

Jag tror inte att man måste vara preferensutilitarist, eller ens preferentialist rörande vad som har ett personligt värde, för att betrakta det som etiskt betydelsefullt huruvida någons preferenser är tillfredsställda eller frustrerade. En majoritet av etiska teorier verkar vara, för att använda Krister Bykvists uttryck, *preferenskänsliga* (2001, s. 18), i bemärkelsen att de ser åtminstone *något* värde i tillfredsställelsen av preferenser.

Av de teorier som uttryckligen har denna preferenskänslighet förutsätter de flesta någon form av idealisering av de preferenser som räknas – varje faktisk preferens är inte värdegrundande. J. C. Harsanyi kräver i en välkänd formulering att vi ska räkna bara med de sanna eller rationella preferenserna:

... en persons sanna preferenser är de preferenser han *skulle* ha om han hade all relevant information, alltid resonerade med största möjliga omsorg, och befann sig i ett sinnestillstånd som var optimalt för träffandet av rationella val. (1982, s. 55)

Många av de kända formuleringarna av ett sådant idealiserings- eller rationalitetskrav på preferenser, delar ett uppenbart problem med Harsanyis formulering: vad är vitsen med att tillfredsställa en rent hypotetisk preferens? Det är inte bara svårt för en persons *faktiska* preferenser att uppfylla kravet; det verkar också tillskriva värde till tillfredsställelsen av sådana preferenser som ingen individ faktiskt har.

Det är sant att man kan ge en snävare och en vidare tolkning till kravet. Å ena sidan kan man förstå det bara som ett negativt krav eller en bortsorteringsmekanism som inte tillerkänner något värde till tillfredsställelsen av sådana faktiska preferenser som inte är rationella. Å andra sidan kan man förstå det som ett positivt krav som tillskriver betydelse även till de preferenser en person skulle ha aktuella om hon var rationell. Jag utgår här från den senare tolkningen. Men jag tror inte att detta är avgörande, eftersom ett liknande problem uppstår i samband med den svaga tolkningen: Vad är vitsen med att avstå från att tillfredsställa en

faktisk preferens med hänvisning till att individen inte skulle haft preferensen i en situation som hon faktiskt inte befinner sig i? Vad gör man med andra ord med de bortsorterade preferenserna?

Denna uppsats behandlar problemet med de hypotetiska preferenserna.¹ Problemet uppstår i samband med traditionella rationalitetskrav, och i synnerhet i samband med det som man på engelska kallar "the full information account of preference rationality" (här översatt till "informationskravet", och som i stora drag motsvaras av Harsanyis² krav). Dessutom vill jag undersöka om de omformuleringar som skett (eller skulle kunna ske) av kravet är bättre rustade att hantera problemet. Jag kommer att argumentera för att de kanske är *bättre* rustade, men att problemet inte heller i dessa formuleringar har fått sin fulla lösning. Detta kan ha betydelse för om man överhuvudtaget bör ställa något informationskrav på preferenser. Det är emellertid inte en diskussion som jag tänker ge mig in i här.

1. ETT EXEMPEL

Tillåt mig utgå från en typ av exempel som jag tidigare använt på annat håll (1990).

Anta att din dotter älskar hårdrock. Du vet att hon mycket gärna vill ha en hårdrocksskiva när hon fyller år. Anta nu att du har skäl tro att hennes svaghet för hårdrock baseras på att hon inte kommit i tillräcklig kontakt med andra musikstilar. Du tror att om hon bara gav den klassiska musiken en chans, så skulle den ge henne djupare och mer intensiva upplevelser.

Här finns, antar jag, en instrumentell och, enligt informationskravet, *irrationell* preferens för hårdrocksskivan. Preferensen uppfyller inte informationskravet. Men vad kan du göra åt det? Det bästa vore att du underkastade din dotter sådan kognitiv träning som gjorde henne medveten om att det är den klassiska musiken som kan ge henne de djupaste upplevelserna. Men detta är inte realistiskt. Du vet att du inte skulle få henne att genomgå en sådan träning. I så fall verkar enda sättet för dig att uppfylla informationskravet vara att köpa den klassiska skivan, efter-

¹ Ett liknande problem har av Robert K. Shope kallats för "the conditional fallacy", 1978, särskilt s. 411–3. Se också Amartya Sen och Bernard Williams, 1982, s. 10, Dan Egonsson, 1990, kap. IV och Bengt Brülde, 1998, ss. 275–8.

² Jag låter här Harsanyi representera en lång tradition filosofer som formulerat rationalitetskravet på preferenser på ungefär samma sätt. Se t.ex. Henry Sidgwick, 1907, ss. 110–1, John Rawls, 1972, s. 416 och Richard B. Brandt, 1979, t.ex. kap VI och s. 247.

som detta är vad du har skäl tro skulle tillfredsställa din dotters rationella preferens. Samtidigt vet du att hon aldrig kommer att lyssna på den skivan.

