

*Kort om spelteori*

Ken Binmore. Översättning: Roland Poirier Martinsson  
 Fri Tanke, 2012, 234 s. ISBN 978-91-86061-39-5

Spelteori är av intresse inom flera akademiska discipliner – t.ex. matematik, ekonomi, statskunskap, filosofi och biologi – och kanske också i andra sammanhang, t.ex. för politiker och pokerspelare. Flera av de personer som fått Ekonomipriset till Alfred Nobels minne har också ägnat sig åt spelteori, t.ex. Alvin Roth och Lloyd Shapley 2012, Robert Aumann och Thomas Schelling 2005 och John Nash, John Harsanyi och Richard Selten 1994. De ”spel” det är fråga om är interaktion i största allmänhet; enligt Ken Binmore spelar människor ”ett spel varje gång de interagerar med varandra” (s. 5), men han säger också att spelteorin ”bara fungerar när människor spelar *rationellt*” (s. 6), så man kan ju undra hur ofta det sker. Vidare kan tydligen inte bara människor bete sig ”rationellt”; spelteorin kan också förklara beteendet hos spindlar och fiskar, ”eftersom de konkurrenter vars gener programmerades för irrationellt beteende nu är utdöda” (s. 6).

Binmore har genomgående en rätt självbelåten attityd och han skriver inte alltid så pedagogiskt. Ibland är det svårt att förstå vad han menar, t.ex. när han säger att samarbete och konflikt ”är två sidor av samma mynt” (s. 11) – men kanske är detta något som beror på översättningen. Flera av tabellerna och diagrammen verkar felaktiga; jag vet inte om det är likadant i det engelska originalet eller om det har tillkommit i den svenska utgåvan. Det hänvisas ofta till olika spel och det hade därför varit bra om man kunde hitta dem med hjälp av index. Men det går inte alltid. Om man t.ex. slår upp ett spel som kallas ”kampen mellan könen” i index, så hänvisas man till sidan 34, men där beskrivs inte spelet (även om det omnämns), utan man får själv leta sig tillbaka till sidorna 18–19 för att få reda på hur det såg ut. Där är emellertid siffrorna i tabellen felplacerade, såvitt jag förstår; övre vänstra och nedre högra rutan borde byta plats. Och för filosofen John Rawls anger index sidan 197, men Rawls kritiseras redan på sidan 16, vilket index inte noterar.

Vad är spelteori för något? Dess syfte tycks enligt Binmore vara att förutsäga personers agerande i spelsituationer med hjälp av observationer av deras tidigare beteenden i olika valsituationer – eller, alternativt, att påstå att deras beteende är ”motsägelsefullt” i den mån det inte skulle överensstämma med förutsägelserna (jfr s. 12–13). Vi kan illustrera detta med hans diskussion av det kanske mest kända av alla de spel som förekommer i den spelteoretiska litteraturen, nämligen *fångarnas dilemma*.

I sin ursprungliga version handlade detta om två fångar som var för sig kan välja mellan att erkänna eller inte erkänna ett brott. Men Binmore

diskuterar också en lite enklare variant och jag ska här hålla mig till den. Även här har två spelare att var för sig välja mellan två alternativ. Var och en kan antingen *ta* 10 kr till sig själv ur en pott med pengar eller *ge* 20 kr ur potten till den andre. Om bägge är givmilda får de alltså 20 kr var, om bägge är själviska får de 10 kr var, om den ene är givmild och den andre självisk, så får den givmilde ingenting och den själviske 30 kr.

	B ger	B tar
A ger	A får 20 kr – B får 20 kr	A får inget – B får 30 kr
A tar	A får 30 kr – B får inget	A får 10 kr – B får 10 kr

Spelteorin anses då säga att det är rationellt för bägge att ta 10 kronor till sig själva (under förutsättning att de bara bryr sig om pengar). ”Dilemmat” består i att de bara får 10 kr var på detta sätt, trots att de ju kunde ha fått 20 kr var, om bägge i stället hade varit givmilda. Men det är likafullt sant att varje spelare får mer pengar om hen betar sig själviskt, oavsett vad den andre gör.

