

WLODEK RABINOWICZ

Värde grundat på preferens

Om två tolkningar av preferensutilitarism*

1. Preferensbaserat egenvärde – två tolkningar

Det som skiljer preferensutilitarism (PU) från andra former av utilitarism är värderingskomponenten. Preferensutilitaristen har ett eget förslag om vad det är som besitter egenvärde: Egenvärde eller, som man ibland kallar det för, *intrinsikalt* värde baserar sig på preferenser. Intrinsikalt värdefulla tillstånd kopplas enligt denna uppfattning till tillfredsställelse av våra önskningar.

Följande bör noteras: (a) Vad som anses vara värdegrundande är "våra" preferenser, och inte bara "mina". (b) Hänsyn tas bara till så k "intrinsikala" preferenser; det intressanta är vad vi föredrar för dess egen skull. (c) Eventuellt kan de relevanta preferenserna antas vara "idealiserade" på ett eller annat sätt. T ex kan det röra sig om de

* Denna korta uppsats baserar sig på en längre förlaga författad tillsammans med Jan Österberg. Som kommer att framgå, är Österberg och jag oeniga vad beträffar tolkningen av preferensbaserad axiologi. Denna oenighet har vi försökt utveckla och klargöra, men inte upplösa, i den gemensamma uppsatsen, "Value Based on Preferences: On Two Interpretations of Preference Utilitarianism", *Economics and Philosophy*, april 1996. I den föreliggande korta versionen tar jag tillfället i akt att i första hand framlägga min syn på saken. Medan jag tar hänsyn till några av Österbergs invändningar och försöker besvara dem, samt tar upp ett par kritiska synpunkter som har framförts av Sven Danielsson och Torbjörn Tännsjö, avstår jag att bemöta ett antal invändningar mot min ståndpunkt som bl a har förts fram av Krister Bykvist (i en opublicerad uppsats "Utilitarian Deontologies? On Preference-Utilitarianism and Agent-Relative Value"), John Broome och Derek Parfit (de senare i korrespondens). Broomes och Parfits synpunkter bemöts i viss utsträckning i den längre förlagan. Diskussionen av Bykvists kritik får anstå till ett annat tillfälle.

(intrinsikala) preferenser vi skulle ha om vi vore rationella, välinformerade om fakta och om vi hade tillfälle att noga överväga saken.

Vad är det nu egentligen som PU tillmäter intrinsikalt värde? Är det

(i) det förhållande *att* våra (intrinsikala) önskningar och preferenser tillgodoses? Eller är det snarare (ii) de förhållanden som är *föremål* för våra (intrinsikala) önskningar och preferenser? Observera att våra preferenser tillgodoses i och med att dessa förhållanden realiserar. De vanliga formuleringarna av PU antyder närmast den förra tolkningen, men även den senare tolkningen är tänkbar.

Här är ett exempel som illustrerar den distinktion som jag är ute efter. När vi önskar, säg, att regnskogen bevaras (inte, eller åtminstone inte bara, för något externt syfte utan för regnskogens egen skull), är det då intrinsikalt värdefullt att vi får som vi önskar eller är det snarare regnskogens fortlevnad som, i ljuset av vår önskan, framstår som något intrinsikalt värdefullt?

Även om vi får som vi önskar om, och endast om, regnskogen bevaras, är det uppenbart att det här är fråga om två olika förhållanden. Den ekvivalens som råder dem emellan är kontingent med tanke på att vi kunde ha andra önskningar.

Min uppsaliensiske kollega, Jan Österberg, har såvitt jag vet varit den förste att uppmärksamma denna skillnad mellan två tolkningar av preferensbaserat intrinsikalt värde (i en opublicerad uppsats "A Defence of the Preference Satisfaction Theory of Intrinsic Value"). Låt oss kalla dem för *preferenstillfredsställelse-* resp. *preferensobjektolkningen*. Österberg pläderar för den förra, medan jag själv vill plädera för preferensobjektolkningen i stället. Åtminstone vill jag göra det villkorligt: Om PU överhuvudtaget är en rimlig ståndpunkt, så är den rimligare givet objektolkningen.

2. *Betraktaren och deltagaren*

Mot dessa två tolkningar av preferensbaserad syn på värde svarar, såvitt jag förstår, två skilda sätt att se på utilitarism: *betraktarmodellen* och *deltagarmodellen*. Enligt den förra förkroppsligas den utilitaristiska inställningen i en opartisk välvillig betraktare som bedömer situationen objektivt och "utifrån". En vanlig människa kan approximera denna inställning genom att distansera sig från sitt personliga engagemang i situationen. Hon uppger, så långt det nu är möjligt, sitt partikulära

perspektiv och försöker anlägga "the view from nowhere", för att använda Thomas Nagels välkända formulering.

Deltagarmodellen uppställer däremot som ett utilitaristiskt ideal en attityd av deltagande med andra: det gäller att se på situationen "inifrån", men inte bara ur mitt eget utan även ur det andres perspektiv. Deltagarmodellen förutsätter att jag i stället för att distansera mig från mig själv identifierar mig med andra subjekt: vad den rekommenderar är inte objektivitet utan universaliserad subjektivitet.

