hegel egentligen sa var att historien slutar nu!

För att i sin tur förstå varför han trodde detta gäller det att erinra sig att historien för Hegel inte i första hand är summan av allt som har hänt i det förflutna, utan det som historikern, och speciellt världshistorikern, beskriver när hon skriver historia. (Dubbeltydigheten hos termen "historia" - att den både betecknar historievetenskapen och dess objekt - är ofta nyttig att hålla i minnet när man läser Hegel. För efter-

kantianen Hegel står dessa saker i ett intimt samband. Det som intresserar honom är inte historien som "ting-i-sig" utan historien som "fenomen" — och denna historia är definitionsmässigt sådan som en ideal historiker uppfattar den. Drag som varje historiker måste tillskriva sitt objekt, tillkommer också apriori detta objekt. De allmänna filosofiska sanningarna om historien är "transcendentala" i Kants mening.)

Att säga att historien har nuet som mål är alltså att säga att historikern med nödvändighet måste beskriva världshistorien som en målinriktad process som slutar i nuet. För den åsikten argumenterar man lämpligen i två steg: 1) Varför måste historikern se historien som målinriktad?, 2) Varför måste målet just vara nuet?

För det första alltså: varför måste man se historien som riktad mot ett mål? Av ett mycket enkelt skäl som vi redan berört. Den historia historikern skriver omfattar inte, och kan inte omfatta, allt som hänt i det förflutna. Historikern skriver bara om de "historiska" händelserna, och ett rimligt sätt att bestämma dem är att de historiska händelserna är de som har betydelse för framtiden, d v s för den fortsatta historien.

För att kunna använda detta kriterium måste man ha en åsikt om hur denna framtid ser ut. Av naturliga skäl kan ingen veta hur "hela" framtiden är beskaffad. Man får välja en punkt i tiden och från den punkten ser man sedan den dittillsvarande historien. Skriver man om det romerska rikets nedgång och fall så blir precis de händelser historiska, som bidrar till - eller på annat sätt påverkar - detta fall.

Det är lätt att se att historien från en sådan punkt kommer att te sig just som en kedja av händelser med denna punkt som mål. (Här är det på sin plats att erinra sig det berömda uttalandet om att "Minervas uggla lyfter först i skymningen" - historisk nödvändighet hos Hegel är nästan alltid "bakåtriktad". Inte: när förutsättningarna var si och så så måste detta resultat följa, utan: detta resultat kunde bara ha uppnåtts på detta sätt.) Svaret på den första frågan är alltså att bara den som ser historien som målinriktad, d v s väljer sina historiska händelser med en viss efterklokhet, har ett kriterium på vad som är en historisk händelse.

Men varför måste detta mål just vara nuet? Vad är det för fel på romerska rikets nedgång och fall? Till detta finns det två saker att säga, en mycket enkel och en något litet djupsinnigare.

Den enkla synpunkten är denna: världshistorien är definitionsmässigt *hela* den hittillsvarande historien, alltså kan ingen som valt sin utgångspunkt någon annanstans än i nuet göra anspråk på att skriva världshistorien.

Den andra synpunkten är denna. Det räcker inte att välja en viss tidpunkt för att ha en utgångspunkt för historieskrivning. Tidpunkten måste också ses ur någon viss aspekt. Den aspekten kan vi beskriva: tag t ex industrialismens framväxt i Sverige fram till 1900. Det är bara det att när vi jämför två olika sådana historier från två olika tidpunkter, så kommer vi att finna att de, även om de har samma ämne, väljer olika i det historiska materialet. Varför?

Svaret är väl att det inte räcker att historikern ser sitt material från en viss utgångspunkt. Han måste också se utgångspunkten själv och bestämma vad det är hos dem som kräver en historisk förklaring. På så sätt kommer hans eget nu tillbaka. Vi ser inte industrialismen på samma sätt nu som vid sekelskiftet, beroende på att den påverkat vår livssituation på sätt som sekelskifteshistorikern inte kunde förutse. Vår samtidsbestämda syn kommer att påverka också vår beskrivning av skeenden som helt och hållet ligger i det förflutna.

Det är dags att övergå till Hegels andra karaktärisering av historiens mål: historien slutar när människoanden inser att hon är fri.

När det gällde Preussen 1830 hade vi bara ett problem att lösa: varför trodde Hegel att just detta samhälle var historiens mål? När det gäller den abstraktare beskrivningen får vi två frågor att besvara. 1) Hur ser världen och samhället ut när människoanden insett sin frihet? 2) Vad finns det för anledning att tro att historien rör sig mot just detta tillstånd?

Vi tar frågorna i tur och ordning, och närmar oss den första genom att ge den en annan formulering med hjälp av Hegels tekniska terminologi: historiens mål är "alienationens" upphörande. Vad betyder det?

Mänskligt tänkande finns för Hegel i två uppenbarelseformer: subjektivt och objektivt. Subjektiv är min tanke så länge den bara finns i min själ, objektiv blir den när den fått ett för mig själv och andra igenkännligt yttre uttryck. Vill jag snickra ett bord så finns tanken på bordet till en början endast subjektivt hos mig, den är "bara" en tanke. När bordet är färdigt så finns, förhoppningsvis, samma tanke förverkligad i t ex trä.

Att forma ett stycke materia i enlighet med sin tanke är dock bara ett sätt att objektifiera den. Konst, litteratur och musik är också objektifierade tankar, "objektiv ande". Lagstiftning och sedelag är objektiva uttryck för ett samhälles ande, dess folkande. Ett religiöst system är också objektifierade tankar - skapade ur subjektiva tankar, men med tiden uppfattade som en form av yttre realitet.

Objektifiering i denna mening är på intet sätt något negativt för Hegel. All andlig utveckling är beroende av objektifiering. Möjligheten att känna igen sin tanke i ett yttre föremål är en av de saker som skiljer Slaven från Herren i den viktiga Herre/Slav-dialektiken, och som ger honom ett försteg på vägen till att bli en fullvärdig människa. Problemet är bara att denna objektifiering (Entäusserung) historiskt sett är nära förbunden med ett annat fenomen: alienation (Entfremdung).

Vad det rör sig om inses lättast i fallet religion. Ett folks religion är för Hegel dess egen skapelse, i själva verket det viktigaste och tydligaste uttrycket för dess ande. "Filosofiskt" sett är det också lätt att uppfatta religionen som den religiösa människans skapelse. Men det är inte så det ser ut från den troendes eget perspektiv. Denne uppfattar sin gud som en yttre makt, en överordnad att lyda och frukta. Den troende människan har blivit slav under sin egen skapelse, därför att hon inte känner igen den *som* sin skapelse.

Ett annat exempel, som vi behöver i det följande, är lagstiftning. Både för sin genes och för sin fortsatta existens som lag är Sveriges Rikes lag uppenbarligen beroende av svenskars tänkande i lagfrågor — den dag ett lämpligt urval av alla svenskar slutar uppfatta en lag som giltig, så slutar den också i realiteten att gälla.

Men lagen möter inte den enskilde svensken som produkt av hans tänkande. Den möter honom som en yttre tvingande kraft - bryter han mot den så möts han med våld.

Ett tredje exempel, mera marxistiskt än hegelianskt inspirerat, är konjunkturcykler. Från ett abstrakt filosofiskt perspektiv är det uppenbart att hög- och lågkonjunkturer är produkter av enskilda människors beslut och handlingar — rörande produktion, konsumtion och sparande i första hand. Men för den enstaka beslutsfattaren ser det annorlunda ut. Han möter den kommande lågkonjunkturen som ett naturfenomen som han måste anpassa sig till för att inte gå under,

samtidigt som han genom just denna "anpassning" är med om att producera fenomenet. (Ett nästan parodiskt exempel på denna tendens att uppfatta trender som objektiva, kausalt verksamma, storheter möter man i användningen av begreppet "valvind" för att förklara röstresultaten på valnatten.)

I sådana exempel möter vi den paradox som är kärnan i Hegels föreställning om historien som en rörelse mot frihetsmedvetande. I en mening är människan av naturen fri, och historien är i sin helhet människans verk. Men samtidigt tror hon sig driven av nödvändiga opersonliga krafter – hon uppfattar inte sin frihet, och just genom detta blir hon faktiskt också ofri. Frihet förutsätter frihetsmedvetande.

Att människoanden blir medveten om sin frihet innebär alltså att alienationen upphör – att vi lär oss att känna igen historien och samhället med dess institutioner som våra egna produkter. Det innebär, för att tala med Marx, "steget från nödvändighetens till frihetens rike". Där vi förut skapat historia utan att veta om det kan vi nu göra vår framtid till föremål för medvetna förnuftiga beslut.

Framställd på det sättet låter Hegel ännu mer som en "idealist" (i den speciella pejorativa marxistiska bemärkelsen) än vad han i själva verket är. Behövs det verkligen inte mer för att avsluta historien och träda in i frihetens rike än en förändring i tanken? Räcker det att dela ut en pamflett med Hegels historiefilosofi?

Naturligtvis inte – och det tror inte Hegel heller. Alienationen är inte bara ett sken. Vad det handlar om illustreras lättast med lagstiftningsexemplet.

Det finns förmodligen i varje historiskt samhälle en mängd lagar och bestämmelser som av många medborgare betraktas som orättfärdiga. Insikten att dessa lagar är produkter av "människoandens" aktivitet minskar på intet vis deras förmåga att tvinga medborgarna till lydnad – bryter jag mot dem så hamnar jag lika fullt i fängelse. Fängslas jag för att ha brutit mot en lag som jag själv uppfattar som orättfärdig, så är min upplevelse av lagen som "tvingande" naturligtvis inte bara en upplevelse – den motsvaras faktiskt av ett yttre tvång.

Hegels poäng är att det finns lagar som fungerar på ett annat sätt. Jag uppfattar t ex, liksom förmodligen de flesta andra svenskar, förbudet mot rattfylleri som en fullt rimlig bestämmelse, och stöder också att detta förbud backas upp med ganska hårda straff. Antag nu att jag

själv kommer att bryta mot detta förbud. Naturligtvis vore jag också i detta fall gladast om jag inte åkte fast - men om jag faktiskt blev gripen, dömd och placerad i fängsligt förvar så skulle jag inte reagera på detta på samma sätt som om jag gripits t ex för vad jag själv betraktar som rimliga politiska aktiviteter. Jag skulle - åtminstone i eftertänksamma stunder - acceptera straffet, och tvärtom betrakta det som orättfärdigt om jag inte straffades. Det är, om man så vill, jag själv som sätter mig i fängelse.

Det finns alltså ur det här perspektivet två sorters lagar. (a) Sådana som jag kan identifiera mig med, och vars tillämpning jag stöder - även om de skulle drabba mig själv. (b) Sådana som jag uppfattar som orättfärdiga, och som man måste tvinga mig att lyda.

I alla historiskt givna samhällen finns det förmodligen lagar eller regler med vilka vissa grupper av "medborgare" inte kan identifiera sig - ens efter lång och noggrann begrundan. Det kan röra sig om kvarlevor av tidigare samhällstillstånd, som nu verkar som onödiga frihetsinskränkningar på alla medborgare, men det kan också handla om bestämmelser som uttrycker ett eller annat *särintresse* i samhället. Någon grupp eller person har haft möjlighet att till sin egen fördel påtvinga samhället som helhet vissa regler.

Slutsats: för att alienationen skall upphöra räcker det inte att vi slutar *betrakta* lagstiftningen som en yttre tvingade kraft. Den måste också sluta *vara* en sådan kraft. Det tillståndet infinner sig i ett samhälle där lagstiftningen inte till någon del uttrycker enstaka gruppers särintressen, utan där varje medborgare, efter reflektion, kan identifera sig med hela lagstiftningen. Rätt och rättsmedvetande befinner sig i fullkomlig samklang — "lagstiftningen uttrycker människans väsen som människa".