Binmore anser att många har missförstått fångarnas dilemma. Som exempel på ett sådant ”förelslut” skriver han: ”I vanligt språk säger Immanuel Kants kategoriska imperativ att det är rationellt att göra det du vill att alla andra ska göra. Om det var sant så skulle det vara rationellt att samarbeta [dvs. att vara givmild] i fångarnas dilemma. Men önsketänkande är aldrig rationellt. Det är en konstant källa till förvåning för mig att Kant aldrig ställs till svars för att ha föreslagit en rationalitetsprincip utan att ge några som helst skäl för varför vi bör ta den på allvar” (s. 205).

Detta är ganska typiskt för Binmores sätt att skriva. Det är tydligt att han ogillar moralfilosofer; han berömmar sig t.ex. av att ha retat gallfeber på ”renläriga moralfilosofer” genom att använda spelteori när han diskuterar etik (s. 8) och han påstår, utan argument, att moralfilosofen John Rawls ”försvarade en sund intuition med fel argument” (s. 197). Men hur Binmore kan vara så säker på att Kant ”aldrig ställs till svars” är ett mysterium; har han verkligen läst hela den minst sagt omfattande Kant-litteraturen? Dessutom kan man ju undra om det Kant föreslår verkligen är en ”rationalitetsprincip” i Binmores mening, snarare än t.ex. en *moralisk* princip. Att moralen ibland kan kräva något *annat* av oss än att göra det vi rent personligen råkar föredra, är ju en föreställning som inte är helt okänd i mänsklighetens historia. Men Binmore ser den nog som ett ”förelslut”.

Binmore anser alltså att Kants morallära är oförenlig med spelteorin. Men det kan knappast stämma. Binmore påpekar själv att att spelteorin

”inte har något substantiellt innehåll” (s. 74). Den har inte ”mer egentligt innehåll än aritmetik eller logik” (s. 82). Det har han nog rätt i; spelteorin tycks bestå av en rad teorem som har bevisats följa logiskt från vissa axiom eller antaganden. Men om spelteorin inte har något substantiellt innehåll kan den inte gärna komma i konflikt med en moralisk princip.

Ibland händer det förstås att folk vill göra det moralen kräver och i fångarnas dilemma kunde man kanske anse att moralen, som Kants princip kan tyckas antyda, kräver att bägge spelarna ska vara givmilda (eller samarbeta, i den ursprungliga versionen). Det visar sig för övrigt också att folk i verkligheten ofta väljer att vara givmilda. Förutsättningen att spelarna bara är intresserade av pengar är ju egentligen väldigt konstig och farligt vilseledande, eftersom alla vet att den är felaktig. Spelteoretikerna har också själva visat empiriskt att den inte stämmer, bland annat i så kallade ultimatumspel (jfr s. 63f).

Binmore menar därför att vi mycket väl kan – och egentligen också bör – slopa antagandet att folk bara bryr sig om pengar. Den ”valuta” som spelteoretiker intresserar sig för är så kallad *nytta* (eng. ”utility”). Inte i den hedonistiska betydelse som klassiska utilitarister som Bentham och Mill tog fasta på, utan i den moderna preferensteoretiska meningen (jfr s. 12). Varje (rationell) individ antas då ha en ”nyttofunktion” som tilldelar ett numeriskt nyttovärde till varje utfall som kan bli resultatet av ett spel. Denna nyttofunktion bestäms av hur individen väljer mellan olika alternativ, inklusive val mellan lotterier. De ”preferenser” som kan beskrivas på detta sätt är så kallade ”avslöjade” (*revealed*) preferenser (s. 12–3). Vad folk anser vara bra och dåligt – eller vad folk skulle må bra eller dåligt av – spelar alltså ingen som helst roll för spelteoretiker; det avgörande är hur individerna faktiskt betar sig i valsituationer.