Deltagarmodellen har en klart Humesk proveniens: den moraliska inställningen vilar på *sympati* – på vår förmåga till medkänsla och identifiering med andra. Som bekant, ansåg Hume att sympati, genom sin solidaritetsskapande karaktär, "produces our sentiment of morals". Den förklarar vår uppskattning av dygderna.

Att deltagarens identifiering med andra och betraktarens opartiska välvilja kan vara svåra att skilja åt framgår av följande citat från Rawls' *Theory of Justice* (s 26f), där de två modellerna betraktas som en:

The most natural way, then, of arriving at utilitarianism ... is to adopt for the society as a whole the principle of rational choice for one man. Once this is recognized, the place of the impartial spectator and the emphasis on sympathy in the history of utilitarian thought is readily understood. For it is by the conception of impartial spectator and the use of sympathetic identification in guiding our imagination that the principle of one man is applied to society. It is this spectator who is conceived as carrying out the required organization of the desires of all persons into one coherent system of desire; it is by this construction that many persons are fused into one.

Jag skulle nu vilja klargöra hur kontrasten mellan de två modellerna förhåller sig till vår distinktion mellan de två tolkningarna av preferentialisk syn på värde.

Deltagarmodellen innebär att det moraliska subjektet identifierar sig med den andre och därigenom övertar hans preferenser. Den andres intrinsikala preferens blir min egen: objektet för hans intrinsikala önskan blir därigenom något jag föredrar för dess egen skull. Deltagarmodellen leder därför på ett naturligt sätt till att intrinsikalt värde tillmäts objektet för preferensen.

Betraktarmodellen innebär däremot att det moraliska subjektet vill den andre väl. Jag vill dig väl, men detta behöver ju inte innebära att jag vill det du vill. Det kan till och med vara så att den distansering som utmärker betraktarperspektivet gör att jag i detta perspektiv förlorar anledning att vilja sådant som vi människor, med vår partikulära situation i världen, brukar vilja. Det kan då vara frestande att göra det *kantianska* steget och dra slutsatsen att det vi råkar vilja – viljans partikulära föremål – saknar ovillkorligt, absolut värde.

Enligt Kant är det inte viljans föremål – dess mer eller mindre kontingenta "materia" – utan snarare själva vårt predikament att vara viljande varelser som är relevant för frågan om det gudas natur. Om man sedan (inte längre i likhet med Kant, för vilken viljans "form" var det väsentliga) lokaliserar det ovillkorliga värdet i viljeuppfyllelse som sådant snarare än i viljans partikulära föremål, då har man nått en ståndpunkt som påminner om PU i dess tillfredsställelsetolkning.

Mycket förenklat kan man alltså säga att valet står mellan två moralfilosofiska traditioner – mellan impulser från Hume och från Kant.

3. Invändningar mot tillfredsställelsetolkningen

Jag skall nu presentera några argument som talar mot preferenstillfredsställelse- och för preferensobjekttolkningen.

(i) *Värdeinternalism*. Det förefaller vara ett rimligt krav på en moralisk teori att de värden den uppställer bör vara relativt intimt kopplade till våra handlingsmotiv. *Hur* intim denna koppling bör vara kan man säkert tvista om, men kanske kan följande villkor uppställas som ett minimikrav på en (rent värdebaserad, "teleologisk") moralteori.

VÄRDEINTERNALISM: När ingen annan än jag själv berörs, och jag är välinformerad, rationell och har tänkt noga på saken, då sammanfaller de intrinsikala värden som jag enligt den moraliska teorin bör förverkliga med de målsättningar som jag själv är angelägen om att nå.

Med andra ord: under idealiska omständigheter och när jag är den ende som berörs, bör intrinsikala värden sammanfalla med de målsättningar som motiverar mig att handla.

PU i dess objekttolkning har inga svårigheter med detta villkor.

Ponera att jag önskar att förstå relativitetsteorin (och min önskan är av den rätta intrinsikala sorten – jag vill nå denna förståelse för dess egen skull, av ren intellektuell nyfikenhet). Anta att situationen är sådan att ingen annan berörs och det är fråga om en övervägd önskan som jag skulle ha kvar om jag verkligen visste vad jag ger mig in på. PU i dess objekttolkning implicerar att det jag är angelägen om, dvs föremålet för denna min önskan, har ett intrinsikalt värde. PU i dess tillfredsställelsetolkning implicerar däremot att det endast är uppfyllelse av min önskan, vilken denna önskan än må vara, som äger intrinsikalt värde. Den senare målsättningen behöver jag emellertid *inte* vara angelägen om (åtminstone inte intrinsikalt, för önskeuppfyllelsens egen skull). Jag *kan* vara det, vid sidan om mycket annat jag eftersträvar, men det är inte nödvändigt att jag har denna önskan ”av andra ordningen”. I synnerhet kan det mycket väl vara så att min önskan att förstå relativitetsteorin är ”kategorisk” i den meningen att objektet för min önskan är att förstå teorin ifråga och inte ”att förstå den förutsatt att jag önskar mig denna förståelse”. Min önskan behöver med andra ord inte vara betingad av sin egen existens. Men i så fall *kan* jag vara likgiltig inför det faktum att jag genom att realisera objektet för min önskan även får min önskan uppfylld. Att min preferens tillfredsställs behöver inte höra till mina motiverande målsättningar.

