

TIDSKRIFT FÖR POLITISK FILOSOFI  
NR 2 2007 | ÅRGÅNG 11

*Bokförlaget* THALES

*Joakim Sonnegård*

... är avunden till nackdel för hela kollektivet: den enskilde som avundas någon annan är beredd att företa sig saker som gör att båda parter får det sämre, så länge skillnaderna dem emellan blir mindre. Därför diskuterar Kant, vars definition jag till ganska stor del har följt, avunden som en av de laster som hör samman med hat mot mänskligheten.<sup>1</sup>

John Rawls<sup>2</sup>

JAG TÄNKER I denna artikel diskutera de rationalitetsantaganden John Rawls ställer upp som grund för sin rättviseteori, och speciellt hans originella antagande om att rationella individer inte är avundsjuka. I artikeln visar jag att Rawls rationalitetsantaganden är oklara, och att denna oklarhet leder till att det inte går att fastställa vilka val parterna i ursprungssituationen faktiskt gör. Avslutningsvis diskuterar jag mycket kortfattat frågan om icke-avundsjuka är ett moraliskt antagande, samt ett möjligt motiv för Rawls att konstruera grundvalen för sin rättviseteori utifrån detta antagande.<sup>3</sup>

\*

I LIKHET MED Immanuel Kant bygger John Rawls sin moralfilosofi på idén att rättvisepprinciper är föremål för rationella val, och att de principer som en människa rationellt kan vilja låta sitt beteende styras av i ett fritt samhälle, är just de principer som rationella och fria individer skulle välja som moralisk grundval för sina liv.<sup>4</sup> Rawls hävdar t.o.m. att den rättviseteori han utarbetat utgör en del – »kanske den allra viktigaste» – av teorin om rationella val.<sup>5</sup>

Varför är nu Rawls så angelägen om att slå fast att teorin om rationella val utgör grunden för hans rättviseteori?

Man kan betrakta teorin om rationella val som en normativ teori, som säger oss *hur* vi bör agera för att uppnå de målsättningar

vi har. Men teorin säger ingenting om *vad* vi ska försöka uppnå.<sup>6</sup> En moralteori grundad på en teori om rationella val vilar alltså på något annat än moraliska målsättningar. Därmed kan man både *förklara* och *rättfärdiga* de moraliska principer man framhåller som värda att efterleva utan att hänvisa till t.ex. intuition, känslor eller andra moraliska principer. Den teoretiska vinsten med denna ansats är betydande: istället för känslor eller gudomliga påbud ligger rationella val till grund för de moraliska uppfattningar man står för. Moralfilosofin blir helt enkelt mindre godtycklig, och förvandlas till en undersökning av hur ett rationellt val utformas och faller ut.

Det är därför viktigt för Rawls att slå fast att parterna i ursprungspositionen – POPsen<sup>7</sup> – är rationella, och att det rationalitetsbegrepp han bygger sin teori på är väl förankrat i teorin om rationella val. Frågan är nu om de antaganden om rationalitet som Rawls gör ligger inom ramen för denna teoribildning, eller om han dessutom lägger till något mer. Rawls skriver i §25:

Det rationalitetsbegrepp som åberopas här är, med undantag för ett väsentligt drag, det normala, som är välbekant från samhällsteorier. Således tänker vi oss, som vanligt, sättet att en rationell person har en koherent mängd preferenser inom ramen för de valmöjligheter som står öppna för honom. Han rangordnar dessa valmöjligheter efter hur väl de främjar hans syften; han följer den plan som kommer att tillfredsställa fler snarare än färre av hans önsknings, och som har störst chans att framgångsrikt verkställas. Det speciella antagande jag gör, är att en rationell individ inte lider av avundsjuka. Han är inte beredd att acceptera en egen förlust bara andra också får mindre. Han blir inte nedslagen av att veta eller se att andras primära sociala nyttigheter ger ett högre indexmått. Åtminstone så länge skillnaderna mellan honom själv och andra inte överskrider vissa gränser, och inte tror att de ojämlikheter som finns har sin grund i orättvisor eller är resultatet av att slumpen får sköta sig själv utan någon kompensation.<sup>8</sup>

Vad är det Rawls här stipulerar? Han säger att det rationalitets-

begrepp han använder är välbekant – med ett undantag, som jag återkommer till – från samhällsteorier, och hänvisar till studier av Amartaya Sen och Kenneth Arrow. Men hur är egentligen den grund han bygger sin rättviseteori på konstruerad?

Rawls lyfter fram tre aspekter i sin specifikation av POPsens rationalitet. Han antar att POPsen

- har en koherent mängd preferenser;
- tycker att mer av primära sociala nyttigheter är bättre än mindre; samt
- inte är avundsjuka, åtminstone så länge skillnaderna mellan POPsens indexnivå för primära sociala nyttigheter ligger inom vissa gränser.

