

Torbjörn Tännsjö

Vad är beslutsteori?

Den 23 mars 1980 beslutade en majoritet av svenska folket att kärnkraften ska byggas ut kraftigt. Samtidigt avvisades förslaget om en snabb avveckling av kärnkraften. Åt detta är nu inte mycket att göra, men en viss eftertanke griper omkring sig i åtskilliga läger. Är det rätt att bygga ut kärnkraften i Sverige, frågar man sig. Var det ett *rationellt* beslut, som folkmajoriteten tog?

Jag ska inte i den här artikeln söka underbygga några bestämda svar på dessa frågor. Ändå är mitt ämne av betydelse för den som vill besvara dem. Jag ska nämligen granska idén om en åtskillnad, av den typ som antyds i frågorna, mellan problem om huruvida handlingar är rätta å ena sidan och problem om huruvida beslut är rationella å den andra. Särskilt vill jag skärskåda den andra typen av frågor närmare, de som rör besluts "rationalitet". Vad betyder det att ett beslut är "rationellt"? Hur ska ett beslut vara beskaffat för att vara rationellt?

Frågan hur ett beslut ska vara beskaffat för att vara rationellt kommer jag inte att fullständigt besvara. Jag ska granska några olika idéer om, vad som är ett korrekt svar på den, men jag ska inte slutgiltigt träffa något val dem emellan. Skälet härför är att jag inte kunnat se hur någon som helst rimlig tolkning ska kunna ges av att ett beslut i en konkret situation av osäkerhet eller risk är rationellt. Och situationen då vi beslutade om kärnkraften tycks ha varit av detta slag. Frågan vad det betyder att ett beslut är "rationellt" har i själva verket inget bra svar i sådana sammanhang. En tes jag driver är därmed, att problem om besluts rationalitet i konkreta situationer av osäkerhet eller risk inte är några genuina problem.

Min tes antyder, om den är välgrundad, att det är något problematiskt med själva ämnet för den omhuldade, delvis filosofiska, delvis ekonomiska, disciplin, som brukar benämnas "beslutsteori". Inom denna disciplin brukar vad det är att fatta rationella beslut under osäkerhet eller risk stå i centrum.

Att vi vet vad det *betyder* att en handling är "rätt" eller "fel" ska jag ta för givet. Däremot ska jag inte förutsätta något bestämt svar på frågan vad det är som *gör* handlingar rätt eller fel. Jag nöjer mig här, för resonemangets skull, med att anta att den frågan *har* ett korrekt svar.

1. Beslutsteori

Standardformuleringen av beslutsteoris problem är, att det gäller att finna metoder att "rationellt" avgöra vad man i en konkret situation bör göra. Att detta problem hänger samman med och kanske t o m för sin lösning förutsätter lösningen av problem om vad som gör handlingar rätt eller fel medges ofta. För att vi ska kunna veta hur vi rationellt ska besluta oss i ett konkret fall så behöver vi veta vad som gör handlingar rätt eller fel. Detta är nödvändigt, men i många fall otillräckligt, har det hävdats. I ett särfall, då vi har kunskap om alla moraliskt relevanta egenskaper hos en handling sammanfaller frågan om en handling är rätt med frågan om förnuftigheten av att utföra den, brukar det hävdas. Då finns inget beslutsteoretiskt problem att lösa. Vi bör då besluta oss för rätt handling. Stundom handlar vi emellertid under osäkerhet eller risk. Vi känner t ex inte alls till konsekvenserna av våra handlingar eller av deras alternativ eller vi får nöja oss med kunskap om sannolika konsekvenser. Det är i sådana situationer, då vi inte vet vad som är rätt eller fel att göra, som beslutsteorin skulle bli ett viktigt komplement till våra teorier om vad som gör handlingar rätt eller fel. Med stöd av våra teorier om vad som gör handlingar rätt skulle vi kunna ta ställning till olika förslag om vad som är en rationell beslutsmetod. Ett rationellt handlande (eller beslut) i ett särskilt konkret fall skulle vara ett handlande (eller beslut) i enlighet med en rationell beslutsprincip. Men vad betyder det att en beslutsmetod är "rationell"? Om det har beslutsteoretiker haft föga att säga. Jag ska också i den här artikeln närma mig den frågan indirekt, genom att granska några förslag om vad som skulle vara rationella beslutsmetoder.