Det är svårt att se på vilket sätt hennes liv skulle bli bättre av att du gav henne den klassiska skivan. Preferensen för att lyssna på klassisk musik är inte aktuell hos din dotter och den kommer heller inte att bli aktuell som ett resultat av dina handlingar. Och än en gång: problemet är inte att din dotter är irrationell i betydelsen att hon inte faktiskt baserar sin preferens på fullständig information och logiskt tänkande. Problemet är att hon inte har, och vad vi vet heller inte kommer att få, en *aktuell* preferens som rent *hypotetiskt* skulle klara rationell kritik. Med andra ord, problemet med hypotetiskheten handlar inte om den *rationella kritiken* men väl om *preferensen*. Att agera utifrån en aktuell preferens som kan uppfylla vissa hypotetiska krav är inte orimligt.

Någon vill kanske påstå att informationskravet både är orimligt och överflödigt. Hårdrocksexemplet visar att det inte är rimligt, men kan också visa att det inte heller behövs. I exemplet har vi att göra med en instrumentell preferens för en viss musik vars uppgift är att skänka positiva upplevelser. Vårt problem skulle försvinna om vad som har värde i samband med preferenser ytterst sett är tillfredsställelsen av *intrinsikala* preferenser.

Koncentrerar du dig på din dotters intrinsikala preferenser, dvs. preferenser rörande objekt som för henne har ett intrinsikalt värde, så kommer du att köpa hårdrockskivan, om du tror att detta kommer att tillfredsställa hennes intrinsikala preferenser för positiva upplevelser i större utsträckning än en skiva klassiskt kommer att göra. (Och du kan resonera så även om du anser att det finns annat än konsekvenser som har betydelse för en handlings status.) På detta sätt kan du undvika att göra henne besviken och göra det bästa av situationen. Informationskravet är överflödigt i den meningen att det är specifikationen av personligt värde i termer av tillfredsställelsen av *intrinsikala* preferenser som verkar ha legat bakom kravet.

Denna position vilar på förutsättningen att dessa intrinsikala preferenser i sin tur inte är irrationella. Låt mig därför säga några ord om intrinsikalitet. Jag tror detta begrepp handlar om fokus och vad den som har preferensen värderar objektet *för*. I en intrinsikal preferens fokuserar man på objektet, dvs. man värderar objektet för dess interna egenskaper. De skäl man har för preferensen grundas på dessa interna egenskaper. Detta betyder inte nödvändigtvis att dina intrinsikala preferenser står opåverkade av fakta rörande externa relationer. Jag tror att det finns en familj av olika slags intrinsikala preferenser. Vissa av dessa är konditio-

nala: man värderar ett objekt för dess interna egenskaper, endast så länge man tror att det står i relation till ett annat objekt.

För att ta ytterligare ett musikexempel: det skulle kunna vara så att jag ser skönheten i ett musikstycke bara så länge jag tror att det är Schubert som skrivit det. Så snart jag inser att det komponerats av ett billigt datorprogram, försvinner det estetiska värde som jag såg i musiken. Och observera att jag inte värderar stycket *för* dess kompositör: fokus ligger hela tiden på musikstyckets interna egenskaper. Skönheten som jag ser i objektet självt är här *konditional* på mina trosföreställningar om de relationer som objektet har till andra objekt.

Men här är det också viktigt att skilja mellan vad vi kan kalla för intern och extern konditionalitet.³ För att börja med det senare. Det skulle kunna vara så att mina fördomar mot datorkomponerad musik hindrar mig från att ägna stycket den tid och uppmärksamhet som krävs för att se skönheten i dess interna egenskaper. På grund av mina förutfattade meningar tar jag inte musiken på tillräckligt mycket allvar för att uppskatta dess rätta värde, dvs. det värde den skulle ha för mig om jag gav den en chans. Här rör det sig om extern eller kausal villkorlighet: en trosföreställning står rent praktiskt i vägen för en intrinsikal preferens. Den konditionalitet som beskrevs i förra stycket är i stället intern. Jag har verkligen pejl på musikstyckets interna egenskaper, men påverkas ändå (eller skulle påverkas) av den nya informationen att det i stället för att ha Schubert som upphovsman har kommit till i en dator. En uppsättning interna egenskaper utgör skäl för att betrakta ett stycke som vackert bara om jag samtidigt tror att Schubert skrivit det.