Men även om vi till nöds tycker oss ha förstått vad Hegel menar med att människoanden kommer till insikt om sin frihet, och kanske t o m instämmer med honom i att ett samhällstillstånd av den art han skisserar vore i hög grad önskvärt - vad finns det för anledning att tro att den historiska utvecklingen kommer att leda till precis detta tillstånd?

Såvitt jag förstår har Hegel inga goda skäl för denna utvecklingsoptimism. Visserligen argumenterar han vidlyftigt för att vi måste se historien som en "logisk" och målinriktad process, men denna

målinriktning är ju, som han själv medger, helt enkelt en produkt av efterklokhet. Det finns ingenting i Hegels syn på historien som motiverar ett anspråk på att kunna förutsäga framtiden.

Men även om Hegel inte ger oss anledning att tro att det alienationsfria samhället i temporal mening är historiens mål, så tror jag att man ganska lätt kan se varför den som ser historien på Hegels sätt måste tillskriva detta samhällstillstånd en *särställning* bland alla möjliga tillstånd. I det alienationsfria samhället uppenbaras en sanning om historien som, trots att den gäller i alla samhällen, bara är synlig med hjälp av föreställningen om samhället utan alienation. Det alienationsfria samhället är *kunskapsteoretiskt* centralt för Hegel.

Vilken är då den sanning om historien som uppenbaras när alienationen upphör? Den sanning vi började med: att historien utspelar sig "på andlig mark" - att det är människor som genom sina handlingar och beslut skapar historien, och på så sätt ger ett objektivt uttryck för sina tankar, sina föreställningar och värderingar.

En analog till det här synsättet finner vi i Marx syn på det kapitalistiska samhället. Det är enligt Marx sant om alla samhällen att de ekonomiska förhållandena är grundläggande både för samhällets struktur och dynamik - det är de som förklarar samhällets utseende i övrigt, samtidigt som de fungerar som utvecklingens "motor". Men det är först i det kapitalistiska samhället som de ekonomiska faktorernas primat blir uppenbar. I tidigare samhällen är den fördold - något som ofta får förklara varför den marxistiska historieteorin utvecklas just under kapitalismen.

Den marxska analogin ger oss kanske också ett skäl för Hegel att identifiera sitt idealtillstånd med det samhälle han själv levde i. Om kunskapen om historien bara är möjlig i det ideala samhället, och jag faktiskt besitter den kunskapen, så måste jag ju leva i det ideala samhället! Vi stöter ånyo på Hegels förnekande av tankens förmåga att "transcendera" det verkliga - att bedöma nuet och det förflutna genom att jämföra dem med ett blott tänkt idealtillstånd.

Litteratur

Collingwood, R G: *An Autobiography,* Clarendon Press 1939.
Dodds, E R: *The Greeks and the Irrational,* University of California Press, 1951.
Feyerabend, P: *Ned med metodologin,* Zenit/Rabén & Sjögren 1977.
Hegel, G W F: *Vorlesungen über die Philosophie der Geschichte,* band XI *Sämtliche Werke,* Jubileumsausgabe in XX Bänden, neu hrsg von Hermann Glockner, Stuttgart 1927–30.
Hegel, G W F: *Phänomenologie des Geistes,* hrsg von J Hoffmeister, Felix Meiner Verlag, sechste Auflage, 1952.
Kojève, A: *Introduction to the Reading of Hegel. Lectures on the Phenomenology of Spirit,* Cornell UP, 1980.
Taylor, C: *Hegel,* Cambridge UP, 1975.

Wlodzimierz Rabinowicz

Om ratifikationismen. Kritik av Jeffreys nya "beslutslogik"

1. Inledning

Denna uppsats behandlar en beslutsteoretisk fråga. När man talar om beslutsteori, är det i själva verket ingen bestämd teori som man normalt åsyftar. Termen ifråga fungerar i stället som en samlande benämning på en hel flora av teorier om beslutsfattande. Av dessa har en del mer eller mindre beskrivande karaktär: deras syfte är att modellera beslutsprocesser och därigenom möjliggöra beslutsförklaringar och beslutsförutsägelser. Andra har i stället en normativ funktion: syftet blir då att tala om hur vi bör fatta våra beslut eller att ange generella villkor som ett beslut måste uppfylla för att vara korrekt eller "rationellt". (Gränsen mellan normer och idealiserade beskrivningar kan emellertid vara ganska flytande.)

Med "ratifikationismen" eller "ratifierbarhetsmaximen" avses här den normativa beslutsprincip som har formulerats av Richard Jeffrey i den senaste utgåvan av hans *Logic of Decision* (1983). Min avsikt är att förklara Jeffreys nya princip och ifrågasätta den.[1]

Först en kort bakgrundshistoria. I första upplagan av sin bok, som kom 1965, försvarade Jeffrey en beslutsprincip enligt vilken agentens handlingsbeslut är korrekt (rationellt, om man så vill) om, och endast om, den förväntade "önskvärdheten" hos den beslutade handlingen är maximal. D v s maximal i jämförelse med andra handlingsalternativ som står agenten till buds. Den förväntade önskvärdheten, eller nyttan, hos en handling definieras som en viktad summa av handlingens önskvärdhetsvärden (nyttovärden) under olika möjliga sakernas tillstånd. Som vikter använde Jeffrey betingade sannolikheter för de olika tillstånden givet handlingen ifråga. De jeffreyska vikterna hade alltså formen

$P(S/A)$,

där S är (beskrivningen av) ett möjligt tillstånd, A är (beskrivningen av) den handling som är föremål för bedömningen, och där P (S/A) - den betingade sannolikheten för S givet A - definieras, som vanligt, som en proportion mellan sannolikheten för S och A och sannolikheten för A:

$$\frac{P(S\&A)}{P(A)}.$$

Observera att $P(S/A)$ är definierad endast i de fall då $P(A)$ är större än noll. Det är också nödvändigt att påpeka att de relevanta sannolikheterna här ges en 'subjektiv' tolkning: de är mått på styrkan i agentens övertygelser. Dessa sannolikheter varierar mellan ett - absolut säkerhet - och noll - absolut misstro (absolut säkerhet på motsatsen). Funktionen P antas representera agentens övertygelser omedelbart före valögonblicket.

Jeffreys ansats från 1965 bör lämpligen kontrasteras med den klassiska approach som utvecklades av Leonard Savage i *The Foundations of Statistics* (1954). Den för oss mest intressanta skillnaden mellan Jeffrey och Savage ligger i att den senare som sina vikter väljer 'obetingade' sannolikheter för tillstånden:

$P(S)$

I fortsättningen skall jag tala om handlingens "Savagevärde" för att beteckna dess förväntade nytta beräknad enligt Savages recept: i termer av de obetingade sannolikheterna.

Jeffrey försvarade sin ansats genom att påpeka att de för utfallet relevanta tillstånden ibland kan vara mer eller mindre beroende av de handlingar som är föremål för bedömning. Och agenten kan vara medveten om detta beroendeförhållande. I sådana fall skulle valet av obetingade sannolikheter leda agenten på avvägar. Betingade sannolikheter däremot tycks utgöra ett adekvat mått på handlingens förväntade kausala inflytande på tillstånden.

Emellertid finns det alla skäl att tro att Jeffrey 1965 skulle ha betraktat den av Savage föreslagna beslutsprincipen som helt oantastlig *om* bara denna princip fick begränsas till de fall i vilka agenten, före valet, är säker på att de olika tillstånden är kausalt oberoende av de

tillgängliga handlingarna. Jeffrey skulle ha uppfattat denna begränsade Savageprincip som ett specialfall av sin egen beslutsprincip. Vid denna tid tycktes han nämligen anta att säkert kausalt oberoende automatiskt implicerar probabilistiskt oberoende:

Oberoendeantagandet. Om agenten, före valet, är säker på att tillståndet S är kausalt oberoende av handlingen A, så

$$P(S/A)=P(S).$$

På senare år har detta oberoendeantagande förkastats av flera filosofer, bl a av Jeffrey själv.[2] Antagandet håller inte i de fall i vilka agenten, före valet, är övertygad om att tillstånden befinner sig utanför hans inflytande men samtidigt betraktar sitt kommande handlingsval som ett mer eller mindre tillförlitligt *tecken* eller symptom på det föreliggande tillståndet. Agenten tillskriver alltså sina handlingar rent evidentiell relevans med avseende på tillstånden – antingen därför att han uppfattar sina handlingar som kausalt beroende av tillstånden (i stället för tvärtom) eller också därför att såväl hans handlingar som tillstånden enligt honom kan hänföras till en gemensam orsak. Jag skall kalla sådana fall för *Newcomblika* eftersom det mest omtalade fallet av denna typ har blivit känt under namnet "Newcombs problem". I Newcomblika fall kan den betingade sannolikheten för ett tillstånd givet en handling – $P(S/A)$ – avvika från den obetingade sannolikheten för samma tillstånd – $P(S)$ – även om agenten är övertygad att tillståndet ifråga är kausalt oberoende av handlingen.

Newcombs problem

"Professor L konfronteras med två lådor, en genomskinlig och en ogenomskinlig. Den genomskinliga lådan kan ses innehålla 1 000 dollar. Han får ta antingen den genomskinliga lådan med allt dess innehåll eller bägge lådorna med allt deras innehåll. En förutsägare, som är mycket duktig på att förutsäga val som människor träffar, har lagt 1 000 000 dollar i den ogenomskinliga lådan om han har förutsagt att professor L skall ta endast den ogenomskinliga lådan, och inget om han har förutsagt att professor L skall ta bägge. Förutsägaren är inte bara mycket duktig i största allmänhet; han är också mycket duktig vad gäller dem som tar endast den ogenomskinliga lådan, d v s för dem som tar enbart den ogenomskinliga lådan är hans andel av kor-

rekta förutsägelser mycket hög. Likaledes vad gäller dem som tar bägge lådorna. [Allt detta känner professor L till.] Under dessa omständigheter tar professor L enbart den ogenomskinliga lådan. I den hittar han 1 000 000 dollar. 'Jag är rik', utropar han. 'Du hade varit 1 000 dollar rikare om du hade tagit bägge lådorna', anmärker professor G" (Skyrms 1981, s 262).[3]

Låt *1M* respektive $\overline{1M}$ representera de alternativa tillstånden: *Det ligger en million dollar i den ogenomskinliga lådan* respektive *Den ogenomskinliga lådan är tom.* A_1 respektive A_2 skall stå för handlingsalternativen: *Professor L tar endast den ogenomskinliga lådan* respektive *Professor L tar bägge lådorna.*

Den tillgängliga informationen om förutsägarens skicklighet gör att professor L, före valet, uppfattar sin kommande handling som en tillförlitlig indikation på det föreliggande tillståndet. Hans betingade sannolikheter för *1M* givet A_1 och för *1M* givet A_2 är bägge mycket höga. De måste alltså bägge vara större än ½. Följaktligen blir deras summa större än ett. Därför måste åtminstone en av dem (och antagligen bägge) vara större än professor L's obetingade sannolikhet för motsvarande tillstånd. Ty hans obetingade sannolikheter för *1M* och $\overline{1M}$ summeras till ett. Samtidigt är professor L säker på att de handlingar som just nu står honom till buds inte kan kausalt påverka innehållet i den ogenomskinliga lådan. Newcombs problem är således Newcomblikt.