Jämför med följande citat ur Stephen Darwalls *Impartial Reason* (Cornell UP, Ithaca and London 1983, kap 3, s 37):

From the fact that any intentional action must be explainable by the agent's desires it does not follow that the agent's reason for acting must itself make reference to his desires. Usually when we act as we desire we are not moved to act by our awareness of the fact that we desire. Rather, a desire itself partly consists in a disposition to be moved by certain facts connected with its *object*. ... Desires are for their objects, and the locus of the desiring person's concern is most typically the object itself and the facts connected with it rather than the fact that she has the desire.

Väsentligen samma tanke förs fram i en inflytelserik uppsats av Philip Pettit och Michael Smith, ”Backgrounding desire”, *Philosophical Review* 99 (1990), 565–92.

(ii) *Butlers problem*. Joseph Butler's berömda kritik av hedonismen tycks i minst lika hög grad träffa värdeuppfattningar som tillmäter in-

trinsikalt värde åt preferenstillfredsställelse snarare än åt objekt för preferenser. Preferenstillfredsställelse kan endast vara möjlig som en sekundär målsättning: om vi hade blott en enda preferens – att våra preferenser blev tillgodosedda – skulle vi ju inte ha några preferenser att tillgodose. I Butler's terminologi, "self-love" – egenkärleken – behöver input. "Take away these affections [= particular passions and appetites], and you leave self-love absolutely nothing to employ itself about; no end or object to pursue..." (Förordet till *Fifteen Sermons*, § 15).

Varför utgör detta egentligen ett problem? Ja, det förefaller att en värdeuppfattning enligt vilken inget annat än preferenstillfredsställelse är intrinsikalt värdefullt å ena sidan "omyndigförklarar" alla våra vanliga intrinsikala strävanden, genom att fränkänna deras föremål intrinsikalt värde, men å andra sidan lever den på dessa omyndigförklarade strävanden. För att uttrycka det drastiskt har vi här att göra med en uppfattning som parasiterar på de attityder som den samtidigt desavouerar. Den skulle ha "absolutely nothing to employ itself about" om vi tog den ad notam och slutade bry oss om annat än att våra preferenser tillfredsställs. (Att vi inte kan ta den ad notam, att sådant är psykologiskt omöjligt, gör väl inte saken bättre?)

Denna uppfattning tycks alltså inte ta på allvar våra önskningar och preferenser samtidigt som den förutsätter att vi fortsätter att ta dem på allvar.

(iii) *Svårbegriplighet*. Att förbehålla allt intrinsikalt värde åt preferenstillfredsställelse ter sig svårbegripligt: Vad kan det finnas för värde i att tillgodose våra strävanden om föremålen för dessa strävanden systematiskt fränkänns ett värde? Varför är det värdefullt att våra önskningar uppfylls om det vi önskar genomgående är värdelöst?

Att folk får det de vill, även i de fall när det de vill saknar värde, kan eventuellt anses vara intrinsikalt värdefullt, vid sidan om mycket annat. Det kan ses som ett värde bland andra. (Många människor tycks lägga en oberoende vikt vid viljeuppfyllelse som sådan och det är inte uteslutet att de skulle behålla denna intrinsikala preferens även om de vore välinformerade och övervägde saken noga.) Men det är inte lätt att begripa hur sådant som preferensuppfyllelse kan framstå som det enda värdet.

(iv) *Hypotetiska preferenser och preferenser avseende hypotetiska fall*. Tänk på en vanlig invändning mot PU: Preferensutilitaristens

värderingar *förefaller* stämma överens med common sense. Han tycks lägga vikt vid sådant som de flesta av oss lägger vikt vid, t ex vid orörd natur, estetiska upplevelser, frihet, kunskap, etc. Men denna överensstämmelse med common-sense värderingar vilar i själva verket på lösan sand. För tänk om vi hade andra preferenser än de vi faktiskt har. Tänk om vi vore likgiltiga inför, säg, naturens rikedom eller inför estetiska upplevelser. Då skulle det enligt preferensutilitaristen inte finnas något värde i dessa tillsynes önskvärda objekt: de skulle ju inte längre tillgodose våra preferenser! Preferensutilitarismens ointuitiva karaktär blir alltså klar när vi betraktar *hypotetiska* fall (i vilka våra preferenser vore annorlunda).