Starkare former av intrinsikala preferenser beror inte på detta (interna) sätt på trosföreställningar hos den som har preferensen (varken i positiv bemärkelse på närvaron av en trosföreställning eller i negativ på frånvaron av en sådan). Jag kan anse att ett musikstycke är vackert utan att ha någon som helst uppfattning om vare sig uppkomsthistoria eller andra externa relationer. Och jag kan fortsätta gilla ett sådant musikstycke även efter det att jag lärt mig att det skrivits av en massmördare eller billigt datorprogram. I så fall skulle jag säga att vi har att göra med en *icke-konditional* eller ovillkorad preferens.

Det är rätt uppenbart att den konditionala preferensen kan vara irrationell enligt informationskravet (låt säga om jag felaktigt tror att det musikstycke jag lyssnar på skrivits av Schubert). Men jag tror också man kan argumentera för att den icke-konditionala och intrinsikala preferensen kan vara irrationell och förändras som ett resultat av att den som

³ Diskussionen av denna skillnad bygger på kommentarer av Ingmar Persson och Wlodek Rabinowicz.

har preferensen får ny information. Än starkare blir argumentet för detta om vi förstår "information" i en rikare betydelse, som inkluderar att man direkt konfronterats med objektet och dess alternativ. Jag kommer emellertid inte att diskutera de icke-konditionala preferenserna här.

Anta att du på falska grunder tror att ett visst slags musik är inne för tillfället och att dina intrinsikala preferenser för dess estetiska kvaliteter delvis är grundade på denna tro. I så fall är dina preferenser irrationella. Översatt till vårt exempel: Anta att din dotter på grund av en längre utlandsvistelse inte haft möjlighet att följa musiktrenderna bland sina vänner, och nu vill du ge henne en present. Ett rykte som du tror på, men däremot inte din dotter, säger att alla i hennes gamla gäng har blivit välartade älskare av klassisk musik. Vad gör du? Än en gång har vi ett fall där du skulle göra din dotter besviken om du bortsåg från hennes irrationella preferens, samtidigt som du vet att du inte kan förmå henne att lyssna på den klassiska skivan.

Det verkar alltså som om informationskravet har ett hypotetiskhetsproblem också i samband med de intrinsikala preferenserna, och om så är fallet kan vi förstås inte, som förslaget löd, lösa detta problem genom att fokusera på de intrinsikala preferenserna och utgå från att problemet endast kommer att störa oss på nivån med de externa och instrumentella preferenserna. I nästa sektion ska jag fråga mig om inte en modifierad version av informationskravet kan lösa hypotetiskhetsproblemet för alla preferenser.

2. SMITHS TVÅ MODELLER

I artikeln "Internal Reasons" har Michael Smith diskuterat vad han kallar för *internalismkravet* på skäl. Han skriver:

Internalismkravet på skäl föreskriver att önskvärldheten av att en aktör ϕ -ar i vissa omständigheter C beror på om hon skulle önska att hon ϕ -ade i C om hon varit fullständigt rationell. Denna idé kan göras mer precis på följande sätt.

Vi kan föreställa oss två möjliga världar: den *värderade* världen där aktören befinner sig i de omständigheter som faktiskt gäller för henne, och den *värderande* världen där vi finner aktörens fullständigt rationella jag. Med denna terminologi så föreskriver internalismkravet att önskvärldheten av aktörens ϕ -ande i den värderade världen beror på om hennes fullständigt rationella jag i den värderande världen skulle önska att hon ϕ -ade i den värderade världen (1995, s. 110).

Poängen är, säger Smith, att det rationella jaget förväntas betrakta sitt mindre rationella jag i den värderade världen, och utifrån vad hon ser forma ett önskemål rörande vad hennes mindre rationella jag bör göra i de

omständigheter hon befinner sig. Smith menar att vi ska föreställa oss att det rationella jaget ger det mindre rationella jaget råd rörande vad hon bör göra och följaktligen kallar han detta för *rådsmodellen* av internalismkravet.

Denna modell särskiljs från *exempelmodellen*, där önskvärldheten av att någon utför en viss handling beror på om hennes rationella jag skulle föredra att utföra handlingen i den värderande världen, dvs. i en värld där hon var fullständigt rationell. I den här modellen ”sätter det fullständigt rationella jaget upp sitt eget agerande i hennes egen värld, den värderande världen, som ett exempel att följas av hennes faktiska jag i den värderade världen” (1995, s. 110). Med andra ord, i den första modellen bestäms aktörens skäl av vad som vore rationellt för henne att göra givet hennes faktiska irrationalitet, under det att de i den andra modellen bestäms av vad som vore rationellt för henne att göra i en rent hypotetisk värld där hon är rationell.

