

# FILOSOFISK TIDSKRIFT

4/92

**HANS ROSING**

*Trettio år med paradigmer*

---

**JAN ERIK LARSSON**

*Kan Searle tänka?*

---

**INGEMAR NORDIN**

*Replik till Lars Bergström*

---

**LARS BERGSTRÖM**

*Slutreplik till Ingemar Nordin*

---



# INNEHÅLL

*Trettio år med paradigmer* 3

HANS ROSING

*Kan Searle tänka?* 11

JAN ERIK LARSSON

*Replik till Lars Bergström* 28

INGEMAR NORDIN

*Slutreplik till Ingemar Nordin* 33

LARS BERGSTRÖM

*Recensioner* 37

*Notiser* 55

Filosofisk tidskrift nr 4 1992 årg 13. Filosofisk tidskrift utges av Stiftelsen Filosofisk tidskrift och Bokförlaget Thales.

*Redaktör och ansvarig utgivare:* Lars Bergström

*Redaktionskommitté:* Aant Elzinga (Göteborg) Göran Hermerén (Lund), Göran Lantz (Uppsala), Sven-Eric Liedman (Göteborg), Åke Löfgren (Karlstad), Giuliano Pontara (Stockholm), Dag Prawitz (Stockholm) och Dag Westerstahl (Stockholm)

*Produktion och prenumerationer:*

Bokförlaget Thales, Box 50034, 104 05 Stockholm, redaktionssekreterare Bitte Malmnäs, tel 08 16 33 58 må, on, fr.

*Prenumeration:* Tidskriften kommer ut med 4 nummer/år, pris 120 kronor, pg 507991-8 Filosofisk tidskrift

*Lösnummer:* 35 kronor

Satt av Ord och Grafik, Stockholm

Tryckt hos Lagerblads tryckeri AB, Karlshamn, 1992

ISSN 0348-7482



*Thomas Kuhn*  
Karikatyr av Kim Michelsen

## HANS ROSING

### *Trettio år med paradigmer*

---

I år har det gått trettio år sedan Thomas Kuhn gav ut boken *The Structure of Scientific Revolutions*, den mest omdebatterade boken inom vetenskapsfilosofin under detta århundrade.

Under dessa tre decennier har idéerna i boken analyserats och kommenterats i hundratals böcker, i tusentals fackartiklar och i oräkneliga mer populära sammanhang. Under de två senaste decennierna har man i otaliga sammanhang stött på idéer som på något sätt, ofta mycket avlägset, har sitt ursprung i denna bok. Speciellt populär har termen *paradigm* blivit. Termen har rönt samma öde som så många andra vetenskapliga facktermer. Den har blivit ett intellektuellt modeord som kan betyda praktiskt taget vad som helst.

En grupp som genast tog Kuhn till sitt hjärta var alla som kämpade för avvikande teorier av alla de slag, parapsykologer, ockultister, spiritister, ufologer, astrologer osv. Det som gjorde dem till hängivna Kuhn-anhängare var att Kuhns teori innebär att den akademiska vetenskapen till sin natur är konservativ. Forskarna försöker tvinga in verkligheten i ett stelt paradigm. De bortförklarar alla sådana fakta som strider mot paradigmat. Detta var just vad alla dessa kvasiforskare alltid hävdade. Inte underligt att de hälsade en etablerad forskare som sade samma sak med stor entusiasm.

Det som kvasiforskarna däremot aldrig har förstått är att de totalt missförstått Kuhn. De tror att Kuhn kritiserar vetenskapen när han hävdar att den är konservativ. De tror att Kuhn menar att den konservativa attityden är ett *hinder* för vetenskapliga framsteg. De fattar inte alls att Kuhn menar att det är tvärtom, att det är genom ett *konservativt fasthållande vid ett enda paradigm som vetenskapliga framsteg blir möjliga*. Enligt Kuhns teori är konservatismen ett tecken på vetenskaplighet. Kvasiforskarna har aldrig förstått att den akade-

miska vetenskapen, enligt Kuhns teori, gör alldeles rätt i att ignorera alla rapporter om mystiska och övernaturliga fenomen.

The *Structure* bygger på en mycket smal bas av vetenskaps-historiska data. Praktiskt taget alla exempel är hämtade från fysik, kemi och astronomi. Det finns inga exempel från de biologiska vetenskaperna, från geologin, från medicinen och allra minst från national-ekonomi och andra samhällsvetenskaper, från psykologin eller från lingvistiska vetenskaper.

Det är därför märkligt att boken väckte ett oerhört stort intresse bland folk som inte visste någonting om fysik, kemi och astronomi. Många tror säkert att förklaringen till detta intresse ligger i originaliteten och djupsinnet i Kuhns idéer. Som motargument till denna hypotes kan man peka på att liknande relativistiska idéer tidigare hade framförts av ett flertal forskare, t ex av Karl Mannheim, Michael Polanyi, N R Hanson och Stephen Toulmin. Man bör inte heller gömma bort den senare Wittgenstein i detta sammanhang.

Att just Kuhns bok fick en enorm uppmärksamhet berodde i främsta hand på att det fanns en s k social beställning på en bok av denna typ i början av 60-talet. Det var en brytningstid inom alla samhällssektorer, ett sökande efter en ny frihet och en ny världsbild. Kuhns insats var att han "bevisade" att den gamla synen på vetenskapen som ett objektivt kunskapssökande som leder till en stadigt växande kunskapsmassa var helt fel.

Att Kuhn väckte uppmärksamhet långt utanför kretsen av fackfilosofer berodde i stor utsträckning på att han uttryckte sin relativism så klart, entydigt och provokativt. Några typiska formuleringar får illustrera detta: "I en mening, som jag inte ytterligare kan förklara, utövar förespråkarna för olika paradigmer sin verksamhet i olika världar". "Just därför att den är en övergång mellan ojämförbara storheter kan övergången mellan konkurrerande paradigmer inte ske steg för steg, framtvingad av logik och neutrala erfarenheter". "Jag vill snarare påstå att i dessa frågor har vi varken att göra med bevis eller misstag" (1970, ss 150-151).

Vid denna tid fanns det, speciellt bland unga forskare, ett utbrett missnöje med den rådande positivistiska vetenskapssynen. Den upplevdes (med goda skäl) som alltför inskränkt och formalistisk. En alternativ vetenskapssyn, som vid denna tid fick allt större uppmärksamhet, var Karl Poppers metodologi. I vida kretsar, och

speciellt bland vänsterintellektuella, uppfattades Popper dock som alltför konservativ och positivistisk. Kuhns synsätt var däremot, tyckte man, precis det som man behövde.

Märkligt nog visade samhällsforskarna och t o m en del humanister stort intresse för Kuhns teori. Tydligen resonerade man ungefär så här: Orsaken till de stora framstegen inom fysiken är att man arbetat just på det sätt som Kuhn beskriver. Härav följer att om man arbetar på samma sätt inom andra forskningsområden så kommer man också där att göra snabba och revolutionerande framsteg.

Detta ledde till ett allmänt letande efter paradigmer inom alla tänkbara område. Termen paradigm började dyka upp i radio- och TV-program och t o m i dagstidningarnas ledarspalter. Inom samhällsvetenskaperna vara detta letande speciellt intensivt. Somliga hittade paradigmer inom t ex sociologin medan andra inte hittade några. Det verkade som om paradigmerna snarare fanns i sökarens medvetande än i vetenskapen själv. Den som var övertygad om att det fanns paradigmer såg dem överallt medan skeptikerna hade svårt att få syn på en enda.

Orsaken till denna märkliga egenskap hos paradigmerna var inte svår att finna. I *The Structure* gav Kuhn aldrig någon klar och entydig definition av vad han menade med termen. Margaret Masterman påpekade i en berömd artikel att Kuhn använde termen i 22 betydelser som något avvek från varandra. Termen kunde betyda allt från "en konkret vetenskaplig prestation" till "en karakteristisk uppsättning åsikter och förutsättningar" (Masterman 1970). Kuhn själv medgav att han varit alltför vårdslös med termen. År 1969 skrev han (mycket träffande): "Jag måste tyvärr dra slutsatsen att en delorsak till dess framgång är att den kan tolkas nästan hur som helst av vem som helst" (Kuhn 1977, s 459).

Kuhn gav 1969 en mer detaljerad analys av paradigmbegreppet. Han menade att man kan dela de ovan nämnda 22 betydelserna i två huvudtyper. Han skrev: "En betydelse av paradigm är global, omfattande alla de förpliktelser som en vetenskaplig grupp delar; den andra avgränsar ett speciellt viktigt slag av förpliktelser och är sålunda en delmängd av den förra". Den senare betydelsen döpte han om till "disciplinär matris". De viktigaste komponenterna i en disciplinär matris är, enligt Kuhn, (a) symboliska generaliseringar, dvs matematiskt formulerade lagar, (b) modeller och (c) exempel.

Kuhns försök till precisering har inte blivit någon succé. Dels blir hans teori inte klarare och dels försvinner den eleganta enkelhet som gjort teorin så enormt populär. Om man skiljer mellan globala paradigmer och disciplinära matriser får man minst två olika typer av vetenskapliga revolutioner. Dels har vi ett byte av globala paradigmer som kan tänkas ske så som paradigmskiftet beskrivs i *The Structure*. Dessa paradigmer är dock precis lika diffusa och undanflugande som tidigare. Dels har vi byten av disciplinära matriser som kan tänkas ske antingen inom ramen för en och samma global paradigm eller från en global paradigm till en annan.

Det som, enligt min mening, är mest märkligt med debatten om Kuhns bok under de senaste två decennierna är att författaren själv inte alls deltagit i den. Efter den heta debatten i slutet av 60-talet koncentrerade han sig helt på vetenskapens historia och överlät åt yrkesfilosofer (och stora skaror av amatörer) att uttolka och utveckla hans idéer.

Kuhn har undvikit intervjuer, men senaste år lät den 68-åriga historikern sig övertalas till en intervju för *Scientific American* (maj 1991). Han beklagar sig, med all rätt, över att han så ofta blivit missförstådd. Han känner sig dock ofta mer frustrerad över sina beundrars missförstånd. "Jag har ofta sagt att jag är mycket mera förtjust i mina kritiker än i mina beundrare". Han berättar om studenter som tacksamt säger: "Tack så mycket mr Kuhn för att ni berättat för oss om paradigmen. Nu när vi vet om dem så kan vi göra oss av med dem". Kuhn är också irriterad över dem som tror att hans teori implicerar att vetenskapen är irrationell. "Om de hade sagt irrationell så skulle jag inte alls ha brytt mig om det".

Det framgår av intervjun att Kuhn trots allt inte är färdig med vetenskapsfilosofin. Han håller på med en bok som antagligen ("probably") kommer att ta upp två centrala frågor i *The Structure*, nämligen det ofta diskuterade begreppet inkommensurabilitet och den likaledes ofta diskuterade likheten mellan evolution i vetenskapen och evolution av levande organismer. Denna skrift kommer säkert att tas emot med stora förväntningar.

Hur skall vi utvärdera Kuhns teori i dag efter trettio år av forskning och debatt? Mitt allmänna intryck är att amatörerna menar att han har rätt och experterna att han har fel.

Med amatörer menar jag filosofer, samhällsforskare, humanister,



intellektuella osv som har mycket begränsade kunskaper i naturvetenskaperna och deras historia. Jag har under årens lopp märkt att amatörerna har en benägenhet att tro att naturvetenskapernas struktur är mycket enklare än vad den i själva verket är.

Larry Laudan som definitivt är expert både på metodologi och vetenskapshistoria skrev 1984: "I själva verket är vi idag i den situationen att vi tämligen entydigt kan säga att Kuhns modell för vetenskaplig förändring, såsom den framläggs i *The Structure* och vidareutvecklas i *The Essential Tension*, innehåller grundläggande fel, inte endast när det gäller speciella detaljer utan beträffande dess grundläggande antaganden" (Laudan 1984, xiii).

Vad säger yrkesforskarna själva, dvs fysiker, kemister och astronomer om den bild Kuhn målat upp av dem? Mitt intryck är att de inte alls känner igen sig. I en intervju i *Scientific American* (nov/91) utdömer Edward Witten, den viktigaste förespråkaren för den s k superstringteorin, av många betraktad som en modern Einstein, Kuhns teori som "silly".

Detta är säkerligen att ta till i överkant, men forskningen under de gångna 30 åren visar att det, lindrigt talat, finns tungt vägande skäl att se på Kuhns idéer med kritiska ögon. Att påpeka detta innebär naturligtvis inget ringaktande av Kuhns insats och betydelse. Kuhn har för all framtid en viktig plats i 1900-talets vetenskapsfilosofi. Att det finns grundläggande fel i hans modell utesluter inte att han betonat sidor av vetenskapen som man tidigare inte alls uppmärksammat. Man kan inte ha någon djupare förståelse av vetenskapen om man inte förstår dess mer arationella sidor. Den debatt Kuhn drog i gång har visserligen producerat mycken emotionell hetta men den har också kastat nytt ljus över den ytterst komplicerade process i vår kultur som vi så gärna tror att vi förstår när vi hör ordet "vetenskap".

En fråga som man sällan ställer sig är: Hur kom Kuhn fram till sin teori? Teorin är en generalisering baserad på historiska data från i främsta hand de senaste trehundra åren. Den (kanske omedvetna) uppfattning om hur det gått till, som de flesta tycks ha är följande: Först har Kuhn omsorgsfullt studerat ett antal exempel på upptäckter i vetenskapens historia. Han har därvid lagt märke till att standardtolkningen av vad som skett (dvs den traditionella positivistiska historietolkningen) inte stämmer. På denna punkt torde alla vetenskapshistoriker idag vara eniga. Den positivistiska historieskrivningen

var onekligen synnerligen förenklad. Men hur kom Kuhn fram till sin teori om en arationell utveckling?

De flesta tycks tro att han på något sätt *härledde teorin från sina historiska data*. Först studerade han i detalj ett antal enskilda exempel på hur ny vetenskaplig kunskap uppstår. Sedan *upptäcker* han att det finns ett antal gemensamma drag hos de enskilda fallen. Han drar då slutsatsen att dessa drag gäller generellt. De gäller inte bara för de fall han studerat utan generellt för all "mogen" vetenskap. Han *inducerar* sålunda ett antal sociala lagar som beskriver vetenskapens sätt att arbeta.

Vilken metod har han sålunda använt? Det är uppenbart att han (enligt denna populära tolkning) använt den traditionella positivistiska metoden, den s k *induktiva metoden*. Först samlar han in fakta. Därefter härleder han induktivt ett antal lagbundheter från dessa fakta. Mot denna bakgrund är det inte förvånande att boken kom ut i en serie som redigerades av positivisterna.

Det tråkiga är att alla experter på vetenskaplig metodologi i dag är eniga om att den induktiva metoden innebar ett mycket förenklat sätt att tolka en komplicerad företeelse. I själva verket kan man bevisa att Kuhn inte kan ha arbetat på detta sätt. Fakta, och speciellt historiska fakta, är inte objektiva och neutrala. Det finns överhuvudtaget inga objektiva historiska fakta från vilka Kuhn kunde ha härlett sina generaliseringar. Alla de "fakta" Kuhn utgår ifrån är i själva verket *tolkningar* som bygger på en bestämd förhandsinställning (s k förförståelse enligt hermeneutisk terminologi).

Jag vill påstå att det som verkligen hände i stället var någonting i stil med följande. Under påverkan av de nya antipositivistiska idéer som låg i luften under slutet av 50-talet kände sig Kuhn otillfredsställd med den positivistiska historieskrivningen. Han började därför söka efter ett nytt sätt att tolka som passade bättre in i tidsandan. I förordet till *The Structure* ger han några uppgifter om varifrån han fått sina grundläggande idéer. Han arbetade sålunda med ett slags vag hypotes, som för övrigt redan finns antydd i hans första bok från 1958 som handlade om den kopernikanska revolutionen.

Kuhn studerade alltså inte historien med kylig neutral blick utan han hade redan en vag hypotes om hur det egentligen har gått till. Han var på jakt efter sådana exempel som lätt kunde tolkas så att de bekräftade denna vaga hypotes. Han fann naturligtvis sådana

exempel. (I historien kan man finna exempel på praktisk taget vad som helst). Med hjälp av ett antal exempel utarbetade han sin hypotes så långt att det var möjligt att publicera den. Men priset för att teorin skulle bli någorlunda elegant och lättbegriplig var att han använde ett synnerligen diffust begrepp, dvs begreppet paradigm. Så länge ett paradigm kunde vara litet vad som helst var det lätt att försvara ett antal generaliseringar om paradigm. Generaliseringarna var så vaga att de var omöjliga att testa.

Det som vi idag torde kunna vara eniga om är sålunda att Kuhn har visat att det är möjligt att tolka vissa episoder i fysikens, kemins och astronomins historia med hjälp av hans vaga modell. Det vore däremot klart fel att påstå att Kuhns tolkning är den enda möjliga. Det finns en mängd andra i någon mening rimliga tolkningsmöjligheter.

En vanlig uppfattning tycks vara att Kuhn inte bara ger en tolkning av vetenskapens historia utan också ger ett *recept* på hur bra vetenskap bör bedrivas. En stor del av amatörernas intresse för hans teori bottnar tydligen i en tro att man kan lära sig något om hur man *borde* gå till väga för att åstadkomma framsteg. Denna uppfattning är emellertid helt felaktig. Kuhn har själv aldrig gjort anspråk på att vara annat än vetenskapshistoriker. Några rekommendationer för hur dagens forskare *borde* bete sig har han vägrat ge. Man kan sålunda, i motsats till vad många tycks tro, inte dra några metodologiska slutsatser ur hans teori.

Det mest ironiska i sammanhanget är att Kuhn faktiskt har gett vissa regler för god forskning. Men dessa regler är gammal välkänd positivistisk skåpmat. I *Postscript-1969* till andra upplagan av *The Structure* (1970) och i boken *The Essential Tension* (1977) försvarar han sig mot beskyllningarna för att vara relativist genom att räkna upp ett antal generella, dvs icke paradigmbundna, regler för god forskning.