I Newcomblika fall kan Jeffreys beslutsprincip från 1965 mycket väl komma i konflikt med den begränsade Savageprincipen. Följer man Jeffreys princip kan det inträffa att man väljer en handling endast på grund av dess höga 'evidentiella värde', d v s endast därför att handlingen ifråga tyder på förekomsten av ett visst fördelaktigt tillstånd. Detta kan vara fallet även om handlingens förväntade *kausala* effekter är sämre än de som agenten kan förväntas uppnå genom att utföra en annan handling i stället. Så, till exempel, i Newcombs problem föreskriver Jeffreys princip från 1965 att professor L bör ta enbart den ogenomskinliga lådan - och därigenom förlora tusen dollar. Ty denna handling tyder på att den ogenomskinliga lådan innehåller en million. Vi skulle kunna säga att Jeffreys förväntade önskvärdhet, genom att den definieras i termer av betingade sannolikheter för tillstånden, inte så mycket mäter handlingens förväntade

kausala värde som dess värde *som nyhet.*[4] När agenten är övertygad att han saknar möjligheter att påverka tillstånden tycks det förväntade kausala värdet hos hans handlingar i stället sammanfalla med deras Savagevärde. Tycker man nu att det är handlingens förväntade kausala värde som bör vägleda valet så följer det att, i Newcomblika fall, de rätta föreskrifterna genereras av den begränsade Savageprincipen och inte av Jeffreys princip. Så till exempel, i Newcombs problem, föreskriver Savageprincipen att professor L utför A_1 - att han tar bägge lådorna. Det är nämligen lätt att inse att A_2 *dominerar* A_1: A_2 ger mer än A_1 under varje tillstånd - såväl under IM som under $\overline{IM}$. Och man kan visa att Savagevärdet hos en dominerande handling med nödvändighet överstiger Savagevärdet hos den handling som är dominerad.

Konfronterad med Newcomblika fall bestämde sig Jeffrey (1983) att revidera sin ursprungliga teori och han framkastade nu en ny beslutsprincip: "ratifierbarhetsmaximen". Därför faller det sig naturligt att undersöka hur denna nya maxim förhåller sig till den begränsade Savageprincipen. Är dessa två förenliga med varandra eller leder de ibland till motstridiga föreskrifter?

2. Ratifikationismen

I den nya utgåvan av *The Logic of Decision* påpekar Jeffrey att handlings*beslut* brukar vara lika tillförlitliga tecken på tillstånd som handlingarna själva. Så till exempel skulle man kunna hävda att redan professor L's beslut att ta enbart den ogenomskinliga lådan ger honom ett lika bra tips om innehållet i denna låda som den handling som åtföljer beslutet. Handlingens evidentiella värde uttöms, med andra ord, i förväg av handlingsbeslutet. Detta förhållande kan vi utnyttja, tänker sig Jeffrey. När vi ställs inför ett val bör vi först göra ett rent hypotetiskt antagande om vårt kommande handlingsbeslut och endast därefter - på grundval av detta antagande - bör vi jämföra de olika handlingar som står oss till buds med avseende på deras förväntade önskvärdhet. Tanken är att genom en sådan 'konditionalisering' av förväntade önskvärdhetsvärden på ett hypotetiskt antagande om vårt handlingsbeslut avskärmar vi, så att säga, de skillnader mellan hand-

lingar som uteslutande beror på deras olika evidentiella relevans. Och en dylik avskärmning är givetvis nödvändig eftersom de rent evidentiella skillnaderna mellan handlingarna inte får tillåtas att vägleda valet.

Låt mig nu beskriva Jeffreys ansats litet mer i detalj. Vi låter '*bA*' stå för påståendet att agentens slutgiltiga beslut blir att utföra handlingen *A*. Observera att för Jeffrey är ett beslut slutgiltigt i och med att det inte kommer att ändras av agenten. Emellertid är misslyckanden alltid möjliga. Det är alltid möjligt att agenten misslyckas att genomföra sitt slutgiltiga beslut och att han utför en annan handling i stället. Därför förutsätter Jeffrey, för varje två handlingsalternativ *A* och *A'*, att den villkorliga sannolikheten för A givet *bA'* alltid är större än noll, även om den kan vara ytterst liten.[5]

Om *A* och *A'* ingår i agentens alternativmängd, kan vi tala om *A:s förväntade önskvärdhet på villkor att bA'*. D v s om det förväntade önskvärdhetsvärde som tilldelas handlingen *A* på basen av ett hypotetiskt antagande att agentens slutgiltiga beslut blir att utföra handlingen *A'*.[6] Detta värde beräknas på samma sätt som Jeffreys ursprungliga förväntade önskvärdhet hos *A* men vikterna är nu annorlunda: alla vikter av formen

P(*S*/*A*)

konditionaliseras nu på det hypotetiska antagandet *bA'*. P(*S*/*A*) ersätts därför med

P(*S*/*A*&*bA'*).[7]

Nu kan vi förklara vad Jeffrey menar med ratifierbarhet. Den intuitiva idén är enkel: ett handlingsbeslut är ratifierbart om det inte berövar den beslutade handlingen dess förväntade värde. Med en mera precis formulering:

Beslutet att utföra *A* är *ratifierbart* om, och endast om, den förväntade önskvärdheten hos *A* på villkor att *bA* är åtminstone lika stor som den förväntade önskvärdheten hos varje alternativ handling bedömd *på samma villkor* (d v s *bA*).

För enkelhets skull kommer jag i fortsättningen att ibland tala om ratifierbara handlingar (och inte bara om ratifierbara handlingsbeslut).

En handling antas vara ratifierbar om beslutet att utföra den är ratifierbart.

Jeffreys ratifierbarhetsmaxim föreskriver nu agenten att fatta ratifierbara beslut. Ratifierbarhet framstår som ett både nödvändigt och tillräckligt villkor för korrekt handlande (jfr Jeffrey 1983, kap 1).[8]

Hur kan denna maxim ta hand om Newcomblika fall? Redan tidigare har jag skisserat Jeffreys svar på denna fråga. Genom att konditionalisera på ett givet beslut kan vi normalt helt och hållet "avskärma" de rent evidentiella skillnaderna mellan olika handlingar. Beslutet att handla utgör med andra ord ett lika pålitligt tecken på det föreliggande tillståndet som handlingen själv. *Om* nu detta "avskärmningsantagande" håller så möjliggör en konditionalisering på ett bestämt handlingsbeslut en rätt sorts jämförelse mellan den motsvarande handlingen och dess alternativ - en jämförelse som uteslutande sker i termer av det förväntade kausala värdet hos de olika handlingarna. Och det är precis en sådan konditionalisering som utgör grunden för vår bedömning om en given handling är ratifierbar eller ej.

Så till exempel, givet avskärmningsantagandet, kommer professor L's betingade sannolikheter för IM givet $A_1\&bA_1$ respektive givet $A_2\&bA_1$ att sammanfalla: bägge kommer att vara mycket höga. Och analogt blir det med hans sannolikheter för $\overline{IM}$ givet $A_1\&bA_2$ respektive givet $A_2\&bA_2$. Att ta bägge lådorna kommer därför att framstå som ett bättre alternativ på basen av varje hypotetiskt antagande om professor L's handlingsbeslut. Att göra så minskar ju inte sannolikheten för en million (ty denna sannolikhet, givet avskärmningsantagandet, bestäms helt och hållet av handlingsbeslutet och påverkas alltså ej längre av den efterföljande handlingen) och samtidigt ger det alltid tusen dollar mer. Ratifierbarhetsmaximen kommer därför att föreskriva A_2 - precis som den begränsade Savageprincipen.

3. Jeffrey möter Savage

Jag skall här inte diskutera den självklara invändningen som kan göras mot Jeffreys avskärmningsantagande: ibland kan agenten uppfatta sitt eventuella beslut att handla som en något sämre indikation på det

föreliggande tillståndet än den som skulle utgöras av handlingen själv. I sådana fall förmår han inte att fullständigt avskärma handlingens rena nyhetsvärde genom att konditionalisera på handlingsbeslutet. Ratifierbarhetsmaximen kommer därför att leda honom på avvägar. Jeffrey själv erkänner styrkan hos denna invändning[9] och han begränsar därför explicit sin ratifierbarhetsansats endast till de fall i vilka avskärmningsantagandet är satisfierat. Jag skall här följa honom i detta avseende.

Följande bör observeras: om vi accepterar avskärmningsantagandet och om vi håller oss till de fall i vilka agenten är säker på att tillstånden står utanför hans inflytande, kan Jeffreys vikter av formen

$P(S/A\&bA')$

förenklas till

$P(S/bA')$.

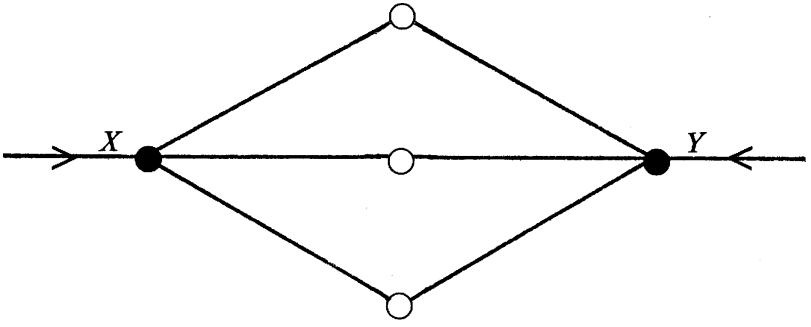
Vi kan alltså undvika referensen till A. Ty i sådana fall är agenten säker på att utförandet av A inte kan ha några kausala effekter på S, och samtidigt (avskärmningsantagandet!) tillmäter han inte utförandet av A någon evidentiell relevans med avseende på S utöver den som redan har tagits om hand genom konditionsaliseringen på bA'.

Vad händer om inget av handlingsalternativen är ratifierbart? Jeffreys maxim kommer då inte att ge oss någon vägledning. Jeffrey gissar att ett sådant problem endast kan uppkomma för en irrationell agent. När inget av handlingsalternativen är ratifierbart finns det enligt Jeffrey fog för misstanke att agentens övertygelser och/eller preferenser vid tiden för valet måste vara patologiska på något sätt. Varför Jeffrey tror att så måste vara fallet är för mig en gåta. I en annan uppsats beskriver jag ett icke-patolgiskt Newcomblikt fall i vilken agenten saknar ett ratifierbart handlingsalternativ (se Rabinowicz 1984; ett liknande fall beskrivs för övrigt så pass tidigt som 1978 av Allan Gibbard och William Harper). Existensen av sådana fall visar redan på en viktig skillnad mellan ratifierbarhetsmaximen och den begränsade Savageprincipen: den senare ger agenten vägledning även i avsaknad av ratifierbara alternativ. Här vill jag emellertid i stället

koncentrera mig på ett Newcomblikt fall i vilket Jeffreys maxim *ger* ett bestämt utslag men där Jeffreys föreskrift kommer i konflikt med den begränsade Savageprincipen.

Vägvalet

Vi tänker oss två agenter, X och Y, som är ute för att mötas. De har startat från olika platser och nu har de kommit fram till var sitt vägskäl. För var och en av dem står valet mellan att gå rakt fram, till vänster eller till höger. Vägkartan ser ut så här:



Punkter anger X och Y:s nuvarande positioner och cirklar visar på möjliga mötesplatser. Som vi ser kommer de att mötas endast om bägge går rakt fram eller också om den ene går till vänster och den andre till höger. Om bägge fortsätter rakt fram blir vägen visserligen något kortre men de blir tvungna att betala höga tullavgifter. (Om vi så vill kan vi tänka oss att den raka sträckan är en fransk motorväg eller att X och Y är två medeltida köpmän.) Vi antar dessutom att vägen till vänster är bekvämare än vägen till höger, något som varje agent känner till. Alla dessa faktorer tillsammans bestämmer deras önskvärdhetsvärden för de olika utfallen. Här presenterar jag endast X:s värdematris. Y:s matris antas vara analog.

		Y		
		vänster	rakt fram	höger
X	vänster	0,1	0,1	1,1
	rakt fram	—0,5	0,5	—0,5
	höger	1	0	0

Det bästa (1,1) är att möta den andre efter att man har tagit den bekväma vänstra vägen. Men även om mötet kommer till stånd efter vissa strapatser är det också mycket bra, sålänge man inte behöver betala tull (1). Att behöva betala tull för att mötas är inte särskilt lockande (0,5) men det är i alla fall bättre än att inte mötas alls – oberoende av om vägen annars är bekväm (0,1) eller strapatsrik (0). Och det sämsta (-0,5) är givetvis om man betalar tull och ändå missar mötet.