Såvitt jag förstår är denna invändning riktad mot PU i dess preferenstillfredsställesetolkning: hade preferenserna varit annorlunda så skulle deras tillfredsställelse kräva att andra tillstånd realiserar. PU i denna tolkning rättar sig alltså efter hypotetiska preferenser i hypotetiska fall. Preferensobjekttolkningen av PU leder däremot till väsentligen annorlunda slutsatser. PU i denna tolkning rättar sig efter våra *aktuella* preferenser beträffande hypotetiska fall. Om vi har en (övervägd) intrinsikal preferens för att, säg, regnskogen bevaras *även* för det hypotetiska fallet i vilken vi inte skulle ha brytt oss om regnskogens fortlevnad, då implicerar objekttolkningen att regnskogens fortlevnad skulle ha ett intrinsikalt värde även om dess existens inte tillgodosåg några preferenser. Hypotetiska preferenser är alltså av intresse för bedömningen av hypotetiska fall endast om våra aktuella preferenser för hypotetiska fall är känsliga för sådana hypotetiska preferenser – något som de inte behöver vara. Den beskrivna invändningen är därför uppenbarligen otillämplig givet objekttolkningen.

Svårighet: Men – när vi bedömer ett hypotetiskt fall – borde vi då inte, givet deltagarmodellen, identifiera oss även med de hypotetiska aktörerna och därigenom låta deras preferenser också påverka bedömningen? Jag är benägen att svara nej på denna fråga, men vet inte riktigt hur man kan argumentera för detta svar. Jag skulle emellertid vilja hävda att deltagarmodellen förutsätter att det verkliga ges ett absolut företräde framför det blott hypotetiska. När vi identifierar oss med andra och övertar deras preferenser, är det följaktligen bara de verkliga andra och deras verkliga preferenser som som räknas. Observera: om vi hade behövt identifiera oss med såväl de hypotetiska som de aktuella individerna vad beträffar bedömningen av hypotetiska fall,

då hade det funnits anledning att göra likadant vid bedömningen av det aktuella fallet. Varför inte betrakta det aktuella fallet från alla tänkbara hypotetiska perspektiv? Men gör vi så, blir deltagarmodellen bankrutt! Det finns för många tänkbara konflikterande perspektiv att ta hänsyn till – perspektiv som mycket väl kan tänkas neutralisera varandra. Det är därför som deltagarmodellen enligt min uppfattning måste ges en aktualistisk utformning: det är bara de aktuella preferenserna som skall utgöra grund för värderingen. För vidare diskussion av aktualism, se nedan, sektion 4, punkt (v).

Ett annat alternativ vore att för varje hypotetiskt fall bara ta hänsyn till de hypotetiska preferenser som skulle ha förelegat i fallet ifråga. Detta skulle emellertid innebära att intrinsikala värden skulle tillåtas variera från en möjlig värld till en annan, i takt med preferensvariation. Regnskogen skulle vara intrinsikalt värdefull i vår värld, men skulle sakna värde i en värld där folk vore likgiltiga. Den avgörande nackdelen med en sådan uppfattning är att den strider mot alla våra föreställningar om intrinsikalt värde: ett sådant värde skall ju *ex definitione* vara oberoende av varierande externa omständigheter: ett objekt är intrinsikalt värdefullt endast om det är värdefullt under alla omständigheter, i varje möjlig situation.

(v) *Värdepluralism*. PU kritiseras ibland för sin monistiska axiologi: preferensutilitaristerna, säger man, ignorerar det uppenbara faktum att värden är heterogena till sin natur. Värdenas mångfald byts ut mot ett enda värde: att folk får som de önskar. Igen tycks PU strida mot vår intuitiva värdeuppfattning. (Samma sorts invändning brukar förresten riktas mot hedonistisk utilitarism.)

Givetvis kan man lätt bemöta denna invändning om man är anhängare av objektolkningen av PU. Givet denna tolkning är det naturligt för PU att utmytna i värdepluralism. Våra intrinsikala preferenser tycks ju ha mycket varierande objekt. Skulle vi behålla dessa preferenser även efter ”moget övervägande”, då följer det att det finns flera oreducerbart distinkta intrinsikala värden.

(vi) *Värden som inte är värden för någon*. Något man ofta betraktar som utmärkande för alla former av utilitarism är värdeindividualism: en uppfattning enligt vilken ingenting kan vara värdefullt utan att vara värdefullt för någon (för någon ”individ”). Hedonistiska utilitarister placerar värdet i lust eller lycka: men om det är bra att människor är lyckliga så är det väl bra därför att det är bra *för* varje människa att

vara lycklig. Likadant förhåller det sig med preferensutilitarismen i dess vanliga tillfredsställelsetolkning: om det är bra att folk får som de vill så är det väl bra därför att det är bra för var och en att få som han eller hon vill.

Dessa individualistiska utformningar av utilitarism kommer därför i konflikt med en vanlig värdeuppfattning enligt vilken det kan finnas mycket som är intrinsikalt värdefullt utan att för den skull vara intrinsikalt värdefullt för någon person eller ens för någon levande varelse: t ex artvariation, personlighetsmångfald, jämlikhet, vackra landskap i öde trakter, osv.