I den senare boken nämner han fem regler. (1) En teori bör vara i god överensstämmelse med existerande experiment och observationer. (2) En teori får inte vara inkonsistent och den måste vara logiskt förenlig med accepterade teorier inom sitt tillämpningsområde. (3) Teorin bör predicera nya företeelser. (4) Teorin bör vara enkel i den meningen att den sammanför tidigare orelaterade företeelser under en gemensam princip. (5) Teorin bör var fruktbar, dvs ge

upphov till nya problem, idéer, experiment osv (Kuhn 1977, ss 321-2). Dessa regler har i själva verket länge varit standard i den positivistiska metodologin. I grund och botten är Kuhn sålunda en moderat och urvattnad positivist.

### *Litteratur*

THOMAS KUHN, *The Structure of Scientific Revolutions*, första upplagan 1962, andra upplagan med ett Postscript 1970, University of Chicago Press. Boken utkom i serien *International Encyclopedia of Unified Science* som utgavs av de logiska empiristerna.

THOMAS KUHN, "Second Thoughts on Paradigms", ingår i Frederick Suppe, utg, *The Structure of Scientific Theories*, andra upplagan, 1977, University of Illinois Press.

THOMAS KUHN, *The Essential Tension*, 1977, University of Chicago Press.

LARRY LAUDAN, *Science and Values*, 1984, University of California Press.

MARGARET MASTERMAN, "The Nature of a Paradigm", ingår i I Lakatos and A Musgrave, utg, *Criticism and the Growth of Knowledge*, 1970, Cambridge University Press.

(Alla översättningarna är gjorda av författaren.)

## JAN ERIC LARSSON

### *Kan Searle tänka?*

---

Kan datorer åstadkomma medvetande? John Searle, professor i filosofi vid University of California, Berkeley, hävdar att så inte är fallet. Han angriper en ståndpunkt som kallas *stark AI*, medan den försvaras av många andra. Searles resonemang är centrerat kring ett exempel som kallas *det kinesiska rummet*, vilket har givit upphov till stor debatt inom filosofi- och AI-kretsar.

Min egen åsikt är att vare sig Searle eller företrädarna för stark AI kan veta att de har rätt. Vi kan för närvarande inte veta huruvida datorer kan åstadkomma medvetande, och med Searles medvetandeteori kan vi aldrig komma att veta det. Searle har emellertid klart deklarerat att han anser sig kunna visa att stark AI är principiellt omöjlig. Hans resonemang är emellertid bristfälligt och jag hoppas att i det följande kunna visa att det inte duger som argument mot stark AI.

#### *1. Searles argumentation*

Låt oss inleda med att beskriva Searles argumentation.

##### *Stark och svag AI*

Searles mål är att angripa en ståndpunkt hos vissa AI-forskare, som säger att man genom att skriva ett program för en dator *de facto* kan skapa ett medvetande.

Ett stort antal forskare inom artificiell intelligens (AI) tror... att de genom att skriva rätt program, med rätt inmatning och utmatning, *de facto* skapar medvetanden, (1990, s 20).

Denna ståndpunkt kallar Searle för *stark AI*, i motsats till *svag AI*, som inte nödvändigtvis anser det möjligt att skapa medvetande, utan hävdar att AI-tekniken har andra värden. Vidare hävdar Searle att

man inom stark AI anser sig ha ett test för att avgöra om man åstadkommit medvetande, nämligen *Turingtestet*.

De tror vidare att de har ett vetenskapligt test för att avgöra sin framgång: Turingtestet, som beskrevs av Alan M Turing, den artificiella intelligensens fader. Turingtestet, i nuvarande tolkning, är helt enkelt: om en dator kan uppträda på så sätt att en expert inte kan skilja den från en människa som har en viss kognitiv förmåga – till exempel förmågan att addera tal eller förstå kinesiska – då har även datorn den förmågan, (1990, s 20).

Så långt Searles distinktion mellan stark och svag AI. Vi skall ha anledning att komma tillbaka till Searles uttalanden, men först skall vi ge en kort beskrivning av de två tänkta experiment som förekommer i argument och motargument.

### *Det kinesiska rummet*

I korthet kan experimentet med det kinesiska rummet beskrivas enligt följande. Låt oss välja en försöksperson, i det aktuella fallet Searle själv, och ett språk som denne inte förstår, till exempel kinesiska. (Vi bortser med Searle från att kinesiska inte är *ett* språk. Den nitiske läsaren kan i det följande ersätta kinesiska med exempelvis mandarin.) Antag att vi stänger in Searle i ett rum i vilket finns en mängd lappar med kinesiska skrivtecken, ett hål genom vilket sådana lappar kan skickas in och ut, samt ett (mycket komplext) bokverk. Detta bokverk innehåller anvisningar om vilka skrivtecken som skall skickas ut till omvärlden, med hänsyn till vad för tecken denna omvärld skickat in. Dessa anvisningar kan mycket väl få innebära att Searle måste genomföra ett långvarigt och komplext arbete i flera led; det väsentliga är att de är fullständigt utförda, så att Searle kan klara jobbet utan att förstå språket.

Antag också att bokverket är så skrivet att, när en kinesisktalande testperson skickar in meddelanden skrivna på kinesiska och Searle genom att mekaniskt följa bokverkets anvisningar skickar ut svarsmeddelanden, testpersonen bedömer att det finns någon eller något i rummet som förstår kinesiska. Detta till trots, menar Searle, finns det inget *medvetande* i rummet som förstår kinesiska. Searle själv gör det inte, medan vare sig bokverket, lapparna eller någon annan fysisk del av rummet besitter medvetande.

Experimentet är, åter enligt Searle, ekvivalent med den tänkta situationen att vi har en dator som tack vare ett listigt skrivet program verkar förstå kinesiska. Bokverket motsvarar programmet, lapparna programmets data och Searle processorn. Inte i någondera fallet uppstår något medvetande till följd av den pågående symbolhanteringen, och alltså, sluter Searle, är stark AI inte korrekt.

### *Turingtestet*

Alan Turing designade sitt test bland annat just för att undvika problem av den art som den här omnämnda debatten behandlar. I stället borde man, enligt Turing, anta ståndpunkter som i mycket liknar vad Searle kallar svag AI. I korthet kan Turings test beskrivas enligt det följande.

Anordna en kommunikationslänk som bara tillåter överföring av skrivna meddelanden, till exempel via två datorterminaler, placerade i var sitt rum. I den första versionen låter vi en kvinna och en man omväxlande sitta vid den ena terminalen, varpå testpersonen vid den andra terminalen ges möjlighet att föra en konversation via kommunikationslänken, i avsikt att utröna huruvida han talar med en kvinna eller en man. Både kvinnan och mannen har att försöka övertyga testpersonen om att de är kvinnan; mannen måste alltså försöka imitera en kvinna för att lyckas.

I den andra versionen har vi i stället en människa och en dator. Datorns uppgift är, på liknande sätt som tidigare, att försöka övertyga testpersonen om att den är människan. Lyckas den har den klarat Turingtestet.

Som synes är den beskrivning av Turingtestet som Searle tillskriver stark AI både felaktig och missvisande. Turing anser inte att testet stöder stark AI eller bevisar att datorn har ett medvetande. Tvärtom är hans avsikt att visa att man bör försöka gå runt de intrikata problemen; det finns intressanta resultat att nå i alla fall.

Det kommer att underlätta för läsaren om jag förklarar min egen åsikt i frågan. Låt oss först ta den mer preciserade frågeställningen. Jag tror att det om cirka femtio år kommer att vara möjligt att programmera datorer... så att de kan spela imitationsspelet så väl att en genomsnittlig frågare inte har mer än 70 procents chans att göra rätt bedömning efter fem minuters utfrågning. Den

tidigare frågan 'Kan maskiner tänka?' anser jag vara alltför meningslös för att förtjäna diskussion, (1950, s 422).

Om mannen i Turingtestets första version skulle lyckas lura testpersonen bevisar detta naturligtvis inte att han är kvinnan, och på samma sätt: om ett program skulle klara testet bevisar detta vare sig att programmet är en människa eller att det har ett medvetande. Däremot kan vi tänkas vara intresserade av att använda programmets goda förmåga att härma en människa.

Searle klargör inte någonstans huruvida han själv omfattar den felaktiga beskrivning av Turingtestet han givit.

### *Searles axiomatiska argumentering.*

Det av Searle föreslagna tankeexperimentet är endast ena halvan av den argumentation som framförs. För att motbevisa den starka AIntes anför Searle även ett formellt resonemang, bestående av fyra axiom och fyra slutsatser. Så här ser axiomen ut:

1. Datorprogram är formella, (syntaktiska).
2. Mänskliga medvetanden har mentalt innehåll, (semantik).
3. Syntax i sig är varken grundläggande eller tillräckligt för semantik.
4. Hjärnor orsakar medvetanden.

Ur dessa premisser följer, enligt Searle, följande slutsatser:

1. Datorprogram är varken grundläggande eller tillräckliga för att orsaka medvetande.
2. Varje annat system som kan orsaka medvetande måste ha kausala förmågor (åtminstone) jämförbara med hjärnors förmåga.
3. Varje artefakt som producerar mentala fenomen, varje konstgjord hjärna, måste kunna duplicera den mänskliga hjärnans specifika kausala förmågor, och den kan inte göra detta enbart genom att exekvera ett formellt program.
4. Det sätt varpå den mänskliga hjärnan producerar mentala fenomen kan inte vara enbart genom exekverandet av ett program.

Vid första påseende kan axiom ett verka okontroversiellt. Searle menar att ett datorprogram är rent formellt till sin natur; dess implementation är i alla avseenden oväsentlig.



Symboler och program är rent abstrakta begrepp: de definieras inte av några väsentliga fysiska egenskaper och kan implementeras i godtyckligt fysiskt medium, (1990, s 21).

Detta gör att det verkar rimligt att tolka det första axiomat som så, att det betyder att datorprogram är *rent* formella.

Det andra axiomat är svårtolkat, eftersom Searle normalt inte antyder vilket medvetandesbegrepp han använder sig av. Av (1984) framgår dock att han med *mentalt innehåll* avser ett subjektivt medvetande, alltså inre upplevelser som sinnesintryck, tankar, känslor, minne, etc. En företrädare för en radikal materialism eller funktionalism skulle troligtvis inte acceptera det andra axiomat.

Vidare är Searles användning av begreppet semantik förvirrande. Att döma av den kommentar som ges till det tredje axiomat är avsikten antagligen endast att säga att det är skillnad på fenomen med och utan mening, och att semantik i det andra axiomat mer eller mindre kan översättas med mentala tillstånd.

Poängen är distinktionen mellan formella element, som inte har någon inneboende mening, och sådana fenomen som har ett inre innehåll, (1990, s 21).

I det tredje axiomat vill Searle antyda att syntax definitionsmässigt inte kan skapa semantik. Huruvida detta är sant i formell mening skall vi inte gå in på här, utan endast konstatera att i Searles tolkning, där semantik är ekvivalent med medvetande, gäller ingen sådan definitionsmässig sanning.

I det fjärde axiomat anger Searle en klassisk ståndpunkt inom kropp- och själproblematiken, nämligen att hjärnan *orsakar* medvetande. Denna ståndpunkt måste dock anses som allt annat än klar.

## 2. Felaktigheter i Searles resonemang

Searle hävdar att de ovan angivna premisserna är evident sanna och att slutsatserna logiskt följer ur dessa premisser. Vi skall emellertid se att dels undgår inte en enda av premisserna svårbemött kritik, dels följer inte självklart, på grund av Searles mångtydiga användning av flera begrepp, slutsatserna ur premisserna. Det är faktiskt anmärkningsvärt hur låg kvalitet Searles argumentation har.

### *Medvetande kontra intelligens*

Den första svagheten i Searles resonemang är hans inriktning mot *medvetande*. AI-forskningens mål har så gott som alltid sagts vara att åstadkomma *intelligens*. Att detta skulle vara liktydigt med att åstadkomma medvetande är långt ifrån uppenbart; snarare verkar det rimligt att intelligens, eller åtminstone intelligent beteende, inte nödvändigtvis kräver ett bakomliggande medvetande.

Stark AI i Searles version har antagligen få försvarare som vet vad de talar om. Detta erkänner Searle själv i (1990, s 20). I artikeln konstaterar Searle att inte ens Churchlands förstått den starka AIns ståndpunkt. Min slutsats är att Searle till stor del skapat ett spöke som han sedan försöker bekämpa.

Redan denna inledande kritik anser jag vara förödande för Searles resonemang. Stark AI handlar om *intelligens*, inte om *medvetande*.

### *Det kinesiska rummet*

Mycken kritik har riktats mot det tänkta experiment som Searle anför: det går inte att skriva ett bokverk med de angivna egenskaperna, Searle skulle vara för långsam för att klara uppgiften, han skulle inte kunna undvika att lära sig kinesiska, etc. All denna kritik är befogad, och kan sammanfattas i det enkla konstaterandet att i praktiken är hela förutsättningen absurd. Searle kan inte rimligtvis ersätta en dator på det sätt som beskrivs, han är alldeles för långsam, han skulle inte kunna följa de komplexa instruktioner som krävs, han skulle aldrig kunna överblicka en tillräckligt stor mängd data och han skulle oundvikligen begå alltför många fel; allt detta skulle med största säkerhet ligga långt bortom varje människas förmåga. Att, som Searle i ett kontraargument föreslår, anta att han skulle kunna memorera allt och klara sig helt utan yttre hjälpmedel är (om möjligt) än orimligare. Därtill kommer att tankeexperimentet har ytterligare en stor brist, nämligen att "processorn," det vill säga Searle, har ett eget medvetande. Detta medvetande förväxlar Searle (mer eller mindre medvetet) med det eventuella medvetande varom debatten rör sig. Hela experimentet är alltså synnerligen bristfälligt.

I stället för att fullfölja dessa argument skall vi emellertid konstatera att Searles rum *i princip* är en rimlig ekvivalent till ett datorsystem, vilket som helst. I samma ögonblick kan vi emellertid också konstatera att därmed försvinner exemplets specifika intresse.

Den principiella möjligheten för ett datorsystem att åstadkomma medvetande är troligtvis lika stor, oavsett om det rör sig om en dator i halvledarmaterial eller Searles föreslagna arrangemang. Vi har lika god eller dålig intuition i båda fallen och alltså kan vi lämna exemplet därhän. Dess praktiska orimlighet gör att vi lämnar det med glädje.

### *Searles medvetandesfilosofi*

Innan vi tar oss an Searles axiomatiska argumentering skall vi undersöka hans teori om medvetandets natur. Denna fråga har notoriskt undvikits i vulgärdebatten.

Searle kallar ståndpunkten att allt som existerar i världen är fysiska partiklar för *naiv fysikalism*, och på samma sätt låter han *naiv mentalism* beteckna ståndpunkten att mentala fenomen verkligen existerar: att det finns mentala tillstånd, några är medvetna, många äger intentionalitet, alla är subjektiva och många har kausala funktioner i den fysiska världen.

Dessa två teorier är korrekta och förenliga, hävdar Searle. De mentala tillstånden är orsakade av processer i hjärnan.

Mentala fenomen, medvetna såväl som omedvetna, visuella som audiella, smärta, kittlingar, klåda, tankar, ja hela vårt mentala liv, orsakas av processer i hjärnan, (1984, s 18).

Samtidigt som mentala tillstånd är orsakade av processer i hjärnan är de också egenskaper hos densamma.

Smärta och andra mentala fenomen är egenskaper hos hjärnan (och kanske resten av det centrala nervsystemet), (1984, s 19).

Searles teori uppvisar stora likheter med en ontologisk teori som beskriver en vardaglig syn på medvetandesproblemet. Vi skall, obekymrade om att termen också används på annat sätt, kalla denna teori för *naiv realism*.

1. Det existerar en värld av fysiska föremål.
2. Det existerar subjektiva, mentala tillstånd.
3. Vi äger kunskap om den fysiska världen genom våra sinnesintryck.
4. Våra upplevelser av fysiska föremål är orsakade av dessa fysiska föremål.

Searle menar att en sådan teori inte behöver innebära dualism. De mentala tillstånden kan uppstå som makroegenskaper ur neurala processer i hjärnan, på samma sätt som till exempel en bordsytas hårdhet uppstår ur atomära gitterstrukturer på mikronivå. Vi noterar emellertid att detta är högst kontroversiellt. Den naiva realismen må vara intuitivt tilltalande, men inte någon av dess satser undgår allvarlig kritik, och teorin har få försvarare. Om man inte accepterar den faller Searles argumentation platt till marken.

### *Axiomens giltighet*

Vi skall nu övergå till att analysera Searles axiomatiska argumentation. Vad det gäller tolkningen av de olika axiomen skall vi komma ihåg att Searle förutsätter den naiva realismens syn på världen. Det finns som bekant flera konkurrerande medvetandeteorier, men inte i någon tolkning är resonemanget hållbart. Låt oss se närmare på Searles egen naiva realism.

De axiom och slutsatser som Searle anför i sin artikel är ganska vaga och oklara. Låt oss därför göra följande tolkning.

1. Exekverande datorprogram är rent formella.
2. Medvetanden innehåller mentala tillstånd.
3. Rent formella system kan inte åstadkomma mentala tillstånd.
4. Hjärnor orsakar medvetanden.

Det första axiomets riktighet är avhängig definitionen av begreppet datorprogram. Ser vi ett program som enbart en instruktionssekvens är axiomat korrekt, men en sådan tolkning är knappast rimlig. En förnuftig definition av ett datorprogram måste omfatta en processorbeskrivning, samt en instruktionssekvens som *exekveras* på en *fysisk* processor. Med en sådan tolkning beror axiomets riktighet på om det är avsett att betyda att vissa representationer av datorprogram besitter syntaktiska egenskaper (vilket är korrekt) eller att datorprogram *enbart* besitter syntaktiska egenskaper (vilket är fel). Enligt den tidigare definitionen är ett program visserligen formellt, men inte enbart syntaktiskt, och detta är den enda rimliga ståndpunkten i en debatt. Ingen representant för stark AI skulle rimligen vilja påstå att en instruktionsföljd i sig (till exempel en programlistning) åstadkommer medvetande. Det är kombinationen av instruktioner och exekverande processor som kommer i fråga. Alltså kan vi konstatera

att Searles resonemang antingen inte gäller i det fall vi faktiskt exekverar ett program på en dator, eller så är det första axiomat felaktigt. Denna invändning är *helt* förödande för Searles resonemang.

