

## JOHAN BRÄNNMARK

### *Deltagarmodellen för preferensutilitarism*

---

I *Filosofisk tidskrift* 4/96 presenterar Wlodek Rabinowicz<sup>1</sup> ett resonemang kring olika varianter av preferensutilitarism. Närmare bestämt handlar det om vad det egentligen är som bör tillmätas egenvärde inom preferensutilitarismen. Exakt vad som gör en preferensutilitarist till just en preferensutilitarist kan få vara osagt, men en viktig beståndsdel torde vara att denne menar att intrinsikalt värde baserar sig på preferenser. Hur skall då detta ”baserar” preciseras? Till att börja med menar Rabinowicz att det inte enbart skall vara subjektets egna preferenser som ses som värdegrundande samt att hänsyn blott bör tas till intrinsikala preferenser (dvs vi intresserar oss för vad det är som vi föredrar för dess egen skull). Han kan dock också tänka sig att man ytterligare kan skärpa kraven på vilka preferenser som skall räknas genom att endast se till de som vi t ex skulle ha om vi vore fullt rationella och informerade; men på den här punkten lämnas dock frågan öppen om vad det mer exakt skulle handla om för krav.

Givet förutsättningarna menar Rabinowicz att det finns två huvudmöjligheter för hur man kan utforma en preferentialistisk axiologi. Den ena är en tillfredsställande-tolkning där vad som tillmäts egenvärde är det förhållande att våra preferenser tillgodoses; den andra är en objekts-tolkning där det är själva föremålen för våra preferenser

<sup>1</sup> Från början bygger dock artikeln i FT 4/96 på en längre förlaga (”Value Based on Preference: On Two Interpretations of Preference Utilitarianism”) som författats tillsammans med Jan Österberg och som publicerats i bl a Rabinowicz (red): *Preference and Value. Preferentialism in Ethics*. Lund: Studies in Philosophy, 1996:1. De sidhänvisningar som föreligger här är till artikeln i FT 4/96. Det sätt att karaktärisera de två positionerna (immanens kontra transcendens) som jag använder mig av är dock hämtat från en av Jan Österbergs passager i den längre förlagan.

som tillmäts egenvärde. Termen tillfredsställelse är måhända en smula olämplig såtillvida att den lätt leder tankarna till någon form av upplevelse, vilket är ett villospår eftersom vad som avses endast är själva realiserandet av det som preferenserna är preferenser för. Men med detta sagt finns det här inte något behov av att föra in någon ny term för den position som avses.

Mot ovanstående tolkningar svarar, enligt Rabinowicz, två olika sätt att se på utilitarism: betraktar- respektive deltagarmodellen. En betraktande preferensutilitarist går in i situationer med en klar övertygelse om vad som bör göras, nämligen att maximera preferensstillfredsställelse, och försöker sedan att avgöra hur detta bäst skall åstadkommas. En deltagande utilitarist däremot går in i situationer med ett slags öppenhet för vad som ytterst sett är av värde. Skillnaden mellan de två ligger i att betraktarutilitaristen arbetar utifrån en substansiell axiologi, medan deltagarutilitaristen arbetar utifrån en formell; med det senare menas att deltagarutilitaristen har visserligen en uppfattning om var egenvärde är lokaliserat, men inte till vilka substansiella entiteter som det kommer att vara knutet. Detta leder till skillnader i vad för slags preferenser dessa två typer av utilitaristiska agenter kommer att bilda för moraliskt relevanta situationer. Beträktaren kommer att bilda ett slags derivat-preferenser, vars bildande närmast skulle kunna gå att åskådliggöra med en form av praktisk syllogism. Deltagaren däremot kommer att "ta över" de inblandade individernas intrinsikala preferenser, dvs att bilda identiska preferenser.

För Rabinowicz är det den axiologiska aspekten som är den centrala, men i den föreliggande texten kommer jag emellertid att närma mig ämnet från det andra hållet. Den axiologiska frågan kommer att till viss del skjutas åt sidan till förmån för en diskussion av i första hand den deltagarmodell som Rabinowicz menar är knuten till objektstolkningen.