Igen är det uppenbart att preferensobjektolkningen klarar av detta problem utan besvär: föremål för (övervägda) intrinsikala preferenser är intrinsikalt värdefulla även i de fall när de inte är intrinsikalt värdefulla för någon. (I den mån som de är föremål för någons intrinsikala preferenser och i förlängningen även föremål för denne någons värderingar måste de ju vara intrinsikalt värdefulla enligt någon; men det är en annan historia.)

Att preferensobjektolkningen av PU lämnar utrymme för värdepluralism och för sådant som är bra utan att vara bra för någon gör att denna ståndpunkt närmar sig common sense vad beträffar värdefrågor. Dessa karaktäristika innebär emellertid samtidigt att ståndpunkten ifråga fjärrar sig från våra gängse föreställningar om vad som utmärker ett utilitaristiskt synsätt. Vi får en värdeuppfattning som saknar de för standardutilitarismen så vanliga reduktionistiska tendenserna: värdeindividualism och värdeindividualism lyser med sin frånvaro.

4. Invändningar mot objektolkningen

(i) *Den rawlska invändningen.* I det citat som har reproducerats ovan (sektion 2), hävdar Rawls att det utmärkande för utilitarism är "organization of the desires of all persons into one coherent system of desire". Vad som händer är att "many persons are fused into one". Den välkända Rawlska slutsatsen blir följaktligen att "[u]tilitarianism does not take seriously the distinction between persons" (*A Theory of Justice*, s 27). Invändningen träffar närmast deltagarmodellen (och därigenom indirekt objektolkningen som denna modell ligger till grund för). Deltagarmodellen involverar ju preferensövertagande. Som deltagare tar jag över andras preferenser och gör dem till mina. Men därigenom reduceras interpersonella preferenskonflikter till *intra-*

personella motsättningar som löses genom en kraftmätning: en svagare preferens ger vika åt den starkare. Det finns därför inget som hindrar att en persons svagare preferens uppoftas för att tillgodose en starkare preferens från en annan persons sida. Inte heller finns det något som hindrar att en persons starka preferens uppoftas för att tillgodose svaga preferenser hos ett stort antal andra personer.

Är denna analys korrekt? Inte nödvändigtvis. Den baseras på ett tvivelaktigt "endimensionellt" synsätt på individuell rationalitet. De olika preferenser hos individen som tillsammans ligger till grund för hans val ses här i analogi med krafter som drar i olika riktningar. Det enda som räknas är deras relativa styrka och antal: den slutliga riktningen – det rationella valet – bestäms genom att de olika krafterna adderas i proportion till deras styrka.

Ett sådant perspektiv på individuell rationalitet är klart överförenklat. Agentens preferenser behöver inte alla vara av samma skrot och korn. Intrapersonella konflikter är inte lätta att lösa när personens olika divergerande målsättningar är heterogena och därför svårligen jämförbara med varandra. I problem som involverar konflikter mellan t ex frihets- och kunskapssträvan, eller mellan personlig ambition och vänskapshänsyn, finns det inget enkelt recept för konfliktlösning. Att bara jämföra preferensstyrka räcker inte (även om sådana jämförelser lät sig göras, vilket inte alls är säkert). Att jämka samman olika dimensioner i ens personliga strävanden är ett komplicerat företag som inte lämnar mycket utrymme för enkel styrketävlan. Inte heller finns det någon garanti för att en sådan sammanjämkning alltid blir framgångsrik; amalgamering av heterogena preferenser kan ibland vara ogörlig.

Om vi accepterar vi en sådan flerdimensionell syn på individuell rationalitet, kan vi försöka värja oss mot den rawlska invändningen: att jag som deltagare övertar andras preferenser behöver inte innebära att distinktionen mellan de olika personerna försvinner. De andras preferenser kastas inte bara i en gryta och blandas ihop med varandra. Även efter övertagande förblir de relativt åtskilda. De uppträder nu som olika dimensioner inom den utilitaristiska agentens preferensstruktur. Interpersonell konflikt transformeras därigenom i en komplex *flerdimensionell* intrapersonell motsättning. Distinktionen mellan personer tas i så måtto fortfarande på allvar.

(ii) *Möjliga världar*. I en möjlig värld i vilken människornas preferenser hade varit annorlunda, i vilken de hade önskat annat än det vi

faktiskt önskar, skulle anhängaren av objekttolkningen ha kommit fram till andra slutsatser vad beträffar intrinsikalt värde än i den aktuella världen. Detta förefaller vara orimligt, med tanke på att påståenden om intrinsikala värden normalt inte antas vara världsbundna. Om sådana påståenden är giltiga, gäller de med nödvändighet: för varje möjlig värld.