Axiom två är avhängigt vilken medvetandeteori man omfattar. I den naiva realismen är det rimligtvis korrekt. Mentala tillstånd äger existens och är på något okänt men fungerande sätt kopplade till de olika processer som försiggår i den mänskliga hjärnan.

Det tredje axiomat kan, på grund av Searles oklara användning av begreppen syntax och semantik tolkas på två olika sätt. Dels kan det antas betyda ungefär att formell syntax är något annat än formell semantik, att syntaxregler inte räcker för att definiera en semantik. I denna tolkning är axiomat logiskt till sin natur. Antagligen är det inte korrekt, men det behöver vi inte diskutera här. (Frågan hänger på existensen av syntaxregler som samtidigt utgör en semantik; och visst bör det finnas sådana urartade fall.) Vi kan nöja oss med att konstatera att om axiomat ges denna tolkning går det inte att dra de slutsatser som Searle önskar, helt enkelt därför att de olika axiomen talar om disparata begrepp.

I den andra tolkningen är axiomat tre i stället avsett att betyda att ren symbolbehandling inte kan åstadkomma medvetande. Denna tolkning är rimligtvis den Searle avsett. Här kan vi med ens konstatera att det inte, som Searle vill antyda, finns någon definitionsmässig motivering att stödja sig på, utan i stället innebär axiomat helt enkelt att Searle förutsätter den slutsats han vill bevisa. Accepterar man inte redan denna, faller det tredje axiomat. Även detta anser jag vara *helt* förödande för Searles resonemang.

Dubbeltydigheten i axiomat tre utnyttjas av Searle på ett i högsta grad oacceptabelt sätt.

På en nivå är denna princip definitionsmässigt sann. Man kan naturligtvis definiera begreppen syntax och semantik på andra sätt. Poängen är att det finns en skillnad mellan formella element, som inte har någon inneboende mening, och sådana fenomen som har inre innehåll, (1990, s 21).

Detta dubbla budskap är i bästa fall en monumental begreppsförvirring, och i värsta fall ett trick, avsett att ge Searles ställningstagande i en avgörande debattfråga en skenbar aura av logisk

nödvändighet. Det sistnämnda stöds tyvärr av att Searle senare hävdar att axiom tre är en logisk sanning.

Det fjärde axiomat bygger helt på att man accepterar den naiva realismen och därmed den lösning på kropp- och själproblemet som brukar kallas *interaktionism*, det vill säga att medvetandet orsakas av fysiska förlopp (och rimligtvis vice versa). Detta är dock inte okontroversiellt. Bland problemen med teorin i fråga brukar bland annat anges följande: kausalitetsrelationer förekommer normalt mellan *fysiska* händelser och det är inte oproblematiskt att anta att de även kan existera mellan fysiska och mentala händelser; en kausalitetsrelation från mentalt till fysiskt skulle dessutom komma att strida mot principen om energins bevarande, etc. Vi nöjer oss med att konstatera att Searle utan närmare motivering väljer en (långt ifrån problemfri) filosofisk teori av många. Accepterar man inte detta val, faller också det fjärde axiomat.

En välvilligare tolkning kan göra det första axiomat mindre kontroversiellt. Antar vi att det skall tolkas som att datorprogram är formellt beskrivbara, blir det plötsligt rimligare.

1. Exekverande datorprogram är formellt beskrivbara.
2. Medvetanden innehåller mentala tillstånd.
3. Formellt beskrivbara system kan inte åstadkomma mentala tillstånd.
4. Hjärnor orsakar medvetanden.

Här måste emellertid axiom tre anses betyda att inget formellt beskrivbart system kan åstadkomma medvetande, en absurd konsekvens som Searle knappast kan önska sig. I så fall försvinner ju allt hopp om att någon gång kunna beskriva ens den mänskliga hjärnan på ett vetenskapligt sätt. Att döma av artikeln i övrigt är det emellertid inte troligt att Searle har avsett denna välvilligare tolkning av det första axiomat.

Tillåter man sig att tänja tolkningen ännu ett steg kan det tredje axiomat anses betyda att formellt beskrivbara system inte kan åstadkomma mentala tillstånd genom sin egenskap att vara just formellt beskrivbara. Searle hävdar också att det är hjärnans specifika kausala egenskaper som orsakar medvetande. Om så skulle vara fallet (vilket är en helt öppen fråga) följer i vilket fall som helst inte att medvetande inte skulle kunna orsakas även på andra sätt, till exempel av datorer. Om detta säger Searle följande.

Finns det något logiskt tvingande skäl för att de inte kunde avge medvetande? Nej. Vetenskapligt sett är det helt uteslutet, men detta är inget som det kinesiska rummet är avsett att bevisa,... Det viktiga är program, och program är rent formella, (1990, s 25).

Searle förnekar alltså inte att datorer i sin egenskap av fysiska föremål skulle kunna avge medvetande, eller, ekvivalent uttryckt, det är alltså enligt Searle fullt möjligt att formellt beskrivbara system i sin egenskap av att vara fysiska skulle kunna åstadkomma medvetande. Att hävda något annat vore knappast rimligt och detta är säkerligen inte den tolkning som Searle avsett.

Låt oss i förbifarten observera hur ohederligt Searle argumenterar. Lika lite som någon annan vet han något om vilka fysiska föremål som kan åstadkomma medvetande, men likväl framställer han möjligheten av att datorer skulle ha denna egenskap som "vetenskapligt sett helt utesluten."

### *Slutsatsernas giltighet*

Vi har nu sett att axiomen inte är problemfria. Men även i det fall att vi antar deras riktighet kan mångtydigheten i de använda begreppen i vissa tolkningar komma att leda till att slutsatserna inte är giltiga. Låt oss återupprepa dessa och kommentera dem.

1. Datorprogram är varken grundläggande eller tillräckliga för att orsaka medvetande.
2. Varje annat system som kan orsaka medvetande måste ha kausala förmågor (åtminstone) jämförbara med hjärnors förmåga.
3. Varje artefakt som producerar mentala fenomen, varje konstgjord hjärna, måste kunna duplicera den mänskliga hjärnans specifika kausala förmågor, och den kan inte göra detta enbart genom att exekvera ett formellt program.
4. Det sätt varpå den mänskliga hjärnan producerar mentala fenomen kan inte vara enbart genom exekverandet av ett program.

Den första slutsatsen skall, enligt Searle, följa ur axiom ett, två och tre. Detta förutsätter dels att axiom ett är betyder att datorprogram är *enbart* syntaktiska, vilket gör resonemanget så gott som ointressant, dels att begreppet semantik enligt axiom två och tre är detsamma. I axiom två betyder semantik ungefär detsamma som mentalt innehåll,

medan det i axiom tre eventuellt kan vara avsett att ha en språk-analytisk innebörd, (den normala). Den första slutsatsen är i så fall inte ens formellt riktig.

Slutsats två verkar, i en rimligt godvillig tolkning, korrekt. Det är väl närmast trivialt sant att varje system som kan orsaka medvetande måste ha förmågor som är i vid mening ekvivalenta med hjärnans. Möjligtvis skulle man kunna fälla den kritiken att slutsatsen är så gott som innehållslös.

I slutsats tre försöker Searle skärpa detta resultat, dock utan att lyckas. Av förutsättningen att hjärnor orsakar medvetande följer givetvis inte att medvetande inte kan orsakas av något annat än hjärnor.

Den fjärde slutsatsen avhänger den första och är på så sätt inte heller klar. Därtill kan den verka något besynnerlig: om man som Searle inte vill kalla den mänskliga hjärnan för en dator, följer trivialt att den mänskliga hjärnan inte producerar medvetande genom att köra ett datorprogram. Searle visar i denna tolkning en enkel sats på ett onödigt komplicerat vis. Detta är emellertid inget fel utan bara en klumpighet.

### *3. Ytterligare kritik*

Det faktum att Searle noggrant begränsar sin argumentation hindrar honom inte från att sedan uttala sig betydligt mer generellt, och påstå att datorer inte kan åstadkomma medvetande, vilket man väl måste tolka som att det här rör sig om vanliga, fysiska datorer och subjektivt medvetande. Axiom ett är inte giltigt i detta fall, axiom två och fyra avhänger icke-triviala filosofiska ställningstaganden och axiom tre förutsätter helt enkelt det som skall bevisas. Följaktligen misslyckas Searle i sitt formella resonemang. Han försöker emellertid en ytterligare form av argument, i vilket exemplet med det kinesiska rummet spelar en viktig roll.

I rummet finns uppenbarligen, enligt Searle, inte något medvetande som förstår kinesiska, fastän Searle med bokverkets hjälp avger för ett Turingtest acceptabla svar. Denna enkla idé går igen i det tredje axiomet: något så formellt som symbolhantering kan inte rimligtvis orsaka medvetande. Detta anser sig Searle fullkomligt säker på, det är äger vardaglig evidens och betraktas av honom till och med som en logisk sanning. Låt oss nu undersöka denna Searles



intuitiva inställning till vad som kan och inte kan tänkas vara medvetet.

### *Vanligt folk och zombies*

Dennett har angivit ett enkelt argument som visar hur svårt det är att uttala sig om ett medvetandes existens eller icke-existens, (1982, s 174). Låt oss anta, säger Dennett, att det finns två sorters människor. Båda har fungerande hjärnor, vilka genom livet tar emot intryck från omvärlden och på något för oss okänt men fungerande sätt lagrar, behandlar och omvandlar dessa intryck, via interna representationer, till olika aktioner. Det som skiljer de två sorternas människor är att hos den ena gruppen åtföljs hjärnaktiviteten av ett subjektivt medvetande, hos den andra saknas detta. Den ena sortens människor (vanligt folk) har medvetande, den andra (zombies) har det inte, och detta är *den enda* skillnaden mellan dem.

Vi kan, enligt Dennett, inte genom någon sorts observationer över huvud taget skilja en vanlig människa från en zombie; det enda medvetande vi kan observera är vårt eget. Världen kan vara full av zombies utan att vi vet om det eller kan veta om det. Inte ens när det gäller människor kan vi alltså veta vem som har ett subjektivt medvetande, utan måste nöja oss med att ställa upp en hypotes, vilken kan stärkas eller försvagas med stöd av observationer av människans beteende, något som mycket liknar Turingtestets idé. Dennetts enkla och, som jag anser, korrekta analys, visar att vi inte kan veta huruvida medvetande kan uppstå ur ett visst fenomen, i det aktuella fallet symbolhantering. Det bör tilläggas att Searle helt och fullt erkänner medvetandets subjektiva karaktär.

Om vetenskapen försöker beskriva hur världen är kommer en av beskrivningens ingredienser att vara att mentala tillstånd är subjektiva, eftersom det är ett enkelt faktum om den biologiska evolutionen att den har åstadkommit vissa sorters biologiska system, nämligen människors och vissa djurs hjärnor, som har subjektiva egenskaper. Mitt nuvarande medvetandestillstånd är en egenskap hos min hjärna, men dess medvetna egenskaper är tillgängliga för mig på ett sätt som de inte är för dig, (1984, s 25).

För den som accepterar existensen av subjektivt medvetande och att denna subjektivitet innebär att var och en endast kan komma i

kontakt med, observera, sitt eget medvetande, bör den rimliga ståndpunkten vara som följer. Vad avser det *egna* medvetandet är den naiva mentalismen en tillfredsställande teori, men vad gäller *andra* medvetanden måste vi nöja oss med en funktionalistisk ståndpunkt. Vi kan tro att andra är medvetna, men det enda sättet att stödja ett sådant antagande är genom iakttagelser av funktion och yttre beteende.

Searle anser sig av någon anledning kunna avgöra huruvida andra medvetanden existerar utan att förfalla till funktionalistiska betraktelsesätt, samtidigt som han accepterar medvetandets subjektiva karaktär. Då han emellertid inte gör någon anstalt att bevisa möjligheten av detta, skall vi temporärt lämna denna suspekta åsikt åt sitt öde.

#### *Vad åstadkommer medvetande?*

Vi kan ställa upp flera fysiologiska förklaringar till vad det är hos en hjärna som på något sätt producerar eller utgör en fysisk motsvarighet till medvetande. Bland annat kan nämnas följande:

1. Närvaron av materia över huvud taget, (animism).
2. Närvaron av vissa kemiska eller biologiska substanser.
3. Att vissa fysiska, kemiska eller biologiska reaktioner äger rum.
4. Närvaron av elektriska eller kemiska signaler.
5. Närvaron av ett informations- eller symbolflöde (i godtyckligt medium).
6. Närvaron av komplexa kopplingsstrukturer mellan nervcellerna.
7. Närvaron av parallellt pågående processer.
8. Etc.

Dessa fenomen kan, bland många andra och eventuellt i kombination, vara förklaringen till hjärnans förmåga att producera medvetande. Hur det faktiskt ligger till vet vi i dagens läge inte. Searles grundläggande fel är att han försöker bevisa att *en* möjlig förklaring, informationsflödet som sådant, inte duger. Det tredje axiomet är helt enkelt en alternativ formulering av detta ställningstagande, och längre än så når inte Searles argument.

#### *Searles "bevis" för det tredje axiomet*

I debatten har huvudsakligen det tredje axiomet angripits, framför allt i det som brukar kallas "the systems reply." Searle själv förstår

måhända inte kinesiska, men systemet bestående av Searle, rum, lappar och bokverk kan mycket väl tänkas ge upphov till ett annat medvetande, som förstår kinesiska.

Bortsett från tankeexemplets praktiska absurditet instämmer vi helt och fullt i detta försvar. Det innebär ett ifrågasättande av Searles tredje axiom. Detta har länge varit helt utan motivering, men i *Scientific American* försöker Searle faktiskt bevisa dess riktighet.

Searle hävdar att axiomet är logiskt nödvändigt, även då begreppet syntax tolkas som symbolhantering och semantik som medvetandes-innehåll. Detta kan tyckas överraskande, men Searle anger ett "motsägelsebevis".

• Som med alla logiska sanningar ser man enkelt att den är sann, eftersom man får inkonsistenser om man försöker anta motsatsen, (1990, s 25).

Låt oss anta, säger Searle, att det kinesiska rummet verkligen åstadkommer ett medvetande, M1, som är en följd av Searles mekaniska användande av bokverk och lappar. Detta medvetande är ett annat än Searles eget, M2, och förstår kinesiska och den konversation som försiggår. Så långt är allt väl, säger Searle, men antag nu att Searle själv, i sitt eget medvetande, beslutar sig för att tolka tecknen som om de beskrev dragen i ett schackparti. En ytterligare person tolkar i stället, i sitt medvetande, M3, symbolerna som förutsägelser om aktiemarknaden. Searle frågar nu vilken semantik som rummet avger, det vill säga hur tecknen tolkas i rummets medvetande, M1. Denna fråga har enligt Searle inget klart svar, utan här föreligger en logisk motsägelse.

Vilken semantik avger systemet nu? Avger det en kinesisk semantik, en schacksemantik eller båda samtidigt? ... Det finns ingen gräns för antalet semantiska tolkningar som kan ges symbolerna, eftersom dessa, än en gång, är rent formella. De har ingen inre semantik, (1990, s 25).

Först konstaterar vi att det verkar ganska orimligt att, som Searle, tänka sig att längre följder av tecken skulle kunna ges oberoende men konsistenta tolkningar. Låt oss emellertid bortse från detta problem och anta att det rör sig om kortare teckensekvenser.

I det beskrivna fallet föreligger ingen motsägelse och Searles

resonemang är grovt felaktigt. De tre inblandade medvetandena, M1, M2 och M3, har *var sin* tolkning av symbolerna. Searles initiala antagande om att M1 tolkar tecknen som kinesiska är fortfarande lika giltigt (eller ogiltigt). Att Searle själv, M2, tolkar symbolerna på ett sätt och den tredje personen, M3, på ett annat omöjliggör inte att rummet kan ha ett medvetande.

Searle tycks tro att i och med att han i sitt eget medvetande, M2, börjar tolka tecknen på ett visst sätt, denna tolkning skulle emanera ut till rummets medvetande, M1, eller på något sätt ersätta rummets tolkning, men naturligtvis finns det ingen anledning att tro att så måste vara fallet.

Återigen är det anmärkningsvärt att en filosof som Searle gör sig skyldig till så synnerligen grova misstag. Om resonemanget vore giltigt hade Searles egen tolkning av symbolerna omöjliggjort inte bara rummets medvetande, utan även den tredje personens medvetande och alla andra tänkbara, tolkande medvetanden över huvud taget.

#### 4. Sammanfattning

Enligt det ovanstående kan vi konstatera att Searles argumentation lämnar mycket övrigt att önska. Sektion 3 illustrerar att Searle inte kan *veta* huruvida symbolhantering är tillräckligt för att åstadkomma medvetande eller ej. De axiom han anger som evidenta är inget annat än åsikter vars riktighet inte låter sig testas eller undersökas. De tekniska delarna av argumentationen har så stora brister att resonemanget i sin helhet knappast kan räddas.

Först och främst rör Searles argumentation begreppet *medvetande*. Det är på inget sätt klart att resultaten också gäller för *intelligens*. Vidare är det angivna tankeexperimentet inte rimligt verklighetstroget, utan är i alla avseenden sämre, oklarare och mindre intuitivt än att direkt resonera om en dator med program.

Searle förutsätter tyst sin egen medvetandeteori, *naiv mentalism*. Den är inte problemfri; rimligtvis kan Searle inte direkt tillämpa den på andra medvetanden än sitt eget. Omfattar man inte denna teori faller hans argument platt till marken.

I sitt första axiom överdriver Searle ett datorprograms formella egenskaper på ett felaktigt vis. Faktiskt finns det skäl att misstänka att han helt enkelt inte förstått vad ett program är. I det tredje

axiomet förutsätter Searle helt enkelt det som skall bevisas, och det motsägelsebevis han anger som stöd är felaktigt på ett anmärkningsvärt grovt sätt. Vidare använder Searle begreppen syntax och semantik på ett högst oacceptabelt sätt; det är närmast att betrakta som en sorts trick eller argumentationsknep.