Man måste därför skilja mellan å ena sidan fångarnas dilemma i den ursprungliga ”fysikaliska” eller ”objektiva” meningen, där man spelar om pengar eller antal år i fängelse, och å andra sidan fångarnas dilemma i det ”preferensbeskrivande” eller ”subjektiva” spel, där de olika utfallen definieras med den *nytta* de har för respektive spelare. Det senare kan se ut så här:

	B ”ger”	B ”tar”
A ”ger”	A får $a$ – B får $b$	A får $c$ – B får $d$
A ”tar”	A får $e$ – B får $f$	A får $g$ – B får $h$

där  $e > a > g > c$  och  $d > b > h > f$ . Vilka numeriska värden dessa åtta bokstäver betecknar beror på vilka nyttoskalor vi väljer. Men tabellen

beskriver ett fångarnas dilemma i den meningen att den undre raden är dominant för A och den högra kolumnen är dominant för B, vilket innebär att den nedre högra rutan ger det utfall som kommer att väljas, trots att den övre vänstra rutan hade varit bättre för båda. Det är den enda Nash-jämvikten i spelet (jfr s. 21 och 26). Men detta spel är inte symmetriskt på samma sätt som i det objektiva spelet. Vi kan t.ex. inte säga att A:s bästa utfall (dvs. *e*) är lika bra som B:s bästa utfall (dvs. *d*), ty vi kan inte göra intersubjektiva jämförelser mellan A:s nytta och B:s. Nyttan, i spelteorins mening, kan bara mätas intrasubjektivt på intervallskalenivå, med godtycklig nollpunkt och enhet.<sup>1</sup>

I tabellen ovan har jag satt citationstecken kring ”ger” och ”tar”. Det är oklart vilka objektiva handlingar dessa beteckningar här kan stå för. Fångarnas dilemma i objektiv mening passar kanske inte så ofta. Men det kan finnas många olika objektiva spel som passar in på det subjektiva spelet.

Man kan också notera att subjektiva utfall är mindre ”självcentrerade” än de motsvarande objektiva. Om B väljer den vänstra kolumnen i det objektiva spelet, så är de spelteoretiskt relevanta utfallen som A kan få inte längre 20 respektive 30 kronor; snarare är frågan vilket A föredrar av de två objektiva utfallen [A får 20 kr och B får 20 kr] å ena sidan och [A får 30 kronor och B får ingenting] å den andra. Och det är ju mycket möjligt att A föredrar det första av dessa, dvs. att A i valet mellan dessa två utfall faktiskt väljer det första.

Kort sagt: det är oklart vad som *är* ett ”fångarnas dilemma”, trots att detta är det mest diskuterade av alla spel. Syftar man på ett objektivt eller ett subjektivt spel? Binmore ger inget tydligt svar på den frågan.

Att folk betar sig så att de tillfredsställer sina preferenser tycks vara en truism, eftersom de preferenser spelteoretiker intresserar sig för är ”avslöjade” preferenser. Men då kan man fråga sig om det överhuvud taget är möjligt att handla altruistiskt i fångarnas dilemma (tolkat subjektivt). Om spelarna skulle göra det, så spelar de väl inte längre fångarnas dilemma! Ty deras preferenser – som ju ”avslöjas” av deras beteende – skulle ju då se annorlunda ut än i den subjektiva tabellen ovan. I själva verket är den tabellen inte en beskrivning av ett spel, utan av ett handlande. Den säger att både A och B ”tar”.

Här kan man tänka sig två olika positioner, beroende på hur spelteorins syfte uppfattas. Som vi har sett tycks syftet enligt Binmore vara att *förutsäga* personers agerande i spelsituationer med hjälp av observationer av deras *tidigare* beteenden i olika valsituationer (jfr s. 12–13). Det skulle alltså innebära att man först skulle fastställa vilken nyttofunktion en

<sup>1</sup>Något som Binmore tycks ha glömt bort i sin diskussion av blandade Nash-jämvikter i kampen mellan könen på s. 34.

individ har för att därefter göra förutsägelser. Men sådana förutsägelser förutsätter då att individens nyttofunktion inte förändras under tiden. Binmore tycks anse att det vore irrationellt av en person att inte agera i överensstämmelse med sina tidigare preferenser, dvs. att inte bete sig i överensstämmelse med hur man tidigare har betett sig. Men han ger inget som helst skäl för detta antagande. Och det vore ju också orimligt. Ett sådant antagande strider ju mot den väletablerade teorin att vi lär oss av våra misstag, så att vi inte upprepar dem i framtiden. Det är klart att vi ofta förutsäger en persons beteende på grundval av tidigare beteenden, men i allmänhet anser vi inte att det är *irrationellt* att bete sig på ett nytt sätt. Men det tycker tydligen Binmore och han kallar sådant beteende ”motsägelsefullt” (se t.ex. s. 13).