Svaret på denna invändning – om det nu finns något svar – måste gå ut på att objekttolkningen förutsätter en subjektivistisk (eller ”projektionistisk”) syn på värde. Värden är inte en del av världen utan något vi projicierar på världen, eller rättare sagt på hela systemet av möjliga världar. I denna bemärkelse kan våra värdepåståenden, som ju görs inom ramen för vår värld, ses som uttryck för ett partikulärt världsbundet perspektiv – ett perspektiv som bestäms av våra faktiska preferenser. Men denna världsboundenhet vad *perspektivet* beträffar gör dem inte världsbundna i deras *applikationsområde*: vi projicierar värden på alla möjliga världar och inte bara på den värld som råkar vara aktuell. Därigenom finns det inget hinder för uppställande av värdeomdömen av intrinsikalt slag (dvs omdömen som tillmäter samma värde åt ett sakförhållande i varje möjlig värld).

Till skillnad från anhängarna av tillfredsställesetolkningen vägrar objekttolkaren att höja sig till betraktarens distanserade perspektiv. Från ett sådant perspektiv, hävdar han, blir inga värden synliga längre. Världen töms på värde. Om värden skall räddas måste vi hålla oss kvar inom det subjektiva perspektivet. Ett sådant är med nödvändighet världsbundet och i denna mening kontingent.

Den aktuella invändningen kan alltså besvaras *om* värdesubjektivism är en försvarbar ståndpunkt (något som givetvis är starkt omdebatterat).

(iii) *Objektolkarens opportunist*. Det skulle kunna hävdas att objekttolkningen inbjuder till opportunist; den tillåter mig att påstå att ett sakförhållande, A, är intrinsikalt värdefullt bara därför att många människor föredrar det för dess egen skull. Men detta är ju kontraintuitivt. Om jag betraktar A som *intrinsikalt* värdefullt, måste jag kunna utpeka några *interna* egenskaper hos A som gör det värdefullt. Intrinsikala värden kan inte grundas på *externa* egenskaper, sådana som egenskapen att vara föremål för preferenser.

Men är det uppenbart att objekttolkningen har denna kontraintuitiva konsekvens? När jag tar över andras intrinsikala preferens för A – när

jag gör den till min egen preferens – blir jag i stånd att ange de interna egenskaper hos A som gör det värdefullt. Dessa är precis de egenskaper för vilka A har föredragits av andra, och för vilka jag nu föredrar detta sakförhållande!

Vi bör skilja mellan å ena sidan de egenskaper på vilka intrinsikalt värde "supervenierar" – de karaktäristika hos objektet som vi pekar på när vi ombeds att rättfärdiga vårt värdeomdöme (egenskaper som vi värderar objektet *för*) – och å andra sidan de egenskaper och tillstånd som "konstituerar" – ger upphov till – intrinsikalt värde. Det är bara de förra som måste vara interna egenskaper hos objektet. Enligt objekt-tolkningen konstitueras intrinsikala värden hos sakförhållanden av något externt – av våra (intrinsikala) preferenser. Men dessa värden supervenierar inte på preferenserna. I stället supervenierar de på de interna egenskaper hos sakförhållanden för vilka dessa sakförhållanden intrinsikalt föredras. En analogi med konventioner kan hjälpa här. Att ett visst drag i schack är tillåtet supervenierar på de interna egenskaperna hos draget tillsammans med de interna egenskaperna hos spel-situationen. Men den konstituerande grunden för att draget är tillåtet står att finna i den sociala konvention som upprätthåller schackreglerna.

Schackanalogin är emellertid inte fullständig, såsom har påpekats för mig av Sven Danielsson. Det intrinsikala värdets beroende av dess konstituerande grunder skiljer sig från ett motsvarande beroendeförhållande i schackfallet. Hade schackkonventionen varit annorlunda, skulle draget ifråga inte nödvändigtvis vara tillåtet. Hade däremot preferenserna varit annorlunda, skulle intrinsikala värden ändå förbli desamma. (Se sektion 4 ovan, punkt *(iv)*.) Denna disanalogi har att göra med de intrinsikala värdenas obegränsade applikationsområde. Samtidigt som de kreeras av de aktuella preferenserna, är de *applicerbara* även på möjliga världar i vilka preferenserna hade varit annorlunda. En bättre analogi har föreslagits av Krister Bykvist: Tänk på vår användning av en term som "en meter". Mitt paraply är en meter lång. Den längd "en meter" betecknar fixeras konventionellt genom en relation till en extern standard (sådan som standardmetern i Paris, eller en viss bråkdel av den väg som ljuset färdas under en sekund). Men om vi sedan funderar på en möjlig värld i vilken *standard* hade varit annorlunda och längre, är vi ju inte alls benägna att säga att i en sådan värld skulle mitt paraply ha haft en annan längd än i den aktuella världen. Det skulle fortfarande ha varit en meter lång. (Även om man

i denna tänkta värld hade sagt att det var kortare än så!) I Saul Kripkes terminologi (från *Naming and Necessity*) är en term som "en meter" en s k rigid designator: en term vilkens innebörd konventionellt fixeras på ett sådant sätt att den plockar ut samma objekt eller egenskap i varje möjlig värld, oberoende av vilka konventioner än skulle råda i världen ifråga. Värdepredikat betar sig analogt: De rådande preferenserna bestämmer vilka sakförhållanden som är intrinsikalt värdefulla på det "rigida" sättet – sakförhållanden ifråga förblir intrinsikalt värdefulla även i de världar i vilka andra preferenser skulle råda.