### *1. Deltagande kontra betraktande*

Termerna "deltagande" respektive "betraktande" tycks innebära att det är fråga om två epistemiska metoder med vilka en agent kan angripa en relevant situation. I det ena fallet handlar det om att genom inlevelse ta till sig de andra personernas preferenser, medan det i det andra fallet handlar om att genom ett slags distanserat kalkylerande bedöma de andra personernas preferenser. Men tanken tycks emellertid

inte vara att det är en skillnad i epistemiska metoder som i första hand skiljer deltagaren och betraktaren åt; snarare handlar det om vad som görs utifrån den kunskap som inhämtats. När väl informationen om de andras preferenser finns framför ögonen på agenten då tar betraktaren ett steg tillbaka och försöker skapa sig en överblick, medan deltagaren tar ett steg framåt och tar till sig, internaliserar, de föreliggande preferenserna – dvs betraktaren strävar mot ett slags transcendens medan deltagaren strävar mot ett slags immanens.

På den här punkten kan man tänka sig två olika varianter av vad processen ifråga är till för: att komma fram till en värdering eller till ett beslut. I fallet med deltagaren kommer de två, givet att idealiska förhållanden råder, att sammanfalla i och med att vad han gör är att han reducerar en interpersonell intressekonflikt till en intrapersonell. Det han vid processens slut värderar högst är också det han kommer att försöka realisera. När betraktaren skall fatta sitt beslut om hur han skall handla står han däremot fortfarande kvar inför en interpersonell konflikt och de inblandade personernas preferenser föreligger som ett slags data som han kan utgå från. Den drivande motivationella kraften är hans benevolens (alternativt hans vilja att göra det rätta), men den tävlar med andra drivkrafter inom honom. Om man ser det i termer av vektorer blir skillnaden att hos deltagaren är det fråga om faktiska, motivationella vektorer, medan hos betraktaren är det fråga om ett slags hypotetiska; detta förhållande kommer också att gälla resultatena, dvs sluttillstånden av övervägandeprocesserna.

Både när det gäller deltagande och betraktande är det osannolikt att dessa idealiseringar kommer att realiseras fullt ut ifråga om verkliga agenter. De sluttillstånd som uppnås kommer att vara "smutsiga": deltagaren kommer inte att fullt ut kunna ta till sig alla de inblandade personernas relevanta preferenser och betraktaren kommer att ta till sig en del av de inblandade personernas preferenser på ett sätt som gör att det beslut han fattar kommer att vara ett där en del av de inblandade personernas preferenser kommer att verka som motivationella krafter inom honom vid sidan av den övergripande preferens som bildats ur hans utilitaristiska överväganden. Men det faktum att vare sig renodlat deltagande eller renodlat betraktande kommer att realiseras i den verkliga världen är emellertid inte något hinder för att tala om dessa två tillvägagångssätt som ett slags idealtyper.

Rabinowicz går i sin uppsats igenom ett antal invändningar mot

objekts- respektive tillfredsställelsetolkningen. Som tidigare nämnts ämnar jag dock här ägna mig åt den andra sidan av saken: deltagande-respektive betraktarmodellen. Jag utgår från att betraktarmodellen är den som utilitarister normalt använder sig av. Det som jag kommer att intressera mig primärt för här är därför deltagarmodellen. Jag kommer att börja med att ta upp frågan om det överhuvudtaget är fråga om en form av utilitarism och sedan kommer jag att ta upp några mer interna problem som den tycks vara behäftad med.

## 2. Är deltagaren överhuvudtaget en utilitarist?

Blotta det faktum att de flesta utilitarister tycks vara anhängare av en betraktarmodell leder till misstankar kring huruvida deltagarmodellen alls är en form av utilitarism. Visserligen verkar det finnas en del snarlika drag, men för att avgöra frågan tycks det som om vi helst skulle vilja ha ett svar på frågan: "Vad är det som gör en utilitarist till utilitarist?"

Tyvärr är frågan inte helt lätt att besvara. Det har visserligen under de senaste seklernas moralfilosofi funnits gott om både filosofer som kallat sig själva utilitarister och sådana som blivit kallade utilitarister av andra. Gemensamt för alla dessa tänkare tycks dessutom vara ett begränsat knippe idéer; men att försöka vaska fram en slutgiltig uppsättning av nödvändiga/tillräckliga villkor med vilken klassen av utilitarismer kan mutas in är en uppgift som jag inte här vågar mig på. Två egenskaper som dock är mycket vanliga inom utilitarismer av olika slag är dels en strävan efter något slags maximering, dels att denna maximering skall vara av något slags välfärd. Om deltagarmodellen avviker på båda dessa punkter betyder det kanske inte automatiskt att den är diskvalificerad som utilitarism, men det innebär åtminstone ett ifrågasättande som deltagarutilitaristen på något sätt bör kunna försvara sig mot om han fortfarande vill göra anspråk på att vara utilitarist.