Ingen kommunikation och inget samarbete mellan agenterna är möjligt: avståndet är för stort och de saknar tillgång till telefon eller radio. För enkelhets skull antar jag också att ingen av agenterna får invänta den andre (de har ont om tid) och att ingen av dem får träffa sitt val genom att kasta mynt eller tärning (kanske blir de hårt bestraffade om de försöker; eller också förbjuder deras religion att viktiga val överlämnas åt slumpen).

X och *Y* antas vara mycket lika varandra psykologiskt sett: de är mer eller mindre psykologiska tvillingar och de känner till att så är fallet. De har dessutom råkat ut för liknande koordineringsproblem flera gånger tidigare och då har de alltid eller nästan alltid handlat på samma sätt: vänster–vänster, höger–höger, eller rakt fram–rakt fram. Emellertid, i de flesta tidigare fall, har de bägge undvikit att gå rakt fram (för att slippa betala tull) och valt att gå till vänster i stället. Att de hade föredragit vänster framför höger beror på att den vänstra vägen varit så bekväm. Bägge agenterna har alltså varit benägna att ta

det säkra före det osäkra: det sämsta möjliga utfallet när man går till vänster (0,1) är bättre än de sämsta utfallen för de två andra handlingsalternativen (-0,5 resp. 0).

Läsaren skulle här kanske vilja påpeka att *X* och *Y* borde för länge sedan ha koordinerat sitt beteende i dylika situationer genom en lämplig överenskommelse. T ex: "I framtiden skall jag, *X*, gå till vänster medan du, *Y*, skall ta den mindre bekväma högra vägen. Kostnaden för dina större strapatser skall vi bära tillsammans". Ja, det borde de ha gjort. Det var dumt av dem att inte komma överens, men nu är det för sent. Nu står de där de står och var och en av dem måste göra det bästa han kan.

Varje agent känner till allt det här och denna kunskap bestämmer hans sannolikhetsbedömningar. I själva verket är agenternas sannolikhetsbedömningar helt analoga. Om vi därför koncentrerar oss på *X* och låter P vara *X*:s sannolikhetsfunktion vid tiden just före vägvalet, så gäller följande:

(1) P(*Y* går till vänster) är hög, säg, högre än 0,5;
(2) P(*Y* går i riktningen *r*/*b*(*X* går i riktningen *r*)) är mycket hög, säg, högre än 0,8.

(1) bestäms av *X*:s kunskap om *Y*:s tidigare vänstervridna beteende och om den benägenhet att ta det säkra före det osäkra som ligger bakom beteendet ifråga. (2) är uppfyllt oberoende av vilken riktning *r* står för: vänster, höger eller rakt fram. "b(*X* går i riktningen *r*)" representerar påståendet att *X*:s slutgiltiga beslut blir att gå i riktningen *r*. Att (2) gäller beror på att *X* uppfattar sitt kommande handlingsbeslut, vilket det än vara månde, som ett pålitiligt tecken på hur *Y*, hans psykologiska tvilling, kommer att handla. Låt oss också, för argumentets skull, förutsätta att avskärmningsantagandet är satisfierat. Här slutar min beskrivning av vägvalsexemplet.

X uppfattar sina handlingar och handlingsbeslut som tecken på tillstånden (*Y*:s handlingar). Samtidigt är *X* säker på att tillstånden är kausalt oberoende av de handlingar som just nu står honom till buds: *X* och *Y* handlar oberoende av varandra. Vägvalsexemplet är alltså Newcomblikt.

Den begränsade Savageprincipen föreskriver att X skall gå *till höger:* den 'obetingade' sannolikheten för att Y kommer att gå till vänster är ju hög (jfr (1) ovan), och om Y går till vänster kommer de bägge agenterna att mötas endast om X går till höger.

Ratifierbarhetsmaximen föreskriver däremot att X skall gå *rakt fram* eftersom att gå rakt fram är det enda ratifierbara handlingsalternativet. Låt mig förklara. Ponera först att X beslutar att gå rakt fram. På basen av detta hypotetiska antagande är det mycket sannolikt (i X:s ögon) att Y kommer att gå den raka vägen. (2) implicerar att denna sannolikhet överstiger 0,8. Och om Y går den raka vägen så kommer de att mötas endast om X gör likadant. I själva verket kan man lätt beräkna att, på villkor att X beslutar att gå rakt fram, blir den förväntade önskvärdheten hos alternativet 'rakt fram' större än 0,3 ($0{,}8\times\frac{1}{2}+0{,}2\times(-\frac{1}{2})$). På samma villkor blir däremot den förväntade önskvärdheten hos 'vänster' mindre än 0,3 ($0{,}8\times0{,}1+0{,}2\times1{,}1$), medan det motsvarande värdet för 'höger' blir mindre än 0,2 ($0{,}8\times0+0{,}2\times1$). 'Rakt fram' framstår alltså som ett ratifierbart alternativ. Ponera nu i stället att X beslutar att gå till höger. På basen av *detta* hypotetiska antagande blir det högst sannolikt att Y kommer att gå till höger (jfr (2)). Och då kommer X, genom att gå till höger, att missa sitt möte med Y. Mötet skulle komma till stånd endast om X, i strid med sitt beslut, gick till vänster i stället. Det är därför klart att alternativet 'höger' inte är ratifierbart. Ett exakt analogt resonemang visar att samma sak gäller alternativet 'vänster'.

Att det föreligger en sådan konflikt mellan de två principerna är lätt att förstå. När agenten är säker på att tillstånden står utanför hans kontroll, föreskriver ratifierbarhetsmaximen att agenten bör träffa sitt val i termer av sina *ex post* sannolikheter. D v s sannolikheter av formen $P(S/bA)$. Valet bör träffas utifrån de (hypotetiska) sannolikheter som agenten skulle tilldela tillstånden efter valet. Den begränsade Savageprincipen rekommenderar däremot agenten att använda tillståndssannolikheter *ex ante.* Valet bör träffas utifrån agentens sannolikhetstilldelningar *före* valet. Och det är uppenbart att i Newcomblika fall kommer dessa bägge typer av sannolikheter att vara olika. Ty i sådana fall antas agentens val vara evidentiellt relevant med avseende på tillstånden. Det är därför inte förvånande att de två beslutsprinciperna ibland kommer att leda till motsatta föreskrifter.

4. Savage på defensiven

Det vore bra att kunna avsluta denna uppsats med ett konklusivt argument till förmån för en av de två beslutsprinciperna eller, eventuellt, till nackdel för bägge. Tyvärr har jag inget sådant argument att komma med. Det enda som jag kan prestera är ett försvar för Savage i det konkreta fall som jag har beskrivit - ett försvar mot två tänkbara invändningar som skulle kunna anföras av en Jeffreyanhängare.

Invändning 1

Enligt Savage bör X gå till höger. Men samma föreskrift måste i så fall gälla även Y. X och Y befinner sig ju i exakt likadan situation gentemot varandra: vägvalsexemplet är helt symmetriskt. Emellertid, om bägge agenterna följer Savages föreskrift, så kommer de inte att mötas. Det vore bättre för var och en av dem om de bägge gick rakt fram i stället. De skulle då visserligen få betala tull men de skulle också träffa varandra. Detta är bättre än att inte mötas alls. Och att gå rakt fram är just det som varje agent bör göra enligt ratifierbarhetsmaximen. Att en beslutsprincip rekommenderar alla agenter ett sätt att handla vars sammanlagda resultat med nödvändighet är sämre för varje agent än det sammanlagda resultatet av ett alternativt handlingssätt - är inte detta något som visar på principens ohållbarhet?

Savage kan bemöta denna invänding genom att påpeka att samma svårighet mycket väl kan uppkomma även för Jeffrey. Låt mig ge ett exempel. Givet avskärmningsantagandet kan det visas att ratifierbarhetsmaximen, i likhet med Savage-principen, alltid rekommenderar agenten att utföra den dominerande handlingen, om en sådan finns att tillgå. (En handling är dominerande, som vi kommer ihåg, om den under varje tillstånd ger agenten mera än de andra handlingsalternativen.) Detta gäller sålänge agenten är säker på att tillstånden står utanför hans inflytande.

Ponera nu att en viss situation involverar flera agenter som får handla oberoende av varandra, att utfallet bestäms av vad de olika agenterna gör (så att ur varje agents perspektiv de andra agenternas alternativa handlingsmönster kan ses som olika möjliga tillstånd som

agenten ifråga måste räkna med), och att det finns ett sätt att handla som är dominerande för varje agent men vars sammanlagda utfall är sämre för var och en än det sammanlagda utfallet hos något alternativt handlingssätt. (Det som man vinner genom att utföra den dominerande handlingen kompenserar en inte för det man förlorar genom att andra handlar likadant.) En situation av detta slag brukar som bekant kallas för "fångens dilemma" och den har varit föremål för omfattande besluts- och spelteoretiska diskussioner. I fångens dilemma kommer ratifierbarhetsmaximen att rekommendera varje agent det dominerande handlingssättet. [10] Den kommer alltså att leda till samma oroande effekt som Savageprincipen leder till i vägvalsexemplet: om alla agenter följer principen blir det sämre för var och en av dem än om de alla hade valt ett annat sätt att handla. [11]

Att hans maxim slår ut på detta sätt i fångens dilemma är givetvis något som Jeffrey är väl medveten om och som han inte alls anser vara någon nackdel hos maximen. Tvärtom (jfr Jeffrey 1983, kap 1). Han själv skulle därför aldrig komma på tanken att kritisera Savage på detta sätt. Desto större vikt skulle han däremot, om jag inte tar fel, lägga vid den invändning som nu följer.

Invändning 2

Om X bestämde sig att gå till höger, såsom Savage-principen föreskriver, så skulle han, efter att ha tagit detta beslut, vara ganska övertygad om att Y, hans psykologiska tvilling, också kommer att gå till höger. Denna övertygelse från X:s sida skulle ju också vara fullständigt befogad under sådana omständigheter. Och om Y går till höger, kommer X genom att gå till höger att gå miste om mötet. Visar detta inte att beslutet att gå till höger måste vara irrationellt? Själva detta beslut skulle ju leda X till en övertygelse i vars ljus den beslutade handlingen måste framstå som klart ofördelaktig. Bör inte X gardera sig mot sådant genom att i stället träffa sitt val i termer av sina hypotetiska *eftervals*sannolikheter (*ex post* sannolikheter)?

Jag gissar att Savages svar skulle lyda så här: Det är riktigt att beslutet att gå till höger skulle leda X till en under dessa omständigheter fullständigt befogad övertygelse om att Y också kommer att gå till höger. *Men,* före valet tror X att Y kommer att gå till *vänster.* Vad

mera är, X är säker på att hans eget val, vilket det än skulle bli, omöjligtvis kan påverka Y:s beteende. Detta innebär att X, före valet, betraktar sin hypotetiska framtida övertygelse om att Y kommer att gå till höger som *falsk*. Befogad, men falsk. Vad finns det då för skäl för X att träffa sitt val i termer av hypotetiska eftervalsövertygelser som han nu, före valet, anser vara falska?

Jag tror att denna fråga är på sin plats. Åtminstone i vägvalsexemplet tycks Jeffreys ratifierbarhetsmaxim leda oss på avvägar.

Noter

1. En längre version av denna uppsats kommer att publiceras i ett annat sammanhang. Se Rabinowicz (1984). Många personer har hjälpt mig med synpunkter och kommentarer. Särskilt vill jag tacka Lars Bergström, Sven Danielsson, Allan Gibbard, Sten Lindström och, inte minst, Howard Sobel.

2. Jfr t ex Nozick (1969), Gibbard och Harper (1978), Sobel (1979) och (1984), Skyrms (1980) och (1982), Lewis (1981), Jeffrey (1983). Oberoendeantagandet försvaras däremot av Eells (1982).