Vi ser lätt hur man kan konstruera ett informationskrav utifrån dessa modeller och att de problem som vi har diskuterat i den föregående sektionen kommer att dyka upp igen i en av modellerna. En rådsmodell av informationskravet kommer att betrakta en aktörs rationella preferenser som de hennes rationella jag skulle önska att hennes irrationella jag tillfredsställer.⁴ Exempelmodellen av kravet, å andra sidan, kommer att betrakta som agentens rationella preferenser de preferenser som hennes rationella jag skulle önska att aktören hade tillfredsställt om hon varit rationell.

De problem som diskuterades i förra sektionen var föranledda av likheten mellan exempelmodellen och de klassiska formuleringarna av informationsvillkoret. Innan vi ser efter om ett informationskrav byggt på *rådsmodellen* skulle kunna lösa problemet med de hypotetiska preferenserna, ska jag kort rekapitulera Smiths eget exempel på skillnaden mellan modellerna:

⁴ Observera att Smith beskriver olika modeller för hur man kan förstå *önskvärldheten* i att en person handlar på ett visst sätt, medan jag använder modellerna som grund för ett rationalitetskrav på de preferenser vars tillfredsställelse har personligt värde. En filosof som på ett mer direkt sätt utvecklar dessa tankegångar som en analys av personligt värde är Peter Railton 2003 (ursprungligen 1986). I en sådan analys är det fullt möjligt att skilja mellan om en preferens är rationell och om det är bra för någon att få en preferens tillfredsställd även när den är irrationell, medan denna skillnad försvinner i min tillämpning av analysen. Anledningen till att jag väljer att diskutera Smith och inte Railton är främst pedagogisk. Smiths modeller presenterar i klartext å ena sidan de klassiska formuleringarna och å den andra det lösningsförslag som Railton formulerar. Jag bedömer vidare att även Railton skulle drabbas av den kritik som presenteras i denna sektion. Se också Wlodek Rabinowicz och Bertil Strömberg, 1996.

Anta att jag just har upplevt en förnedrande förlust i squash och är så uppbragt att jag har god lust att drämna till min motståndare i huvudet med min racket. Om jag varit rationell hade jag inte haft en sådan önskan, utan hade snarare velat visa god sportsmannaanda genom att helt vänligt gå fram och ta motståndaren i hand och gratulera till segern. Smith frågar: ”följer det av detta att vad jag har skäl att göra *i mitt uppbragta och upprörda tillstånd* är att gå fram till honom och skaka hand?” (1995, s. 111). I exempelmodellen blir svaret ja, men i rådsmodellen som Smith själv förespråkar blir svaret nej. I stället bör jag le vänligt och lämna planen så snabbt som möjligt, eftersom detta är enda sättet att få kontroll över mina känslor. Och, säger Smith, det är vad mitt rationella jag vill att mitt irrationella jag ska göra i situationen.

Kanske har Smith rätt när han slår fast att rådsmodellen ger det korrekta svaret i hans exempel. Jag tvivlar emellertid på att rådsmodellen också passar vårt exempel om din dotters preferens för hårdrock (med hänvisning till den diskussionen vi tidigare förde kan vi anta att den preferensen åtminstone delvis är intrinsikal och final). Vi kan diskutera om denna preferens är rationell och bör tillfredsställas enligt rådsmodellen. Vad skulle din dotters rationella jag önska att hennes irrationella jag gör med sin preferens för hårdrock? Så vitt jag kan se skulle hennes rationella jag mycket väl kunna önska att hennes irrationella jag inte tillfredsställer eller får sin preferens för hårdrock tillfredsställd.

Jag tror att man kan nå denna slutsats utifrån två olika argument.

(1) Låt oss föreställa oss att det rationella jaget har fullständig information i den meningen att hon kan jämföra upplevelserna av att lyssna till olika slags musik och väga dem mot varandra. Det rationella jaget vet – och har livliga representationer av – vad för slags upplevelser som är inblandade i lyssnandet till t.ex. Schuberts stråkkvartetter. För argumentets skull antar vi också att dessa upplevelser är rikare och djupare än de som kommer av att lyssna på hårdrock, och att detta avspeglar sig i de intrinsikala preferenser som musiken blir föremål för. Dessutom vet det rationella jaget att främsta skälet till detta är att de estetiska kvaliteterna i Schuberts musik, att döma av människor med extraordinär musikalisk talang och kunskap, är högre än kvaliteterna i hårdrocken.

Det här argumentet grundar sig på möjligheten att någons rationella jag föredrar att hennes irrationella jag är frustrerat framför att inlåta sig på aktiviteter och glädjemen hon betraktar som undermåliga. Men hur vet man om denna möjlighet kommer att aktualiseras? Det ser jag som en rent psykologisk fråga.

(a) Vi kan först behandla det som en *allmänpsykologisk* fråga. Hur skulle en person normalt reagera om hon genomgick kognitiv terapi och där-

efter frågade sig om hon föredrog att hennes irrationella jag var frustrerat framför att ägna sig åt lägre rangens glädjeämnen?