Om vi börjar med maximering innebär deltagarmodellens immanenta form av övervägande att något verkligt ställningstagande för maximering inte görs. Tanken är dock den att i och med att alla relevanta preferenser tas in i agentens övervägandeprocess blir resultatet samma handling/slutvärdering som om det hade varit fråga om en betraktare och därmed skulle det finnas grund för att åtminstone se deltagarmodellen som ett slags *de facto*-utilitarism. Det är emellertid

värt att betona att *de facto*-momentet ligger just i vilken handling/ slutvärdering som deltagaren kommer fram till, inte i ett slags *de facto*-maximering av preferenstillfredsställelse. I både fallet med betraktaren och deltagaren bygger resultatet av deras överväganden på den situation som fanns innan de började överväga – men i fallet med deltagaren innebär själva övervägandeprocessen att situationen förändras (i och med att deltagaren tillägnar sig vissa preferenser) på ett sätt som inte är fallet med betraktarens övervägande. Därmed kan det heller inte, givet att resultat av de båda typerna av överväganden är detsamma, vara så att deltagarmodellen mynnar ut i en form av *de facto*-maximering. Det här behöver inte nödvändigtvis vara en invändning mot deltagarmodellen i och med att den fungerar fortfarande som en modell för övervägande, men för frågan om huruvida det är ett slags utilitarism talar dessa avvikelser från maximering för att frågetecken måste sättas för dess status som utilitarism.

När det gäller välfärd är måhända problemen ännu större. Utilitarismen ses som en nytto-moral. Men begreppet "nytta" tycks sakna plats inom ramen för deltagarmodellen. Anledningen till det är dock att dess immanenta form skalar bort den meta-nivå på vilken betraktarutilitarister rör sig. Att *tala* om nytta blir därför poänglöst inom deltagarmodellen, men vad man gör är ju att faktiskt tillägna sig andra personers nyttofunktioner. I båda fallen blir det alltså fråga om ett slags aggregeringar av nyttofunktioner. Det kan dessutom påpekas att för deltagaren kan det räcka med att röra sig på en rent psykologisk nivå; det krävs inte att han för sig själv lyckas matematiskt representera de relevanta motivationella krafterna. Deltagarmodellen innebär inte att man helt slipper de problem med interpersonella nyttojämförelser som kan ställa till det för en utilitarist, men den öppnar för en möjlighet att släppa det besvärliga kvantifieringsideal som betraktarmodellen har att dras med; i stället kan man nå fram till resultat genom ett slags immersion i de motivationella rå-data man samlat in.

Vidare finns det mycket i deltagarmodellen som trots allt kan få en att misstänka att den bör klassas som en form av utilitarism. Deltagarmodellen diskriminerar inte på något direkt sätt mellan olika intrinsikala preferenser. Den kännetecknas av samma slags instrumentalism som de flesta sorters utilitarismer; med detta menar jag att inom icke-utilitaristisk etik kan man visserligen rent allmänt sätta värde på att preferenser uppfylls men samtidigt mena att det finns vissa "onda"

preferenser där redan själva tillfredsställandet av dem i sig har negativt värde. Utilitarister menar däremot att det i och för sig generellt ligger ett negativt värde i att uppfylla "onda" preferenser, men att detta negativa värde är en resultant i en kalkyl där den "onda" preferensens tillfredsställelse har ett positivt värde, som dock är mindre än det negativa värde som beror på att den för med sig att en rad andra preferenser frustreras. Att inte särskilja mellan det "högre" och det "lägre" är ett av nyttomoralens kardinal-drag.

En annan sak som talar för att deltagarmodellen bör ses som en del av det utilitaristiska lägret är att den, precis som de flesta utilitarismer, inte tar åtskillnaden mellan olika individer på allvar. Måhända kan det tyckas en smula underligt att ta upp denna rawlska invändning mot utilitarism som ett argument för tesen att deltagarmodellen är en form av utilitarism. Egentligen är det dock bara en fråga om en variant på föregående synpunkt, dvs deltagarmodellen har utilitarismens icke-diskriminerande karaktär. Om man känner sig tvekan till huruvida deltagarmodellen uppfyller utilitarismens maximeringstanke kan man möjligen se det som att denna odiskriminerande totalitetssyn på de hänsyn som måste tas fångar just de centrala elementen i den tanke på maximering som tenderar att finnas i "vanlig" utilitarism.