Oavsett vilken medvetandeteori man omfattar och vilken tolkning man vill ge Searles axiom och slutsatser är resonemanget felaktigt. Det går alltså inte att motbevisa stark AI på det sätt Searle försöker. Accepterar man en naiv mentalism, följer rimligtvis att varken stark AI eller Searle har rätt. Den starka AIns företrädare kan inte *veta* att ett symbolbehandlande system verkligen besitter subjektivt medvetande, och Searle kan inte *veta* att det inte gör det.

Ända sedan Searle presenterade sina argument om det kinesiska rummet år 1980 har debatten rasat, framför allt inom AI-kretsar i USA. Hela tiden förbises så gott som alla grundläggande teorier och resultat som finns i de flesta grundkurser i teoretisk filosofi. Kontrahenterna anger till exempel sällan eller aldrig sina åsikter om medvetandets problem eller kropp- och själproblemet. Det naturliga vore givetvis att man började med att klargöra sina ställningstaganden vad gäller vanligt, mänskligt medvetande, och först när detta klarats av, kastade sig över det artificiella medvetandets problem. Debatten visar emellertid att inte ens välkända, amerikanska filosofer orkar vara så metodiska.

### Litteratur

- DENNETT, DC, *Brainstorms*, Bradford Books, Montgomery, Vermont, 1978.
- DENNETT, DC, "How to Study Human Consciousness Empirically or Nothing Comes to Mind," in Hintikka, J, (utg.), *Synthese*, Vol 53, 1982.
- SEARLE, JR, "Is the Brain's Mind a Computer Program?" *Scientific American*, Januari, 1990.
- SEARLE, JR, "Minds, Brains, and Programs," in Hofstadter, DR and DENNETT, DC, *The Mind's I*, Bantam Books, Toronto, 1981.
- SEARLE, JR, *Minds, Brains, and Science*, Harvard University Press, Cambridge, Massachusetts, 1984.
- TURING, AM, "Computing Machinery and Intelligence," *Mind*, Vol 59, 1950.

## INGEMAR NORDIN

### *Replik till Lars Bergström*

---

En viktig distinktion i den (nyliberala) rättighetsetiska diskussionen är den mellan positiva och negativa rättigheter. De negativa rättigheterna är sådana som innebär att andra personer, inklusive dem som stiftar och genomdriver lagar, bör avhålla sig från att med tvång eller beslagtagande av inkomster och egendom förhindra människor från att leva sina egna liv. De positiva rättigheterna är tvärt om ett slags anspråk eller krav som människor anses kunna ställa på varandras liv.

I sitt svar till mig i *Filosofisk tidskrift* (1/92) reflekterar Lars Bergström över det faktum att välfärdsstaten i hög grad baseras på att medborgarna tilldelas positiva rättigheter mot att staten beskattar och reglerar oss. Välfärdsstaten är med andra ord en organisation som ofta kommer i konflikt med just de negativa rättigheter som nyliberaler/libertarianer vill slå vakt om. Om, å andra sidan, de negativa rättigheterna skulle respekteras fullt ut så skulle detta medföra svårigheter för alla de sociala och tekniska system som vi vant oss vid att välfärdsstaten tillhandahåller; vägnät, bibliotek, sjukvård, barnteater, universitet, socialbidrag, etc.

Om emellertid personer som jag och andra libertarianer insisterar på att människor har rätt till liv och egendom och vill att denna rätt respekteras, kan vi då få lov att vara ifred? Jo, faktiskt, det verkar inte helt orimligt, menar Lars Bergström, att vi skulle kunna få lov att hoppa av välfärdsstaten och ingå ett frivilligt avtal med "resten av samhället". Men i ett sådant avtal bör då ingå att avhoppare får klara sig helt på egen hand utan allt detta goda som staten ger!

Detta är givetvis en intressant tanke. Och den är faktiskt inte alls så avskräckande som Lars Bergström tycks tro. Allt som behövs för att uppfylla libertarianens blygsamma önskan att få vara ifred är ett litet tillägg till grundlagen, där den enskilde medborgaren tillåts

hoppa av välfärdsstaten. Vi skulle kunna lämna välfärdsstaten på ungefär samma sätt som man idag kan lämna svenska kyrkan.

Välfärdsstaten i Sverige hade därmed i mina ögon genast förvandlats till en privat skyddsorganisation vilken som helst och slaveriet hade förvandlats till en civiliserad marknad. Och alla välfärdsutilitarister fick lov att stanna kvar om de så önskade. Mycket liberalt!

Situationen blir då ungefär denna. Välfärdsstaten låter mig fortsättningsvis vara ifred (sjävklart utan att jag behöver lämna landet – Sverige är inte identiskt med den statliga maktorganisationen) med mitt liv och min privategendom: Inga rättighetskränkande inskränkningar vad beträffar val av näringsverksamhet, sjukvård, utbildning, försäkringar, mat- och dryckesvanor, etc. Inga pålagor eller restriktioner såsom tullar, skatter eller byråkrati. "Staten" har naturligtvis då heller inga som helst positiva förpliktelser gentemot mig – utöver vad som kan bedömas såsom en rimlig kompensation för genomlidna stölder och trakasserier. Allt eventuellt utnyttjande av den välfärdsstatliga service- och varuproduktionen får jag naturligtvis vara beredd att betala för, liksom den får betala för ett eventuellt utnyttjande av min. Och såvida inte organisationen skulle få för sig att helt förbjuda sina enskilda medlemmar (privatpersoner, firmor, etc) att göra avtal eller att handla med mig så har jag en del att erbjuda dem som jag tror kan intressera, och omvänt. Samma förhållande kommer att gälla andra stater och dess medborgare världen över. Därutöver kommer jag förstås att tillsammans med andra eventuella dissidenter ingå i ett civilt samhälle med frivilliga organisationer och en fri marknadsekonomi.

Den negativa bild som Lars Bergström antyder får mig att tänka på en gammal valfilm från 50-talet där sossarna försökte skrämmas med frågan: Hur skulle det gå om inte den goda och givmilda staten fanns och om undersåtarna skulle vara tvungna att klara sig på egen hand? Jo, det hade blivit dyrt minsann. Medelsvensson hade först blivit tvungen att köpa sig en "flygande tunna" (50-talets JAS) till sitt försvar. Bara det hade kostat flera miljoner. Sedan blir han ju tvungen att fixa med privat sjukvård (vad kostar inte en operation?), och skola till barnen, och ... Ja, vem som helst kan ju räkna ut att det vore en ekonomisk omöjlighet att leva utan en stat som tillhandahåller allt sådant där. – Budskapet gick tydligen fram ty något regeringsskifte blev det inte den gången.

Ändå borde rimligtvis varje förnuftig åskådare till filmen efter en stunds eftertanke inse hur fullkomligt absurt hela propagandaspektaklet var: Det finns som bekant inga fria luncher utan allt måste betalas av någon. Och när det gäller den välfärdsstatliga servicen så har jag en stark känsla av att det inte är politikerna personligen som betalar kalaset. Det betyder i klartext att eftersom omfördelnings-effekten i det svenska samhället är ytterst liten så betalas det mesta av just den skattebetalande medelssvensson. Rimligtvis är det då också så att allt som medelssvensson har råd med inom välfärdssystemet via skatten (polis, försvar, sjukvård, skola, motorvägar etc) har han rimligtvis också råd med utanför detta (minst). Men produktionen av dessa varor och tjänster får i det senare fallet givetvis organiseras på ett annat sätt.

Jag vill inte på något sätt hävda att Lars Bergström använder sig av ett så vulgärt och lättgenomskådligt argument som man gjorde i denna valfilm. Men om hans reflektioner kring detta ämne syftar till att påvisa problem för min nyliberala hållning så måste han precisera kritiken. Olika välfärdstjänster kräver olika lösningar inom ett privat system: Sjukvården kräver privata försäkringar och sjukhus, utbildning kräver terminsavgifter och privata skolor, kommunikationer kräver rimligtvis vägavgifter eller färdbiljetter i vanlig ordning, skydd kräver avgifter och privata vaktbolag (typ Securitas), avloppssystem kräver kanske privata abonnemang (eller varför inte mulltoor?), osv. Det är möjligt att skyddsorganisationen Sverige AB skulle vålla vissa problem för oss dissidenter initialt, p g a sin storlek och traditionella bufflighet. Men man får anta att konkurrensen med andra privata alternativ, samt den ständiga risken för ytterligare avhopp bland de kvarvarande medlemmarna, så småningom skulle leda till en någorlunda harmonisk samexistens.

Låt mig så ta upp frågan om rättigheternas ursprung. Om jag förstår Lars Bergströms kritik av mitt argument här, så menar han att det är så tunt att det skulle duga till att bevisa i stort sett vad som helst, t ex att utilitarismen är riktig.

Mitt argument lyder i korthet: Antag att vi inte kan bevisa riktigheten i någon speciell moralisk eller religiös syn på livet (civilmoralen). Om nu någon vill göra gällande att han eller hon utifrån sin civilmoral mot min vilja bör tvinga mig att böja mig för Den Enda Sanna Lärans, så bör denna Lära enligt min mening först



kunna bevisas. Eftersom detta inte går, så bör jag och alla andra som inte tror på den inte heller tvingas att handla i enlighet med den. Alltså bör vi behandlas som om vi hade rätt till oss själva.

Bergström förtydligar sin kritik genom att ge ett, som han hävdar, helt analogt argument för utilitarismen: Om någon vill göra gällande att han eller hon utifrån sin icke-utilitaristiska Lära har rätt att *handla mot utilitarismens krav*, så bör denna Lära först kunna bevisas. Eftersom detta inte går *så bör man handla i enlighet med utilitarismens krav*.

Jag har kursiverat det som främst skiljer argumenten åt. Enligt min mening så är argumenten inte särskilt analoga. Det finns visserligen en rent formell likhet, men det är högst tveksamt om den går på djupet.

Den för argumentets sundhet avgörande skillnaden är rimligheten och de interna sambanden i böra-premisserna. I mitt argument så finns det en koppling mellan å ena sidan kravet på bevisbarhet och kravet på att avhålla sig från tvång å den andra. I mitt förra inlägg så antydde jag faktiskt en argumentation för denna koppling genom att säga: "Våldet, särskilt då det angripande våldet, är förnuftets motsats. Bland människor som gör anspråk på att vara förnuftiga och intellektuellt hederliga så är det ett rimligt minimikrav att allt fysiskt tvång bör kunna visas vara legitimt." Det faktum att vi står utan bevis för vad som utgör ett gott liv och ett moraliskt beteende är således relevant för avståndstagandet från allt tvång. I Bergströms argument däremot finns så vitt jag kan se ingen påtaglig koppling mellan kravet på bevisbarhet och utilitarism. Varför i all världen skulle vi vara tvungna att handla just utilitaristiskt då vi inte med förnuftiga argument kan visa att något annat är korrekt?

Det Lars Bergström inte vill se, eller i vart fall inte berör, är att frågan om tvång eller ej i detta sammanhang är speciell och inte kan behandlas på samma nivå som andra handlingar.

Antag att Lars Bergström befinner sig i en extrem situation där den utilitaristiska kalkylen ger vid handen att han bör tillgripa tvång för att få en motsträvig person – t ex mig – att handla på ett visst sätt (t ex att betala skatt). Jag har inte stulit, eller på något annat sätt varit aggressiv mot Lars Bergström eller andra. Och jag låter mig inte bli övertalad om utilitarismens välsignelser. Antag vidare att Bergström inte har något avgörande bevis för att hans utilitarism är etiskt korrekt. Dvs han vet – i sträng mening – inte huruvida hans egen

övertygelse är riktig. Vad göra? Uppenbarligen duger det inte att under dessa omständigheter hänvisa till utilitarismen för att avgöra frågan. Det vore att gå runt i en cirkel.

Givet den negativa kunskapssituationen så finns det ingen moralisk legitimation för att vara sig tvinga eller för att låta bli – inte utan ytterligare någon värdepremiss. Frågan blir då om det går att hitta någon sådan premiss som är gemensam för Lars Bergström och för mig. Detta skulle ju lösa vår konflikt.

Låt mig börja med frågan: Behövs det någon legitimation överhuvudtaget? Det är sant att hur Bergström än gör i detta läge så handlar han. Det är bara det att det tycks finnas en moraliskt relevant skillnad mellan tvångshandlingen ifråga och avhållsamhetshandlingen. I det första fallet handlar det om att sätta sitt eget moraliska förnuft – eller förmåga – före en annan persons dito. Det gör det inte i det andra fallet. Att påtvinga någon sin subjektiva trosuppfattning, trots att man inte kan bevisa att man har rätt, är kort och gott att handla dogmatiskt ("våldet är förnuftets motsats"). Vad jag menar nu är att om Lars Bergström gör anspråk på intellektuell hederlighet – och det hoppas och utgår jag ifrån att han gör – så bör han undvika att handla just dogmatiskt. Dvs, han bör undvika att tvinga andra moraliska subjekt att handla i enlighet men hans moraliska bedömning då hållbara skäl härför saknas.

Visst, icke-handlingen "görs" också på basis av enbart en subjektiv övertygelse. Men eftersom en sådan handling inte gör våld på någon annans moraliska förnuft så strider det inte mot någon princip om intellektuell hederlighet. Den typen av handling behöver med andra ord ingen strikt moralisk legitimation!

Om, å andra sidan, Lars Bergström mot förmodan inte hyllar principerna för en förnuftig samvaro människor emellan, så duger förstås inte mitt argument ovan. Grunden för det rationella samtalet bryter samman och vi får väl då göra upp med knytnävarna i stället förmodar jag. Men jag har svårt att riktigt tro på denna möjlighet. Lars Bergströms bedömning att det är rimligt att låta mig och andra likasinnade hoppa av från den välfärdsstatliga organisation tyder på att kanske även han i sitt hjärta trots allt accepterar att människors eget moraliska förnuft kommer före hans egen utilitaristiska trosuppfattning. Detta inger hopp om framtiden.

## LARS BERGSTRÖM

### *Slutreplik till Nordin*

---

Ingemar Nordin vill hoppa av välfärdsstaten. Han är nöjd om grundlagen bara ändras så att den tillåter detta. Han anser också att ett sådant avhopp inte medför att han måste lämna landet.

Det sista är inte så självklart. Men låt oss för resonemangets skull anta att Nordin, när han lämnar den svenska staten, ändå tillåts uppehålla sig på den mark han äger i Sverige (om han nu äger någon mark), och att han också juridiskt tillförsäkras denna rätt. Därutöver har han emellertid ingen rätt till de varor, tjänster, naturresurser och investeringar som svenska medborgare har rätt till. Nordin är som sagt nöjd med detta.

Jag undrar om han riktigt inser vad det innebär. Som jag tidigare påpekat har han då ingen rätt att utnyttja svenska vägar, sjukhus, bibliotek, vattenledningar, flygfält, osv. Naturligtvis står det honom fritt att försöka ingå avtal med det svenska samhället (dvs staten, landsting och kommuner), för att få tillgång till något av detta. Vissa varor och tjänster kan han kanske också köpa av privata producenter – men hur skall han alls kunna komma i kontakt med dessa om han inte tillåts utnyttja samhällets resurser som vägar, telenät, osv?

Det är naturligtvis tänkbart att samhället tillåter honom att utnyttja dessa resurser gratis eller för en relativt billig penning, men betydligt troligare är väl att samhället kommer att ta mycket bra betalt för detta. Ty nu har vi ju en marknadssituation, och Nordins förhandlingsposition på denna marknad är uppenbarligen mycket svag. Samhället klarar sig ju utmärkt utan Nordin och det han har att sälja, och samhället har inte heller något intresse av att underlätta situationen för den som på ett så demonstrativt sätt tar avstånd från dess målsättning och ställer sig utanför dess gemenskap. Tvärtom. Det troliga är väl att om samhället alls ingår några avtal med Nordin, så kommer

han att få betala mycket dyrt för detta – betydligt mer än han slipper betala i skatt genom att ställa sig utanför. Detta bl a för att avskräcka andra från att följa Nordins exempel.

Nordin undrar kanske varför samhället skulle vilja avskräcka folk från att hoppa av. Skälet är förstås att majoriteten av samhällets medborgare anser att samhället har en moralisk förpliktelse att bidra till allmän välfärd och att garantera landets medborgare de rättigheter de har enligt FN-stadgan, samt att avhopp i Nordins stil motverkar detta. Att avskräcka folk från att hoppa av kommer därför att uppfattas som moraliskt legitimt.

Men är det då moraliskt legitimt att ta ut stora avgifter av Nordin för att avskräcka folk från att följa hans exempel? Enligt Nordin själv kan man inte veta något om detta, men tydligen har han ingenting att invända så länge han inte utsätts för fysiskt våld eller fysiskt tvång. Här är det ju inte heller fråga om sådant. Snarare är det fråga om en ren marknadstransaktion. Säljaren-samhället kräver ett högt pris av köparen-Nordin för att det skall bli någon affär av. Det är allt.

Låt mig till sist också säga något om en annan fråga som spelar stor roll för Nordin: har vi någonsin moralisk rätt att med våld hindra någon från att handla på ett visst sätt? Nordin tycks svara nej på denna fråga. Enligt min åsikt är svaret ja. Det kan t ex vara moraliskt fullt berättigat att med våld (i lagom dos) hindra en berusad person att köra bil eller en auktoritär man att misshandla sin fru. Detta är ingalunda detsamma som att ”påtvinga någon sin subjektiva trosuppfattning”. Det innebär bara att man väljer det minst onda av två alternativ, det som kan antas leda till minst skada för de inblandade.

Men kanske skulle Nordin ändå gå med på att tvång kan vara berättigat i de fall där det krävs för att större skador skall undvikas? Nej, det kan han inte gärna göra, ty i så fall skulle han resonera precis som en utilitarist, och det vill han inte. Dessutom menar han tydligen att tvång endast är berättigat om man kan ”bevisa” att det är berättigat – samt att detta villkor aldrig är uppfyllt. Alltså kan han inte gärna mena att tvång är berättigat ens för att undvika de mest extrema katastrofer.