3. Skyrms beskriver också ett antal andra Newcomblika fall. Newcombs problem har först presenterats i tryck av Robert Nozick (1969). Jeffreys favoritexempel på ett Newcomblikt fall är en version av det s k "fångens dilemma", i vilken fångarna antas vara (och känna till att de är) psykologiska tvillingar. Följden blir att varje fånge uppfattar sitt eget handlingsbeslut som en indikation på att den andre fången kommer att handla likadant. Se Jeffrey (1983), kap 1. Jfr Lewis (1979) och Sobel (1983).

4. 1965 ansåg Jeffey att dessa bägge värden sammanfaller med varandra. Jfr Jeffrey (1965), ss 73 f. Konstigt nog återkommer samma påstående i Jeffrey (1983), s 84, fast hans åsikter på denna punkt har under tiden undergått väsentliga förändringar. Jag skulle gissa att det är fråga om ett förbiseende från Jeffreys sida.

5. Den nämnda förutsättningen är oundgänglig för Jeffreys teori (se nedan, not 8) men den är samtidigt inte alls okontroversiell. Det är ju en sak om agenten erkänner att han alltid kan misslyckas i genomförandet av sitt slutgiltiga beslut. Därigenom behöver han inte erkänna att ett sådant misslyckande i princip kan leda till vilken annan alternativ handling som helst. Kan jag verkligen misslyckas att genomföra mitt slutgilitga beslut att stanna hemma ikväll på ett sådant sätt att jag går på bio i stället? Och detta, märk väl, utan att jag 'ändrar' mig, ty om jag ändrade mig, så hade mitt slutgiltiga beslut att stanna hemma inte varit slutgiltigt.

6. Ett specialfall föreligger när $A=A'$. Normalt kommer emmellertid A och A' att vara olika handlingar.

7. Egentligen borde vi också tillämpa samma sorts konditionalisering vad gäller de viktade önskvärdhetsvärdena. Men för enkelhets skull antar vi här att denna komplikation inte behövs. Vi antar alltså, för varje S och A, att A:s värde under S är detsamma oberoende av vilket som var agentens slutgiltiga beslut. Med andra ord: utfallet bestäms uteslutande av vad agenten faktiskt gör och inte av vad han beslutar sig att göra.

8. Som nämndes ovan förutsätter Jeffrey att $P(A/bA')>0$, för alla handlingsalternativ A och A'. Detta är detsamma som att anta att $P(A\&bA')$ alltid är större än noll. Att denna förutsättning är väsentlig för Jeffrey kan nu lätt inses. Om $P(A\&bA')=0$, blir alla vikter av formen $P(S/A\&bA')$ odefinierade. Därför kan den förväntade önskvärdheten hos A på villkor att bA' inte heller definieras, och det blir följaktligen omöjligt att tillämpa ratifierbarhetsmaximen på handlingen A'. A' kommer inte att kunna jämföras – enligt Jeffreys recept – med den altenativa handlingen A och beslutet att utföra A' kommer att framstå som varken ratifierbart eller oratifierbart.

9. Han hänvisar i detta sammanhang till Bas van Fraassen (se Jeffrey 1983, kap 1). Samma argument framförs också av Sobel (1984).

10. Följande värdematris kan exemplifiera fångens dilemma:

		Y: gör si	Y: gör så
X	gör si	1 / 1	0 / 3
	gör så	3 / 0	2 / 2

(I varje cell anger den nedre siffran X:s värdering av utfallet; den övre siffran anger utfallets värde för Y.) Att göra si dominerar att göra så för varje agent. Men om bägge gör si får var och en mindre än om de bägge gör så i stället. Givet avskärmningsantagandet föreskriver ratifierbarhetsmaximen, liksom Savageprincipen, varje agent att göra si.

11. Denna oroande effekt uppkommer för övrigt även i andra typer av situationer, bland annat sådana som, till skillnad från fångens dilemma, saknar dominerande handlingsätt och i vilka, återigen till skillnad från fångens dilemma, råder det en fullständig överensstämmelse mellan agenternas värdering

av de olika utfallen. Ett exempel: Anta att X och Y ställs inför likadana val (mellan tre handlingar: h_1, h_2 och h_3), att de får handla oberoende av varandra och att utfallet bestäms av vad de bägge gör, att det vad den ene väljer saknar evidentiell relevans för vad den andre kommer att göra (situationen är alltså *inte* Newcomblik), och att X och Y har exakt samma värdematris:

		Y		
		h_1	h_2	h_3
X	h_1	10	0	0
	h_2	0	9	0
	h_3	0	0	10

Förutsatt att varje agent anser det vara mest sannolikt att den andra utför h_2 (något som väl är att vänta om agenterna saknar ytterligare informationer), kommer h_2 att föreskrivas av såväl Jeffrey som Savage. Men om bägge agenterna följer föreskriften blir det sämre för var och en av dem än om de bägge hade valt ett annat sätt att handla (d v s om de bägge utförde h_1 eller om de utförde h_3).

Litteratur

Ellery Eells, 1982, *Rational Decision and Causality* Cambridge University Press, Cambridge.

Allan Gibbard och William L. Harper, 1978, 'Counterfactuals and Two Kinds of Expected Utility', i A. Hooker, J.J. Leach and E.F. McClennen (utg.), *Foundations and Applications of Decision Theory,* vol. 1, Reidel, Dordrecht, Holland; omtryckt i W.L. Harper, R. Stalnaker, G. Pearce (utg.), *Ifs,* Reidel, Dordrecht, Holland, ss. 153-190.

Richard C. Jeffrey, 1965, *The Logic of Decision,* Mac Graw-Hill, New York.

Richard C. Jeffrey, 1983, *The Logic of Decision,* andra utgåvan, University of Chicago Press, Chicago and London.

David Lewis, 1979, 'Prisoner's Dilemma is a Newcomb Problem', *Philosophy and Public Affairs* 8, ss. 235-240.

David Lewis, 1981, 'Causal Decision Theory', *Australian Journal of Philosophy* 59, ss. 5-30.

Robert Nozick, 1969, 'Newcomb's Problem and Two Principles of Choice', i Nicholas Rescher (utg.), *Essays in Honour of Carl G. Hempel,* Reidel, Dordrecht, Holland.

Wlodzimierz Rabinowicz, 1984, 'Ratificationism without Ratification: Jeffrey meets Savage', antagen till publicering i *Theory and Decision.*

Leonard J. Savage, 1954, *The Foundations of Statistics,* Wiley, New York; andra omarb. uppl., Dover, New York, 1972.

Brian Skyrms, 1980, *Causal Necessity,* Yale Univ. Press, New Haven.

Brian Skyrms, 1981, 'The Prior Propensity Account of Subjunctive Conditionals', W.L. Harper, R. Stalnaker, G. Pearce (utg.), *Ifs,* Reidel, Dordrecht, Holland, ss. 259-265.

Brian Skyrms, 1982, 'Causal Decision Theory', *The Journal of Philosophy* 79, ss. 695-711.

J. Howard Sobel, 1979, *Probability, Chance and Choice: A Theory of Rational Agency,* opublicerad.

J. Howard Sobel, 1983, 'Not Every Prisoner's Dilemma is a Newcomb Problem', opublicerad.

J. Howard Sobel, 1984, 'Adequate Partitions of Circumstances and Dominance Arguments in a Causal Decision Theory', antagen till publicering i *Synthese.*

Hans Rosing

Till den psykoneurala identitetens försvar

Det har varit glädjande för mig att min bok *Medvetandets filosofi* (Akademilitteratur/Schildts 1982) har uppmärksammats och diskuterats både i Sverige och Finland. Recensenterna har, med ett undantag, uttalat sig positivt om boken. I FT har boken recenserats av Peter Gärdenfors i nr 3:83. Ingmar Persson kommer i FT 2:84 med djuplodande kritik av kapitlet om psykoneural identitet.

Persson tycker att boken förtjänar ett bättre öde än att bemötas med tystnad. För att föregå med gott exempel analyserar han i detalj min framställning av identitsteorin (IDT). Jag för min del finner Perssons kritik så stimulerande och utmanade att jag vill fortsätta diskussionen. För övrigt är IDT i dag så allmänt accepterad inom psykologin och neurovetenskaperna och t o m av massmedia och lekmän att det måste vara en viktig uppgift för filosofin att noggrant analysera den. Uttryck som "vi tänker med hjärnan", "tankarna finns i huvudet" ingår i dag i vardagsspråket. De kan visserligen tolkas på olika sätt men vanligen tycks de uppfattas mycket konkret och bokstavligt.

Nedan skall jag bemöta en del av Perssons kritik. (1) Persson skriver (s 23): "förbluffande nog stannar Rosing för en klassifikation av epifenomenalismen som en variant av materialism, även om han lägger i dagen en viss tvehågsenhet och på ett ställe karakteriserar den som en 'svag form av dualism' ". Enligt Persson är epifenomenalismen "tveklöst en dualistisk teori". Jag håller med om att epifenomenalismen *formellt sett* är en dualistisk teori. Enligt teorin är de psykiska processerna av helt annan art än de fysiska. Men *historiskt* sett hör epifenomenalismen ihop med materialismen. Epifenomenalismen är helt främmande för den klassiska platonsk-cartesianska dualismen. Att jag tog upp epifenomenalismen i samband med materialismen beror på att mitt sätt att bedriva filosofi alltid går ut på att försöka analysera och förstå teorier i deras historiska sammanhang. I sak torde vi vara eniga på denna punkt.

(2) Persson skriver: ". . . Rosing vacklar mellan att betrakta det *tillstånd* eller *faktum* vilket består i att en varelse förnimmer smärta som identiskt med K och att betrakta det *objekt* som förnimmes, smärtan, som identisk med K". (K används här som benämning på ett komplex av neurala processer). Han påpekar att jag ofta talar och tänker som om själva smärtan var identifikationsobjektet, och därför leds jag till ". . . diverse egendomliga krumbukter, som påståendet att det är en illusion att smärtan från en fotskada är lokaliserad till foten" (25). Något senare höjer Persson ett varnande pekfinger och skriver: ". . . vi får inte tillåta oss att säga att det förnumna i själva verket inte har de kvaliteter som vi förnimmer att det har, utan helt andra som vi aldrig kan förnimma att det har" (25).

Det är sant att jag inte beaktat skillnaden mellan smärta som ett tillstånd och smärta som objekt. Smärta är förstås inte ett objekt i samma mening som en bok eller en stol, men den har objektkaraktär därför att vi kan känna efter hurudan den är. Den har tydliga kvaliteter, den är lindrig, svår, outhärdlig, stickande, brännande, malande, bultande, pulserande. Den är också mer eller mindre tydligt lokaliserad, till en tand, till högra stortån, till magen. Man kan däremot inte säga att upplevelsen av smärta har dessa kvaliteter. Det bör heta "han upplevde en stark, skärande smärta i högra stortån", inte "han hade en stark, skärande upplevelse av smärta i högra stortån".

Enligt min tolkning av IDT innebär den att både smärtan som objekt och tillståndet att uppleva smärta måste identifieras med kroppsliga processer. (Vilka dessa kroppsliga processer kan tänkas vara är en ytterst komplicerad fråga som vi inte kan gå in på här. Jag talar i fortsättningen bara om fysiologiska processer utan att försöka precisera dem närmare). Jag tycker att det är alldeles onödigt att komplicera IDT genom att försöka visa att allt tal om smärta i själva verket är tal om tillstånd. (Använd Ockhams rakkniv!) Varför hävdar jag att vår subjektiva lokalisering av smärtan till en viss kroppsdel, t ex höger stortå, är en illusion? (Detta påstående har för övrigt väckt en hel del kritik från filosofiskt håll men inte från psykologiskt).

Enligt IDT är smärtan i stortån (S) identisk med vissa fysiologiska processer (K). I stortån finns inga fysiologiska processer som kan komma ifråga som smärtprocesser. De enda tänkbara processerna finns i hjärnan. Men om S finns i stortån och K finns i hjärnan så är det

omöjligt att S=K och IDT måste vara falsk.