Det bör slutligen påpekas att den intuitiva distinktionen mellan interna och externa egenskaper hos ett objekt, vilken ligger till grund för den aktuella invändningen, inte alls är lätt att klargöra. I svaret på invändningen har den emellertid tagits för given, för argumentets skull.

(iv) *Värden enligt någon*. Objekttolkningen tillåter inga "oupptäckta" intrinsikala värden: allt som är intrinsikalt värdefullt måste intrinsikalt föredras av någon. Detta verkar kontraintuitivt. Kan det inte finnas sakförhållanden som är av värde fast de inte har uppmärksamats av någon, eller åtminstone inte tillräckligt uppmärksamats för att kunna intrinsikalt föredras?

Som den står kan invändningen bemötas genom en *idealiserad* tolkning av intrinsikal preferens (se ovan, sektion 1). Objekttolkningen kan mycket väl tillåta värdefulla sakförhållanden som inte föredras av någon. Det räcker att vi *skulle* föredra dem om vi noga reflekterade över deras (interna) beskaffenhet.

Men det måste ändå erkännas att objekttolkningen i sista analys endast godtar värden som som uppskattas av någon eller skulle uppskattas under lämpliga omständigheter. "Oupptäckbara" värden godtas inte. Är detta erkännande katastrofalt? Eller är det något vi kan leva med?

(v) *Hypotetiska perspektiv*. Deltagarmodellen och objekttolkningen kan kritiseras för utpräglat "aktualistisk" eller "realistisk" orientering: Det ideala utilitaristiska subjektet förväntas identifiera sig med andra aktuellt existerande subjekt, men däremot inte med blott hypotetiska personer. Även när jag betraktar ett hypotetiskt fall antas jag identifiera mig med aktuella subjekt och ta över deras aktuella preferenser beträffande fallet ifråga. Jag förväntas inte ta över de hypotetiska preferenser som skulle föreligga i detta hypotetiska fall. Varför det? Hur kan man rättfärdiga en sådan favorisering av det verkliga och ak-

tuella på bekostnad av det blott möjliga?

Det är en svår fråga att besvara, främst av det skälet att favoriseringen av det verkliga är en så fundamental inställning hos oss människor. Det är som om någon undrade varför vi endast bryr oss om sådant som händer i den aktuella världen men tar med ro alla de hemskheter som utspelar sig i blott möjliga världar. Varför är jag nöjd sålänge den aktuella världen är tillfredsställande? Varför bryr jag mig inte om de tänkbara världar som är fulla av obeskrivligt lidande och smärta? Det enda svar som jag kommer på är: Varför skulle jag bry mig om sådant som inte är fallet?

Anhängaren av objektolkningen kan svara likadant. Varför skulle jag bry mig om subjektiva perspektiv som inte är verkliga? Varför skulle jag ta hänsyn till preferenser som ingen hyser (eller ens skulle hysa efter moget övervägande)? Gränsen mellan det verkliga och det överkliga tycks vara av helt central vikt, även om vikten i denna gränsdragning kanske bara är ett "brute fact", som inte låter sig ytterligare rättfärdigas.

(vi) *Preferensoberoende?* Men i vilken mening är egentligen objektolkningen preferentialistisk i så fall? Den innebär ju att samma sakförhållanden skulle förbli intrinsikalt värdefulla även om folk hade andra preferenser! Hur kan man då hävda att den låter värdet vara beroende av (aktuella) preferenser? Beroendeförhållanden mellan två faktorer borde ju innebära att om en faktor varierar blir även den andra faktorn förändrad. Inget beroende utan hypotetisk variation! (Invändningen har formulerats av Sven Danielsson.)

Såvitt jag förstår bör objektolkaren ta denna invändning på allvar och i viss utsträckning modifiera sin ståndpunkt. Hypotetisk variation kan ges utrymme via *värderelativism*. Hittills har vi rört oss med en ansats i vilken det aktuella preferensperspektivet har förutsatts som givet. Därför var det möjligt för oss att inte låtsas om värdets relativisering till detta perspektiv. Så länge perspektivet hålls konstant, är det möjligt att tala om intrinsikalt värde i absoluta termer, utan relativisering. Men om man så vill, kan man säga att relativiseringen till det aktuella perspektivet har förelegat implicit. En sådan implicit relativisering kan göras synlig om man i stället laborerar med en modellering i vilken vi avstår från att förutsätta ett bestämt aktuellt preferensperspektiv, utan istället tillåter det aktuella perspektivet att hypotetiskt variera. Relative *ett* sådant perspektivval blir då vissa sakförhållanden