En nyliberal som accepterar tvång i vissa lägen skulle däremot (till skillnad från Nordin) kunna hävda att det kan vara rätt att (med tvång) hindra folk från att skada andra, men att det däremot aldrig kan vara rätt att tvinga folk att hjälpa andra. En sådan nyliberal

kunde sedan ta detta till intäkt för att staten inte har rätt att tvinga honom att betala skatt. (Det är ju detta som det till sist är fråga om!)

För resonemangets skull kan vi här anta att man verkligen hjälper andra genom att betala skatt: man bidrar i någon mån till att andras välfärd ökar, delvis kanske genom att en känsla av solidaritet breder ut sig och befästs. Men i så fall skadar man övriga medborgare genom att inte betala skatt – de får det något sämre än de annars skulle ha det, bland annat kanske genom att skattesmitningen bidrar till minskad solidaritet eller till välfärdsstatens sönderfall. Därför skulle vi ändå ha rätt att (med tvång) hindra folk från att undvika att betala skatt. Och det tycks vara detsamma som att tvinga folk att betala skatt.



## RECENSIONER

Jean Grimshaw, *Feminist Philosophers. Women's perspectives on Philosophical Traditions*, London, 1986

---

Jag läser schemat för praktisk och teoretisk filosofi vid Stockholms universitet inför våren 92 och ser, att bland lärarna finns inte en kvinna. Hösten 91 var situationen densamma, förutom i ämnet feministisk filosofi. Detta förhållande är ett gott skäl, tycker jag, att fundera över om kön påverkar vårt tänkande, så att kvinnor inte kan tänka filosofiskt, eller om de inte haft möjlighet att studera filosofi, eller om de helt enkelt tänker annorlunda. Feministisk filosofi har utvecklats som ett försök till svar på dessa frågor, och har idag fått visst utrymme på de akademiska institutionerna.

En bra introduktion i ämnet ger Jean Grimshaws bok *Feminist Philosophers. Womens perspectives on Philosophical Traditions*. (London, 1986.) Som filosof av facket har hon studerat några anglosaxiska feministerna. Dels för att skärskåda deras oenigheter, vilka tycks ha uppstått ur svårlösta motsättningar i feminismens grundvalar. Dels för att granska den kritik som framförts gentemot filosofins historia, att den präglats av ett specifikt manligt perspektiv.

### *Den manliga filosofin*

En del av de manliga filosofernas fördomsfulla yttranden om kvinnor kan förstås ses som marginella företeelser. Vad Bertrand Russell haft för åsikter om äktenskapet bör inte påverka sanningshalten av hans kunskapsteori, säger Grimshaw. Men i vissa fall har dock kvinnosynen betydelse för hela det filosofiska bygget. T ex hos Aristoteles, som helt uteslöt kvinnor. Eller hos Locke, vars äganderättsteori grundas på att en fri man skall äga det han producerat, vilket också förutsätter att kvinnor genom äktenskap underställs och är beroende av mannen. Ty på så sätt får han också kontroll över reproduktionen, dvs vem som skall ärva hans egendom.

Men feministerna menar ofta mer än så i sin kritik av filosofin, nämligen att det manliga sättet att tänka är livsfientligt och att det i sin förlängning leder till militarism och krig. En av dem, Jane Flax, utgår från Nancy Chodorows teori om manlig resp kvinnlig identitetsutveckling. Chodorow spårar en skillnad mellan könen i det lilla barnets relation till modern. Pojkar separerar från den allsmäktiga modern mer definitivt än flickor, eftersom de har en könsolikhet att bygga sin sär-skildhet på. Med en otydlig bild av den mestadels frånvarande

fadern, finner pojkar sin identitet i en defensiv motsatsställning gentemot modern. Flickor däremot identifierar sig med moderns kön och 'ärver' på så sätt förmåga till inlevelse, empati och omsorg, karaktärsdrag i modersrelationen.

Jane Flax 'psykolanalyserar' filosofihistorien enligt Chodorows modell. Hon tycker sig se den lilla pojkens sätt att handskas med sina ambivalenta känslor till modern i de vuxna filosofernas behov av kontroll över det de mest fruktar och längtar till: kroppslighet, känslor, sexualitet, kvinnan – representerad av modern. Män tänker därför defensivt och fientligt, betonar individens ensamhet och oberoende, och ställer förnuft och abstrakt tänkande i opposition mot kropp och känsla, menar Flax och andra feminister.

Grimshaw är skeptisk till dessa generaliseringar om manligt resp kvinnligt tänkande. Flax uttalar sig om filosofin genom historien, där psykologiska och sociala villkor varierat stort, medan Chodorows teori utgår från nutida uppväxtförhållanden. Att inringa ett specifikt manligt sätt att filosofera förbigår dessutom det faktum att det alltid finns någon filosof, som säger emot vad en annan sagt. Men på *ett* sätt kan det sägas att hela filosofins historia tänker "manligt", medger Grimshaw. Det "kvinnliga", hur det än definierats, har nästan alltid nedvärderats. Om man sett det som "equal but different", har man ändå

i praktiken inrangerat det "kvinnliga" på en lägre nivå än det "manliga". Mannen har varit normen för mänskligheten, medan kvinnan utgjort ett sämre alternativ. Framförallt har de områden och aktiviteter som traditionellt varit kvinnornas inte ansetts särskilt angelägna. Och det är bland annat här, i en omvärdering av vilka frågor som är viktiga, som feministisk filosofi är värd att lyssna till.

### *Teorier och språk*

Som protest mot manliga forskares teorier och 'förklaringar' av kvinnors erfarenheter (t ex manliga läkare som föreskrivit hur kvinnor bör föda sina barn eller förklarat hur deras sexualitet är konstruerad etc) har man från feministiskt håll velat betona den unika, personliga erfarenheten. Emellertid begår den feministiska vetenskapskvinnan samma misstag som de manliga teoretikerna om hon tror sig kunna säga något om en generell kvinnlig erfarenhet. Grimshaw är som en hök på varje sådan generalisering av manligt resp kvinnligt. Man kan inte komma ifrån, att varje erfarenhet är individuellt upplevd. Och söker man övergripande teorier om kvinnlig erfarenhet, riskerar man att förbise kulturella, etniska, klassmässiga och andra skillnader mellan kvinnor.

Den feministiska kritiken har också påpekat, att språket är genomsyrat av manlig terminologi och manligt tänkesätt samt att kvinnlig



erfarenhet saknar akdekвата begrepp. Dessutom vimlar det av pejorativa begrepp för kvinnlighet, vilket skapar nedvärderande attityder mot kvinnor. Grimshaw instämmer, men frågan för henne är, om man inte kan använda ord i överförd mening och omdefiniera gamla uttryck. Kan språket inte reformeras efter våra syften?

Grimshaw är mycket optimistisk vad gäller dessa möjligheter. T ex kan man använda uttryck från gemensamma erfarenhetsområden, som att ligga på sjukhus, för att beskriva hur det känns att ligga på BB och föda barn. Troligen menar inte Grimshaw att man med hennes exempel skall kunna förmedla upplevelsen av att föda barn. Men ser hon de verkliga svårigheterna när hon väljer ett så fatalt exempel som barnafödande? Hon säger "it is not impossible in principle", och så långt har hon kanske rätt. Men jag tror att det är *mycket* svårt att sätta sig in i en så komplex erfarenhet som att föda barn. "In principle" må det vara möjligt, men hur ofta är det inte omöjligt i praktiken! Det finns all anledning att understryka språkets och vetenskapens tillkortakommanden, något den feministiska kritiken påminner oss om.

De feminister som verkligen dragit konsekvenserna av det ofrånkomliga gapet mellan erfarenhet och språk, har dock hamnat i det relativistiska (postmodernistiska) läget. Kvinnors perspektiv kan på så sätt uppvärderas och bli lika

'sant' som ett manligt. Men en sådan subjektivism blir ytterst tandlös och förmår inte verka för reella förändringar i kvinnors liv, påpekar Grimshaw. Allt blir som sagt lika 'sant', även att säga att kvinnor bör förtryckas. Denna uppgivenhet rimmar illa med den argumentationsglöd som de flesta feminister uppvisar när sådana uttalanden gjorts. Grimshaw visar i sin bok, att här har feminismen hamnat i ett gammalt filosofiskt dilemma.

### *Självförverkligande eller nära relationer*

Kvinnor har som bekant svårt att hävda sina egna intressen. För många kvinnor har därför feminismen kunnat erbjuda stöd för rätten till ett mer självständigt liv. Detta är nog den vanligaste uppfattningen om feminister, att de förmår lämna man och barn utan skuldkänslor, gå ut i arbetslivet och göra vad de själva har lust till, utan att behöva tänka på andra.

Många kvinnor (feminister såväl som anti-feminister) har skrämts av denna individualism och de militanta kraven på eget självförverkligande. Vem skall nu ta hand om barnen, de sjuka, de gamla, de svaga? Det manliga svaret har oftast varit att kvinnorna skall göra det. De feminister som betonat kvinnors självbestämmande och autonomi, som exempelvis Shulamith Firestone, Mary Daly och Simone de Beauvoir, har inte riktigt kunnat ge ett bra alternativt svar. En del tycks

tro, att om alla människor, kvinnor som män, blir "fria" och "förverkligar sig", så kommer alla att leva i harmoni med varandra och behoven automatiskt bli tillfredsställda. Det är en mycket naiv och blåögd föreställning, menar Grimshaw.

Andra feminister anser att kvinnor, på grund av sina specifika erfarenheter såsom mödrar, döttrar, och vårdarinnor, själva företräder ett sådant alternativ till individualismen. Återigen har man använt sig av Chodorows psykologiska teori om att kvinnor förmår "leva sig in i" och "känna med" en annan människa i högre grad än män. Naomi Scheman hävdar t o m att kvinnor har svagare jagkänsla och att kvinnor därför är mer altruistiska än män. Grimshaw uppmärksammar här en oklarhet vad gäller förutläggningarna för empati, förståelse etc, som vidhäftar både Chodorows och framför allt Schemans resonemang. Man kan inte, som Scheman, sammanblanda jag-svaghet med omsorg och förståelse för andra. Verklig förmåga att förstå och hjälpa en annan människa förutsätter tvärtom en viss distans, till sig själv och andra, för att kunna urskilja vad som är projektioner av det egna jaget och vad som är den andres verklighet.

Denna föreställning om kvinnlig hjälpsamhet och förståelse för andra kan också bli ett hinder för kvinnlig frigörelse. Det finns kvinnor som känner sig så 'nära' sina män, att de helt ger upp egna livsmål. Samma

fenomen kan uppstå i kvinnogrupper, där man slutit sig samman för att utveckla och kämpa för egna (feministiska) intressen. För att inte förråda sin "kvinnlighet" förnekas konkurrens och aggressivitet och individualiteten kommer inte över tröskeln. I stället uppstår ett nytt tvång att uppträda inlevelsefullt och känslösamt. Grimshaw vill därmed inte försvara individualism eller psykologisk egoism – tvärtom. Men en renodlad altruism är inte ett oproblemiskt alternativ. Speciellt inte för kvinnor, som ännu inte lyckats göra sin individuella stämma hörd.

#### *En kvinnlig etik*

Genom nästan hela historien har kvinnor varit uteslutna från den 'stora', offentliga världen och inneslutna i den 'lilla', privata. Deras uppgift har varit att ta hand om livets dagliga omsorger, om hem och barn, etc, och de har därmed utvecklat ett annat förhållningssätt till sin omgivning än män. De har ansvarat för föda och klädsel, ibland även försörjningen, och de har ofta lärt sig hur man åstadkommer en harmonisk atmosfär i hemmet, med goda, livsbefrämjande relationer för allas välbefinnande. Detta kvinnliga "omsorgsperspektiv" vill den kvinnliga etiken lyfta fram, som alternativ till den gängse moralen i den offentliga världen bland konkurrensinriktade och aggressionsstyrda män.

Företrädarna för en kvinnlig etik vill i allmänhet hävda, att kvinnor

resonerar annorlunda än män i moraliska frågor; att män tänker mer abstrakt och principiellt, medan kvinnor ser mer till sammanhanget och den unika situationen. En av de empiriska studier som brukar åberopas är Carol Gilligans undersökning av barns sätt att resonera i moraliska dilemman. Den visade, att fler pojkar än flickor kunde uppställa principer för sitt resonemang. Är flickor därför mindre moraliskt mogna än pojkar i samma ålder (som Kohlberg, Freud m fl menar)? Nej, säger Gilligans undersökning. I stället för att lita till principer, utgår flickor från att de inblandade är beroende av varandra för sin existens. De försöker därför finna lösningar som ökar förståelsen och kommunikationen *mellan* dem som är inblandade.

Grimshaw vill emellertid inte gå med på att kvinnor tänker principlöst i moraliska frågor. De har emellertid ofta *andra* principer än män, menar hon. En sådan kvinnlig princip är att prioritera goda relationer med andra. Och eftersom mänskliga relationer alltid är komplexa och mångfasetterade, så kan det kanske krävas större insikt i den konkreta situationen, känedom om individernas bakgrund etc, för att våga fälla ett omdöme.

Grimshaw är inte negativ till möjligheten att i moralfilosofin tillföra ett sk kvinnligt perspektiv, men ser vissa problem och svårigheter i detta. En kvinnlig etik riskerar att idealisera kvinnan som god

och dygdig, medan mannen anses ond och krigisk. Som om kvinnan opåverkad skulle kunna frälsa världen med sitt kvinnliga tänkande. Som om inte män är offer för en kapitalistisk ekonomi, där konkurrens och makt premieras.

Här ligger det en unken, traditionell uppdelning mellan manligt och kvinnligt, som motarbetar den som vill reformera traditionella föreställningar om könen, påpekar Grimshaw. Antifeministerna utnyttjar den, när de anvisar kvinnan rollen som "ängeln i huset". Med en så god och ljus karaktär är kvinnan inte lämpad för en hård och tuff manlig offentlighet, menar de. Än en gång varnar Grimshaw för sådana generaliseringar, som så lätt spelar i händerna på traditionalisterna. Det finns både kvinnor och män som inte passar in i mönstret. Hon menar, att man i stället för denna opposition mellan könen bör se det som en skillnad mellan privat och offentlig moral.

Men i den privata världen är relationer långvariga och livsnödvändiga, och omsorgsperspektivet relativt självklart. I den 'stora' världen däremot, där människors möten är flyktiga, måste omsorgsetiken transformeras på något sätt, säger Grimshaw. En allvarlig svaghet med omsorgsperspektivet är att det inte kan användas där två intressen står emot varandra. Särskilt olämpligt blir det förstas när en kvinnas önskan att t ex börja

arbeta, står i konflikt med hennes mans eller barns önskan att ha henne hemma dygnet runt. Omsorgsperspektivet kan faktiskt användas som täckmantel för en hel rad olika syften. Man har från abortmotståndarhåll använt omsorgsargumentet mot kvinnor som väljer abort – man säger att kvinnan försummar omsorgen om det mänskliga (potentiella) liv som tagit plats i hennes livmoder. Likaså kan det användas i fackliga och politiska tvister, där en grups krav på exempelvis högre lön eller bättre arbetsförhållanden stävjas med motiveringen att man bör se till det allmännas bästa. I reklamen känner vi till hur man talar om "omsorg" om kundens bästa, när man egentligen bara är ute efter att tjäna pengar.

Grimshaw tar upp dessa och andra exempel och visar att omsorgsperspektivet inte är tillämpligt i en lång rad av moraliska konflikter, att det kan användas *mot* kvinnors anspråk på makt och inflytande (de visar inte tillräcklig omsorg om andras intressen), och att det kan missbrukas för att dölja orättvisor och förtryck. Ändå anser hon att omsorgsperspektivet i praktiken bör beaktas i en rad sammanhang, utan att vara det enda riktmärket för mänskligt handlande. Hon menar, om jag förstår henne rätt, att omsorgsperspektivet tar hänsyn till människors upplevelser i högre grad än en moral som syftar till att uppnå goda resultat.

Grimshaw ger tyvärr bara ex-

empel från sjukvården, vilket gör det svårt att förstå hur hon tänkt sig omsorgsperspektivet praktiserat i andra sammanhang. Sjukvårdens själva syfte är ju redan att sörja för patienterna bästa och sammanfaller nästan med omsorgsperspektivet. (Därför är majoriteten av dem som jobbar på sjukhus kvinnor!) Att det sjukhus där hennes pappa blivit ovärdigt behandlad borde bättra sig är ju självklart, och är kanske heller inte så svårt att åstadkomma. Men i många andra moraliska dilemman (även inom sjukvården), vilket Grimshaw själv påpekar, står en människas välbefinnande i konflikt med en annans. I sådana fall tycks ett omsorgsperspektiv inte ha så mycket att tillföra, mer än att *det är just det perspektivet – en önskan att minska mänskligt lidande – som är bakgrunden till det moraliska dilemmats uppkomst*. I Grimshaws exempel, med hennes stackars pappa som blir ovärdigt behandlad (som om han vore ett spädbarn eller 'grönsak') när han ligger på sjukhus för infarkt, skulle man kunna säga att sjukhusets "omsorgsperspektiv" endast ägnade sina omsorgsfulla åtgärder åt hans kropp, medan ett vidgat omsorgsperspektiv, som ser till *hela* människan, upptäcker hans 'sjäsliga' lidande och motiverar därför en annan sjukhusrutin.

Iris Murdoch har, utan att vara feminist, influerat flera feministiska moralfilosofer, särskilt dem som vill ställa moderligheten och moderskapet som modell för kvinnlig etik.

I motsats till den förhärskande föreställningen inom moralfilosofin sedan Kant menar Murdoch, att genom att vända sin totala *uppmärksamhet* mot omvärlden uppdragas det goda och det rätta. Med "a just and loving gaze directed upon an individual reality", ödmjukt och "self-less", dvs utan att lägga in egna fantasier eller önskemål, kommer det rätta och det goda att stå självklart inför en. Människan är fri så till vida att hon kan välja att lyda eller inte lyda "verklighetens" krav. Men den avgörande normativa instansen ligger så inte inom individen, utan i verkligheten själv. Sara Ruddick menar, att Murdoch här tangerar den etik mödrar handlar efter. Ser de sitt barn lida eller behöva dem på något sätt, följer deras handlande automatiskt.