Vad kan Rawls mena med att preferenserna är *koherenta*? Möjligen att POPsens preferensordning är *fullständig*, vilket i så fall betyder att en POP vet hur han ska ställa sig i varje valsituation där det för honom gäller att välja mellan två alternativ.<sup>9</sup> Men han kan inte bara mena det: att enbart anta fullständighet räcker inte för att försäkra sig om att andra nödvändiga rationalitetsvillkor är uppfyllda. Snarare tänker sig nog Rawls att preferensordningen karakteriseras av någon typ av inre konsistens. Och då kan han med *koherent mängd preferenser* avse det som Arrow (1951) betecknar som en *svag* preferensordning. I så fall innebär antagandet följande:<sup>10</sup>

Det finns en binär relation,  $R$ , som är definierad över mängden möjliga val,  $S$ , som POPsen kan göra.<sup>11</sup>  $R$  har då följande egenskaper:

- (1) Reflexiv:  $\forall x \in S : xRx$ ;
- (2) Fullständig:  $\forall x, y \in S : (x \neq y) \Rightarrow (xRy \vee yRx)$ ; och
- (3) Transitiv:  $\forall x, y, z \in S : (xRy \ \& \ yRz) \Rightarrow (xRz)$ .<sup>12</sup>

En reflexiv, fullständig och transitiv preferensordning utgör en tillräcklig grund för att kunna anta existensen av en valfunktion.<sup>13</sup> Och Rawls behöver försäkra sig om att det existerar en valfunktion baserad på POPsens preferensordningar, eftersom existensen av en valfunktion är en förutsättning för att POPsen ska kunna komma fram till beslut och göra rationella val.<sup>14</sup>

Rawls andra antagande är lättare att tolka. POPsen försöker helt enkelt skaffa sig själva ett så högt indexmått på primära sociala nyttigheter som möjligt: det är ju det effektivaste sättet att befrämja det de anser är eftersträvansvärt, oavsett vad det nu än är. Det handlar helt enkelt om standardantagandet inom ekonomisk teori: *more is better*.

Innan jag närmare går in på Rawls tredje antagande ska jag göra en kort utveckling om hur man inom en annan del av teorin om rationella val – ekonomisk teori – formellt kan hantera fenomenet avundsjuka och de rationalitetsantaganden som angetts ovan. Jag gör det för att ha en analytisk ram för den fortsatta undersökningen.

\*

I EN EKONOMISK-TEORETISK analys, publicerad 1973, försöker ekonomen Geoffrey Brennan besvara frågeställningen: *Kan en Pareto-sanktionerad omfördelning i en ekonomi där individerna är avundsjuka på varandra, förklaras och motiveras utifrån individernas preferenser?* Brennans analys visar att man inte behöver anta att individerna i en ekonomi är altruistiskt lagda för att Paretosanktionerade omfördelningar av inkomsterna i ekonomin ska vara möjliga och önskvärda. Speciellt visar han att omfördelningen i en sådan ekonomi kommer att ske från rika till fattiga. Den drivande kraften i processen är ett antagande om att individerna upplever en ökande (marginell) olust när inkomstskillnaderna ökar. Det gör Brennans analys till en relevant jämförelsepunkt med Rawls ursprungssituation: Rawls hävdar ju att avundsjuka bl. a. har att göra med hur stora skillnaderna är mellan individerna i ursprungssituationen.<sup>15</sup>

Brennans analys görs inom ramen för traditionell mikroekonomisk teori, och grundar sig därmed på de tre antaganden om individernas preferenser som diskuterats ovan.<sup>16</sup> Hans diskussion utgår ifrån följande enkla modell:

Vi antar att det finns tre individer A, B och C, med inkomsterna  $Y_A$ ,  $Y_B$ ,  $Y_C$ .<sup>17</sup> Varje individ antas ha en nyttofunktion (som man kan betrakta som en valfunktion) med följande utseende:<sup>18</sup>

$$U_i = u(Y_i, Y_j, Y_k) \text{ för alla } i, j, k.$$

Brennan antar vidare att individerna upplever:

(a) negativ marginell nytta av de andra individernas inkomst, vilket betyder att individ  $i$  upplever olust bara pga. att de andra individerna  $j$  och  $k$  har en positiv inkomst. Detta kallar Brennan *illvilja* («malice»);

(b) en ökande marginell negativ nytta av de andra individernas inkomstökningar, dvs. olusten ökar med inkomstskillnaderna. Detta kallar Brennan *avundsjuka* («envy»).<sup>19</sup>

Brennan gör dessutom antagandet att individerna över något inkomstintervall är beredda att acceptera en egen förlust så länge de andra individerna i ekonomin också får en lägre inkomst. Ett motsvarande villkor finns med i Rawls specifikation av avundsjuka.

Det intressanta, ur det perspektiv jag har här, är hur Brennan rent modelltekniskt behandlar avundsjuka. Jag återkommer till det längre fram i artikeln.

\*

NÄR DET GÄLLER Rawls tredje antagande – om att POPsen inte »lider av avundsjuka» – är läget däremot inte lika klart. Det är därför svårt att göra en entydig positionsbestämning av Rawls grundantaganden. I §25, där rationalitetsantagandet specificeras, slår Rawls fast att POPsen »inte lider av avundsjuka», åtminstone så länge skillnaderna mellan individer inte »överskrider vissa gränser». Men både före och efter detta parti i texten talar Rawls då och då om att POPsen besitter en »ömsesidigt ointresserad rationalitet». Och det är ju inte samma sak!