intrinsikalt värdefulla. Låter vi i stället ett annat preferensperspektiv fungera som det aktuella – dvs låter vi en annan möjlig värld fungera som den aktuella – får vi en annan uppsättning av intrinsikalt värdefulla sakförhållanden. När värderelativiseringen görs explicit på detta sätt synliggör vi den hypotetiska variation som krävs för ett beroendeförhållande mellan intrinsikala värden och aktuella preferenser. Det bör emellertid understrykas att för varje val av ett aktuellt preferensperspektiv blir de intrinsikala värden som existerar relative detta perspektiv fortfarande invarianta över olika möjliga världar: ett sakförhållande som är intrinsikalt värdefullt i en värld relative ett givet preferensperspektiv förblir intrinsikalt värdefullt relative detta perspektiv även i de världar där preferenser skulle ha varit annorlunda! Ty varje sådant preferensperspektiv bestämmer ju en uppsättning *intrinsikala* egenskaper hos sakförhållanden som intrinsikalt värde (relativiserat till detta perspektiv) supervenierar på: de egenskaper för vilka sakförhållanden föredras för deras egen skull, oberoende av externa omständigheter, utifrån det valda preferensperspektivet.

De läsare som är bevandrade i möjligvärldssemantik för modallogik med rigida designatorer är väl förtrogna med samma dubbelhet vad beträffar relativiteten hos nödvändig sanning: I gängse möjligvärldsmodeller antar man den aktuella världen (eller mera generellt: den aktuella kontexten) som given och detta på förhand gjorda val, som fixerar referensen hos rigida designatorer, gör det möjligt att tala om nödvändiga sanningar i absoluta termer, även vad beträffar påståenden som innehåller rigida designatorer. Men en modell med en fix aktuell värld kan alltid inbäddas i en modell där den aktuella världen inte fixeras: i stället laborerar man där med ett sanningsbegrepp som relativiserar till olika val av aktuella världar (kontexter). I en sådan modell blir olika påståenden nödvändigt sanna (dvs invariant sanna över alla möjliga världar) relativt olika aktualitetsval.

Normalt finns det ingen anledning att göra värderelativiseringen explicit; vi tar ställning till värdefrågor i absoluta termer, utifrån de preferenser vi faktiskt hyser. Men objektolkaren kan alltid falla tillbaka på en sådan relativisering när beroendet mellan värden och preferenser sätts ifråga.

(vii) *Preferentialistisk axiologi?* Men innebär inte det ovan sagda att objektolkningen i själva verket inte alls är en form av preferensutilitarism utan snarare får ses som en s k objective list-teori om intrin-

trinsikalt värde kombinerad med en projektionistisk preferensbaserad metaetik? För objektolkaren spelar preferenserna en viktig roll endast vad värdekonstitution beträffar men inte vad beträffar värdeinnehållet: i princip behöver de inte omnämnas alls när vi i en "objektiv lista" specificerar vilka sakförhållanden som är intrinsikalt värdefulla. Objektolkarens axiologi, till skillnad från hans metaetiska värdeteori, behöver därför inte alls vara preferentialistisk. (Invändningen ifråga har framförts av Torbjörn Tännsjö i ett bokmanuskript med den preliminära titeln *Classical Hedonistic Utilitarianism: A Defence*, kap 6.)

Även om denna invändning är bestickande bör den ändå tillbakavisas. Den bygger på ett antagande att det går att dra en skarp gräns mellan metaetik och substantiell etisk teori. Men kan en sådan skarp gränsdragning upprätthållas? Även om vi ignorerade Quines välkända argument och insisterade på en strikt åtskillnad mellan substantiella frågor och frågor om mening, förblir gränsen mellan metaetik och axiologi oklar. Visserligen kommer rent semantiska teser om meningen hos värdepåstående tillhöra den förra, medan rena värdekataloger kommer att räknas till den senare. Men teser av den typ som görs inom objektolkningen förblir svåra att kategorisera. Att värde konstitueras av preferenser är en tes om värdets natur och inte en utsaga om meningen hos värdepåstående. Som sådan är den en filosofisk tes med viktiga substantiella implikationer. Den tycks överbrygga gränsen mellan det metaetiska och det axiologiska.

Dessutom, även om objektolkningen ändå kategoriserades som primärt en metaetisk ståndpunkt, kan invändningen fortfarande bemötas genom hänvisning till rådande praxis. Det förefaller som om såväl filosofer som icke-filosofer i sina diskussioner av preferensutilitarism ofta (om än ej alltid) har i tankarna en eller annan version av objektolkningen, med värden tillkommande objekten för preferenser. (Jämför t ex Sen och Williams' inledning till deras välkända antologi *Utilitarianism and Beyond*, 1982, eller Dan Egonssons lundsiska avhandling om preferensutilitarism *Interests, Utilitarianism, and Moral Standing*, 1993.) Dessutom bör man ju inte glömma att den deltagarmodell som ligger till grund för objektolkningen har en klart utilitaristisk karaktär. Den metaetiska karaktären hos objektolkningen skulle därför inte diskvalificera den från att vara en version av preferensutilitarism. I stället skulle preferensutilitarismen själv visa sig vara en metaetisk uppfattning, i en av sina vanliga tolkningar.