Grimshaw instämmer med Murdoch om att fakta och värderingar inte kan skiljas åt. Men hon invänder: Vi vet alla att det ofta inte räcker med ödmjuk uppmärksamhet för att handla rätt. Inte sällan är vår blick ideologiskt eller politiskt färgad utan att vi vet om det. Det kan krävas mer generell kunskap och överblick för att genomskåda en förtryckande ideologi, som påverkar mitt eget handlande såväl som någon annans. Att enbart hänvisa till bättre uppmärksamhet är otillräckligt och i många fall politiskt enfaldigt, om inte rentav mystiskt.

Inte heller hur en "god moder" bör uppträda är självklart, säger

Grimshaw. Visserligen kan man utifrån empiriska studier säga en hel del om vad ett barn behöver, men det är också mycket omtvistat hur en god moder bör handla – av mödrarna själva, av dottrarna och av experterna. Ja, hela moderskapet är indraget i ett ideologiskt spänningsfält, betingat av befintliga termer och föreställningar, vilka indirekt uttrycker värderingar av såväl köns- som klassmässig art, menar Grimshaw.

Det är bl a sådana värderingar, tidigare självklara eller ignorerade, som feministisk filosofi drar fram i ljuset och granskar. Grimshaws bok visar att dessa rymmer problem och motsättningar, vilka belyser filosofins gamla frågor på ett delvis nytt sätt. Feministisk filosofi angår därför inte bara kvinnor, utan alla som är intresserade av filosofi. Men frågan varför kvinnliga filosofer är så få vid Stockholms universitet är nog mindre av filosofisk art, än en viktig kvinnopolitisk angelägenhet.

*Lena Forssén*

Jörgen Hermansson, *Spelteorins nytta – Om rationalitet i vetenskap och politik*, Uppsala 1990, Acta Universitatis Upsaliensis, Skrifter utgivna av Statsvetenskapliga föreningen i Uppsala, 109, 385 sid.

Spelteori är den del av beslutsteorin som studerar sk *strategiska beslutssituationer*, dvs situationer där utfallet av vad en aktör gör beror på vad andra aktörer gör. Alla aktörer

i ett sådant "spel" antas vara sådana att de i varje situation har vissa bestämda målsättningar/preferenser. De antas vidare vara rationella i så måtto att de gör sitt bästa för att maximera tillfredsställelsen av de egna preferenserna. Härvidlag kan man skilja dels mellan spel där aktörerna genom olika former av hot har möjlighet att ingå "bindande" överenskommelser och sådana där denna möjlighet inte erkänns (sk icke-kooperativa spel), dels mellan spel där aktörerna känner till allt om spelet och sådana där de saknar information om de andra aktörernas preferenser (sk spel med ofullständig information).

Statsvetare och politiska teoretiker har under de senaste trettio åren i ökande utsträckning sökt använda sig av spelteoretiska modeller. Försök har gjorts dels att ge ny precision åt äldre politiska idéer (t ex Marx historiesyn eller Hobbes idéer om samhällsfördraget), dels att förklara olika empiriska politiska fenomen (vanliga exempel är kapprustningen mellan USA och Sovjet respektive Kuba-krisen). I *Spelteorins nytta* genomför statsvetaren Jörgen Hermanson (hädanefter JH) en mycket ambitiös diskussion om i vad mån dessa ansatser har fog för sig (s 13). Mer precist gäller frågan om spelteorin kan hjälpa statsvetaren att antingen konstruera hållbara *förklaringar* till förekomsten av olika politiska fenomen, eller *rättfärdiga* existerande eller tänkta politiska ordningar (s 17).

"Framställningen bör", enligt JH, "uppfattas som ett argument för spelteorins nytta, men däri ligger också ett försök att ange vissa av dess begränsningar" (s 39). Vid sidan av detta har JH också velat att hans bok ska kunna fungera som en svenskspråkig lärobok i spelteori.

Tyngdpunkten i JHs framställning ligger på de icke-kooperativa spelen. Dessa behandlas i kapitlen 2-7. Det första av dessa kapitel behandlar spel med ofullständig information, medan det i de övriga förutsätts att aktörerna känner till allt om spelet. I kapitlen 8 och 9 diskuteras kooperativ spelteori och några av dess tillämpningar inom politisk teori och filosofi. Avslutningskapitlet ägnas åt en summering av de tidigare slutsatserna samt en kortfattad och ganska svårgenomskådlig diskussion av förhållandet mellan rationalitet och moral, som jag upplever som ganska malplacerad. JH hävdar tidigare i boken att han här ska föra fram tesen att "moral förutsätter rationalitet" (s 15, not 8). Tyvärr blir det dock aldrig klart vari denna tes består och på vilket sätt den är av intresse i sammanhanget.

En central diskussion i boken rör huruvida spelteorin kan tillhandahålla bestämda besked om resultatet av alla typer av icke-kooperativa spel – om det för alla sådana spel finns en unik *lösning*. JH har ansträngt sig att föra denna diskussion utan, som han själv uttrycker det, "alltför mycket mate-

matisk sofistikaion". Samtidigt som det mesta som sägs är referat av andra författare (Barry, Elster, Brams, Riker m fl) presenteras här på ett förhållandevis lättillgängligt vis olika spel-typer och centrala begrepp som dominans, jämvikt och pareto-optimalitet. Såvitt jag kan bedöma är JHs framställning korrekt, även om det ibland kan vara svårt att uppnå en önskvärd grad av precision med den jämförelsevis vardagliga prosa han valt. Inte minst är det tillfredsställande att läsa en statsvetenskapligt vinklad framställning av spelteori, där det klart framgår att de olika försök som gjorts att undvika resultatet i det välkända spel som benämns *fångarnas dilemma*, enbart innebär att spelet förändras så att det inte längre är ett fångarnas dilemma.

JHs slutsats är att det lösningsbegrepp som kan abstraheras ur behandlingen inte är vare sig så tydligt eller så heltäckande som man skulle kunna önska. Det existerar en hel rad av spel för vilka det inte är möjligt att peka ut någon unik lösning och för en del av de spel som har en lösning kan det vara oklart vilken den är. (ss 348 ff.) Denna slutsats är också ett av JHs skäl *mot* spelteorins nytta. Utan ett heltäckande och entydigt lösningsbegrepp kan inte spelteorin göra anspråk på vare sig rollen som *den* adekvata förklaringsmodellen för politiska skeenden, eller som den form i vilken fruktbara politisk-filosofiska argument och teorier bör stöpas.

"Idén om rationellt agerande aktörer kan", menar JH, "därmed aldrig bli generellt tillämpbar" (s 351). Även om det uttalandet förefaller en aning tillspetsat – vem vet? kanske ett såväl heltäckande som okontroversiellt lösningsbegrepp utvecklas i framtiden – så är det en viktig slutsats vad gäller de tillämpningar av spelteori som idag görs inom såväl statsvetenskapen som andra discipliner.

En viktig del i denna slutsats utgörs av tesen att spel-situationer där aktörerna saknar information om varandras preferenser har en unik lösning endast om samtliga aktörer kan utföra någon strategi som är *dominerande* (dvs den bästa strategin oavsett hur de övriga aktörerna betar sig). Därmed avvisar JH det lösningsförslag för denna typ av spel som innebär att aktörerna väljer sin *försiktiga strategi* (eller maximerar minimum respektive sin säkerhetsnivå, som det ibland också kallas). Denna strategi är den vars sämsta utfall inte är sämre än andra strategiers sämsta utfall. Fördelen med detta lösningsbegrepp (vilket kan benämnas FS) är dels att det inte kommer i konflikt med kravet att en rationell aktör alltid väljer sin dominerande strategi (då han saknar information om andra aktörers preferenser), dels att det är tillämpbart på alla typer av spel med ofullständig information.

JH anför två argument mot FS. Det första argumentet illustreras med hjälp av ett spel kallat *Chicken*:

Denna sk spelmatris illustrerar en situation där två aktörer, X och Y, har att välja mellan två strategier vardera. Det vi frågar oss är vilket val som är rationellt för X. Därför återges endast Xs nyttovärden (Ys nyttovärden är ju okända för X och kan därför inte tjäna som vägledning för honom). Vilken strategi som faktiskt skänker X mest preferenstillfredsställelse beror förstås på vad Y väljer. Men, hävdar JH, "[eftersom X inte har någon information som kan vara vägledande... [dvs han vet ingenting om vilken strategi Y väljer], så är det ena valet *lika välgrundat* som det andra. Slutsatsen blir att det därmed inte finns något tillämpligt kriterium för rationalitet i en beslutssituation av detta slag" (s 53, min kursivering).

Vad menar JH med "lika välgrundat"? En närliggande tolkning är att han underförstått hänvisar till den sk *principen om otillräckliga skäl* och utifrån denna hävdar att det är *lika sannolikt* att Y gör  $y_1$  som att Y gör  $y_2$ . Men i så fall borde han ha pekat på att detta ger underlag för X att beräkna den förväntade nyttan för sina olika strategier. I just *detta* spel är visserligen denna förväntade nytta lika för Xs båda strategier (2,5 vardera). Men om vi tillåter X att göra en mer nyanserad värdering av utfallen (vilket JH antyder är möjligt på s 71), t ex så att 4 i figuren ovan byts ut mot 5, så kan principen om att maximera den förväntade nyttan peka ut en unik strategi som *den*

rationella (med min modifikation blir det  $x_2$ ).

Men är det lika välgrundat att tro att Y gör  $y_1$  som att Y gör  $y_2$ ? Är det inte snarare så att det inte är välgrundat alls att tro vare sig det ena eller andra? Om det är detta som är Xs strategiska förutsättning kan jag inte se att det vore uppenbart irrationellt av X att gå in för sin försiktiga strategi ( $x_1$ ). Just *därför* att X inte har en aning om vad Y tar sig till förefaller det mig kunna vara en högst rimlig strategi att i första hand söka undvika det sämsta utfallet. Samtidigt bör det påpekas att *hur* rimligt detta är förefaller bero lite på hur mycket aktören ifråga har att vinna/förlora. Om det värsta utfallet inte är *så* dåligt och det bästa är väldigt bra kan det kanske vara värt att chansa på att kamma hem full pott. Om vi t ex förändrar JHs *Chicken* på så vis att 4 byts ut mot 10 000, 3 mot 1000, 2 mot 950 och 1 mot 900, så förefaller det inte uppenbart irrationellt att gå in för  $x_2$ . I stället för att maximera minimum verkar det således kunna vara rationellt att maximera *maximum*.

JHs andra argument mot FS är att det inte går att definiera något motsvarande *jämviktsbegrepp* (s 54.) I många fall är nämligen det utfall som blir resultatet av FS sådant att det finns ett annat möjligt utfall som är en *pareto-förbättring* jämfört med det uppnådda utfallet (att ett utfall är en *pareto-förbättring* av ett annat innebär att det är bättre för



någon av aktörerna och inte sämre för någon). Endast om det utfall som blir resultatet av FS också är sådant att det inte är möjligt att utifrån det göra pareto-förbättringar, så kan JH gå med på att FS pekar ut en sorts jämviktspunkt (s 55). Frågan är ju emellertid om begrepp som jämvikt och pareto-optimalitet alls är relevanta i detta sammanhang. Visserligen kan en FS-lösning ibland lämna utrymme för pareto-förbättringar, men samtidigt känner ju inte *aktörerna* till detta. Även om vi kan tala om jämvikt och pareto-optimalitet är det således högst tveksamt att med hjälp av dessa begrepp bedöma vad som är rationellt för olika aktörer utifrån deras strategiska förutsättningar i situationer där de inte känner varandras preferenser och därför inte kan agera utifrån det faktum att det i spelet finns någon typ av jämviktspunkt.

En andra faktor som minskar spelteorins nytta rör, enligt JH, tillämpningen av olika lösningsbegrepp (bl a JHs eget) då antalet aktörer är fler än två-tre men samtidigt inte väldigt många. Invändningen är att de stora tekniska svårigheter detta för med sig "kraftigt reducerar spelteorins användbarhet som analysinstrument" (s 351). Vad som avses med detta är emellertid inte helt klart. Att tillämpningen av något lösningsbegrepp på t ex spel med sex aktörer är tekniskt komplicerad visar ju inte i sig självt vare sig att lösningsbegreppet har svag-

heter eller att fruktbara tillämpningar av spelteoretiska resonemang på statsvetenskapliga problemställningar (såväl empiriska som normativa) är omöjliga. Möjligen vill JH antyda att det ligger något absurt i att det skulle krävas såpass avancerade teknikaliteter (t ex avancerade dataprogram) för att statsvetare ska kunna besvara många av sina frågeställningar.

Ett tredje problem rör, enligt JH, sådana situationer där aktörerna kan sägas vara inbegripna i flera spel samtidigt: dels ett "vanligt" spel om vilken strategi som skall väljas, dels ett spel på "metanivå" där aktörerna överväger strategier som kan förändra det "vanliga" spelets förutsättningar. JHs exempel rör en situation där aktörerna, vid sidan av de olika strategierna i någon "sakfråga" (t ex vilka förslag som i riksdagen ska läggas fram om det svenska kärnavfallets slutförvaring), också överväger huruvida de ska söka ytterligare relevant information, för att på så vis förbättra sitt beslutsunderlag. Problemet är, enligt JH, att den senare frågan inte kan besvaras utifrån det lösningsbegrepp han fört fram: "För att kunna besvaras kräver nämligen frågan att man redan vet vad slags information man kan få genom att söka vidare, och i sådant fall upphör själva grunden för den ursprungliga frågan" (s 352). Exemplet gäller en situation där aktörerna har kunskaper, om än osäkra sådana. Utifrån dessa kan en viss

strategi (t ex att föreslå förvaring av kärnavfallet i urberget) framstå som rationell. Men hindrar verkligen detta att det kan vara rationellt att förfina eller inte förfina kunskaperna? Antag att man kan visa att ingen förfining av tillgängliga sannolikhetsuppskattningar skulle påverka det ursprungliga spelets förutsättningar så att vi fick ett nytt spel med samma strategier men vars lösning var bättre för var och en av aktörerna än det första spelets lösning. I sådana fall kan jag inte se att det vore uppenbart absurt att förorda underlåtet av ytterligare informationsinhämtning. (Besluts-teoretiska idéer i denna riktning har presenterats av Per-Erik Malmnäs, "Numeriskt oprecisa sannolikheter och ändliga beslutssituationer", i Christian Munthe & Folke Tersman (red), *Osäkerhet och beslut. Rapport från ett seminarium om beslut under osäkerhet i anknytning till kärnavfallsfrågan*, SKN Rapport 45, Stockholm 1991: Allmänna Förlaget.)

Ett viktigt hinder för spelteoriens nytta som förklaringsmodell sammanhänger med svårigheten att tillämpa spelteoretiska modeller på verkliga aktörer. JHs egen diskussion av hur USAs respektive Sovjets agerande under Kuba-krisen kan förklaras utifrån spelteori (ss 43 ff) illustrerar här väl vilken typ av spekulationer och förenklingar som ofta krävs för att det ska vara möjligt att beskriva verkliga händelser som resultatet av ett spel

mellan rationella aktörer. Detta för i sin tur med sig att de prediktioner och förklaringar som görs utifrån tillämpningar av spelteori på den politiska verkligheten inte sällan bygger på falska antaganden och/eller leder till felaktiga resultat. JH exemplifierar detta förhållande med Downs "medianväljartelem" (ss 93 ff.). JH vill i detta sammanhang göra poängen att spelteori som sådan inte drabbas. Spelteori är en formell teori och bör, enligt JH, "kunskapsteoretiskt sett likställas med begrepp, dvs som verktyg konstruerade av forskaren i syfte att formulera empiriskt prövbara påståenden om verkligheten..." (s 99). Spelteori gör därmed inte "i sig själv [...] anspråk på att påstå något om verkligheten. Det är då över huvudet taget inte relevant att ställa frågan om den kan falsifieras" (ss 99 f.).

I mina ögon är detta resonemang en aning förenklat. När spelteori används för förklaringar och prediktioner är den mycket riktigt en formell del av en större teori som också innehåller diverse empiriska hypoteser. Av detta följer dock inte att vi, i de fall dessa hypoteser inte "stämmer", inte kan ha anledning att göra förändringar av den formella delen av teorin. Den vetenskaps- och kunskapsteoretiska diskussion som förts i Quines och Kuhns efterföljd visar ju i själva verket att en vetenskapsman som finner att hans teori inte stämmer med empiriska data alltid kan välja mellan att antingen formulera nya hypoteser

med hjälp av de begrepp han tidigare använt, *eller* att formulera nya hypoteser genom att förändra den tidigare begreppsbildningen. När JH insisterar på att "rågången bör göras tydlig mellan å ena sidan begreppsbildning och å andra sidan empiriska resonemang eller normativ argumentering" (s 99) bortser han således ifrån att detta kanske inte alls är möjligt. JH försöker värja sig mot detta genom att påstå att förändringar av spelteorin utifrån empiriska motexempel "vore att jämföra med att de empiriska beläggen mot Newtons fysik även skulle drabba det system av algebraiska ekvationer varpå denna fysik grundar sig" (not 45, s 100). Ingen som kritiserat spelteorin utifrån empiriska observationer har väl emellertid hävdad att detta drabbar dess matematiska förutsättningar. En mer passande analogi är att beläggen mot Newtons fysik kan ge oss anledning att omformulera vissa av dess definitioner, vilket inte förefaller uppenbart orimligt.

Men även om spelteorin "som sådan" inte skulle drabbas av att verkliga strategiska beslutssituationer inte är beskaffade på ett sätt som gör att de kan beskrivas som spel, så kan man fråga sig vad detta innebär för spelteorins fruktbarhet inom empirisk forskning. JHs egen inställning förefaller här vara en aning kluven. Samtidigt som han erkänner spelteorins bristande realism hävdar han t ex att det finns "många verkliga spel som befinner

sig ganska nära 'idealet'" (s 102). Exemplet som åberopas är voteringsringar i parlament (ss 102 f). I dessa fall kan, enligt JH, aktörerna betraktas som mycket välinformerade om hur andra aktörer rangordnar de olika möjliga utfallen. Vad JH glömmer är emellertid att ställa sig frågan om det verkligen finns några sådana rangordningar att vara informerad om! Eftersom aktörerna inte är säkra på i vilken mån olika strategier och utfall tillmötesgår mer grundläggande målsättningar förefaller verkliga politiska aktörer oftast sakna den typ av *bestämda* preferensordningar över möjliga strategi-kombinationer, som krävs för en tillämpning av de spelteoretiska modellerna.