Funderar man över vad begreppet »ömsesidigt ointresse» har för extension och ställer det mot att »inte vara avundsjuk», står det klart att begreppen delvis omfattar olika fenomen. Ett »ömsesidigt ointresse» utesluter t. ex. alla former av altruism, medan det är fullt möjligt att inte känna någon avundsjuka och ändå samtidigt vara altruistisk.<sup>20</sup> (Det är förmodligen det normala förhållnings sättet för altruistiskt lagda personer.) Begreppens omfång är helt enkelt olika stora. Begreppet *inte avundsjuk* anger *ett* av de möjliga sätt som *öm-*

*sesidigt ointresse* kan ta som uttryck. *Ömsesidigt ointresse* utesluter däremot alla former av interdependence i POPsens valfunktioner, t. ex. alla former av altruism. Och det är väsentligt för Rawls att han kan utesluta åtminstone en viss form av altruism hos POPsen.

I §30 noterar Rawls att POPsen inte kan vara *fulländade altruister*, eftersom det skulle betyda att de inte kan nå fram till en överenskommelse om vilka rättvisepprinciper som ska styra utformningen av samhällets grundstruktur.

Rawls skriver:

Den fulländade altruismen har dock ett märkligt kännetecken som förtjänar att nämnas. En fulländad altruist kan uppfylla sin önskan endast om någon annan har oberoende önsknningar, eller önsknningar av första ordningen. För att belysa detta faktum antar vi att när man ska besluta om vad som ska göras, så röstar alla på det som alla andra vill göra. Uppenbarligen blir ingenting avgjort: i själva verket finns det inget att avgöra. Ty för att ett rättviseproblem ska uppstå måste åtminstone två personer vilja göra något annat än vad alla andra vill göra, oavsett vad detta är. Det är således omöjligt att förutsätta att parterna *enbart* (min *kursiv*) är fulländade altruister.<sup>21</sup>

Men Rawls tycks här bara utesluta den fulländade altruismen. Övriga former av altruism lämnar han (medvetet?) utrymme för.

För att kasta ytterligare ljus över Rawls sätt att behandla avundsjuka, ska jag nu återknyta till den analys av Brennan som skisserats ovan.

Brennans individer upplever positiv nytta i den egna inkomsten, vilket är ett standardantagande i ekonomisk teori. Man kan jämföra det med Rawls antagande om att POPsen tycker att mer av de primära sociala nyttigheterna är bättre än mindre («more is better»).

När Brennan för in avundsjuka i sin modell, gör han det genom att låta värdet hos en individs nyttofunktion också bestämmas av andra individers inkomster. Överfört till Rawls ursprungssituation betyder det att POPsen inte bara skulle fästa vikt vid nivån på det egna indexet för primära sociala nyttigheter, utan även bry sig om

vilken nivå de andra POPsen uppnår. I tekniska termer kan man uttrycka det som att POPsens valfunktion innefattar de andra POPsens indexmått.

Uttryckt i Brennans modelltermer innebär Rawls antagande om att POPsen inte är avundsjuka, att POPsen inte har några negativa upplevelser med avund som grund, som på något sätt påverkar deras val av rättvisepprinciper, och som beror av vilken indexnivå de andra POPsen uppnår. Åtminstone så länge skillnaderna mellan POPsen ligger inom vissa gränser.

Rawls har helt enkelt konstruerat POPsens valfunktion så att de andra POPsens indexmått inte ger en viss typ av negativ effekt på en enskild POPs valfunktion. Men han har i och med sitt antagande inte uteslutit att valfunktionen kan konstrueras så att andras indexmått har t. ex. en positiv effekt på densamma. Tvärtom antyder det han säger om altruism raka motsatsen.

Om Brennan skulle vilja hantera altruism i sin modellvärld, kan han mycket enkelt göra det genom att ändra tecknet framför de antaganden han gör om avundsjuka.<sup>22</sup> Han kan dessutom utöka sin modell så att den även tillåter olika former av altruism, även om en sån modell förmodligen inte skulle ge entydiga analysresultat.

Min poäng är att Rawls inte har uteslutit att POPsen t. ex. är altruister. Vad Rawls explicit utesluter är endast det fall som Brennan behandlar i sin avundsjukemodell. Rawls har genom de olika, och ibland motstridiga, formuleringarna om vad POPsen fäster vikt vid, skapat en oklarhet om vilka drivkrafter det är som styr POPsen första viktiga val av rättvisepprinciper. Och det är ett problem för hans teori.

Vad innebär nu detta för de beslut POPsen fattar i olika valsituationer?