2. Några beslutsmetoder

Låt oss hålla fast vid kärnkraftsexemplet, som presenterades inledningsvis. I en artikel i *Filosofisk tidskrift* nr 1 1980 har Dag Prawitz just i anknytning till den svenska kärnkraftsdebatten diskuterat några olika beslutsprinciper. Prawitz formulerar tre principer av vilka han förkastar två och accepterar en.

Enligt principen om *maximering av det förväntade värdet* bör vi söka uppskatta de olika sannolika utfallen av de två alternativen, att bygga ut och att avveckla kärnkraften. Vi bör också värdera dessa olika möjliga konsekvenser. I valet mellan att bygga ut och att avveckla kärnkraften bör vi välja det alternativ för vilket summan av produkterna av sannolikheter och värden för de olika möjliga utfallen är störst.

Enligt *maximinprincipen* bör vi i stället undersöka vilka utfall som är de värsta möjliga av de två alternativen. Har ett alternativ några möjliga konsekvenser, hur osannolika dessa än är, som är outhärdligt dåliga, så bör detta alternativ undvikas.

Enligt principen om en *lexikografisk ordning* mellan olika värden, slutli-

gen, bör de olika alternativen först undersökas med avseende på vissa överordnade värden, särskilt värdet av att mänskligheten lyckas överleva ännu en tid framöver. Det alternativ, som med största sannolikhet medför mänsklighetens överlevnad bör väljas. Endast om alternativen står lika i detta avseende bör vi gå vidare och jämföra dem med avseende på andra värden.

Prawitz menar, med rätta tror jag, att maximinprincipen lider av en allvarlig svaghet. Den pekar i praktiken aldrig ut några handlingsalternativ framför några andra. *Varje* alternativ kan ju leda till mänsklighetens undergång om vi godtar att också det mycket osannolika är *möjligt*. Men vilken av principerna om maximering av förväntat värde och en lexikografisk ordning är rationell? Den senare, menar Prawitz. "Hela tanken att spela roulett med mänsklighetens undergång som ett möjligt utfall och något ökad välfärd som ett annat är naturligtvis absurd", skriver han. Jag är frestad att säga emot. Att ta stora risker för att vinna litet välfärd framstår givetvis inte som särskilt rationellt. Men varför inte gå med på tvärt emot vad Prawitz' princip tycks kräva, att sannolikheten för att mänskligheten ska gå under får öka något litet (i ett läge där den är mycket låg) om vi därmed kan byta ut en på det hela taget ganska eländig tillvaro för de flesta mot ett liv i välmåga för alla? Prawitz' princip hindrar oss för övrigt inte endast från att *förbättra* villkoren drastiskt för alla i lägen där förbättringen måste köpas till priset av något större risk för mänsklighetens undergång. Principen tycks också kräva att vi i vissa situationer drastiskt ska *försämra* villkoren, om därmed också en liten risk för hela mänsklighetens undergång något kan minskas. Lars Bergström har i en diskussion illustrerat detta med ungefär följande exempel. Antag att en av de två stora kärnvapenmakterna får s k förstaslagsförmåga. Ett försök att utnyttja denna skulle sannolikt lyckas. Därmed skulle halva mänskligheten utplånas men risken för att *hela* mänskligheten går under i en total kärnvapenkatastrof elimineras helt. Utnyttjandet av förstaslagsförmågan är dock förknippat med vissa risker. Sannolikheten är en på tusen att attacken ska upptäckas i tid av motparten, med total kärnvapenkatastrof som följd. Om attacken *ej* företas är sannolikheten en på niohundra tionio för total kärnvapenkatastrof. Prawitz' princip bjuder kärnvapenmakten med förstaslagsförmåga att slå till. Visserligen är priset högt. Minst utplånas halva mänskligheten, i värsta fall hela. Men *risken* för att hela mänskligheten ska utplånas reduceras från en på niohundra tionio till en på tusen!

Framstår Prawitz' princip som rimlig? Varför inte i stället tolerera en något större risk för total katastrof och rädda halva mänskligheten?

Ja, varför inte? För att veta hur vi alls ska värdera olika beslutsprinciper måste vi nog vara på det klara med vad principerna är tänkta att besvara för frågor. Vad *betyder* det att ett beslut som jag fattar i en konkret situation är "rationellt"?

3. Rationalitet

I beslutsteoretiska sammanhang är det vanligt att tala om rationalitet hos *beslut* men om att *handlingar* är rätt. Beslut är dock ett slags handlingar. Det förefaller därför enklast att se såväl uttrycken "rätt" och "fel", som "rationell" och "irrationell", som tillämpliga på handlingar. Vad är det då att en handling är "rationell"?