Vi förändrar normalt sett vår musiksmak under uppväxtåren och anser nog också normalt att det är en förändring från det som har lägre kvalitet till det som har högre. Men skulle vi föredra att ha våra musikaliska ungdomspreferenser *frustrerade* framför att tillfredsställa dem med undermålig musik? Förmodligen inte, av två skäl. För det första kan vi betrakta lyssnandet på dålig musik som ett trots allt nödvändigt men övergående led i utvecklingen av vår musiksmak. För det andra genomgår vi vanligtvis inte så pass genomgripande förändringar som våra rationella jag gör (i transformeringen från sitt irrationella tillstånd). Frågeställningen om vi skulle föredra att ha våra musikaliska ungdomspreferenser frustrerade ställs kanske inte mot bakgrund av en radikal förvandling av vår musikaliska smak. Om inte annat tror jag att de flesta av oss kan ha kvar en rent nostalgisk svaghet för sina musikaliska ungdomsförälskelser.

Men det finns vad jag kan förstå inget som hindrar att ett fullständigt rationellt jag ställs inför ett sådant radikalt val – hon väljer från en position där hon haft livliga upplevelser från lyssnandet till den yppersta musiken med maximalt tränade öron. Dessutom vet hon förmodligen hur det är att ha nostalgiska upplevelser av de flesta musikstilar. Och kan vi verkligen med tillförsikt utgå från att en person i denna position normalt inte skulle välja frustrationen framför att ha glädje av dålig musik? Det tycks inte helt otänkbart att hon skulle kunna avvisa en sådan glädje. Och det låter heller inte absurt att tänka sig att detta slags person som förälder skulle kunna föredra att hennes barn var musikaliskt frustrerat framför att vara frälst på hårdrock – det vill säga *av ren omsorg om barnet*.

(b) Vad skulle *denna särskilda individ* föredra givet att hon var maximalt rationell? Att vara fullt rationell betyder bl.a. att vara lugn och sansad, men jag kan inte se varför det också betyder att man måste göra sig av med alla de egenheter och preferenser som gör oss till de personligheter vi är.⁵ Vissa av oss är mer elitistiska än andra på så sätt att vi identifierar oss väldigt mycket med den musik, litteratur osv. som vi tror har hög kvalitet. Jag har själv hört människor säga att de föredrar att ha tråkigt framför att bli underhållna av dålig tv. Anta att detta är ett drag som du också har upptäckt hos din dotter, t.ex. när hon säger sig föredra tyst-

⁵ Den här frågeställningen ligger väl i linje med Railtons analys (se förra noten) som tycks utgå från att en individs särskilda personlighet är av avgörande betydelse för vad som blir slutresultatet i en rationaliseringsprocess (2003, s. 60). Något liknande tycks däremot inte gälla för Smith, som avvisar en relativistisk syn på interna skäl, till förmån för ett synsätt enligt vilken de olika individerna i en sådan process tenderar uppnå samsyn.

naden framför att lyssna på Schubert, även om tiden skulle gå snabbare om hon lyssnade. Pekar inte detta på att hon skulle kunna resonera på ett liknande sätt om hon var rationell och då föredra att inte ha någon musikglädje alls i sitt irrationella tillstånd framför att njuta av någon annan musik än Schuberts? Jag kan inte se att en sådan preferens på något sätt skulle vara utesluten. Därför drar jag slutsatsen att rådsmodellen mycket väl kan föreskriva att du ignorerar din dotters preferens för hårdrock, även om det inte rör din dotter i ryggen vad hennes rationella jag skulle säga om hennes preferens för den musiken. Och om så är fallet tror jag att vi har en invändning mot rådsmodellen.

Detta är visserligen bara en mjuk variant av elitism, eftersom vi baserar vårt omdöme på de attityder som vi tror att *din dotter* skulle ha om hon vore rationell. Vi baserar det inte på vad vi tror att en oberoende elit skulle säga om saken, annat än som en indikation på vad *dottern* skulle säga om hon haft tillgång till elitens kunskap och omdöme. Men man kan ändå säga, att rådsmodellen vill att vi ska lyssna på handlingsagentens råd under förutsättning att hon var mer lik den absoluta eliten än vad hennes faktiska jag är. Och hur nära står en sådan person agentens faktiska jag?

Jag tror med andra ord att vi har ett kontroversiellt drag av elitism också i rådsmodellen.