I analysen av nedanstående matriser ska endast de rationalitetsantaganden som Rawls gör ligga till grund för POPsens ställningstaganden. Det är alltså *inte* frågan om att pröva situationerna mot någon av hans rättvisepprinciper, t. ex. differensprincipen. Här handlar det *endast* om hur en POP i ett stadium före det att han nått fram till en möjlig rättvisepprincip för hur samhällets grundstruktur



ska se ut, ställer sig till olika situationer där han får olika mycket av de primära sociala nyttigheterna. Det är alltså bara POPsens preferenser och den struktur som Rawls stipulerat att dessa har, som styr de val som görs. Siffrorna i matriserna står endast för indexmättet för primära sociala nyttigheter. POPsen har en valfunktion som baseras på deras preferenser, och vars värde beror av bl.a. hur de värderar mängden primära sociala nyttigheter, men – och det är viktigt att komma ihåg – även eventuella interdependencer POPsen emellan kan spela en roll för hur valen faller ut.

*Figur 1:*

	POP <sub>1</sub>	POP <sub>2</sub>
Alternativ 1	9	9
Alternativ 1'	7	9
Alternativ 1''	7	7

I *Figur 1* framstår alternativen 1' och 1'' som lika dåliga för POP<sub>1</sub> i förhållande till alternativ 1. Fäster han bara vikt vid nivån på sitt eget indexmätt, så är 7 alltid sämre än 9. Men skulle man tillåta avundsjuka, i alla fall av det slag som Rawls uttryckligen utesluter i sina antaganden, skulle dock 1'' vara bättre än 1'; man skulle ju kunna tänka sig att POP<sub>1</sub> kan gå med på ett lägre indexmätt under förutsättning att också POP<sub>2</sub> får ett lägre indexmätt. Men det utesluter som sagt Rawls antaganden.

Ett antagande om ömsesidigt ointresse ger samma analysresultat av valsituationen i *Figur 1*: 7 är alltid sämre än 9 för POP<sub>1</sub>, oavsett vad POP<sub>2</sub> uppnår för indexmätt.

Men om man antar att POP<sub>1</sub> inte är avundsjuk, men en aning altruistisk, dvs. också sätter ett personligt värde på att POP<sub>2</sub> får behålla sina 9 enheter av primära sociala nyttigheter, så framstår 1' som ett bättre alternativ än 1'' för POP<sub>1</sub>! Och det tycks Rawls antagande om att POPsen inte är avundsjuka medge. Det han skriver om altruism pekar dessutom i samma riktning. För utgår man ifrån att parterna

inte »enbart (min kursiv) är fulländade altruister», så öppnas möjligheten för att parterna *också* är altruister, vilket för övrigt är troligt om man har antagit att de inte lider av avundsjuka.

Figur 2:

	POP <sub>1</sub>	POP <sub>2</sub>
Alternativ 1	9	9
Alternativ 2	9	20

Vad säger Rawls rationalitetsantaganden om det fall som illustreras i *Figur 2*? Tycker POP<sub>1</sub> att alternativ 1 är ett lika bra alternativ som 2? Svaret tycks bero på om skillnaden mellan 9 och 20 »överskrider vissa gränser» eller ej. Det är viktigt att notera att den oklarheten inte uppstår om man istället antar att POPsen är *ömsesidigt ointresserade*, dvs. inte fäster någon som helst vikt vid absoluta eller relativa skillnader mellan det egna indexmättet och övriga POPs indexmätt. I det senare fallet framstår alternativ 1 som ett lika bra (eller dåligt) alternativ som 2 för POP<sub>1</sub>. Och det beror på att POP<sub>1</sub> endast är intresserad av hur hög nivå på indexet för primära sociala nyttigheter *han själv* får. Vad POP<sub>2</sub> får fäster han ju ingen vikt vid.

Hur förhåller det sig då i det fall som visas i *Figur 3* nedan?

Figur 3:

	POP <sub>1</sub>	POP <sub>2</sub>
Alternativ 1	9	9
Alternativ 3	10	20

Rimligtvis bör POP<sub>1</sub> föredra alternativ 3 framför 1. Men helt säker kan man inte vara. Rawls har visserligen stipulerat att en POP inte blir nedslagen av att se andra POPs nå ett högre indexmätt på primära sociala nyttigheter, men han lägger ju till brasklappen att en POP förhåller sig så endast »så länge skillnaderna mellan honom själv och andra inte överskrider vissa gränser ...». Har denna gräns

överskridits i *Figur 3*? Och blir slutsatsen en annan om POP<sub>2</sub> får 100 istället för 20? Det går inte att avgöra utifrån de antaganden om rationalitet som Rawls stipulerat.

Ett sätt att rädda Rawls ifrån denna oklarhet är att helt bortse från hans formuleringar om avundsjuka, och istället enbart utgå från de avsnitt i texten där han tycks luta sig mot att POPsen är *ömsesidigt ointresserade* och ingenting annat. Rawls skriver t. ex.:

Antagandet om en ömsesidigt ointresserad rationalitet innebär således detta: personerna i ursprungssituationen försöker hitta principer som i största möjliga utsträckning befrämjar just deras system av mål. Detta gör de genom att försöka skaffa sig själva ett så högt indexmått på primära sociala nyttigheter de kan ... Parterna söker inte ge varandra fördelar och inte åsamka skada; de drivs varken av tillgivenhet eller av hätskhet. Inte heller försöker de vinna i förhållande till varandra; de är varken avundsjuka eller fåfänga. Uttryckt i speltermen skulle vi säga att de strävar efter så hög absolut poängsumma som möjligt.<sup>23</sup>

Om man tar fasta på vad Rawls skriver här, blir analysen av *Figur 3* entydig: POP<sub>1</sub> föredrar alternativ 3 framför 1. Och den slutsatsen står sig oavsett vilken indexnivå POP<sub>2</sub> uppnår i alternativ 3!