För att rädda IDT måste vi övertygande kunna visa att S i själva verket inte finns i tån utan i hjärnan. Vi måste alltså kunna visa att den smärta vi känner i tån egentligen finns i hjärnan. Allmänt uttryckt: vi måste kunna visa att smärtan objektivt sett inte befinner sig där som vi subjektivt upplever den, d v s att den subjektiva smärtlokalisationen är en illusion. Smärtan som sådan är förstås ingen illusion.

Vi bör lägga märke till att ordet "illusion" har två användningar, en logisk och en psykologisk. I sin logiska roll används ordet på samma sätt som ordet "misstag". I psykologin används ordet om en upplevelse (syn-, hörsel- etc) som är helt normal men som inte stämmer överens med verkligheten. Ett enkelt exempel: Vi kan alla se att gestalterna i en film rör sig. Men vi vet att i verkligheten rör de sig inte, utan det är frågan om stillbilder som rör sig över en duk med en hastighet av ca 20 bilder/s. Vi förnimmer sålunda en rörelse som inte existerar. Det är inte frågan om ett misstag i ordets egentliga mening. Vi upplever verkligen rörelse. På samma sätt är det med smärta. Vi upplever verkligen att smärtan finns i stortån, men i själva verket finns den i hjärnan.

Det är ett elementärt psykolgiskt faktum att vi ofta förnimmer kvaliteter som föremål inte har - psykologiböckerna är fulla av exempel - liksom att vi ofta inte förnimmer kvaliteter som föremål har. Huruvida en upplevelse är en illusion eller inte är ett problem för den empiriska vetenskapen inte för filosofin. Nedan ger jag några exempel på empirisk evidens för att smärtlokalisationen är en illusion:

1. Om vi kapar av högra benet på en person så är det fortfarande möjligt att personen känner att det gör ont i höger stortå (s k fantomsmärta). Eftersom det är absurt att smärtan skulle vara lokaliserad i en lem som inte finns måste vi anta att den är lokaliserad någon annanstans.

2. Det är möjligt att koppla om nervbanorna så att nervimpulserna från högra stortån når det område i hjärnan som normalt tar emot signaler från vänstra stortån. När en spik tränger genom en persons högra stortå känner han en svår smärta i *vänstra* stortån. Han lokaliserar smärtan till fel stortå. Lokalisationen är alltså inte beroende av från vilken lem nervimpulserna kommer utan av *vilken del av hjärnan* som stimuleras.

3. Om man genom elektrisk stimulering av hjärnan åstadkommer komplex K så upplever personen en svår smärta i höger stortå trots att den är helt oskadd.

4. Om det relevanta området K i hjärnan bedövas så känner personen ingen som helst smärta i höger stortå vad man än gör med den.

Dessa empiriska data utgör en helt övertygande evidens för att den subjektiva smärtlokalisationen är en psykologisk illusion. Vi behöver sålunda inte alls tillgripa den adverbiella analys, som Persson talar om (s 25) för att försvara IDT. Det räcker med att påpeka att smärtlokalisationen är en subjektiv illusion och att S alltså i själva verket mycket väl kan vara lokaliserad till samma plats som K. Den empiriska evidens som finns tyder i själva verket på att så är fallet och kan sålunda tolkas som stöd för IDT.

(3) Jag skall avsluta med några synpunkter på Perssons diskussion av den version av IDT som jag kallar "strikt" i motsats till reduktionistisk. Enligt Persson är det jag kallar strikt IDT i själva verket"... snarast en förtäckt form av epifenomenalism" (s 26). Vidare säger han: "Denna teori gör en eftergift åt dualismen: den medger existensen av mentala egenskaper som är artskilda från fysiska egenskaper" (s 26). På s 28 skriver han: "Det verkar således som om det skulle vara omöjligt att hålla isär den 'strikta' identitetsteorin och epifenomenalismen: båda är egenskapsdualistiska och båda måste därför laborera med psykoneurala lagar".

Persson har rätt i att jag behandlade den strikta IDT alltför ytligt i boken. Ärligt talat hade jag svårt att få ett grepp om denna version av IDT. Fortfarande känner jag mig osäker, men jag skall göra ett försök att dra en gräns mellan strikt IDT och epifenomenalism.

Enligt IDT (både strikt och reduktionistisk) är alla mentala tillstånd eller processer (eller objekt) identiska med materiella tillstånd eller processer. Enligt epifenomenalismen är det psykiska något ickemateriellt som orsakas av materiella processer. Men enligt strikt IDT är det psykiska verkligen materiellt.

Jag skall försöka förklara vad jag menar genom en analogi. Det finns en kvalitativ skillnad mellan en levande organism och död materia. Ändå har vi inga problem att betrakta livsprocesserna, rörel-

se, tillväxt, reproduktion o s v som materiella processer. (Det är förstås inte så länge sedan man trodde på en speciell livskraft som skötte livsfunktionerna. Många tror ju ännu att det finns en speciell själskraft som sköter de mentala funktionerna.) Vi säger inte att förökningen är något som korrelerar med materiella processer utan att den *är* en materiell process. Men det är långtifrån självklart att livsprocesserna kan reduceras (i någon intressant mening) till icke-livsprocesser. (Livsprocesserna innebär t ex ett upphävande av termodynamikens första lag *inom* organismen). Det är sålunda möjligt att hävda att liv är ett högre, emergent tillstånd hos materia. Men det gör inte livsprocesserna mindre materiella.

På samma sätt kan det mentala betraktas som ett högre, emergent tillstånd hos liv, och följaktligen hos materia. Vi kan använda predikat om levande entiteter som inte är tillämpbara på icke-levande. Men vi gör oss därmed inte skyldiga till någon egenskapsdualism. Predikatet "reproduktion" betecknar inte en egenskap som korrelerar med ett visst komplex av materiella processer. Det betecknar i själva verket detta komplex av emergenta materiella processer. De flesta människor som använder predikatet har ingen aning om vilka dessa materiella processer är. De avser inte dessa processer när de använder termen, men det hindrar inte att det som de avser i själva verket är identiskt med dessa processer. Detta torde vara självklart och det är ju det som vetenskapen i grund och botten handlar om: vad saker och ting verkligen är.

I analogi härmed används mentala predikat, t ex "smärta" inom ramen för den strikta IDTn som benämning på, inte en egenskap som korrelerar med någon komplex materiell egenskap, utan som benämning på just det materiella komplexet. Lekmannen använder visserligen ordet "smärta" om det som han upplever, men enligt IDT är det som han upplever ett högre, emergent materiellt komplex. Lika litet som en förklaring av reproduktionen i materiella termer innebär ett förnekande av att reproduktionen är ett reellt fenomen innebär förklaringen av smärta i materiella termer ett förnekande av smärtan som reell, mental process.

Den strikta IDTn är sålunda verkligen monistisk och skiljer sig därmed från epifenomenalismen. Men den är inte reduktionistisk i och med att den accepterar emergenta tillstånd. Den är interaktionistisk i

så måtto att den hävdar interaktion mellan materiella tillstånd på olika nivåer, t ex mellan mentala processer och livsprocesser.

Denna beskrivning är naturligtvis mycket skissartad, men det är inte här möjligt att gå in på detaljer. Det är emellertid viktiga och fascinerande problem vi här har att göra med och jag hoppas diskussionen skall fortsätta och fördjupas. □

Recension

Anders Gullberg, *Det fängslande planeringstänkandet och sökandet efter en verklighetsutväg,* Lund 1984 (distr.: Hb SALG, Gullberg, Råggatan 1, 116 59 Stockholm).

Denna avhandling i sociologi innehåller, förutom en granskning av teorier och tankar om samhällsplanering, en filosofisk diskussion om orsaksrelationers natur och hur vi bäst skaffar oss kunskaper om orsaker och effekter. Syftet är främst att ge en vetenskapsteoretisk grund för utvärdering av planeringsansatser.

Gullberg går till angrepp mot den sedan Hume vitt utbredda uppfattningen att orsakssamband inte är annat än korrelationer mellan typer av händelser. Han menar att en sådan kausalitetsuppfattning inte tar verkligheten på allvar utan blir fångad i en *ontologi* som projicerats från en i vid mening empiristisk och positivistisk *metodologi.* En sådan ontologi ser verkligheten som uppbyggd av "de isolerade och iakttagbara händelser som kan beskrivas som värden på variabler".

Som ett fruktbart och lovande alternativ - en "verklighetsutväg" - ansluter sig Gullberg till "den vetenskapsteoretiska realismen" som den förespråkats främst av Roy Bhaskar. Enligt denna realism finns nödvändiga samband i naturen, orsakssamband som vi finner dem är grundade i produktiva mekanismer.

Avslutningsvis vill Gullberg peka ut några metodologiska konsekvenser. Främst skall realismen ge en grund till en friare metodologi, byggd på "kontextuerande" fallstudier, jämfört med den gängse som ser det naturvetenskapliga experimentet som ideal.

De filosofiska delarna av *Det fängslande planeringstänkandet* är mycket ofullgångna. Kritiken mot den tradition Gullberg vill ta avstånd från blir alltför ofta bara en svepande polemik. Det realistiska alternativet presenteras i en form som bäst kan kallas ett kommenterat citatcollage. Det finns många intressanta tankespår i avhandlingen. Jag hoppas Gullberg kommer igen och försöker renodla sina *egna* tankegångar.

Bengt Molander

Sven Danielsson

Blooms Tes

Blooms Tes, "No sport from a miss", brukar ju antas innebära att inget (icke analytiskt) påstående om Molly kan härledas ur en premissmängd som inte innehåller åtminstone något påstående om Molly. Tesen är tilltalande, men det finns starka invändningar mot den:

Betrakta följande påståenden: (1) Molly är mullig. (2) Stephen är slank. (3) Stephen är inte slank. (4) Molly är mullig, eller också är Stephen slank. Påståendet (1) är uppenbarligen en utsaga om Molly, medan (2) och (3) lika uppenbart *inte* handlar om Molly. Om nu (4) handlar om Molly, så är Blooms Tes felaktig, därför att (4) följer ur (2). Och om (4) inte handlar om Molly, så är Blooms tes felaktig därför att (1) följer ur (3) och (4). Alltså måste Blooms Tes överges, åtminstone i denna form.

Den kunde förstås försvaras i formen: Inget påstående, som handlar bara om Molly, kan härledas ur premisser som handlar bara om andra.

En annan version vore: Inget påstående om Molly kan härledas ur en mängd *sanna* premisser som inte innehåller åtminstone något påstående om Molly. Frågan om (4) handlar om Molly görs beroende av sanningen hos (2) och (3). Om Stephen är slank, så handlar (4) inte om Molly. Om Molly är mullig, så handlar den inte heller om Stephen.

(För uppslaget till denna version som, tror jag, i en mer generaliserad form skulle kunna innebära betydande nyvinningar i den semantiska teoribildningen – man har ju brukat anta att en sats sanningsvärde beror på vad den handlar om – står jag i tacksamhetsskuld till Lars Bergströms resonemang på sidan 9 i Filosofisk tidskrift nr 4 1984.)

Recension

Derek Parfit: *Reasons and Persons,* Clarendon Press, Oxford, 1984.

Oxfordfilosofen Derek Parfits sedan länge förebådade bok *Reasons and Persons* utgör ett otvetydigt belägg för att Parfit är en av de mest idérika filosoferna i samtiden. Boken innehåller ett överdåd av snillrika tankar - ja, så många att Parfit har måst ge avkall på kompositionens krav för att få med dem alla. Det är ingen stor överdrift att karakterisera verket mer som en redovisning av vad Parfit har tänkt än som en tematiskt enhetlig konstruktion.

Bokens huvudsyfte är att kritisera omhuldade och grundläggande föreställningar rörande praktisk rationalitet, moral och personlig identitet. Det är bara ifråga om personlig identitet denna kritik växer ut till en positiv teori. Vad moralen beträffar lämnar Parfit medvetet läsaren i sticket, men även för rationalitetens vidkommande gäller att Parfits positiva förslag framstår som högst fragmentariska. Icke desto mindre är boken fängslande läsning, dels därför att den drivs fram av en aldrig sinande ström briljanta infall, dels därför att ett otal fantasieggande och science-fictionartade illustrationer förlänar framställningen charm och åskådlighet.