I det sjätte kapitlet diskuterar JH möjligheten av att utveckla de spelteoretiska modellerna så att de blir mer tillämpliga på verkliga aktörer. Diskussion rör här om det spelteoretiska betraktelsesättet kan utvidgas till situationer där aktörerna har osäker (snarare än säker respektive ingen alls) information om de övrigas preferenser. För att föra denna diskussion stipulerar JH en mycket förenklad situation med ett begränsat antal strategier och aktörer, där de senare antas ha bestämda rangordningar av de möjliga utfallen. JH nämner också möjligheten av att "i tur och ordning låta den substantiella tolkningen av *utfallen* variera under spelets gång, aktörernas *preferenser* rörande utfallen genomgå förändringar, uppsättningen av

alternativa *strategier* ändras och uppsättningen *aktörer* växla" (not 2, s 186). Kanske skulle sådana modifierationer öka spelteorins empiriska relevans, men i och med att dessa möjligheter inte demonstreras kvarstår i alla fall mina tvivel på denna punkt. Dessutom förefaller det mig som att situationen för spelteorins del är värre än vad JH tänker sig. För även om man kan konstruera spelmodeller där de faktorer JH nämner "ändras i tur och ordning", så tar detta inte hänsyn till vare sig att vi i verkligheten sällan har någon bestämd beskrivning av spelet att utgå ifrån (varvid det är obestämt vad som ska förändras på vilket sätt), eller att (i den mån förändringarna kan beskrivas) det nog ofta är av yttersta vikt att kunna fånga in vad som sker när samtliga nämnda faktorer förändras *samtidigt*.

Vid sidan av dessa problem lyfter JH också fram två fördelar hos spelteorin. Det spelteoretiska perspektivet är, enligt JH, "en idé som hjälper oss att ställa helt nya frågor och som bidrar till preciseringar av gamla invanda problemställningar" (s 346). Pga spelteorins bristande empiriska relevans torde detta främst gälla den normativa delen av statsvetenskapen. En andra fördel ligger mer på det vetenskaps-sociologiska planet. Eftersom spelteorin är "någorlunda ideologiskt neutral" är den "ett redskap som de flesta samhällsvetare kan nyttja". Därmed bidrar spelteorin

till "en ökad kommunikation mellan samhällsvetare från skilda läger". (s 347)

JHs slutsats blir till slut att spelteorin gör mest nytta inom den politiska filosofins område: "...den moderna rationalistiska teoribildningens främsta bidrag till statsvetenskaplig forskning kan visa sig vara, att den gör den politiska filosofins klassiska problemställningar tillgängliga för förnyade och mer preciserade analyser" (s 366). Detta är en slutsats som det är lätt att sympatisera med, men samtidigt är det också lätt att fråga sig huruvida spelteorin gör *någon som helst* nytta någon annan stans (vid sidan av att vara kommunikationsbefrämjande, vilket ju inte nödvändigtvis måste ses som en fördel). Inom den empiriskt inriktade samhällsvetenskapen verkar spelteorin på sin höjd kunna tillhandahålla diverse tillspetsade och idealiserade illustrationer, som kan tjäna som inspiration under studiet av den politiska verkligheten.

*Spelteorins nytta* är en på många sätt imponerande prestation. JHs problem är också ett i högsta grad intressant sådant, vilket även gäller hans slutsatser. Samtidigt finns det dock ett problem med bokens komposition, som jag tror kan återföras på JHs dubbla syften; att skriva en bok som är både en lärobok och en partsinlaga i en bestämd debatt. För den som är intresserad av JHs argumentation om spelteorins nytta blir det en aning tradigt att ta sig igenom de pedagogiskt utformade refe-

raten av diverse spelteoretiska idéer, vilket tyvärr bidrar till att göra JHs argumentation något svåröverskådlig. En kort inledande sammanfattning av de viktigaste argumenten som gav besked om var de kan återfinnas kombinerat med ett index hade här varit till stor hjälp.

Vad gäller bokens värde som lärobok för statsvetare är jag något kluven. Kombinerad med lämplig handledning kan den utan tvivel vara av stort värde. Men frågan är om ett så rikt utbyte som möjligt inte kräver att studenterna först är bekanta med grundläggande beslutsteori och därmed bl a preferensbegreppets speciella användning inom detta område. Ur pedagogisk synvinkel hade det således varit önskvärt om de rationalistiska utgångspunkterna förklarats något mera ingående. En novis på området torde t ex ha en aning svårt att inse de problem som sammanhänger med spelteoriens empiriska relevans med ledning av JHs framställning på sid 13–17. Kanske hade det också varit fruktbart att inkludera en mer detaljerad och systematisk framställning av de villkor som en aktör måste uppfylla för att de spelteoretiska modellerna ska kunna användas för att förklara hans beteende.

Dessa anmärkningar förtar dock inte intrycket av *Spelteoriens nytta* som något som närmast kan liknas vid ett svenskt monumentalverk på området. Boken kan förmodligen bidra till att föra det svenska studiet av politik framåt på flera olika sätt

– inte minst genom att återföra ”spel-glada” statsvetare till den empiriska verklighet de är tänkta att utforska. *Christian Munthe*

Torsten Thurén, *Vetenskapsteori för nybörjare*, Runa förlag, 1991

---

Vid våra universitet och högskolor förekommer många småkurser med ett vetenskapsteoretiskt innehåll. Att hitta lämplig litteratur till sådana kurser är inte alltid lätt. Den ska vara kortfattad och enkel – men ändå med ett substantiellt innehåll – saklig och helst avfattad på svenska. Några utmärkta arbeten finns i genren. Jag tänker på böcker av Molander, Mårtensson och Nordenfelt. Med de passar inte i alla kurser. Marknaden är långtifrån mättad. Därför hälsade jag – och jag antar många med mig – med glädje nedkomsten av en bok med den lovande titeln *Vetenskapsteori för nybörjare*, författad av Torsten Thurén, högskolelektor vid institutet för journalistik och kommunikation vid Stockholms universitet.

Tyvärr är den nog inte den guldklimp vi väntade på. Att den uppfyller några av kraven ovan är uppenbart. Den är kortfattad (148 sidor) och den är skriven på svenska. Dispositionen är lyckad och de problem och begrepp som Thurén väljer att ta upp, är i min mening lämpliga och väl valda. Så långt är allt sålunda tämligen gott. Språket är genomgående livfullt och färgstarkt.

men också genomgående oprecist. Och här börjar problemen.

De oprecisa formuleringarna gör att man ibland dels har svårt att förstå vad författaren menar, dels måste fråga sig om författaren själv överallt har förstått det han talar om. Jag tror säkert att Thurén har ett genuint intresse för vetenskapsteoretiska problem och en ambition att popularisera och gestalta sitt material för nybörjare på området. Men de suddiga formuleringarna – vare sig de nu beror på slarv eller okunnighet i ämnet – gör boken mindre lämplig som kursbok och omöjlig för självstudier för den som inte redan är litet bekant med ämnet. Definitivt skulle Thurén vunnit på att slipa något på sina formuleringar och kanske ta litet råd av vetenskapsteoretiker och filosofer av facket innan han lämnade materialet för publicering.

Vilken sakkunskap på området som Thurén kan ha, kan jag inte bedöma. Men att Føllesdal-Walløe-Elsters *Argumentasjonsteori, språk og vitenskapsfilosofi* är en av hans huvudkällor framgår av texten. Han lånar också – helt juste – exempel från denna snart klassiska lärobok i ämnet. Att Føllesdal-Walløe-Elster i sin tur ofta hämtat dessa exempel från vetenskapsteoretiska originaltexter, t ex Hempel tycks dock inte vara bekant. Men detta är en bisak. På en punkt avviker dock Thurén helt från Føllesdals framställning och det gäller s k funktionsförklaringar i samhällsvetenskap. Følles-

dal redovisar här många problem och ifrågasätter användningen av funktionsförklaringar utanför biologin. Inte med ett ord kommenterar Thurén att det skulle finnas några svårigheter på denna punkt.

Nu till min huvudinvändning som gäller de frekventa oprecisa formuleringarna och de direkta felaktigheterna. Jag ger några smakprov på Thuréns text, så får läsaren själv döma. Den hypotetisk-deduktiva metoden får följande presentation innan ett exempel tar vid. "I den hypotetiskt-deduktiva metoden ställer man upp hypoteser, gissningar som premisser. Sedan gör man en deduktiv slutledning, och till slut undersöker man om premisserna stämmer med verkligheten" (sid 25). Hur hypoteserna ser ut, vilken slutledning det är fråga om, vad slutsatsen kan tänkas uttala sig om, osv och hur allt detta hänger samman med undersökningen av om premisserna överensstämmer med verkligheten, får man inga upplysningar om. Exemplet – Semmelweis – ger naturligtvis "kött på benen", men nog kunde väl författaren kostat på sig en begriplig formulering som visar resonemangets struktur. På samma sida i boken menar författaren förresten att man bör "vara litet misstänksam mot alltför logiska resonemang", hur nu dessa kan se ut.

Deduktionen ges följande beskrivning: "Deduktion innebär att man drar en logisk slutsats som

betraktas som giltig om den är logiskt sammanhängande. Däremot behöver den inte nödvändigt vara sann i den meningen att den överensstämmer med verkligheten" (sid 23). För att få sammanhang i denna formulering måste ordet "den" ibland åsyfta själva slutledningen – nämligen när det gäller giltigheten och vad som är logiskt sammanhängande. ("Logiskt sammanhängande" är för övrigt en formulering som är mindre väl vald här. Den för ju tankarna till koherens och motsägelsefrihet och det är ju inte det som avses.) Men när det talas om sanningen är det i stället slutsatsen i slutledningen som åsyftas. Ja, den som redan är bekant med begreppet deduktion kan nog leta sig fram till meningen. Men hur ska nybörjaren gå tillväga som vill lära sig något av Thuréns text?

Vad säger vi om följande sammanfattning av Pythagoras sats? "Det finns inte någon risk att man någonstans i universum hittar en rätvinklig triangel där summan av kvadraten på kateterna inte är lika med summan av kvadraten på hypotenusan" (sid 17). Alldeles oavsett vad som gäller universums och det krökta rummets geometri, har också den som bara intresserar sig för vanlig euklidisk geometri svårt att känna igen sig. Hur får man summan av "kvadraten på kateterna"? Och framför allt, vad är "summan" av kvadraten på hypotenusan?

Att Kuhns paradigmbegrepp är mångtydigt, behöver vi inte bry oss

om, menar Thurén aningslöst och anför också en egenartad motivering för detta. "Men egentligen spelar det nog inte så stor roll om begreppet är vagt i konturerna. Paradigmen är ju ändå bara tankekonstruktioner" (sid 73).

Thurén tar så upp förhållandet mellan "verklighets- och värdeomdömen". Man kunde tänka sig att man skiljer mellan dessa på grund av att de på något väsentligt sätt är olika. Här kunde vara rum för litet filosofisk argumentering och analys. Men icke! Det finns framför allt två skäl, säger Thurén. "För det första hamnar diskussioner ofta i dödläge om man utgår från värderingar" (sid 37). Ett annat "gott skäl för att skilja mellan verklighetsomdömen och värdeomdömen är att man då inte så lätt faller i farstun för experter" (sid 38). Detta är skälen. Ja, man häpnar.

Men vad är då en värdering? Studenten får sig följande formuleringar till livs: "Egoismen är en värdering, lika väl som människokärleken, grymheten lika väl som medlidandet. Dessa värderingar ligger utanför vetenskapen, de handlar om känslor och motivation, inte om fakta, förklaringar och tolkningar" (sid 41). Att få någon reda i detta är inte lätt. Exempelen på värderingar förefaller vara helt bakvända. Egoism, grymhet, människokärlek kanske kan vara uttryck för vissa värderingar eller förutsätta vissa värderingar eller motsvarande. Men att de skulle vara värderingar är

åtminstone för en filosof något magstarkt. Jag kan förvisso hålla med om att värderingar inte handlar om fakta, förklaringar och tolkningar. Men handlar de om känslor? Handlar de för övrigt om någonting alls? Menar Thurén att värderingen påstår något om känslor eller ...? Och på vad sätt "ligger" värderingarna utanför vetenskapen?

Avsnitten om förklaringar hör till de jobbigaste. Vad sägs om följande slutkläm på en diskussion om voodoopsykologi och psykoanalys? "Kanske varken voodoons gudar eller libidons energi existerar utanför respektive samhällsammanhang. Men i dessa sammanhang existerar de. Haitiska landsbygdsbor är verkligen besatta av demoner, och New York-intellektuella är verkligen neurotiska" (sid 88). Det går kanske att få fram vad Thurén menar. Men inte är det väl det som står i texten?

Thurén skiljer sedan friskt mellan "tillräckliga" och "nödvändiga orsaker" (sid 111). Är det tillräckliga och nödvändiga betingelser som avses? När han ska förklara skillnaden mellan en orsaksförklaring och en intentionalförklaring låter det så här: "... den ger svar på frågan Varför? Varför reser jag på semester? Jo dels för att jag är trött (orsaksförklaring) dels därför att jag vill resa utomlands för att se mig omkring (ändamålsförklaring)" (sid 106). Så långt är ju allt gott och väl. Men när Thurén nu ska förtydliga sin

tankegång trasslar han till det. "Min trötthet förklaras kausalt och min vilja att se mig om intentionellt." Men hur ska den läsande nu förstå detta? Är det min trötthet som ska förklaras? Det kan säkert göras kausalt. Men var det inte här så att det var tröttheten som skulle förklara – orsaksförklara – varför jag reser? På samma sätt med viljan. Det är i och för sig riktigt att det som används i en förklaring ofta självt måste förklaras. Men här får vi ta ett led i taget och hålla isär vad det är vi förklarar och genom vad vi förklarar det. Det är viktigt. Lärobokstexten måste vara tydlig.

Ja, jag slutar min exempelsamling här. Den är inte uttömmande, men jag tror att den är tillräcklig. Det smärtar mig på sätt och vis att jag måste gå fram med storsläggan. Det kan synas orättvist med tanke på den blygsamma ambition som Thurén redovisar i förordet. "Det är fråga om en grov skiss över en komplicerad och svår terräng." Visst är terrängen svår. Men vi kan inte bortse från att syftet också måste vara att boken ska kunna användas som lärobok, kanske vid akademiska kurser, och som en introduktion till ämnet för den som vill läsa på egen hand. De blygsamma formuleringarna i inledningen får därför inte avhålla oss från plikten att kritiskt granska. Vi längtar ju, som sagt, efter en liten enkel och kortfattad bok i ämnet. Tyvärr vad det inte denna.

*Bo Petersson*



## NOTISER

¶ Bo Sellstedt, som är docent i företagsekonomi vid Handelshögskolan, har utgivit en bok med titeln *Samhällsteorier. Vad har Giddens, Habermas m fl att säga ekonomer?*, Studentlitteratur 1992.

¶ Ingemar Nordin, som är högskolelektor i filosofi vid tema hälso- och sjukvård vid Linköpings universitet, har utgivit *Etik, teknik och samhälle*, Timbro 1992.

¶ Hans Nystedt, docent i religionsfilosofi och under många år lektor i Visby, har givit ut *Uppgörelse med Hedenius. Vad Ingemar Hedenius förstod och inte förstod av religion*, Proprius förlag, Stockholm 1992.

I inledningen till boken berör han bland annat Hedenius doktorsavhandling, som handlade om Berkeleys filosofi. Nystedt påpekar att Hedenius hade mycket stor respekt för Berkeley, och att han vänder sig mot den vanliga populäruppfattningen att Berkeleys filosofi kommer i konflikt med sunt förnuft.

"Vad Berkeley vill försvara mot verklighetsfrämmande spekulationer är just det sunda förnuftets övertygelse om de omgivande tingens handfasta realitet. Berkeley var en framstående fysiker, men hans fiende är främst den mekanistiska världsbilden. Vi tänker oss ett barn med ett äpple i handen. Farbror Locke frågar: 'Min lilla vän, hur tror du det där äpplet är?' 'Jo, det

är rött och gott.' 'Nej, nej. Ser du, det finns två äpplen, ett överkligt och ett verkligt. Det röda och goda äpplet finns bara i din fantasi. Det verkliga äpplet har varken färg eller smak.' Barnet förstår ingenting utom att farbrorn är tokig. Enligt Berkeley är det barnet som har rätt. Det finns bara ett äpple, och det har färg och smak. Vad Berkeley förnekar är bara att det dessutom har något konstigt som kallas substans, som självt inte har några egenskaper utan bara är bärare av sådana. Det är detta som är själva poängen i den berömda satsen 'esse est percipi', att vara är detsamma som att uppfattas. Lika självklart riktig är sunda förnuftets övertygelse att tingen existerar oberoende av vår uppfattning och att vi kan ta fel. Det är skillnad mellan dröm och verklighet. Ur dessa två självklarheter följer rent logiskt att världen uppfattas av ett överindividuellt väsen." Hoppsan!

¶ En samling uppsatser i estetik, *Understanding the Arts – Contemporary Scandinavian Aesthetics*, har utgivits av Jeanette Emt och Göran Hermerén, Lund University Press 1992. Sexton författare från Sverige, Norge och Finland deltar.

¶ Johann Gottlieb Fichte (1762–1814) förkastade som bekant Kants antagande om ett ting i sig. Anders Wedberg skriver i sin mästerliga filosofihistoria att Fichte "därmed gjorde medvetandet till en självförsörjande fabrik i evighetens värld. Ja, medvetandet blir dessutom hos Fichte en fabrik som bygger sig själv". Kanske kan detta vara ett tips till den krisdrabbade svenska industrin. Är felet helt enkelt att den i onödan släpar runt på tinget i sig? Nej, tvärtom. Den bör nog snarare hålla sig till Søren Kierkegaards mer kritiska yttrande: "Fichte kastade den empiriska barlasten överbord – och kantrade".

¶ Christian Munthe har disputerat på en avhandling i praktisk filosofi, *Livets slut i livets början. En studie i abortetik*, Thales 1992. Fakultetsopponent var Thomas Anderberg från Lund.