Men Rawls vacklar mellan att tala om »en ömsesidigt ointresserad rationalitet» och att framhålla att POPsen »inte lider av avundsjuka ... åtminstone så länge skillnaderna inte överskrider vissa gränser». Frågan är varför han gör det? Det hela blir än mer gåtfullt i ljuset av hans noggranna diskussion av vad som döljs av »okunighetens slöja».

Rawls skriver i §80:

Jag har hela tiden förutsatt att det finns vissa psykologiska böjelser som personerna i ursprungspositionen inte drivs av (§25). En rationell person drabbas inte av avund, åtminstone inte när skillnaderna mellan honom själv och andra ... inte överskrider vissa gränser. Inte heller skiljer sig parternas inställning till risker och osäkerhet, eller deras benägenhet att dominera, underkasta sig och dylikt. Sådana speciella psykiska drag har jag tänkt

mig ska döljas bakom okunnighetens slöja.<sup>24</sup>

Varför låter han inte också avund döljas av okunnighetens slöja? Om han hade gjort det skulle han undvikit (åtminstone) två problem. Det första problemet är en kritik som riktar in sig på att hans teori nu inte bara baseras på teorin om rationella val, utan faktiskt utgår ifrån något som ser ut som ett moraliskt ställningstagande; att kräva av ett rationellt val att det inte formas av avundsjuka, är att lägga en moralisk restriktion på vilka val individen får göra. Det andra problemet är det som illustrerats ovan i de tre figurerna, dvs. att hans rationalitetsantaganden ger upphov till oklarheter som gör att man inte med säkerhet kan avgöra hur POPsen i den inledande fasen av värdering av olika rättvisepprinciper faktiskt ställer sig till de alternativ de har att välja mellan. Visserligen ska slöjan av okunnighet dölja de psykiska drag POPsen senare kommer att begåvas med när väl slöjan har fallit, dvs. när de förvandlats från POPs till POLs<sup>25</sup>, men det räcker inte för att lösa det problem Rawls förutsatt sig att lösa: de val POPsen gör bakom slöjan av okunnighet i ursprungspositionen, styrs av POPsens preferenser och vi vet inte om preferenserna utmärks av ett ömsesidigt ointresse, eller t.ex. tillåter någon form av altruism. Och det kan, som jag visat ovan, påverka POPsens val.

För att ytterligare illustrera de problem som Rawls rationalitetsantaganden ger upphov till, ska jag i ett sista exempel diskutera begreppet svartsjuka.

Rawls skiljer mellan avundsjuka och svartsjuka: det handlar inte om samma sak, säger han, även om de på något (oklart vilket) sätt hör samman. Rawls skriver:

Vi bör också lägga märke till de icke-moraliska känslor som hör samman med avund, men som inte ska misstas för den. I synnerhet svartsjuka ... En person som har det bättre ställt kan önska att de som är mindre lyckligt lottade än han själv ska hålla sig på sin plats. Han bevakar svartsjukt sin förmånligare position och missunnar de mindre gynnade de fördelar som skulle placera dem på samma nivå som honom själv.<sup>26</sup>

Har Rawls uteslutit att POPsens preferenser är lagda åt svartsjuka?

Om han slagit fast att POPsens preferenser karakteriseras av en »ömsesidigt ointresserad rationalitet», hade man kunna svarat ett bestämt »ja» på den frågan. Men som det nu är gör hans oklarhet om vilka preferenser POPsen har, att det inte går att fastställa vad POP<sub>2</sub> väljer i följande situation (där  $\epsilon$  är ett mycket litet positivt tal):

*Figur 4:*

	POP <sub>1</sub>	POP <sub>2</sub>
Alternativ 3	10	20
Alternativ 4	20+ $\epsilon$	20+ $\epsilon$

Att POP<sub>1</sub> föredrar alternativ 4 i figuren ovan, råder det inget tvivel om. Men hur förhåller det sig med POP<sub>2</sub>? Om hans preferenser präglas av ömsesidigt ointresse uppstår inga oklarheter: även han föredrar då alternativ 4. Är han däremot inte avundsjuk, som Rawls stipulerar, men svartsjuk, skulle han faktiskt föredra alternativ 3 framför 4.

Det här är ett problem för Rawls. Åtminstone om han vill bygga sin rättviseteori på att POPsen har preferenser som uppfyller de krav som ställs av teorin om rationella val, för då måste dessa preferenser vara entydigt definierade. Det krävs dels för att rationalitetsantagandet ska vara uppfyllt, men även för att Rawls andra anspråk – att hans moralfilosofi grundar sig på rationella val och inget annat – ska gälla.