I slutändan blir förmodligen svaret på frågan om huruvida deltagarmodellen är en form av utilitarism beroende på vem frågeställaren är. För en utilitarist är måhända avvikelserna alltför stora för att deltagarmodellen skall kunna betraktas som renlärig (eller ens som utilitarism överhuvudtaget), medan för en icke-utilitarist är förmodligen likheterna alltför stora för att det skall kunna vara fråga om något annat än utilitarism. Självt kommer jag att här behandla deltagarmodellen som ett slags utilitarism, om än av ett litet udda snitt, men möjligen har det huvudsakligen att göra med att jag ser på den med icke-utilitaristiska "glasögon". I vilket fall som helst har jag svårt att se att det skulle vara uppenbart att den inte är fråga om ett slags utilitarism.

### *3. Är immanens att föredra framför transcendens?*

Oavsett om deltagarmodellen är en form av utilitarism eller ej är det möjligt att det finns aspekter hos dess immanenta sida som gör att den är att föredra framför betraktarmodellen. Frågan om huruvida en objektstolkning av preferentialistisk axiologi är rimligare än en tillfredställelsetolkning vilar ju inte bara på frågan om huruvida den ena

tolkningen är mer utilitaristiskt renlärig än den andra. Jag kommer här att ta upp tre punkter som jag menar talar för att deltagarmodellen är att föredra framför betraktarmodellen just på grundval av dess immanenta karaktär.

(i) *Deltagaren tar de aktuella preferenserna på allvar.* Anta att jag har en preferens för att regnskogen skall räddas. Vad är det då jag värderar? Jo, det är just regnskogens räddning. Visserligen är det så att det faktum att jag har en preferens för att regnskogen skall räddas gör att om regnskogen faktiskt räddas då realiserar även min preferens. Men det är inte för min preferens' skull som regnskogen bör räddas utan för regnskogens skull. Deltagaren klarar av att göra rättvisa åt dessa aspekter av mitt värderande. En deltagare skulle likt jag värdera just regnskogens räddning, utan att reducera det till ett instrumentellt värde. Beträktaren behandlar däremot preferenser som om deras innehåll är irrelevant (annat än för graden av deras realiserbarhet); men detta innebär att han inte tar de inblandade personernas värderingar på allvar. Genom det "från ovan"-perspektiv som hans transcendentala hållning innebär gör betraktaren ett slags anspråk på att veta bättre än de berörda personerna. Samtidigt förutsätts att vi fortsätter att ha de preferenser som vi faktiskt har. Rabinowicz (s 26) uttrycker detta som att betraktarmodellen i det sätt på vilket den tillmäter intrinsikalt värde parasiterar på strävanden som den omyndigförklarar. Möjligen är detta inte ett avgörande problem om man bara ser på den moraliske agenten som just moralisk agent. Men den moraliske agenten är inte bara moralisk agent; även han har en personlighet, även han ägnar sig åt det immanenta värderande som styr hans vardagliga liv. Det här innebär att en och samma individ å ena sidan, *qua* moralisk agent, skall ha övertygelsen att det enda som har intrinsikalt värde är preferenstillfredsställelse; men å andra sidan skall han, *qua* mänsklig varelse, ha en mängd "intrinsikala" preferenser som utgör hans vardagliga personlighet. Leder då inte detta antingen till ett slags mycket underlig schizofreni eller till att han uppfattar dessa vardagliga preferenser som endast instrumentella? Men om det senare blir fallet så slutar det med att han endast har en intrinsikal preferens: nämligen att tillfredsställa intrinsikala preferenser. Han har alltså en preferens vars objekt är sin egen tillfredsställelse. Hur tillfredsställer man en sådan?

(ii) *Deltagande innebär ett ställningstagande färre.* Det sätt på vilket deltagaren kommer fram till sitt resultat tycks garantera att han

också kommer att vara motiverad att handla därefter. Detta eftersom den interpersonella konflikten reduceras till en intrapersonell; därmed blir de relevanta motivationella krafter som skall balanseras i situationen till krafter inom honom själv och balanserandet sker genom ett slags intern motivationell dragkamp. Eftersom betraktaren däremot behandlar de andra personernas preferenser som data betyder det att hans preferens för att göra det rätta kommer sedan att vara en i mängden av hans totala mängd preferenser. Deltagarens egoistiska böjelser kommer att strida på lika villkor med de utifrån den givna situationen internaliserade preferenserna, medan det finns en allvarlig risk för att betraktarens egna preferenser dubbelräknas: först ingår de som data i den utilitaristiska kalkylen och sedan deltar de som motivationella krafter vid sidan av preferensen för förverkligande av det utilitaristiskt sett riktiga resultatet.