Vill vi förstå hur Prawitz besvarar den frågan måste vi införa ännu ett begreppspar, nämligen det om *riktiga* och *oriktiga* handlingar (eller beslut). Det är med hjälp av det begreppsparat Prawitz definierar sina begrepp om rationalitet och irrationalitet.

Det gäller alltså nu att klara ut sambanden mellan och innebörden av följande begreppspar:

rätt—orätt
riktig—oriktig
rationell—irrationell

Låt oss först se på förhållandet mellan rationella och riktiga handlingar. Prawitz skriver att han vill se de rationella handlingarna som en "underart" av de riktiga. Jag tror han tänker sig att de rationella handlingarna är de, om vilka vi *vet*, att de är riktiga. Men jag tror inte att han skulle ha något emot att, som jag tycker är rimligare, vidga klassen av rationella handlingar till att innefatta samtliga de handlingar om vilka vi har *goda skäl att anta* att de är riktiga. Det är då möjligt att vissa rationella handlingar faktiskt är oriktiga.

Hur förstår då Prawitz, att en handling är riktig? Först när vi har förstått detta kan vi värdera olika argument för och emot den ena eller den andra principen för rationellt beslutsfattande (eller handlande).

Den mest näraliggande tanken är väl, att se "riktiga" handlingar som handlingar om vilka vi *vet*, eller åtminstone har goda skäl att anta, att de är rätt. Så skulle nog *jag* vilja använda ordet "riktig" eller "rationell", om jag alls ska bruka dem. (Det blir nu ingen intressant skillnad mellan riktighet och rationalitet varför vi helt kan avstå ifrån att bruka ordet "riktig". Vi slipper då paradoxen att vissa handlingar kan vara på en gång rätt och oriktiga!) Så har emellertid inte Prawitz använt dessa uttryck. Ty i *denna* mening *kan* vi inte ge några välgrundade svar på frågan vad som är ett rationellt eller riktigt handlande under osäkerhet eller risk. Då saknar vi ju per definition goda skäl att anta att något bestämt handlingsalternativ är rätt.

Kanske är då en "riktig" handling en handling som *sannolikt* är rätt? (Och en rationell handling en handling vi har goda skäl att anta, att den sannolikt är rätt.) Nej, så kan inte Prawitz heller ha tänkt sig begreppen. Ty det är fullt möjligt att det *ex är* rätt att bygga ut kärnkraften i ett socialistiskt Sverige. Det är nog *mera* sannolikt att det är rätt att bygga ut kärnkraften i ett

socialistiskt Sverige än att det är rätt att avveckla den. (Sannolikt inträffar ingen allvarligare olycka). Och ändå skulle många med Prawitz sannolikt invända, att ett beslut att bygga ut kärnkraften också i ett socialistiskt Sverige vore oriktigt. Det vore riktigt att undvika stora risker och i stället göra vad som sannolikt är fel, nämligen att avveckla kärnkraften!

Ska då detta med handlingars riktighet uppfattas på ett generaliserat sätt? Att en handling är "riktig" betyder, att den är av en typ, som om den utförs allmänt, sannolikt ger upphov till de bästa konsekvenserna? Om detta är vad Prawitz förstår med "riktighet" måste han emellertid acceptera, att principen om maximering av förväntat värde är ett kriterium på riktighet. Ty om denna princip *vet* vi, att den i det långa loppet sannolikt *ger* det bästa utfallet. Men Prawitz protesterar mot detta synsätt. Han medger att principen om maximering av förväntat värde i det långa loppet *ger* det bästa utfallet:

Rationaliteten i att följa principen om att maximera det förväntade värdet vid beslut under risk tycks ligga i att denna beslutsprincip i det långa loppet ger ett bättre samlat resultat än varje annan beslutsprincip (s 10).

Inte desto mindre är det tveksamt, menar Prawitz, om det i ett enskilt konkret fall är rationellt att maximera det förväntade värdet:

... förutsättningen är i så fall att principen följs i ett stort antal fall, och den förutsättningen kan bortfalla om man spelar roulett med livet som insats. Det är svårt att se att det skulle finnas något förnuftigt skäl för en agent att följa principen om maximering av det förväntade värdet, om det eventuellt bara är ett enda beslut som han skall fatta (s 10).