I den första av verkets fyra delar undersöks om vissa allmänna antaganden rörande rationalitet och moral är självupphävande i en begränsad mening. För moralens del består detta bl a i en granskning av den konsekvensetiska tesen att man bör handla så, att resultatet blir så gott som möjligt. Parfits slutsats är att detta syfte motverkas om man låter sitt handlande direkt styras av denna princip. Bättre konsekvenser uppnås om ens beteende i stället regleras av ett detaljerat regelsystem i stil med det sunda förnuftets moral. Parfit kommer här nära R M Hares dubbeldäckarteori i *Moral Thinking* (Oxford, 1981).

Den rationalitetsapproach som skärskådas är egenintresseteorin, S, vilken går ut på att det för var och en finns ett rationellt mål framför andra: att det egna livet förlöper så väl som möjligt. Uttrycket "så väl som möjligt" kan specificeras på olika vis, t ex i termer av maximal tillfredsställelse av ens viljeattityder livet igenom. Så tolkad blir S ingen egoistisk teori, eftersom det inte utesluts att ens viljeattityder gäller andras välfärd. Parfits argument är här att om en agent ständigt har sitt eget bästa för ögonen, så tenderar resultatet att bli sämre för honom än det annars kunde ha blivit. Personligen är jag inte övertygad av detta resonemang, men även om det hade ägt riktighet, hade denna "självupphävande" effekt inte varit förödande för S, ty - som Parfit också påpekar - S fordrar inte att ens beteende i livets alla skeenden direkt kontrolleras av egenintressets kalkyler, lika lite som en konsekvensetisk grundsyn hindrar att man för det mesta följer färdiggjorda tumregelsnormer i stället för att hänge sig åt komplicerade konsekvensberäkningar.

S är således inte vederlagd, och det är därför befogat att fortsätta den kritiska granskningen av den i del II, vilken ägnas åt hur tidsaspekten kommer in i rationalitet. Parfits strategi är här att försöka visa hur svårt det är för en anhängare av S att förklara varför man inte skall vara partisk gentemot vissa *tidpunkter* - t ex det innevarande ögonblicket - om man skall vara partisk gentemot en viss *person,* sig själv. Termerna "jag" och "nu" är semantiskt likartade, varför då privilegiera den ena? Som en rival till S ställer Parfit upp en teorityp, P, vilken gör gällande att en rationell person bör söka den maximala tillfredsställelsen av sina *nuvarande* viljeattityder. I kontrast till P favoriserar S inte nuvarande attityder på bekostnad av attityder i det förflutna eller framtiden.

Denna temporala neutralitet hos S kan emellertid betyda två skilda saker som Parfit inte håller isär med erforderlig klarhet. Enligt en mer beskedlig uttolkning innebär den (1) att om man hyser två viljeattityder, vilka är identiska i alla avseenden förutom att den första är nuvarande och den andra framtida (eller förfluten), har man inget skäl att föredra stillandet av den första attityden (eller tvärtom). Den andra, mer långtgående utläsningen är (2) att om man hyser två attityder som är av samma styrka, men som har olika objekt, och en är nuvarande och den andra framtida (eller förfluten), har man inget skäl att föredra stillandet av den förra framför den andra (eller tvärtom). Uppenbarligen går (2) utöver (1) i det att den diskvalificerar hänsyn rörande en attityds *objekt* som skäl för att föredra tillgodoseendet av en nuvarande viljeattityd. Det står också utom tvivel att Parfit förstår S så, att den omfattar även (2) (se t ex s 154).

Parfit avlossar (s 130–2, 149–58) förödande invändningar mot tesen (2). Helt klart är det möjligt att en rationell person, som förutser att han kommer att utvecklas i en viss riktning om han inte vidtar mått och steg, kan fördöma denna utveckling som degeneration, göra sitt bästa för att förhindra den och se till, att om han misslyckas med detta, så skall hans förfallna jags vilja inte kunna tillfredsställas. Däremot är de av Parfits argument som berör (1) - d v s hans diskussion huruvida våra spontana böjelser, att bry oss mer om den nära framtiden än den mer avlägsna och mer om framtida välbehag och smärta än om de förflutna, är rationella eller ej (s 158–86) - inkonklusiva, trots all sin genialitet. Detta är också anledningen, tror jag, till att Parfit till slut medger (s 188–9) att ett krav på temporal neutralitet kan infogas i hans version av P, den kritiska versionen, CP. Förmodligen avses här bara neutralitet i betydelsen (1).

Resultatet av del II blir att S är definitivt vederlagd endast om den inkorporerar (2). Men i så fall överdriver Parfit vikten av sin slutsats när han gör anspråk på att ha kullkastat en doktrin som "de flesta människor har trott på i mer än två årtusenden" (s 194). Det tycks mig som om en så speciell tes som (2) med säkerhet bara kan tillskrivas ett fåtal nutida filosofer. Kanhända har det varit ett stående antagande under ett par årtusenden att människor är i grunden självviska och att det är oförnuftigt att offra långsiktiga större fördelar

till förmån för kortsiktiga mindre. Men detta rör (1) snarare än (2), förutom att det är ett misstag att förväxla egenintresse i S's form med självviskhet. Och även om S viker för P så garanteras inte att altruistiskt beteende är rationellt, såvida inte Parfit lyckas bevisa sin version av CP, enligt vilken altruistisk motivation är *intrinsikalt* rationell - en tanke som han leker med (s 121, 133, 194), men aldrig försöker föra i hamn.

I viss mån fortsätter prövningen av S i del III där ämnet är personlig identitet, för detta har relevans för en bedömning av S-teorins betonande av jaget. Parfits mästerliga behandling av föreställningen om en persons identitet genom tiden innehåller tre huvudteser. (a) Personer är ingenting utöver sina "hjärnor och kroppar och skilda interrelaterade fysiska och mentala händelser" (s 216). Personlig identitet måste följaktligen bestå i psykiska och fysiska processers kontinuitet i tid och rum - Parfit framhäver speciellt den psykiska kontinuiteten. (b) Det finns inte alltid ett icke-godtyckligt svar på frågan om personlig identitet föreligger eller ej. Det vill säga, den psykiska kontinuitet som föreligger mellan två personer, vilka existerar vid olika tidpunkter, kan vara mer eller mindre omfattande - de kan ha fler eller färre gemensamma minnen och karaktärsdrag - och det finns en gränszon där kontinuiteten varken är tillräckligt stark för att man skall kunna hävda att vederbörande är samma person eller tillräckligt svag för att detta skall kunna förnekas med bestämdhet. (c) "Personlig identitet är inte vad som spelar roll. Det som spelar roll är i grunden . . ./psykisk kontinuitet/ . . ., med vilken orsak som helst" (s 217); fysisk kontinuitet, att samma (delar av en) kropp fortlever, är nästintill betydelselöst för ens känslomässiga inställning till en person. Själv kan jag skriva under på mycket i teserna (a)-(c), men på en kärnpunkt vill jag ifrågasätta Parfits position: är inte psykisk kontinuitet, bevarandet av minnen och karaktärsdrag, ur emotionell synvikel lika oväsentlig som fysisk kontinuitet? Låt oss emellertid först följa hur Parfit når fram till sin ståndpunkt(c), att det är psykisk kontinuitet snarare än personlig identitet som spelar roll för en rationell persons känslor.

Gör detta antagande: om *corpus callosum,* förbindelselänken mellan hjärnans två hemisfärer, skärs av, är var och en av de två isolerade hjärnhalvorna nog för att vidmakthålla full psykisk kontinuitet. Antag vidare att jag är en individ i en grupp identiska trillingar. I en olycka skadas vitala organ i min bål, men min hjärna förblir oskadd. För mina två bröder är det precis tvärtom: deras hjärnor totalförstörs, men för övrigt är deras kroppar intakta. Neurokirurgin har utvecklats till den grad att man är i stånd till att klyva hjärnor och transplantera delarna. Min hjärna tudelas, och mina bröders kroppar får var sin halva. Föreställ er nu två fortsättningar på den här historien: (1) bara en av hjärnhalvstransplantationerna lyckas eller (2) båda transplantationerna lyckas. Om jag före operationen förutser utfallet (1), skulle jag då inte behandla den resulterande personen som om han var identisk med mig själv (se till att han får tillgång till min egendom etc)? Kvalitativt sett är han ju precis lik mig. Men kanske är den resulterande personen verkligen identisk med mig (vilket

är Parfits åsikt), och i så fall indikerar inte förekomsten av mina omsorger om honom att vad som i realiteten spelar roll är något annat än personlig identitet. Detta uppenbaras först när man skärskådar versionen (2). Här kan jag, menar Parfit (s 260), inte vara identisk med någon resulterande person, för jag kan omöjligen vara identisk med båda, och det vore godtyckligt att identifiera mig med den ena. Icke desto mindre, kan jag annat än att bry mig om dessa två dubbelgångare lika mycket som den ende i (1)? Ingenting skiljer ju dem åt. Därför måste den relation som bär upp nämnda komplex av känslor vara en annan relation än personlig identitet. Förslagsvis psykisk kontinuitet, ty de två personerna i (2) är fullt kontinuerliga med mig i detta hänseende, men fysiskt kontinuerliga med mig bara i begränsad utsträckning (i form av en hjärnhalva).

Parfits uppfattning – att psykisk kontinuitet, men inte fysisk, är väsentlig – ter sig emellertid egendomlig när det går upp för en att ifrågavarande psykiska kontinuitet förutsätter en fysisk. Observera att vi här inte talar om mental kontinuitet i betydelsen en ström av upplevelser utan avbrott, ty i så fall skulle den kärvänliga inställningen till det egna jaget inte kunna överleva perioder av medvetslöshet och drömlös sömn. Nej, vad som här åsyftas är en kontinuitet som utgörs exempelvis av att man minns vad en person vid en tidigare tidpunkt tänkte och varseblev. Men att minnas en tidigare upplevelse är inte bara att ha en "bild" som troget återger denna upplevelse; det innebär också att denna "bild" står i en viss orsaksrelation till den föregående upplevelsen. Normalt består den kausala bakgrunden av att den första upplevelsen lämnar ett "spår" i hjärnan, ett spår som existerar kontinuerligt sålänge man kan erinra sig upplevelsen. Parfit föreställer sig att psykisk kontinuitet kan bibehållas även om de bakomliggande kausala mekanismerna är andra än de gängse – som jag redan har noterat talar han om psykisk kontinuitet "med vilken orsak som helst". T ex tänker han sig en maskin som exakt registrerar tillståndet hos en levande kropps alla celler och neuroner, förstör kroppen och därefter återskapar en perfekt kopia av ny materia. Även detta kausala förlopp, resonerar Parfit, upprätthåller psykisk kontinuitet.

Det sista påståendet är inte självklart sant, men jag är beredd att acceptera det för resonemangets skull. Tesen, att psykisk kontinuitet vilar på fysisk, står ändå oemotsagd, för i detta, liksom i Parfits andra science-fiction exempel, hänvisar orsaksförklaringen till kontinuerliga materiella processer. Om man eliminerar denna referens till fysisk kontinuitet, reducerar man inte då psykisk kontinuitet till enbart en kvalitativ likhet mellan tidigare och senare upplevelser? Jag kan inte annat än se att detta blir följden. Men om nu appellen till fysisk kontinuitet utgör den enda skillnaden mellan psykisk kontinuitet och psykisk likhet, och fysisk kontinuitet är känslomässigt oväsentlig, följer det att på sin höjd psykisk likhet kan ha emotionell signifikans.