(iii) *Den idealiske deltagaren är en mer sympatisk gestalt.* Utilitarister blir ofta utsatta för beskyllningar som går ut på att de är omänskliga, känslokalla etc. Visserligen är det så att dessa beskyllningar ofta framförs på ett sätt som gör dem osakliga och orättvisa; men det förefaller mig ändå som om det finns något i denna strävan efter att lyckas uppnå ett slags "point of view of the universe" som är en smula obehaglig. Problemet är förmodligen att medan det visserligen är så att sett från universums synvinkel har jag inte större anledning att verka för min egen välgång än för en annans, tycks det samtidigt som att från universums synvinkel finns det inte någon anledning att verka för någons välgång *överhuvudtaget*. Man kan uttrycka det som Rabinowicz att fullständigt uppnådd transcendens leder till att "världen töms på värde" (s 31). Men min poäng här är att *även om* värdena fortfarande är synliga, så har det inte längre någon betydelse; utifrån universums synvinkel är det inte intressant – eller som Nils Ferlin uttryckte det: "Stjärnorna kvittar det lika om någon är född eller död". Engagemanget och subjektiviteten spelar en komplex roll inom moralen; å ena sidan kan de leda oss till snävhet i vår omsorg och partiskhet i vårt handlande, men å andra sidan är det de som driver oss till omsorg och handling. Detta är anledningen till att det ligger något obehagligt i betraktarutilitaristens vilja till transcendens: i sin strävan efter att rensa bort den ena sidan av subjektiviteten och engagemanget röjer man även den andra ur vägen. Kvar finns bara tomhet. Deltagarmodellen med sitt ideal av universaliserad



subjektivitet, i stället för distanserad objektivitet, tycks klara sig bättre på denna punkt.

#### *4. Problemet med reduktionen av subjekt till objekt*

Många av de preferenser vi har går ut på att vi skall få eller uppnå något visst objekt; det kan vara ett föremål, en position, en summa pengar eller något annat. Om Andersson vill ha ett visst jobb då tänker han inte "Andersson skall få jobbet", förmodligen tänker han inte ens "Jag skall få jobbet"; och anledningen är uppenbar, nämligen att han behöver inte tänka i dylika termer. Sådana självorienterade preferenser behöver inte indexeras eftersom Andersson är Andersson.

Men om våra preferenser nu ofta ser ut på detta sätt (och den här typen av självorienterade preferenser är onekligen ganska vanliga), vad händer då vid det preferensövertagande som deltagaren försöker sig på? Ingmar Persson (1996, s 67) har tagit upp den här typen av situationer som en svårighet för deltagarmodellen. Anta att Pettersson sätter sig in i Anderssons situation; tycks inte då det rimliga vara att Pettersson sedan kände glädje ifall Andersson fick det eftertraktade jobbet? Men om Pettersson verkligen tagit över Anderssons preferens kommer han ju själv att eftertrakta jobbet ifråga! Pettersson och Andersson kommer att bli konkurrenter och om Andersson får jobbet kommer Pettersson att känna sig missnöjd. Problemet uppstår eftersom för Andersson gäller det att Andersson är subjekt, medan för Pettersson är Pettersson subjekt och Andersson objekt.

Men argumentet tycks dra väl stora växlar på det självorienterade elementet i vissa preferenser. Vi får inte glömma bort att det handlar om just preferenser. Någon närmare definition av dessa kommer jag inte att försöka mig på, men även taget tämligen vagt handlar preferenser om saklägen och med de flesta saklägen är det så att de utesluter andra saklägen; eller för att fortsätta med det ovanstående exemplet: både Pettersson och Andersson kan inte få jobbet – Anderssons preferens utesluter alltså det sakläge som innebär att Pettersson får jobbet. Pettersson kan alltså inte sägas ha tagit över Anderssons preferens om deltagandet mynnar ut i att han själv vill ha det jobb som Andersson eftertraktar.

Dessutom är det viktigt att här komma ihåg att det finns två steg i övervägandeprocessen. Det första är det epistemiska: att klarlägga alla relevanta fakta i situationen. Det andra är det beslutsmässiga: att

utifrån den faktamängd som föreligger välja det handlingsalternativ som är det riktiga. Det tycks inte rimligt att anta att faktamängden kommer att innehålla icke-indexerade preferenser av typen "Jag vill ha X" – hade så varit fallet skulle även betraktaren ha haft svårt att pussla ihop vad det är han bör göra. Den enda skillnaden ligger i att betraktar-Pettersson kommer att ha en deriverad preferens för att Andersson skall få jobbet, medan deltagar-Pettersson kommer att ha en direkt preferens för det.