(2) Det finns ett annat sätt att resonera för den som funderar över vilken ställning dotterns preferens har enligt rådsmodellen. Är hennes preferens rationell? Enligt rådsmodellen bestäms detta av om hennes rationella jag skulle önska att hennes irrationella jag tillfredsställde den (se också Railton, 2003, t.ex. s. 11). Som vi sett formulerar Smith denna tanke också på följande sätt: "jaget i den värderande världen ger jaget i den värderade världen råd om vad som ska göras" (1995, s. 110). Rationaliteten hos din dotters preferens bestäms alltså av vad hennes rationella jag skulle önska att hon gjorde eller alternativt av vad hennes rationella jag skulle råda henne att göra med preferensen.

I första argumentet för möjligheten att hennes rationella jag inte skulle önska att hon tillfredsställde preferensen för hårdrock, diskuterades alternativet att hennes rationella jag skulle insistera på detta, även om det skulle resultera i känslor av frustration. Det skulle betyda att hennes preferens också är irrationell enligt rådsmodellen. Anta nu att hennes rationella jag *inte* skulle föredra att hennes faktiska jag upplevde otillfredsställelse framför att hon fick sin preferens för hårdrock tillfredsställd. Anta, med andra ord, att det rationella jaget trots allt vill tillskriva ett positivt värde till tillfredsställandet av denna preferens. Skulle dotterns faktiska preferens i så fall vara rationell, dvs. bör den tillfredsställas?

Så vitt jag kan se riskerar vi få samma resultat här som i vårt första

argument. Det verkar naturligt att utgå från att det rationella jaget i den här situationen skulle *föredra* att det irrationella jaget inte tillfredsställde hårdrockspreferensen, utan i stället underkastade sig själv den kognitiva träning som skulle möjliggöra för henne att tillfredsställa preferensen för klassisk musik. Om möjligt så verkar det än naturligare att utgå från att detta är det råd som hennes rationella jag skulle ge till sitt irrationella jag. Varje annat råd skulle vara ett dåligt råd, med tanke på intensiteten och kvaliteten hos upplevelserna av att lyssna på klassisk musik.

Här kan någon invända att det fullt rationella jaget också har skäl att tro att det irrationella jaget *inte* kommer att underkasta sig denna nödvändiga träning och därför inte heller kommer att utveckla en preferens för klassisk musik. Detta stämmer, men om vi frågar oss vad det rationella jaget skulle *önska* (jfr Smith) att hennes irrationella jag gjorde för att förändra de faktiska preferenserna, så är det tämligen uppenbart att hon *skulle* önska att det irrationella jagets preferenser byttes ut.

Trots allt skulle det inte vara omöjligt för det irrationella jaget att åstadkomma detta – hon skulle lyckas om hon bara ville göra det. Detta är som jag ser det tillräckligt för att kunna tala om en önskan eller preferens. Det faktum att du (som fullständigt rationell) har goda skäl att tro att du faktiskt *inte* kommer i besittning av ett objekt, kan inte vara ett hinder för dig att önska eller föredra det.⁶ Jag utgår således från att din dotters rationella jag kan önska att hennes irrationella jag får sin preferens för hårdrock utbytt mot en preferens för klassisk musik.

Vid första anblick kan man misstänka att det här resonemanget passar också för den formulering av modellen som använder sig av det begrepp som fått ge namn åt modellen. Om det rationella jaget ska *råda* sitt eget irrationella jag vad hon ska göra med sina faktiska preferenser, så kommer hennes rationella jag (om jag har rätt i vad jag förut sagt) att ge råd som speglar hennes (det rationella jagets) egna preferenser. Det rationella för din dotter att göra i den situation hon befinner sig, och som innebär en total frånvaro av musikintresse för annat än hårdrock, är trots allt att försöka intressera sig för klassisk musik.

⁶ Vi kan förstås diskutera vad för slags kunskap som ska inkluderas i begreppet fullständigt rationell. Inkluderas även kunskap om vilken handling som agenten *faktiskt* kommer att utföra? Det är en metafysisk fråga om denna typ av förutsägelser verkligen är möjliga. De verkar åtminstone inte vara möjliga i första person och inte heller i andra fall där personen som är föremål för förutsägelsen får kunskap om denna. Av denna anledning följer heller inte förutsägbarhetens i en deterministisk värld. Men detta är inte avgörande för vår fråga, eftersom den komplikation som diskuteras uppstår också när vi talar om kunskap om sannolikheter.

Men saken är den att det irrationella jaget inte kommer att fråga efter råd, eftersom, som sagt, dottern inte ger ett dyft för om hon framstår som rationell i dina eller i hennes eget rationella jags ögon. Och frågan är vad för slags råd – om det överhuvudtaget är aktuellt att ge något – som en rationell person skulle ge till någon som inte ens lyssnar på henne.