Den andra möjliga positionen skulle innebära att man accepterar att folks nyttofunktioner ofta förändras över tiden och att det inte är något irrationellt med det. I så fall ska vi i fångarnas dilemma hålla oss till spelarnas nyttofunktioner just vid den tid då spelet spelas. Men då blir det också omöjligt att spela det altruistiska ”ger”. Skulle man göra det, så spelar man inte fångarnas dilemma. Man kan därför inte heller säga att det ena beteendet är mer rationellt än det andra.<sup>2</sup> Spelarna gör bara det de gör och deras nyttofunktioner bestäms av vad de gör (och vad de skulle göra). Inte tvärtom.

Binmore diskuterar en stor mängd olika spel, mer eller mindre komplicerade. Ofta kan det vara lärorikt för den oinitierade. Men de teoretiska grundvalarna är delvis otydliga, bland annat på det sätt som jag här har försökt illustrera i samband med fångarnas dilemma. Låt mig avslutningsvis nämna ytterligare ett av de ”felslut” som Binmore tar upp i samband med ett spel som, i likhet med fångarnas dilemma, diskuterats flitigt av filosofer, nämligen *Newcombs paradox*.

I Newcombs paradox får man välja mellan att ta en eller två lådor, där den första, som är stängd, innehåller antingen en miljon dollar eller ingenting och den andra, som är öppen så att man ser vad som finns i den, innehåller tusen dollar. En mycket duktig förutsägare har på förhand lagt en miljon i den stängda lådan om och endast om hen har förutsagt att man bara kommer att ta den lådan. Man vet att förutsägaren har spelat

<sup>2</sup>Såvitt jag vet säger de spelteoretiska teoremen aldrig någonting om vad som är ”rationellt”. Det enda som kanske kunde kallas ”rationellt” i sammanhanget är väl i så fall det att en person uppfyller de antaganden som krävs för att hen alls ska ha en nyttofunktion och i sitt handlande maximera förväntad nytta. Binmore säger (s. 13) att dessa antaganden är ”blygsamma”, men det kan nog diskuteras – och det har också diskuterats, dock inte av Binmore i denna bok. Kanske har de flesta av oss inga nyttofunktioner alls, eftersom de ”blygsamma” villkoren är alltför krävande!

spelet med olika personer väldigt många gånger och nästan alltid har lyckats göra korrekta förutsägelser. Ska man då bara ta den stängda lådan (och därmed troligen få en miljon dollar) eller ska man ta bägge lådorna (och därmed troligen bara få tusen dollar)?

Jag sa en gång till Binmore, när vi bägge deltog i en konferens om spelteori, att jag för min del bara skulle ta den stängda lådan. Han förklarade då att jag inte var riktigt klok. I samma anda säger han i boken att den som säger att man bara ska ta den stängda lådan gör sig skyldig till en ”motsägelse” (s. 213). Och han säger att ”det blir [...] aldrig optimalt att spela en starkt dominerad strategi” (s. 214).<sup>3</sup>

Om saken vore så enkel som Binmore säger, så skulle det vara obegripligt att Newcombs paradox har diskuterats så mycket och att flera bedömare anser att det vore rationellt att bara ta den stängda lådan. Visserligen är det *dominanta* alternativet att ta bägge, ty oavsett om det ligger en miljon i den stängda lådan eller ej – vilket man inte kan påverka när man väljer, ty man kan (troligen) inte påverka det förflutna – så får man mest om man tar bägge. Men å andra sidan tycks den *förväntade nyttan* av att endast ta den stängda lådan vara större än den förväntade nyttan av att ta bägge. Ty den evidens man har tillgång till pekar ju på att sannolikheten för att det ligger en miljon i den stängda lådan är mycket stor om man bara väljer den. Poängen med Newcombs paradox är just att den tycks visa att två av den vanliga spelteorins principer kan komma i konflikt.