### *5. Problemet med sammanhangslösa preferenser*

Människor har de mest märkliga preferenser. De har en del som är mindre märkliga också. Men en sak brukar de preferenser som en enskild individ besitter ha, nämligen något slags sammanhang. Det är t ex ont om personer som är enormt intresserade av Haydns Paris-symfonier, men samtidigt fullständigt ointresserade av både Haydns övriga produktion och all annan musik. Men anta nu att man uppfann ett slags apparat med vilken man genom ett slags elektrokemisk stimulans av hjärnan kunde inplantera preferenser. Vad skulle bli deras preferensers öde? Förmodligen skulle de försvinna nästan lika snabbt som de planterades in, och det av den enkla anledningen att det inte finns något sammanhang i vilket de kan fogas in. Mänskliga agenter är inte ett slags automater som man kan stoppa in viss data i och sedan förvänta sig att denna data skall överleva intakt oavsett vad det är fråga om för typ av data.

Det här skapar problem för deltagarutilitaristen. Om en deltagare t ex internaliserar en preferens för att rädda regnskogarna och strikt bara internaliserar just denna preferens vad kommer den då att få för status i hans överväganden? Det tycks som om den kommer att bli ett slags främmande element och med tanke på att den inte alls är rotad i agentens inre liv förefaller det osannolikt att den skall kunna fungera som en motivationell kraft av den magnitud som krävs.

De preferenser som skall tas över måste bäddas in i ett slags historier som ger dem mening och sammanhang. Egentligen är detta bara naturligt eftersom det vi refererar till med termen "preferens" är inte några faktiskt existerande mentala entiteter med utseendet  $xPy$  utan ett slags knippen av erfarenheter, åsikter, associationer (etc) rörande  $x$  och  $y$ . Det vi benämner preferens kan betraktas som en motivationell resultat av det mentala material vi samlat på oss rörande  $x$  och  $y$ . Men en

resultant måste vara en resultant utifrån något, och ett fullständigt övertagande av en viss preferens skulle alltså innebära att vi tog över det mentala material utifrån vilket den s k preferensen kan ses som ett slags konstruktion. Vad deltagaren skall ta över är alltså ett slags paket utifrån vilket den givna preferensen/resultanten är meningsfull.

På ett sätt är detta problem alltså lätt att lösa såtillvida att det är bara att stipulera att vad man tar över inte är enskilda preferenser, fritt hängande i luften, utan preferenser inbäddade i ett slags menings- skapande sammanhang. Nu är det emellertid så att i verkligheten kommer vi sällan att ha tillgång till särskilt rika historier i fråga om de preferenser som vi har att ta hänsyn till. Dylika epistemiska problem kommer dock alla former av utilitarism att dras med så detta är något som jag väljer att inte diskutera vidare här.

Värre är i så fall att det kan uppstå problem med att om vi faktiskt lyckas internalisera allt det mentala material som ligger bakom de i den givna situationen relevanta preferenserna så är det tänkbart att det i detta samlade material kommer att finnas mot varandra stridande trosföreställningar och i så fall kan det finnas en risk för att vi be- roende på våra trosföreställningar i övrigt kommer att sortera bort vissa preferenser – och då kommer vårt resulterande handlande inte att vara en återspeglning av samtliga relevanta preferenser. Möjligen skulle detta kunna lösas genom att man endast tar hänsyn till de preferen- ser som skulle innehas givet vissa, stränga krav på rationalitet och information. Men även om en sådan manöver skulle kunna rensa ut en del av problemen skulle en del andra fortfarande vara kvar och dessutom skulle alltför långt drivna krav på rationalitet och in- formation göra att modellen skulle bli litet väl *ideal observer*-artad. För att deltagarmodellen skall kännas intuitivt rimlig krävs att man tar de personer man har deltagande med på allvar. Frågan är dock om det är möjligt att åstadkomma fullt ut med det subjektiva elementet bibehållet.

## *6. Problemet med upprepning av preferenser*

Många moraliska problem består i interpersonella intressekonflikter: olika personer har olika preferenser och det är omöjligt att till- fredsställa dem alla. En betraktare kommer att försöka lösa situationen såtillvida att han uppskattar styrkan och riktningen i de berörda personernas preferenser och sedan utför han den handling som kommer

att leda till att så stor preferenstillfredsställelse som möjligt kommer till stånd. En deltagare kommer däremot att *internalisera* de berörda personernas preferenser; på så sätt reduceras en interpersonell konflikt till en intrapersonell. Men en sådan reduktion behöver inte på ett oproblematiskt sätt innebära att alla relevanta preferenser summeras inom en individ.