Kanske bör vi därför utforma rådsmodellen i enlighet med det råd som det rationella jaget skulle ge det irrationella jaget, allt under förutsättning att det irrationella jaget *bad* om ett råd och att det inte var meningslöst att avge. I så fall verkar det, än en gång, naturligt att utgå från att det rationella jaget skulle råda det irrationella att göra sig av med preferensen för hårdrock. Om det irrationella jaget verkligen bad om råd skulle hon förmodligen också kunna förändra sin musiksmak. Men i så fall kommer man att bygga in problemet med de hypotetiska preferenserna i själva modellen, i sina försök att använda rådsbegreppet i formuleringen av den.

Det verkar alltså som om rådsmodellen har att hantera samma problem som de klassiska formuleringarna av informationskravet. Vi kan i alla fall inte utesluta det.

3. ROSATIS ”TVÅLAGERSINTERNALISM”

Connie S. Rosati diskuterar, i artikeln ”Internalism and the Good for a Person”, vissa av de problem vi tagit upp, och förespråkar mot bakgrund av det något som kan uppfattas som en utvecklad version av Smiths rådsmodell, som hon kallar *tvålayersinternalism*. Modellen definieras på följande sätt:

Den starkare form av internalism som vi slutligen hamnar i säger att någonting X kan vara gott för personen A endast om två villkor är uppfyllda:

1. Om A befunnit sig i omständigheterna C och begrundat de förhållanden som hennes aktuella jag befinner sig i, såsom någon som kommer att inta hennes aktuella jags position, så skulle A måna om X för hennes aktuella jags skull;
2. omständigheterna C är sådana att fakta rörande vad A skulle måna om för hennes aktuella jags skull, är viktiga för A också när hon befinner sig i ordinärt optimala omständigheter. (1996, s. 307)

Det finns mycket i denna definition som behöver diskuteras. Jag vill bara visa att om omständigheterna C motsvarar det traditionella informationskravet, dvs. om C innebär att vara fullständigt informerad och rationell, så tror jag att det finns ett problem med hypotetiska preferenser

som kan slinka igenom också i Rosatis modell.⁷ I själva verket tror jag att problemet riskerar att dyka upp i någon form hur vi än specificerar C, så länge C skiljer sig från A:s faktiska omständigheter.

Vi ser att det första villkoret är mycket likt Smiths rådsmodell, eftersom det kräver att vi lyssnar på de attityder som den rationella A skulle ha visavi X, under förutsättning att hon befann sig i den faktiska, dvs. mindre rationella, A:s position. Någoting X kan vara gott för A endast om A skulle bry sig om det, om hon varit fullständigt rationell och begrundat hur det skulle vara att hamna i den aktuella A:s position. ("Måna" och "bry sig om" ska förstås i en vid betydelse som inte nödvändigtvis motiverar, menar Rosati.)

Vad som är speciellt med Rosatis modell är i stället det andra villkoret. När hon formulerar detta, verkar hon vilja försäkra sig om att de kontra-faktiska omständigheterna har något slags auktoritet över den faktiska A och att A så att säga kommer att respektera vad hennes rationella jag föreskriver. Detta är i sin tur nödvändigt för att hon ska slippa känslan av alienation och godtycklighet i förhållande till vad som betraktas som gott för henne.

Det väsentliga i definitionen är inte att A faktiskt och obetingat fäster vikt vid vad hennes rationella jag bryr sig om, eftersom hon kan vara hypnotiserad eller drogad eller på annat sätt manipulerad till att bry sig om eller inte bry sig om olika saker och ting. Det är därför som Rosati introducerar dessa "ordinärt optimala omständigheter" och skriver: "Dessa kommer att inkludera att en person inte sover, är drogad, eller hypnotiserad, att hon tänker redigt och rationellt, och att hon inte förbiser lättillgänglig information ... Ordinärt optimala omständigheter är helt enkelt de som vi redan accepterar att en person måste uppfylla för att på ett någorlunda sansat sätt överhuvudtaget kunna bedöma vad som är gott för henne själv" (1996, s. 305, se också s. 301 f). Rosati nämner hjärntvätt som en allvarlig icke-optimal omständighet, men med tanke på vad hon för övrigt säger om de ordinärt optimala omständigheterna utgår jag från att hon bland dessa också skulle inkludera frånvaron av indoktrinering och liknande fenomen, eftersom även detta är en form av manipulation.

I ett avseende sympatiserar jag med det andra villkoret, eftersom jag i vårt exempel utgick från att det var en relevant omständighet huruvida det gjorde intryck på din dotter att hennes preferens för hårdrock var irrationell. Men vad jag hade i åtanke var om den *faktiska* handlingsagenten

⁷ Observera att Rosati anger ett nödvändigt villkor eller vad hon själv kallar för ett internalismkrav rörande personvärdet. Om man använder hennes förslag som en explikation av rationalitetskravet, skulle man också få betrakta det som tillräckligt.

brydde sig om huruvida hennes preferenser uppfyllde informationskravet, och inte om den faktiska agenten under *ordinärt optimala omständigheter* skulle bry sig om detta. Min misstanke är därför att introducerandet av de ordinärt optimala omständigheterna kommer att vålla samma eller liknande problem som Smiths rådsmodell för de personer som faktiskt inte uppfyller dem.