Frågan är nu varför Rawls gör sitt »speciella antagande» om att parterna inte drivs av avundsjuka? För det går inte att bortse ifrån att Rawls tycks mena att avundsjuka är ett centralt problem i den analys han gör. Han ägnar bl. a. två hela paragrafer (§§80 & 81) åt att diskutera begreppet. Andra former av ömsesidigt intresse, t. ex. altruism, diskuterar han däremot mycket sparsamt. Och det förhållandet är lite underligt med tanke på att boken omfattar mer än femhundra sidor.

\*

I §§80–81 diskuterar så Rawls avund och dess roll i hans rättviseteori. Han hävdar att människor i allmänhet anser att avund är något man har anledning att akta sig för eftersom den är till nackdel för samtliga indi-

vider i ett samhälle. Rawls instämmer i detta, och säger sig dela Kants uppfattning att avund är en av de laster som hör samman med »hat mot mänskligheten».<sup>27</sup> Den uppfattningen visar sig vara en central utgångspunkt för Rawls när han konstruerar sin ursprungssituation.

Rawls skriver:

... eftersom avund i allmänhet anses vara något man bör undvika och frukta, åtminstone då den blir intensiv, verkar det önskvärt att valet (i ursprungssituationen), om möjligt inte påverkas av den. Därför har jag, både för enkelhetens skull och av moralteoretiska skäl, förutsatt att det inte förekommer någon avund och att ingen vet något om psykologiska särdrag (i ursprungssituationen).<sup>28</sup>

Vad är det för »moralteoretiska skäl» som ligger bakom att POP:s rationalitet definieras så att de inte är avundsjuka?

Rawls dröjer några sidor med att själv ge ett svar på den frågan. Och det går att förstå varför. Till att börja med diskuterar han olika former av avund, och på vilket sätt avund kommer in i prövningen av olika rättvisepprinciper. Sedan slår han, lite oväntat, fast att avund inte är någon moralisk känsla! Såvida den inte uppkommit pga. »orättvisa institutioner, eller ett orättfärdigt beteende». Och i det fallet är det dessutom inte frågan om avund, utan om »harm», vilket, enligt Rawls, *är* en »moralisk känsla».<sup>29</sup>

Därpå följer ett resonemang kring frågan om den avund som ändå uppstår i ett samhälle grundat på de rawlska rättvisepprinciperna på något avgörande sätt skulle hota samhällets stabilitet. Rawls finner att den inte skulle göra det.

Och nu är Rawls framme vid ett problem han snuddat vid tidigare i texten: de möjliga sambanden mellan avund och jämlikhet, och vad de betyder för hans rättviseteori. Uppenbarligen känner Rawls ett behov av att förekomma en specifik typ av kritik mot de rättvisepprinciper han kommit fram till. Han skriver:

Många konservativa författare har påstått att jämlikhetstenden-  
ser i de moderna sociala rörelserna är ett uttryck för avund ...  
Men att insistera på en jämlikhet som den de båda rättvisepprin-

principerna definierar är inte att ge utlopp för avund ... Det framgår också tydligt av hurudana parterna i ursprungspositionen är: rättvisepuffattningen väljs under förhållanden där, enligt hypotesen, ingen drivs av hätskhet och illvilja (§25). De krav på jämlikhet som ställs av dem som säger ja till principerna kan ibland uttrycka indignation, men detta är som vi sett en annan sak.

För att visa att rättvisepuffattningarna delvis baseras på avund skulle man behöva slå fast att ett eller flera av ursprungspositionens villkor har sina rötter i denna böjelse. Eftersom frågan om stabilitet inte tvingar fram någon omprövning av det gjorda valet, måste argumentationen för avundens inflytande utföras med hänvisning till teorins första del. Men var och en av stipulationerna för ursprungspositionen grundar sig på skäl där avund över huvud taget inte nämns ... de båda rättvisepuffattningarna ... väljs under förutsättning att avund inte förekommer.<sup>30</sup>

Ett skäl till att Rawls stipulerar att rationella individer i ursprungspositionen inte är avundsjuka, är alltså att han vill slå undan grunden för en kritik som går ut på att sammankoppla jämlikhetssträvanden med avundsjuka. Rawls delar uppenbarligen puffattningen att avund är en last, som om den tar för stora proportioner kan hota samhällets grundvalar. Men är inte det en moralisk puffattning? Och om det är det: varför har Rawls i så fall lagt ett moraliskt begrepp till sina rationalitetsantaganden? Förmodligen har även Rawls ställt sig samma frågor, hur som helst antyder hans lite märkliga påstående att avund inte skulle vara någon moralisk känsla något åt det hållet.<sup>31</sup>

Räcker det inte med att anta att även avund döljs för POPsen av okunnighetens slöja? Det borde göra det, men Rawls har av någon anledning dragit en annan slutsats. Och det är inte lätt att förstå varför. För han skriver ju själv att man på samma sätt som med avund, måste pröva hur *andra* psykologiska särdrag påverkar de rättvisepuffattningar som POPsen fastnar för, i syfte att undersöka om den valda grundstrukturen är robust mot de krafter som POLsens riskattityd, vilja att dominera andra, etc. utsätter samhället för.<sup>32</sup> Vad är det då som gör att inte även avund kan läggas till de psykiska särdrag som okunnighetens slöja döljer för POPsen?