Den process med vilken man tillägnar sig berörda personers preferenser är av naturliga skäl utsträckt över tiden; först sätter man sig in A:s situation, sedan i B:s osv. Anta nu att det skäl som både A och B har för att vilja handlingen *f* är samma. Är inte då denna situation identisk med en där man har en övervägandeprocess som sträcker sig över tiden och där man om vartannat tar upp ett visst antal olika skäl till övervägning; under sådana omständigheter är det ju inte så att bara för att skälet tagits upp vid ett flertal tillfällen så blir det proportionellt starkare (åtminstone är det så att *blotta* upprepandet av ett visst skäl gör inte att det skälets tyngd blir större). Borde det då inte fungera på ett likartat sätt när deltagaren upplever samma skäl?

Ta följande exempel: en familj står framför två stycken insamlingsbössor. Den ena insamlingen går till att försöka få politiska fångar frigivna och den andra insamlingen går till ett försök att rädda regnskogarna. Familjen har endast en sedel och den ägs av modern. Hon vill gärna ge till välgörenhet, men är indifferent inför frågan om till vilken av de två insamlingarna hon bör ge sin sedel. Hennes make har en preferens för befria politiska fångar, medan de bägge barnen har var sin preferens för att rädda regnskogarna. Faderns preferens är den enskilt starkaste, men taget tillsammans är de bägge barnens preferenser starkare än faderns. Utifrån dessa förhållanden tycks det som om en preferensutilitarist skulle svara att kvinnan, *ceteris paribus*, bör ge till insamlingen för att rädda regnskogarna. Men vad händer i denna situation om modern är en "deltagare"? För att avgöra frågan sätter hon sig alltså in i de övriga familjemedlemmarnas situationer och tar till sig deras preferenser. Från fadern hämtar hon en preferens för att befria politiska fångar. Från det första barnet hämtar hon en preferens för att rädda regnskogarna. Och från det andra barnet? Har hon överhuvudtaget något att hämta från denne? Visserligen håller det andra barnet en relevant preferens, nämligen att rädda regnskogarna, men det är en preferens som modern redan har internaliserat.

Tanken här är, som nämnts ovan, att en parallell kan dras med

vanligt beslutsfattande i vilket vi går igenom våra skäl för respektive emot att utföra en viss handling. Vi kan då återkomma till samma skäl flera gånger. I sådana fall leder som sagt knappast det blotta uppreandet av det skälet till att det stärks eller försvagas. Det faktum att det hålls upp i förgrunden kan innebära att det i just det ögonblicket är starkare. Men några sådana effekter behöver vi inte ta hänsyn till här. Det kan också vara så att när det hålls upp i förgrunden kommer dess styrka att förändras till följd av den granskning som det i just det ögonblicket utsätts för. Man kan naturligtvis också tänka sig att ett uppreande av en viss preferens av en rad andra skäl faktiskt stärker den; men även om det kan finnas sådana effekter är det ytterst tveksamt att det skulle vara fråga om ren addition, vilket är vad som skulle behövas här.

Vad som krävs är att både det första och det andra barnets respektive preferenser blir till separata krafter i moderns övervägande. Jag har svårt att se någon enkel lösning på dessa svårigheter, men misstänker att om en sådan lösning finns bör den gå i linje med det som sades ifråga om problemet med sammanhangslösa preferenser, dvs att vad som internaliseras är inte blott en enkel preferens utan en preferens inbäddad i ett meningsskapande paket. Dessa meningsskapande paket kommer knappast att vara identiska och då kan man tänka sig att de preferenser som är inbäddade i dem kommer att hållas isär och bli till separata motivationella krafter.

### *7. Problemet med utveckling över tiden*

Det kan förekomma hopp i tiden mellan olika inlevelsituationer och under den tiden kan internaliserade preferenser från tidigare situationer rota sig. I så fall är det ju fullt möjligt att man hos andra stöter på preferenser som egentligen härstammar från en själv. Anta att vi har två deltagande utilitarister, Andersson och Pettersson. I situation 1 är Andersson den som skall handla medan i situation 2 är det Pettersson. I båda situationerna står valet mellan  $f$  och  $-f$ . Pettersson har en preferens för  $f$ . Inför sin handling i situation 1 internaliserar Andersson denna preferens. Men vad händer i situation 2 då Pettersson skall handla? Skall han då internalisera den preferens som Andersson i situation 1 har hämtat från honom själv?