¶ En viss uppståndelse i filosofiska kretsar åstadkom universitetet i Cambridge genom att utse Jacques Derrida till hedersdoktor i maj i år. Bland annat skrev nitton kända filosofer från ett tiotal länder ett brev till *The Times* där de beklagade förslaget. De nitton påpekar att Derrida – trots att han gör anspråk på att vara filosof – i allmänhet inte tas på allvar av filosofer, utan nästan enbart av folk från andra ämnen (t ex filmvetenskap, litteraturvetenskap, osv). Brevskrivarna hävdar vidare att Derridas skrifter till största delen är obegripliga: "... his works employ a written style that defies comprehension.... When the effort is made to penetrate it,...., it becomes clear, to us at least, that, where coherent assertions are being made at all, these are either false or

trivial". Bakom denna rätt gruvliga salva står Barry Smith, Hans Albert, David Armstrong, Ruth Barcan Marcus, Keith Campbell, Richard Glauser, Rudolf Haller, Massimo Mugnai, Kevin Mulligan, Lorenzo Peña, W V Quine, Wolfgang Röd, Edmund Runggaldier, Karl Schumann, Daniel Schulthess, Peter Simons, René Thom, Dallas Willard och Jan Wolenski. Trots protester vann Derrida omröstningen i Cambridge med 336 röster mot 204. Efteråt gjorde filosofiprofessorn Hugh Mellor, som var en av motståndarna till förslaget, bl a följande uttalande: "We do not bear M Derrida any ill will. We are happy to congratulate him on what we think is a quite undeserved honour" (se *The Times*, 18 maj 1992).

¶ Edmund Husserls *Cartesianska meditationer*, med undertiteln "En inledning till fenomenologin", har nyligen utkommit på Daidalos förlag (1992) i översättning av Daniel Birnbaum och Sven-Olov Wallenstein.

¶ Kaj B Hansen från Uppsala har utgivit en elementär lärobok i logik, som riktar sig till "studenter i ADB eller filosofi" samt "vem som helst som vill lära sig logikens grunder": *Grundläggande logik*, Studentlitteratur, 1992.

¶ En svensk översättning av de viktigaste av David Humes religionsfilosofiska skrifter har nyligen utkommit under titeln *Om religion*, Bokförlaget Nya Doxa 1992. Översättningen är gjord av Joachim Retzlaff och Jan Landgren, och den har möjliggjorts genom bidrag från Erik och Gurli Hultengrens fond för filosofi. Boken innehåller ett förord av Richard Wollheim samt följande skrifter: "Religionens naturhistoria" (1757), "Dialoger om naturlig religion" (1779), "Om underverk" (1748), "Om en särskild försyn och om livet efter detta" (1748), "Om vidskepelse och svärmeri" (1741), "Om självmord" (1777) samt "Om själens odödlighet" (1777).

¶ Vissa filosofiska skribenter uttrycker sig tyvärr på ett sätt som är totalt obegripligt för den oinvigde. Men dessutom kan man misstänka att det obegripliga av vissa läsare ses som ett symptom på djup- och skarpsinne. Det kan man komma att tänka på när man läser följande uttalande av den framstående och spirituella konversatören Oscar Wilde: "Jag tycker om att höra mig själv tala. Det är ett av mina käraste nöjen. Jag för ofta långa samtal med mig själv, och jag är så begåvad att ibland förstår jag inte ett ord av vad jag säger".

Oscar Wilde studerade i sin ungdom i Oxford, och där läste han också filosofi. Bland hans anteckningar från denna tid kan man exempelvis läsa följande: "Berkeley förintade icke-jaget. Hume gjorde

detsamma med jaget, och när i deras efterföljd lagen om orsak och verkan sönderföll och blev enbart till en samling subjektiva begrepp, då trodde allt gott folk att världens ände var nära".

Wilde var ju också (en sorts) socialist, och med tanke på det kan det vara av ett visst intresse att citera ytterligare ett uttalande av honom från hans studietid: "Bacons hänfulla ord om filosofin är dess högsta beröm. Liksom en jungfru helgad åt Gud bär den ingen frukt. Dess plikt är att förstå världen, inte att förbättra den".

Det bör kanske tilläggas att Wilde inte skrev på svenska; citaten ovan är hämtade ur den svenska översättningen av Richard Ellmanns stora biografi om honom.

¶ Gunnar Svensson, docent i teoretisk filosofi i Stockholm, har utgivit en bok med titeln *Wittgenstein om kunskap och visshet*, Thales 1992.

¶ En ny avhandling i ämnet vetenskapsteori i Göteborg har titeln *Djurförsök. Forskningsetik, politik, epistemologi*, Almqvist & Wiksell International, 1992. Författare är Birgitta Forsman.

¶ I Stockholm har Claudio Tamburrini disputerat i praktisk filosofi på en avhandling med titeln *Crime and Punishment?* Fakultetsopponent var docent Lars Lindahl från Lund.

¶ I början av oktober 1992 gästades filosofiinstitutionerna i Stockholm och Uppsala av den brittiske filosofen Timothy Williamson. Ett av hans föredrag handlade om den s k "vetbarhetsparadoxen". Den består i ett bevis för att antagandet att vi *kan* veta allt som är sant – vilket låter optimistiskt men inte helt orimligt – leder till slutsatsen att vi faktiskt också *vet* allting som är sant – vilket däremot är uppenbart orimligt. Men kanske ligger det ingenting paradoxalt i detta när allt kommer omkring. Vid närmare eftertanke kan man nog inse att det måste finnas sanningar som vi inte kan veta. Följande kunde vara exempel på en sådan sanning:

(A) Antalet blad just nu på trädet utanför mitt fönster är udda, men ingen kommer någonsin att veta detta.

Det som påstås i (A) kan mycket väl vara sant. Och om inte (A) är sant, så kan vi lätt åstadkomma en variant av (A) – nämligen den där vi byter ut ordet "udda" i (A) mot "jämnt" – som i så fall är sant. Det

är nämligen så att trädet utanför mitt fönster just nu har väldigt många blad, och jag känner mig på goda grunder övertygad om att ingen någonsin kommer att räkna dem. Även om någon mot förmodan skulle räkna trädets blad i framtiden, så kan vi knappast ta för givet att de blad som då räknas är exakt de som just nu finns på trädet.

Låt oss alltså utgå från att (A) är sann. Härav följer ju att ingen någonsin kommer att veta att (A) är sann, ty (A) säger just i sin andra halva att ingen någonsin kommer att veta att den första halvan är sann, och för att veta att hela (A) är sann måste man ju bland annat veta att första halvan är sann.

Det är med andra ord lätt att visa att vi inte kan veta allt som är sant. Alltså är vetbarhetsparadoxen helt oförarglig.

Somliga filosofer, s k "anti-realistier", tycks emellertid lägga stor vikt vid tesen att vi (i princip) kan veta allting som är sant. Ty det att en sats är sann är enligt anti-realistisk uppfattning detsamma som att det är möjligt att "bevisa"

den, eller kanske till och med att det i någon mening "finns" ett bevis för den. Det som är sant måste vara sant i kraft av något – enligt en realist är det sant i kraft av ett faktiskt föreliggande sakförhållande, enligt en anti-realist i kraft av möjligheten eller existensen av ett bevis. Och det verkar ju rimligt att påstå att om man kan bevisa en sats, så kan man därmed också veta att den är sann.

Men detta är inte särskilt övertygande. Det verkar exempelvis uppenbart att (A) kan vara sann, trots att vi inte kan veta att (A) är sann. Detta förutsätter förstas att vi kan förstå innebörden av (A). Det

verkar också oproblematiskt. Men i anti-realistisk anda kunde man hävda att vi förstår en sats om och endast om vi vet hur ett bevis för satsen skulle se ut. På sätt och vis är det naturligtvis svårt att veta hur ett bevis för (A) skulle se ut, eftersom det inte kan finnas något bevis för (A). Men det naturliga är väl här att tänka sig att vi vet vad (A) betyder genom att vi vet vad de olika komponenterna i (A) betyder, och i fråga om komponenterna kan vi väl någorlunda veta hur bevis för dessa skulle se ut. Alltså borde även en anti-realist gå med på att det finns satser som är sanna, trots att vi inte kan veta att de är sanna.

*Lars Bergström*

¶ Vid filosofiska institutionen i Stockholm ges under innevarande läsår en seminariserie om *Mening och tolkning*, som riktar sig till lärare och doktorander vid hela den humanistiska fakulteten. Varje sammanträde inleds av en medarbetare vid filosofiska institutionen, enligt följande lista:

1. Lars Bergström: Empirism, mening och tolkning
2. Gunnar Svensson: Wittgenstein om mening och språkspel
3. Daniel Birnbaum: Fenomenologi och hermeneutik
4. Dag Prawitz: Michael Dummetts språkfilosofi
5. Hans Ruin: Heideggers hermeneutik
6. Peter Pagin: Moderna meningsteorier
7. Lars O Ericsson: Mening i bildkonst
8. Sören Häggqvist: Daniel Dennetts heterofenomenologi
9. Folke Tersman: Donald Davidsons tolkningsteori
10. Fredrik Stjernberg: Mening och objektivitet
11. Dag Westerståhl: Mening och logik
12. Dagfinn Føllesdal: Husserl om mening
13. Jan Österberg: Nykritik och nystrukturalism
14. Staffan Carlshamre: Derrida och dekonstruktion

¶ Svante Bohman har utgivit *What is concept analysis?*, Almqvist & Wiksell International, 1992.

¶ Kanske finns det vissa speciella sidor hos verkligheten, som vi inte kan veta något om, trots att vi vet en hel del om andra områden. Den biologiska utvecklingen kan ha trimmat oss på ett sådant sätt att vi klarar vissa vetenskapliga problem mycket bättre än andra. Colin McGinn är inne på denna möjlighet i en recension i *London Review of Books*, nr 17, 1992. Recensionen handlar om en bok av Nicholas Humphrey om medvetandets historia. McGinn skriver: "There is no empirical or *a priori* reason to

suppose that our capacity to understand nature extends to all things that puzzle us; it would be amazing if it did. Consciousness may be one of the subjects that our biology has not equipped us to understand. This should be regarded at least as a live possibility by anyone who takes the biology of mind seriously. For all his vaunted naturalism, Nicholas Humphrey is, like so many others, unwilling to take his naturalism the whole way. The reason he can't produce a good theory may be that his brain won't let him".

¶ Den extremt flitige Torbjörn Tännsjö har återigen publicerat en bok, nämligen *Populist Democracy: A Defence*, Routledge, London and New York 1992.

¶ Att planeekonomi kan ha obehagliga för att inte säga katastrofala konsekvenser är något som många har insett på senare år. Att även marknadsekonomi har brister har exempelvis höstens valutakris lärt oss. I DN den 7 oktober kunde man bl a läsa följande: "Varje enskild finanschef i varje företag handlar rationellt när hon eller han enligt sitt uppdrag försöker gardera företaget mot kursförluster. Men helheten kan bli något annat än summan av delarna. Om många inför farhågor om devalvering omplaceras

från svensk till utländsk valuta blir följden ett valutautflöde och en räntestegring som på längre sikt ger alla ökade kostnader. [...] Summan av alla enskilda rationella beslut kan alltså bli en kollektiv irrationalitet. Alla förlorar i längden på det som för varje enskilt företag verkade ekonomiskt riktigt". Detta är i och för sig ingen ny insikt, men man kan notera att den här har formulerats av Nils-Eric Sandberg, en person som för övrigt i hög grad brukar sjunga marknadens lov.

¶ Det är i år 100 år sedan filosofen och folkbildaren Alf Ahlberg föddes (1892–1979). Han disputerade i Lund 1917 och var bl a rektor för Brunnsviks folkhögskola 1932–59. Hans insats bestod kanske framför allt i att föra ut filosofiska teorier till en bredare allmänhet. Han skrev själv över 50 egna böcker, översatte omkring 200 och skrev ett oräkneligt antal artiklar.

(Oräkneligt?) Åtminstone två av hans böcker finns för närvarande i handeln: *Humanismen* och *Det ondas problem*, bägge på Dualis förlag 1991.

¶ Daniel Dennetts bok *Brainstorms* från 1978 har nu, med stöd av Erik och Gurli Hultengrens fond för filosofi, utkommit i svensk översättning: *Hjärnstormar*, Nya Doxa 1992. För den som är intresserad av modern medvetandefilosofi är detta en mycket givande och ovanligt underhållande bok.

¶ Georg Christoph Lichtenberg (1742–1799) var en tysk fysiker och kulturarbetare som i sin samtid var känd för bl a inträngande analyser av Hogarths' berömda etsningar som skildrar livet i London. Under hela sitt vuxna liv gjorde Lichtenberg nästan dagligen anteckningar om allt möjligt, inklusive en del filosofi. Anteckningsböckerna publicerades efter hans död och är ofta en intressant och rolig läsning, eftersom L var en mycket kvick och klyftig person.

Nu har ett rikt urval kommit på svenska, med titeln *Kladdböcker* på förlaget Propexus. Översättningen, av Peter Handberg, är stilistiskt osäker, t ex "I de allra finaste ramifikationerna i våra vetenskaper ...". Översättarens efterord är informativt men också ett tragikomiskt exempel på en sorts uppstyld skönandlighet som Lichtenberg ofta

gisslar. Hursomhelst är boken en nöjsam läsning, här en träffande beskrivning på något vi alla känner igen:

"Jag tror att man i våra dagar bedriver vetenskapshistorien för minutiöst, till stor nackdel för vetenskapen själv. Man läser det gärna och huvudet förblir visserligen inte tomt, men sannerligen utan egentlig kraft; just därför att det blivit så fullproppat. Den som någonsin känt viljan inom sig att inte fylla huvudet utan stärka det, utveckla förmågorna och anlagen och breda ut sig, han har funnit att det inte finns någonting kraftlösare än samtal med en så kallad vetenskapslärd, där denne inte själv tänker men känner till tusentals historiskt-litterära omständigheter. Det är nästan som en föreläsning från kokboken när man är hungrig."

*Hans Mathlein*

Medarbetare i detta nummer av *Filosofisk tidskrift*: *Hans Rosing* är docent i teoretisk filosofi vid Åbo universitet, *Jan Eric Larsson* är tekn lic vid Lunds Tekniska Högskola, *Ingemar Nordin* är docent i teoretisk filosofi i Linköping, *Lars Bergström* är professor i praktisk filosofi i Stockholm, *Lena Forssén* är fil kand och arbetar i Hedengrens bokhandel i Stockholm, *Christian Munthe* har disputerat i praktisk filosofi och *Bo Petersson* är docent i praktisk filosofi i Uppsala.

# Filosofi från THALES

Våren 1992

**Ludwig Wittgenstein**

*Tractatus logico-philosophicus*

övers. Anders Wedberg. 147 sidor. Ca 132:-

*Filosofiska undersökningar*

övers. Anders Wedberg. 278 sidor. Ca 177:-

*Essäer om Wittgenstein*

utgiven under redaktion av Lars Hertzberg

296 sidor. Ca 177:-

**Gunnar Svensson**

*Wittgenstein om kunskap och visshet*

208 sidor. Ca 167:-

**Filosofin genom tiderna**

Texter i urval av Konrad Marc-Wogau

*Första bandet – Antiken Medeltiden Renässansen*

463 sidor. Ca 247:-

*Andra bandet – 1600-talet 1700-talet*

422 sidor. Ca 222:-

**Thales avhandlingsserie**

**Christian Munthe**

*Livets slut i livets början*

440 sidor. Ca 342:-



# Filosofi från THALES

Hösten 1992

Ständigt efterfrågad, åter tillgänglig:

## *De vetenskapliga revolutionernas struktur*

THOMAS KUHN

övers. Örjan Björkhem. 192 sidor. Ca 187:-

## *Om visshet*

LUDWIG WITTGENSTEIN

övers. Lars Hertzberg. 103 sidor. Ca 122:-

*"Dessa anteckningar har ett stort egenvärde, men är av intresse också därför att de på ett djuplodande sätt föregriper vissa moderna kunskapsteoretiska idéer, t.ex. idén om vetenskapliga paradigmm."*

Lennart Nordenfelt

Nya tryckningar

### *Språk och påverkan*

Jan Andersson och Mats Furberg

### *ABC i symbolisk logik*

Dag Prawitz

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

Tidskriften står öppen för olika filosofiska inriktningar, men den vill undvika bidrag som man inte kan tillgodogöra sig utan speciella förkunskaper eller tekniska färdigheter. Utöver längre artiklar på omkring 10-15 sidor, tar tidskriften gärna emot även kortare bidrag och inlägg av notiskaraktär.

#### MANUSKRIFT TILL FILOSOFISK TIDSKRIFT

- sändes till redaktören, Lars Bergström, Stora Änby allé 24, 161 54 Bromma
- skall vara försedda med namn och adress
- skall som regel vara skrivna på svenska, men bidrag på norska och danska accepteras också
- skall vara maskinskrivna på A4-papper med skrift endast på arkets ena sida; bör ej innehålla mer än 1 500 nedslag per sida och skrivas med bred vänstermarginal (räkna med 50 nedslag/rad och med 30 rader/sida = 1 500 nedslag)
- skicka gärna med manus även på diskett för Macintosh (disketten kommer naturligtvis att returneras)
- noter och litteraturhänvisningar inarbetas i texten
- särskild litteraturförteckning upprättas i alfabetisk ordning och placeras sist i manus
- för icke beställt material ansvaras ej
- korrektur läses i regel endast av redaktören
- ändringar mot manuskript bekostas icke av förlaget; författaren debiteras kostnad för sådana ändringar
- införda bidrag honoreras ej för närvarande
- i stället för särtryck erhåller författaren gratis 10 exemplar av det nummer av tidskriften i vilket bidraget varit införd

Tidigare årgångar av Filosofisk tidskrift kan köpas direkt från förlaget. Adressen är Box 50034, 104 05 Stockholm.

För årg 1/1980 — 9/1988 är priset 25 kr/årg, lösa ex 10 kr

För årg 10/1989 — 12/1991 är priset 100 kr/årg, lösa ex 30 kr

Porto tillkommer