Prawitz' invändning visar att han inte uppfattar "riktighet" på här skisserat generaliserat manér. Hans koncentration på ett enstaka konkret fall visar detta. Ty det är närmast en självklarhet, att synpunkten hur det går i det långa loppet om man gör handlingar av viss *typ* inte är något tillräckligt skäl för att handla på det ena eller andra sättet i ett konkret fall. Det är emellertid möjligt att Prawitz på denna punkt fjärrnar sig från det problem som besluts-teori *normalt* har tänkts handla om. Kanske är det just ett svar på frågan, vad som är rätt beslutsmetod, *förutsatt att en och samma metod ska brukas i alla fall*, som beslutsteoretiker i allmänhet sökt? Nå, Prawitz vill dock uttala sig om det partikulära fallet. Han skriver också explicit om detta när han formulerar sitt problem:

... har de allmänna filosofiska teorierna om rationalitet någonting att erbjuda, när det verkligen gäller att i en viktig konkret beslutssituation säga något om vad det innebär att gå fram rationellt? (s 1).

Men vad *betyder* det att handla "rationellt" eller "riktigt" i en konkret situation under osäkerhet eller risk?

Prawitz ger oss inte något som helst svar på denna fundamentala fråga. Vi får veta *vissa* saker om begreppet om riktighet. En handling som är fel kan vara riktig, tycks han tänka sig. Riktighet och rationalitet är något "positivt". Jag tror också att Prawitz tänker sig, att ett riktigt beslut "tar hänsyn"

till faktiskt förekommande sannolikheter. Det senare är inte så klart uttryckt och det kan förstås på endera av två sätt. Antingen så att det skulle finnas *objektiva* sannolikheter att ta hänsyn till eller så att olika utfall är mer eller mindre sannolika i relation till någon bestämd, historiskt föreliggande, bästa tänkbara *evidens*. Men också den som känner till sannolikheter av olika utfall kan rimligen reagera mycket olika på dem. Den ene tar stora risker, för att realisera betydande värden, den andre brukar en "försiktighetens" strategi. Vilket är riktigt? Begreppet är inte klart nog för att vi ska kunna ge något välgrundat svar på den frågan.

4. Slutsats

Prawitz lyckas aldrig göra klart vad det *betyder* att en handling är riktig. Han menar att en handling kan vara riktig fastän fel, att dess rationalitet är något positivt (ett skäl för oss att göra den?). Begreppet ska ha tillämpning på handlingar i konkreta fall, under risk eller osäkerhet. Min hypotes om *varför* begreppet blir så oklart som det blir, i Prawitz' framställning, är att det *inte finns* något rimligt begrepp om riktighet av den typ Prawitz är på spaning efter! Det ena svaret är inte bättre än det andra på frågan vad vi ska göra i ett konkret fall av osäkerhet. Några universellt giltiga skäl för vad vi där ska göra har vi inte, ty *per definition* saknar vi goda grunder att anta att det ena eller det andra handlingsalternativet är rätt. Olika personer reagerar då helt enkelt olika på de insikter de får om sannolikheter och möjliga utfall. Det som är ett skäl för den ene lämnar den andre oberörd. Olikheter i temperament och kynne får ett stort spelrum. (I detta konstaterande ligger ingen ringaktning för statistiskt uppskattade sannolikheter. Sådan kan vara av intresse då vi fattar beslut, kunskap om dem kan *få* oss att fatta ett beslut snarare än ett annat o s v. Vidare ska, som Prawitz påpekar, resonemang om sannolikheter ofta förstås i relation till bestämda hypoteser. En *tillräckligt* låg sannolikhet för en viss händelse kan t ex ses som en evidens för hypotesen att denna händelse ej kommer att inträffa. Känner vi till sådana sannolikheter och vet vi vad det innebär att de är *tillräckligt* höga eller låga så har vi faktiskt goda grunder i vissa fall att anta att vissa beslut snarare än andra är rätt.)

Att det ena beslutet inte kan vara mera rationellt än det andra i en konkret situation av osäkerhet eller risk kan förefalla som en omtumlande slutsats. Dess betydelse är dock begränsad. I politik är det inte ovanligt att sammanhang framställs som krångligare än de är. Ofta nog *har* vi faktiskt goda skäl att anta att något visst alternativ är rätt. (Så var enligt mitt förmenande fallet vid kärnkraftsomröstningen, då inte endast säkerhetsresonemang utan dessutom ekonomiska och politiska skäl talade starkt mot kärnkraften.) För

övrigt är det nog mera sällan vi verkligen har tillgång till statistiska data, som kan underbygga välgrundade omdömen om sannolika utfall. I dessa fall av risk *har* vi dock intet att säga om besluts rationalitet eller riktighet. Det förblir min övertygelse intill dess att någon lyckats göra reda för ett begrepp om rationalitet eller riktighet, som har tillämpning på dessa fall.