En första reaktion skulle kunna vara att sådana preferenser på något sätt borde uteslutas ur en övervägande. Annars tycks det som om vi

skulle kunna få ett slags kedjereaktioner där en del preferenser pumpas upp till abnorma proportioner av den enkla anledningen att de figurerar i en process där två deltagare utför serier av interrelaterade handlingar. Samtidigt är det förstås så att i varje given situation har den andre personen preferensen ifråga och då borde man väl ta hänsyn till den? Vägen ut ur dessa svårigheter borde för deltagarutilitaristen ligga i linje med det som tagits upp tidigare. För att klara sig från problemet med icke-additiviteten hos identiska preferenser tycktes deltagarutilitaristen behöva förlita sig på idén om att man internaliserar preferenser som är inbäddade i stället för nakna; givet denna lösning tycks det rimligt att problemet med icke-additivitet kan uppstå igen givet att vi rör oss med identiska inbäddningar också. Det senare torde vara fallet när man stöter på egna preferenser hos andra och i så fall blir problemet med icke-additivitet här en lösning i stället.

Hur skulle det då se ut om man hade en blandad population, dvs en värld med både deltagande och betraktande utilitarister? Anta, för att fortsätta i enlighet med exemplet ovan, att istället för den deltagande utilitaristen Pettersson blir det i situation 2 den betraktande utilitaristen Lundström som står inför valet om han skall utföra f. Lundström skulle till skillnad från Pettersson ta hänsyn till Anderssons preferens för f. Det tycks alltså som om i en värld där det finns både deltagande och betraktande utilitarister, skulle de två modellerna kunna leda till olika handlingsrekommendationer, såtillvida att det kommer att finnas vissa preferenser som vissa deltagare inte kommer att ta hänsyn till.

## 8. *Slutord*

Jag har diskuterat deltagarmodellen för preferensutilitarism. Tre resultat tycker jag mig kunna skönja. Det första är att deltagarmodellen, trots en del oortodoxa drag, är en form av utilitarism. Det andra är att vid en första anblick tycks det som en del aspekter hos deltagarmodellen gör den mer attraktiv än betraktarmodellen. Det tredje är att deltagarmodellen dras med en del interna svårigheter som inte har sina motsvarigheter hos betraktarmodellen. De lösningsförslag som skisserats här är tämligen hårt uppbundna till en idé om att den internalisering som deltagaren gör inte är av enskilda preferenser utan av ett slags paket där de relevanta preferenserna ses som ett slags resultat utifrån det ingående materialet; ett material som dessutom fungerar meningsskapande åt preferensen ifråga.

Med detta sagt kan jag konstatera att en viss fadd eftersmak in-  
finner sig så här vid slutet såtillvida att den verkligt centrala termen,  
”preferens”, kvarstår som ytterst besvärlig även vid textens slut.  
Många av de problem som tagits upp bottnar i just hur ”preferens”  
skall uppfattats. Vad jag försökt är att ringa in termen genom att se  
den som ett slags resultat, men detta sätt att uttrycka saken gör att  
preferensbegreppet snarast ses som ett slags bekväm ”shorthand”; allt  
detta inom ramarna för en diskussion om värde grundat på preferens.  
Hur uttunnat kan man egentligen göra preferensbegreppet utan att sam-  
tidigt gör det poänglöst att tala om *preferentialistiska* axiologier? Det  
här är emellertid en typ av problematik som det inte här finns något  
utrymme för att gå vidare in på. Den som dock verkligen vill förfäktas  
en tanke om värde grundat på preferens kan dock i längden förmodli-  
gen inte ducka för nödvändigheten i att precisera preferensbegreppet.<sup>2</sup>

### *Litteratur*

- PERSSON, INGMAR (1996): "Benevolence, Identification and Value" i Lindström,  
Sliwinski & Österberg (red): *Odds and Ends. Philosophical Essays Dedicated to  
Wlodek Rabinowicz on His Fiftieth Birthday*. Uppsala Philosophical Studies 45.  
RABINOWICZ, WLODEK & ÖSTERBERG, JAN (1996): "Value Based on Preference:  
On Two Interpretations of Preference Utilitarianism" i Rabinowicz (red.): *Pre-  
ference and Value. Preferentialism in Ethics*. Lund: Studies in Philosophy, 1996:1.  
RABINOWICZ, WLODEK (1996): "Värde grundat på preferens" i *Filosofisk tidskrift*  
4/96.

2 Jag vill rikta ett stort tack till Högre seminariet i Praktisk filosofi vid  
Filosofiska institutionen i Lund för de värdefulla synpunkter som där fördes  
fram på en tidigare version av föreliggande text. Vidare har jag haft stor  
hjälp av konversationer med Wlodek Rabinowicz och Ingmar Persson.