Ett tankeexperiment kan belysa kontinuitetens perifera ställning. Antag att en slumpmässig händelse av detta slag äger rum då och då: de elementarpartiklar som konstituerar en persons kropp förskingras, men en bråkdel av en

sekund senare bildar andra elementarpartiklar en konstellation som är en fulländad kopia av denna kropp. Här finns ingen kontinuitet mellan de två kropparna, men ändå kan jag inte förstå hur en rationell person skulle kunna frukta denna händelse eller hysa andra känslor för kopian än för det framtida jag den kopierar. I förlängningen av detta resonemang finns en konsekvens som kan hända är ännu mer omskakande. Föreställ er att resultatet av den slumpmässiga händelsen inte är en kopia av den person man faktisk *är* utan en materialisering av den person man *skulle vilja vara,* av ens ideal- eller drömjag - skulle en rationell person då inte hoppas och önska att den slumpmässiga händelsen inträffar och att han "efterträds" av denna idealperson snarare än att han fortsätter att existera i sin medelmåttighet? Detta vore ett grundskott mot S.

I fjärde och sista delen tar Parfit upp några moraliska problem som uppstår i vårt förhållande till kommande generationer. Man kan tycka att det är självklart att en handling är moraliskt fel endast om den förvärrar situationen för någon. Parfit vill kasta tvivel över denna princip genom följande typ av exempel. Antag att vi har att besluta hur en viss naturtillgång skall exploateras. Om vi hårdexploaterar den,kommer mänskligheten att åtnjuta en behaglig välfärd 300 år framöver, men därefter kommer resursen att vara nästan uttömd, och de därpå följande generationerna kommer att få genomlida en drastisk sänkning av levandsstandard - dock inte så drastisk att deras liv inte blir bättre än icke-existens. Med alternativet en mer begränsad exploatering under de närmsta 300 åren åsamkas de då levande generationerna en viss förlust i välstånd jämfört med vad hårdexploateringen skulle medföra, men denna policy innebär betydligt bättre levnadsvillkor för de följande generationerna. Det är naturligt att tycka, att om vi väljer hårdexploatering, handlar vi moraliskt förkastligt: för att tillskansa oss en mindre fördel förvärrar vi situationen högst betydligt för dem som kommer att leva om 300 år. Men det finns här ett förbiseende som Parfit sätter fingret på. Det är sannolikt att vårt val av policy får genomgripande effekter på samhället i stort. T ex kommer andra män och kvinnor att träffas om vi verkställer den ena policyn snarare än den andra. Resultatet blir att de personer, som existerar om 300 år om vi väljer hårdexploatering, inte är identiska med dem som skulle existera om vi fastnade för det andra alternativet. Valet av hårdexploatering förvärrar därför inte situationen för någon person, eftersom livet om 300 år fortfarande är bättre än icke-existens.

Jag är inte beredd att sträcka vapen inför detta argument. Om vårt val i dag av hårdexploatering är moraliskt fel, kan det inte vara för dess konsekvenser om 300 år, för dessa konsekvenser inställer sig bara om de mellanliggande generationerna ansluter sig till vår policy, och vårt val förblir fel även om de bryter med den och introducerar en mer restriktiv policy. Anledningen till att vårt beslut är förkastligt tycks snarare vara att det i ett visst avseende förvärrar situationen för *nästa* generation: det krävs större uppoffringar av dem än av oss för att de skall kunna iscensätta en hushållspolicy som är lika effektiv

som den vi hade kunnat initiera en generation tidigare. Om våra barn i sin tur skyggar för detta offer och fortsätter på den inslagna vägen, försämrar de situationen på ett analogt sätt för sina barn o s v. Är detta synsätt - som förklarar policyvalets felaktighet med hänvisning till dåliga konsekvenser för nästa generation - korrekt, glider man undan Parfits problem. Det är nämligen inte rimligt att hävda att valet av policy påverkar identiteten av nästa generation, ty denna har redan fötts vid valögonblicket.

Därmed har jag eliminerat ett av de problem Parfit anser en adekvat moralteori måste komma tillrätta med. Ett annat krav han ställer på en dylik teori är att den skall undvika "den motbjudande konklusion", vilken går ut på att en ökning av en populations storlek kan kompensera en sänkning av välståndsgenomsnittet. Ponera att vi har två populationer, A och B, där B är dubbelt så stor som A. Alla varelser i en och samma population åtnjuter samma grad av välfärd: i A är den +100 och i B +60. Det är inte orimligt att säga att B är bättre - eller åtminstone inte sämre - än A. En välkänd form av utilitarism säger just precis detta. Men i så fall bör också en population C, vilken är tre gånger A's storlek och vilken har en välfärdsnivå på +45, vara bättre än A. På så vis erhåller man den motbjudande konklusionen att ett enormt stort antal individer med liv som är precis värda att levas bildar en helhet som är bättre än en mindre population med avsevärt bättre levnadsvillkor.

Naturligtvis har det framlagts många teorier som inte leder till denna motbjudande slutsats, t ex den variant av utilitarism som vill höja välståndsgenomsnittet så mycket som möjligt. Men Parfit argumenterar bestickande för att de kan belastas med andra kontraintuitiva konsekvenser. Så kanske man i stället bör fråga sig om den motbjudande konklusionen verkligen är motbjudande i den bemärkelse som är relevant vid testningen av moralteorier, nämligen *moraliskt* motbjudande. Att en sänkning av vars och ens välfärd är motbjudande för de drabbade ur ett *egoistiskt* perspektiv är oomtvistligt, men också ointressant. En omfördelning av jordens tillgångar, vilken består i att en stor del av I-ländernas överflöd överflyttas till U-länderna, ter sig givetvis motbjudande för I-länderna ur deras självviska synvinkel. Likväl kan åtgärden vara moraliskt motiverad. I ljuset av denna tankegång förefaller det inte alltför långsökt att tänka sig att moralen också kräver, att därest alla existerande varelser lever liv som är mer än väl värda att levas, bör de, om så är möjligt, reproducera sig och fördela sitt värdeöverskott över sin avkomma tills genomsnittslivet blir endast minimalt bättre än icke-existens. Observera att gränsen för moralens krav på uppoffringar ligger vid ett liv som är precis *värt att levas* och inte vid ett liv på existensminimum eller dylikt. Moralen måste tillerkänna en rätten att leva ett liv värt att levas, annars undermineras fundamentala förpliktelser som den att rädda eller bevara liv. Genom att ställa stora fordringar på ett liv värt att levas kan man således hålla moralens pockande på uppoffringar i schack.

Parfit söker en moralteori som dels löser det tidigare nämnda problemet om kommande generationers identitet, dels undviker denna motbjudande konklusion, men tvingas erkänna sitt sökande som förgäves. Om min diskussion är riktig, framstår detta misslyckande som lättbegripligt: Parfit har ställt felaktiga adekvanskrav. Men med tanke på de många intrikata problem och insiktsfulla observationer som hopas i Parfits bok verkar ändå en tillfredsställande moralteori mer avlägsen än någonsin.

Parfit har, sammanfattningsvis, skrivit en utomordentligt stimulerande och uppslagsrik, men något ostrukturerad, bok – en bok som knappast någon med ett seriöst intresse för filosofi kan unna sig att ignorera, ty den kommer säkerligen att för lång tid framåt prägla den anglosaxiska diskussionen av moralteori, rationalitet och personlig identitet.

Ingmar Persson

Notiser

Vid filosofiska institutionen i Lund arrangerades 29 oktober 1985 en Temadag, som syftade till att tala om för omvärlden vilka intressanta saker som filosofer idag håller på med, även ur samhällets synvinkel. Det övergripande temat "Filosofi—Samhälle" ville understryka att filosofin behandlar angelägna problem. Åtta forskare av olika generation presenterade forskning om vitt skilda saker som medicinsk etik, samhällsfilosofi, vetenskapsteori, människokännedom, argumentationsanalys, kvalitetskriterier m m. Inbjudna från HSFR, BFR, näringsliv och kommunala förvaltningsorgan samt allmänheten lyssnade i Carolinasalen från 9—17.00 med en avslutande allmän diskussion om filosofin i samhället. Med över sextio inbjudna deltagare blev Temadagen en framgång, som befäste intrycket att Humaniora idag möter en allt bredare respons. Temadagen har stötts av Crafoordska Stiftelsen, för vilket arrangörerna är mycket tacksamma. Även Rektorsämbetet stödde initiativet på ett mycket aktivt sätt. Syftet är att följa upp med en Proceedingsskrift av mindre format.

Bertil Mårtensson

Stellan Welin har redigerat en antologi med uppsatser av svenska vetenskapsteoretiker: *Att förstå världen,* Doxa 1984.

Tore Strömberg har utgivit *Rättsfilosofins historia,* andra upplagan, Studentlitteratur 1985.

På Tidens förlag har utkommit *Hur skall vi ha det med jämlikheten?* av S O Hansson m fl, 1984.

Erik Ryding, som är filosofidocent i Lund, har fått besked av läkare om att han kommer att dö inom ett år. Med anledning härav har han nu utgivit en mycket läsvärd bok med titeln *Karon i luren tutar,* som behandlar frågor om liv och död och livets mening. Doxa 1985.

Ytterligare nyutkomna böcker är: Kåre Elgmork, *Videnskapelig metode,* Universitetsforlaget, Oslo 1985. Jens Allwood och Erland Hjelmquist (utg) *Foregrounding Background,* Doxa 1985. Tore Nilstun och Göran Hermerén, *Utvärderingsforskning och rättsliga reformer. Analys av orsaker och effekter,* Studentlitteratur, Lund 1984. Arno Werner (utg) *Filosofi och kultur 2, Etiken och estetiken,* Filosoficirkeln, Lund 1985. Rättsfonden har utgivit *Manipulation med människan,* Liber 1985, från ett seminarium i februari 1984.

Årets Hägerströmsföreläsare vid filosofiska institutionen i Uppsala har varit professor Saul Kripke från Princeton, U.S.A. Hans föreläsningar handlade om identitet och tid.

Sven Wedar, som tidigare varit docent i praktisk filosofi vid Lunds universitet, har nu utgivit en bok med titeln *Realism and validity, Studies in the legal theory of Alf Ross,* Lund 1985.

Ingmar Persson har utgivit *The Primacy of Perception, Towards a neutral monism,* Library of Theoria No 16, Gleerup 1985.

Professuren i praktisk filosofi i Stockholm efter Harald Ofstad har nu ledigförklarats. Ofstad skall avgå i december 1986.

Medarbetare i detta nummer av *Filosofisk tidskrift*: *Staffan Carlshamre* är doktorand i filosofi i Göteborg, *Wlodek Rabinowicz* är docent i praktisk filosofi i Uppsala, *Hans Rosing* är lärare i filosofi vid Åbo Akademi, *Bengt Molander* har doktorerat i teoretisk filosofi i Uppsala, *Sven Danielsson* är docent i praktisk filosofi i Uppsala, och *Ingmar Persson* är forskarassistent i praktisk filosofi i Lund.

Manuskript till Filosofisk tidskrift

- sändes till redaktören, Lars Bergström, Stora Ängby Allé 24, 161 54 Bromma
- skall vara försedda med namn och adress
- skall som regel vara skrivna på svenska, men bidrag på norska och danska accepteras också
- skall vara maskinskrivna på A4-papper med skrift endast på arkets ena sida
- bör ej innehålla mer än 1 500 nedslag per sida och med bred vänstermarginal (räkna med 50 nedslag/rad och med 30 rader/sida = 1 500 nedslag)
- noter och litteraturhänvisningar bör inarbetas i texten
- särskild litteraturförteckning upprättas i alfabetisk ordning och placeras sist i manus
- för icke beställt material ansvaras ej

Korrektur

- läses i regel endast av redaktören
- ändringar mot manuskript bekostas icke av förlaget; författaren debiteras kostnad för sådana ändringar

Honorar

- införda bidrag honoreras ej för närvarande

Särtryck

- I stället för särtryck erhåller artikelförfattaren, gratis, 10 fullständiga exemplar av det nummer a tidskriften i vilket bidraget varit infört

25:— BOKFÖRLAGET THALES

ISSN 0348-7482