I en undersökning bland över 3200 filosofer svarade 21,9 procent att de skulle välja – eller lutade åt att välja – två lådor medan 19,5 procent föredrog en låda. Men de flesta, 58,6 procent, valde något ”annat” (*PhilPapers Surveys*)<sup>4</sup>. De som svarade något ”annat” avsåg främst någon variant av ”vet ej”, men övriga 3,6 procent gav svar som möjligen kunde antyda att själva spelet på något sätt är omöjligt. Och kanske är det detta Binmore menar när han säger att Newcombs paradox rymmer en motsägelse. Motsägelsen skulle bestå i att förutsägaren på förhand ganska säkert vet vad man kommer att välja, trots att man har mer än ett handlingsalternativ.

Men är detta en motsägelse? Det verkar inte så. Att någon vet vad man *kommer* att göra utesluter väl inte att man *kan* göra något annat? Och förutsätter verkligen spelteorin att vi har ”handlingsfrihet” i en så stark mening att det är omöjligt att göra mycket goda förutsägelser av våra handlingar? Många filosofer – 51,5 procent enligt den undersökning

<sup>3</sup>Han anger också två misslyckade försök att undvika ”motsägelsen”, men de är konstiga och såvitt jag kan se innehåller de minst en felaktig siffra.

<sup>4</sup>Bland de filosofer (inklusive studenter och ev. en del icke-filosof) som räkade beslutsteori (antagligen inklusive spelteori) som en specialitet var siffrorna helt annorlunda. 53,8 procent av dessa valde två lådor, 21,2 procent valde en låda och 25 procent svarade något ”annat”.

jag nyss nämnde – anser exempelvis att handlingsfrihet är förenlig med determinism; och om de har rätt, så borde det ”i princip” vara möjligt att förutsäga alla våra handlingar, även i de fall då vi kunde ha handlat annorlunda. Här finns olösta problem som Binmore verkar helt omedveten om.

LARS BERGSTRÖM

### *Kort om döden*

Lennart Lundmark

Fri Tanke, 2012. 148 s. ISBN 978-91-86061-44-9

Fri Tanke förlags *Kort om*-serie är – med inspiration från Oxford University Press-serien *Very Short Introductions* – avsedd att erbjuda initerade introduktioner till angelägna, ofta filosofiska problem. De hittills publicerade titlarna uppvisar en skimrande mångfald i både ämnesval (såsom ångest, virus, yttrandefrihet och spelteori) och kvalitet. Lennart Lundmarks *Kort om döden* är det första svenska originalverket i serien.

Till största delen är boken en redogörelse för några historiskt inflytelserika skolors syn på döden. Lundmark refererar epikuréernas två huvudargument för att döden inte kan vara något ont för den som dör: för det första finns inte personen längre när hon är död och är därför utom räckhåll för att skadas av den; för det andra är döden bara en spegelbild av den harmlösa period under vilken personen ännu inte blivit född. En lite mer kritisk diskussion av dessa viktiga men omstridda tankegångar skulle ha höjt temperaturen i Lundmarks framställning. Han uppehåller sig främst vid stoicismen, med dess rekommendation att med förnuftet vid rodret finna sig i naturens ordning. Denna lära lockade bland andra Cicero, som skrev att den vise inte brinner ”av törstigt begär eller flyger hojtande iväg i larvig munterhet” (s. 24) och att “[e]ländig är den åldrade man som inte under ett långt liv har lärt sig att strunta i döden” (s. 28). Den stoiker som ägnas störst utrymme är Seneca, inte minst dennes rekommendation att träna bort skadlig dödsskräck genom att reflektera över i vilka omständigheter det vore rationellt att ta sig själv av daga. Sådana omständigheter befinner sig enligt Seneca exempelvis en förtryckt individ i: ”Galning, varför klagas du! . . . Vart du än vänder ögonen finns en utväg ur plågorna . . . Frågar du efter frihetens väg? Vilken åder som helst i din kropp!” (s. 39) Lundmark skildrar även den längtan efter martyrdöden – uppfattad som en biljett till himmelriket – som utmärkte kristenhetens första generationer. ”Måtte jag få glädje av vilddjuren som väntar mig”, skrev biskop Ignatios av Antiokia förväntansfullt (s. 55). De