\*

DET KANSKE FINNS ett annat skäl till att Rawls gör ett medvetet avsteg från teorin om rationella val. Man skulle kunna tänka sig att han vill undgå den kritik som riktats mot t. ex. spelteori och ekonomisk teori för att vara alltför långtgående idealiseringar av mänsklig rationalitet, vilket enligt kritikerna gör dessa teorier olämpliga att ha som utgångspunkt för en analys av verkliga människors val och vedermödor.<sup>33</sup> Rationalitet förutsätter, enligt dessa kritiker, något mer än bara »konsistenta val» mellan olika alternativ. Det krävs därutöver att de mål som beslutsfattaren försöker uppnå verkligen är eftersträvansvärda.<sup>34</sup>

Rawls skulle i denna tolkning helt enkelt försöka föra teorin om rationella val *närmare människor av kött och blod* och det samhälle vi lever i. Möjligen är hans utgångspunkt följande: människor kan bara göra rationella val om de låter sig vägledas av sina känslor; känslorna har nämligen egenskapen att bringa de val individen gör i harmoni med den mänskliga naturen.<sup>35</sup> Eftersom Rawls tycks utgå ifrån att det ligger i människans natur att sträva efter att samarbeta och skapa ett samhälle som är till ömsesidigt utbyte för dess medlemmar, förutsätter ett rationellt val att individerna inte är avundsjuka. Åtminstone så länge skillnaderna mellan individerna inte hotar att spränga sönder det samarbete som samhället bygger på.

Utifrån detta synsätt är det naturligt att jämställa antagandet om icke-avundsjuka med traditionella rationalitetsantaganden – som t. ex. transitivitet – när man konstruerar en teoretisk grundval för rationella val. Rawls teori skulle i så fall vara en *realistisk förfining* av den allmänna teorin om rationella val. Men hur det förhåller sig med den saken, går det inte att reda ut inom ramen för denna artikel.

→

*Joakim Sonnegård är fil. dr. i nationalekonomi och chef för sekretariatet till Kommissionen för hållbar utveckling*



## Noter

- 1 Rawls (1999), s. 492.
- 2 Jag kommer i denna uppsats genomgående att använda mig av Annika Perssons översättning från 1999 – Rawls (1999) – utgiven på Daidalos förlag.
- 3 Jag vill rikta ett stort tack till Lars Ö Ericsson för inspirerande samtal och uppmantran att skriva denna artikel. Artikelns brister är endast jag själv ansvarig för.
- 4 Rawls (1999), s. 245.
- 5 Rawls (1999), s. 37.
- 6 Elster (1986), s. 1.
- 7 Jag följer här den terminologi som Hare (1973) introducerade: POP = People in the Original Position.
- 8 Rawls (1999), s. 149.
- 9 Se Sen (1988), kapitel 1, fotnot 4.
- 10 Nedanstående resonemang bygger på Sen (1988), kapitel 1 och 1\*, vilket är de kapitel Rawls hänvisar till i en fotnot: se Rawls (1999), s. 149 resp. s. 141.
- 11 Vi kan låta *R* stå för relationen »åtminstone lika bra som, eller bättre än». För en diskussion av rationalitet och binära relationer, se Sen (1987).
- 12 (1) kan man betrakta som ett sundhetsvillkor snarare än som ett rationalitesvillkor; (2) är nödvändigt att anta, i annat fall existerar det inte någon valfunktion som i handling uttrycker POPsens preferenser; (3) är ett rent rationalitesvillkor.
- 13 Se Sen (1988), kapitel 1 och 1\*. Möjligen ligger det ytterligare antaganden dolda i Rawls ursprungssituation. Arrow (1987a) hävdar t.ex. att det implicit i Rawls differensprincip ligger ett antagande om att POPsens preferensordningar är *oberoende av irrelevanta alternativ*.
- 14 Se Sen (1988), s. 16.
- 15 Rawls (1999), s. 149.
- 16 Dvs. preferenserna antas vara reflexiva, fullständiga och transitiva. Möjligen kan man lägga till ytterligare ett (eller flera) antagande(n) om preferensernas egenskaper i Brennans modell, men det förhållandet spelar ingen roll för diskussionen i denna artikel.
- 17 Eftersom det inte spelar någon roll för analysen om man antar att det finns två eller flera individer i den värld som analyseras, följer jag här Brennans modellspecifikation.
- 18 För en diskussion av valfunktioner resp. nyttofunktioner, se Sen (1987).
- 19 Jämför denna klassificering med Nozick (2004) diskussion av avundsjuka mm.; se s. 301–316, speciellt vad Nozick säger i fotnot på s. 309; se även Hammond (1987b)
- 20 Man kan dessutom se altruism som ett exempel på en viss *klass* av osjälviskhet. Hammond (1987a) citerar en studie där altruism kategoriseras som »... one of the ›modalities‹ of unselfishness».
- 21 Rawls (1999), s. 189–190.
- 22 Ett enkelt sätt att modellera altruism skulle vara att formulera om antagande (a) i Brennans modell, så att individen upplever positiv nytta av andras inkomster. Och det kanske är ett rimligt antagande. De flesta går nog med på att ett samhälle blir mer drägligt att leva i om den allmänna inkomstnivån (eller för den delen, nivån på indexet för sociala primära nyttigheter) är så hög som möjligt, även om det knappast kan betraktas som ett tillräckligt villkor för att det ska vara så. Men det finns även andra sätt att modellera altruism, och dessutom andra möjliga former av interdependence mellan individerna i modellen.