5. Sammanfattning

På två punkter är jag oenig med Dag Prawitz.

För det första är det klart, att vi i vissa konkreta fall av handlande under osäkerhet eller risk skulle handla *olika*. Prawitz visar enligt mitt förmenande en överdriven "försiktighet" i fall då livet står på spel. *Jag* är t ex beredd att acceptera en något ökad risk för hela världens undergång om därmed t ex massfattigdomen kunde bytas mot välfärd för alla. Prawitz motsätter sig detta. *Jag* är beredd att med näbbar och klor motsätta mig halva mänsklighetens utplåning också om därmed risken för *hela* mänsklighetens utplåning ökar något litet. Prawitz anbefaller en sådan politik (den följer åtminstone av hans beslutsprincip om en lexikografisk ordning av värden). Till grund för vår oenighet ligger nog en ganska olika syn på död och utplåning. För Prawitz tycks det vara något i sig alldeles förskräckligt att dö. Själv ser jag det som en fråga uteslutande om *hur* ett liv gestaltar sig, om det är värt att leva. Att vara död är bättre än att leva ett eländigt liv. Just risken för utplåning är inte mera ryslig än risken för ett liv i plågor. Ja, till *skillnad* från ett liv i plågor så är döden som sådan värderingsmässigt neutral, skulle jag säga.

Att jag avvisar Prawitz' beslutsprincip betyder inte att jag i alla lägen är beredd att söka maximera det förväntade värdet. Om jag, likt Prawitz, frågar mig vad jag ska göra i konkreta situationer av risk, så är det ju inget tillräckligt skäl att tillämpa principen om maximering av förväntat värde att denna princip *i det långa loppet* sannolikt ger det bästa utfallet. Ofta är det nog *faktiskt* ändå denna princip jag tillämpar. Men jag kan tänka mig fall, där jag säkert lever ett någorlunda hyfsat liv, då jag skulle, på maximinmånér, välja bort alla alternativ som hade som ett *realistiskt* utfall att jag dog. Jag kan också tänka mig desperata situationer, där jag lever ett eländigt liv utan mycket hopp om förbättring, där jag skulle tillämpa en princip om maximax. Led jag t ex av svår ångest, som något kunde lindras genom samtalsterapi men helt botas endast med en operation, så skulle jag kanske välja operationen, också om risken för misslyckande var mycket stor och om ett misslyckande skulle leda till ännu värre kval än min nuvarande ångest. Jag skulle alltså välja, mot alla odds, det alternativ som *om* det osannolika inträffade, skulle ge de bästa konsekvenserna. Den *enda* kända beslutsprin-

cip, som jag *inte* kan tänka mig att tillämpa (då den är i konflikt med övriga principer), är i själva verket Prawitz' idé om en lexikografisk ordning av värden med värdet av överlevande på topp!

För det andra är Prawitz och jag uppenbarligen oense om *arten* av den ovan skisserade oenigheten. Jag menar att denna oenighet endast avspeglar skillnader i temperament oss emellan. Prawitz menar att dess kärna utgörs av oförenliga åsikter om vad som är ett rationellt eller riktigt handlande under osäkerhet eller risk. Men *finns* det något sådant som rationalitet eller riktighet under osäkerhet eller risk? Prawitz har förvisso *inte visat* detta. Och mycket talar emot att så är fallet.

Representerar inte i själva verket beslutsteorins ambition att säga något om rationalitet och riktighet under risk eller osäkerhet en fåfång önskan om att överskrida den begränsning vårt vetande nu en gång är underkastat? Ja, insmyger sig inte t o m ett slags *moralism* i dessa teorier? Det blir något positivt *i sig* att tänka på ett visst sätt (förnuftigt). Men *har* vi verkligen någon befogad kritik att komma med, om vi skulle stöta på en person som alltid handlade rätt, fastän mot alla odds och efter eländiga resonemang? Inte kan det väl vara så att vi har något slags (*prima facie*) *plikt* att tänka rationellt?

På *mig* skulle en person som alltid handlade rätt mot alla odds göra ett outplånligt starkt intryck. Snarare än att kritisera denna person skulle jag vara benägen att ta upp en applåd!