Ta en titt på hårdrocksexemplet igen och låt mig än en gång göra vissa förändringar av det. Anta att din dotter tillhör ett gäng som uteslutande lyssnar till hårdrock. Vilka är realistiskt sett mekanismerna bakom din dotters intresse för hårdrock under dessa omständigheter? Jag tror att gruppträck och vad som lättast kan beskrivas som indoktrinering har en inte helt oviktig roll. Jag skulle i alla fall inte säga att det finns ett fritt val mellan gruppens musiksmak och en radikalt annorlunda smak. Så fungerar det bara inte, och under dessa omständigheter tror jag inte att dottern uppfyller de ordinärt optimala omständigheterna eller villkoren. Därför skulle man heller inte i Rosatis modell behöva bekymra sig om huruvida hon fäster vikt vid vad hennes rationella jag säger.

Och anta, på samma sätt som vi gjorde i diskussionen av Smiths exempel, att hennes rationella jag skulle föredra att hon lyssnade på klassisk musik och även föredra att uppleva besvikelse framför att lyssna på hårdrock. Anta också att detta faktum skulle göra intryck på din dotter under ordinärt optimala omständigheter, men inte i hennes nuvarande situation. Vad skulle du i så fall göra om du visste att hon önskade sig en hårdrocksskiva? Avstår du från att köpa den och lämnar henne besviken med hänvisning till *två* hypotetiska personer som inte är identiska med din dotter? Och är inte i så fall detta ett problem?

4. SLUTSATS

Jag drar slutsatsen att vi har ett problem med hypotetiska preferenser i informationskravet, även i Smiths och Rosatis modeller. Problemet är kanske mindre i Smiths modell än i de mer klassiska formuleringarna. Än mindre är det möjligen i Rosatis modell. Men jag hävdar att problemet inte har sin fullständiga lösning heller i hennes modell. Hur och om problemet går att lösa inom ramen för informationskravet verkar alltså vara en öppen fråga.⁸

⁸ Denna text har diskuterats på seminarier vid Lunds och Göteborgs universitet. Ett stort tack till alla deltagare! Ett särskilt tack till David Alm, Åsa Andersson, Johan Brännmark, Mats Johansson, Andreas Lind, Ingmar Persson, Björn Petersson, Wlodek Rabinowicz, Toni Rønnow-Rasmussen, Caj Strandberg, Daniel Svensson och Anders Tolland.

LITTERATUR

- Brandt, R. B. 1979. *A Theory of the Good and the Right*. Oxford: Clarendon Press.
- Brülde, Bengt. 1998. *The Human Good*. Acta Universitatis Gothoburgensis 6.
- Bykvist, Krister. 2001. "What is Wrong with Past Preferences?", i Wlodek Rabinowicz (utg.). 2002. *Value and Choice*, Lund Philosophy Reports, 1, 17–44.
- Egonsson, Dan. 1990. *Interests, Utilitarianism and Moral Standing*. Lund: Lund U.P.
- Harsanyi, J. C. 1982. "Morality and the Theory of Rational Behaviour", i Amartya Sen och Bernard Williams (utg.): *Utilitarianism and Beyond*, Cambridge: Cambridge U. P., 39–62.
- Rabinowicz, Wlodek och Bertil Strömberg. 1996. "What if I were in his shoes? On Hare's argument for preference utilitarianism", *Theoria*, Vol. LXII, 95–123.
- Railton, Peter. 2003. *Facts, Values, and Norms*. Cambridge: Cambridge U. P.
- Rawls, John. 1972. *A Theory of Justice*. Oxford: Oxford U. P.
- Rosati, Connie S. 1996. "Internalism and the Good for a Person", *Ethics*, Vol. 106, Nr 2, 297–326.
- Sen, Amartya & Bernard Williams. 1982. "Introduction: Utilitarianism and Beyond", i Amartya Sen och Bernard Williams (utg.), *Utilitarianism and Beyond*. Cambridge: Cambridge U. P., 1–21.
- Shope, Robert K. 1978. "The Conditional Fallacy in Contemporary Philosophy", *The Journal of Philosophy*, Vol. LXXV, No. 8, 397–413.
- Sidgwick, Henry. 1907. *The Methods of Ethics*. London: Macmillan & Co.
- Smith, Michael. 1995. "Internal Reasons", *Philosophy and Phenomenological Research*, Vol. 55, Nr 1, 109–131.