23 Rawls (1999), s. 150.

24 Rawls (1999), s. 490.

25 Se Hare (1973): POL = People in Ordinary Life.

26 Rawls (1999), s. 493.

27 Rawls (1999), s. 492.

28 Rawls (1999), s. 490.

29 Rawls (1999), s. 492–493.

30 Rawls (1999), s. 497–498.

31 Att Rawls stipulerar att avund är en icke-moralisk känsla, gör det möjligt för honom att definiera POPsen som icke-avundsjuka och ändå klara sig ifrån att föra in – vad som konventionellt betraktas som ett – moraliskt begrepp i grundvalen för sin rättviseteori. Men Rawls behöver även hantera den avund som uppstår då »skillnaderna överskridit vissa gränser»; genom att definierade denna som »harm» kan han hävda att de utjämningsskrav som uppstår om skillnaderna mellan POPsen ser ut att bli för stora, *inte* har avund som grund.

32 Rawls (1999), s. 499.

33 För en diskussion av denna frågeställning: se t.ex. Arrow (1987b) eller Sen (1977) för en konceptuell analys av frågeställningen; för en empirisk diskussion, se Roth och Kagel (1995).

34 Sen (1987) tar upp detta problem med teorin för rationella val, och citerar Patrick Suppes som påpekat att: »... the standard normative model of expected utility can be satisfied by cognitive and moral idiots ... Put another way, the consistency of computations required by expected-utility models does not guarantee the exercise of judgement and wisdom in the traditional way ...» (s. 73–74).

35 Detta är en tankefigur som jag fått ifrån Næss (2005). Han skriver: »Med »människans natur» menar inte Spinoza alla aspekter hos människan, utan något som är grundläggande hos oss. Detta grundläggande kommer upp till ytan i form av det han kallar *ratio*s röst. Den konventionella översättningen av *ratio* är »förnuftet». Men det vi idag kallar förnuft, är väldigt olikt det filosofer på 1600-talet karakteriserade som *ratio*. När vi står inför viktiga handlingsval, är det *ratio* som säger oss vilket val som är i harmoni med människans natur. Människans natur är enligt Spinoza sådan att handlingsvalen som utpekas av *ratio* som de riktiga, är val utifrån aktiva, positiva känslor, inte av passiva, negativa känslor. På så sätt hjälper *ratio* oss att transformera negativa känslor i riktning mot positiva.» (s. 20–21). Och Spinoza menar, enligt Næss, att avundsjuka är en passiv affekt i förhållande till vår natur genom att den inte låter något av vår essens komma till uttryck (s. 92).

## Referenser

ARROW, K. J., *Social Choice and Individual Welfare*, New York: Wiley, (1951).

ARROW, K. J., »Arrow's Theorem», *The New Palgrave Dictionary of Economics*, New York: The MacMillan Press Limited, (1987a).

ARROW, K. J., »Economic Theory and the Hypothesis of Rationality», *The New Palgrave Dictionary of Economics*, New York: The MacMillan Press Limited, (1987b).

BRENNAN, G., »Pareto Desirable Redistributions: the Case of Malice and Envy», *Journal*

- of *Public Economics*, vol. 2, s. 173–183, (1973).
- ELSTER, J., *Rational Choice*, Oxford: Basil Blackwell, (1986).
- HAMMOND, P. J., ›Altruism›, *The New Palgrave Dictionary of Economics*, New York: The MacMillan Press Limited, (1987a).
- HAMMOND, P. J., ›Envy›, *The New Palgrave Dictionary of Economics*, New York: The MacMillan Press Limited, (1987b).
- HARE, R. M., ›Rawls Theory of Justice I›, *Philosophical Quarterly*, vol. 23, s. 144–155, (1973).
- KAGEL, J. H. och ROTH, A. E., *The Handbook of Experimental Economics*, Princeton: Princeton University Press, (1995).
- NÆSS, A., *Livsfilosofi*, Stockholm: Natur & Kultur, (2005).
- NOZIK, R., *Anarki, Stat och Utopi*, Stockholm: Timbro, (2004).
- RAWLS, J., *En teori om rättvisa*, Göteborg: Daidalos, (1999).
- SEN, A., ›Rational Fools: A Critique of the Behavioural Foundations of Economic Theory›, *Philosophy and Public Affairs*, vol. 6, s. 317–344, (1977).
- SEN, A., ›Rational behaviour›, *The New Palgrave Dictionary of Economics*, New York: The MacMillan Press Limited, (1987).
- SEN, A., *Kollektiva beslut & Social välfärd*, Stockholm: Thales, (1988).