

Föreläsning 7: Gruppbeslutsteori

Litteratur:

Resnik, Choices, kap. 6

1# Gruppbeslutsteori

Gruppbeslutsteorin handlar om hur enskilda människors önskemål kan vägas samman till ett socialt beslut. Resultat i gruppbeslutsteorin har på ett avgörande sätt påverkat de senaste decenniernas utveckling inom ekonomi, statskunskap, sociologi och politisk filosofi.

En tidig tänkare inom denna tradition var Jean Antoine Nicolas Caritat de Condorcet (1743–94), fransk matematiker och politiker (hade en ledande roll under Första Republiken, men dog under Skräckväldet). Charles Lutwidge Dodgson (1832–1898), brittisk matematiker och författare (till bl.a. *Alice i Underlandet* under pseudonymen Lewis Carroll). Under 1900-talet har dock två böcker kommit att bli genrebildande: *Social Choice and Individual Values* (1951) av den amerikanske nationalekonomen Kenneth Arrow (f. 1921; Riksbankens ekonomipris 1972) och *Collective Choice and Social Welfare* (1970) av den indiske nationalekonomen/filosofen Amartya Sen (f. 1933; Riksbankens ekonomipris 1998).

Filosofers intresse för gruppbeslutsteorin grundar sig delvis i kopplingarna till etik och politisk filosofi samt till (det kollektiva och det individuella) rationalitetsbegreppet. Området har dessutom blivit något av en lekplats för filosofer med tycke för predikatlogik. Bland svenska filosofer har bl.a. Bengt Hansson (f. 1943; professor i Lund), Stig Kanger (1924–1988; professor i Uppsala) och Wlodek Rabinowicz (f. 1947; professor i Lund) erhållit internationell uppmärksamhet för sitt arbete inom gruppbeslutsteorin.

2# Social choices

Ett enkelt exempel för att illustrera ett problem inom gruppbeslutsteorin är följande: Tre vänner bor i samma hus och måste fatta ett gemensamt beslut; vilken färg ska huset målas i?

- Ahmed föredrar blått framför rött och rött framför gult.
- Betty föredrar rött framför gult och gult framför blått.
- Carl föredrar gult framför blått och blått framför rött.

3# Preferensprofil

Ovanstående situation kan åskådliggöras på följande sätt (Tolkningen av bokstäverna borde vara självklar):

$$\begin{aligned} b &>_{Ar} >_{Ag} \\ r &>_{Bg} >_{Bb} \\ g &>_{Cb} >_{Cr} \end{aligned}$$

Detta är ett enkelt exempel på den berömda *Condoretparadoxen*; en majoritet föredrar b framför r och en majoritet föredrar r framför g , men en majoritet föredrar g framför b .

4# Preferensprofiler

En *preferensprofil* kännetecknas i gruppbeslutsteorin av en alternativmängd A (i detta fall $\{b, r, g\} = \{\text{"blått"}, \text{"rött"}, \text{"gult"}\}$), en mängd av individer I (i detta fall $\{A, B, C\}$).

= {"Ahmed", "Betty", "Carl"}) och en preferensordning. Det sistnämnda innebär att någon preferensrelation (indifferens eller strikt preferens) ska gälla för *varje individ* och för *varje par av alternativ*.

Det antas här att varje *individuell* ordning är en s.k. "svag ordning", dvs. att reflexivitet (för alla x i A och för alla i i I : $x \succsim_i x$), transitivitet och fullständighet i alternativmängden A gäller för varje relation $\succsim_i$, ifall i är ett element (en person) i mängden av individer I . För alla x, y i A och för alla i i I gäller alltså exakt ett av följande: (1) $x \succsim_i y$ och $y \succsim_i x$, dvs. indifferens, (2) $x \succsim_i y$ men inte $y \succsim_i x$, dvs. strikt preferens för x , eller (3) $y \succsim_i x$ men inte $x \succsim_i y$, dvs. strikt preferens för y .

5# Majoritetsregeln

En relativt enkel *regel* för att transformera en preferensprofil till en uppsättning samhälleliga preferenser är följande: Samhället föredrar (strikt) x framför y om och endast om antalet individer som föredrar x framför y är fler än antalet individer som föredrar y framför x . (Med tillägget att "samhället är indifferent mellan x och y om samhället varken föredrar x framför y eller y framför x " blir preferensuppsättningen komplett.)

Men som vi såg tidigare (i exemplet Ahmed, Betty och Carl) kan detta ge upphov till cykliska preferenser. Majoritetsregeln är alltså otillfredsställande i sådana fall.

6# Social welfare function

En SWF är en transformation som, för varje möjlig preferensprofil i en given mängd I av individer och en given mängd A av alternativ, ger en samhällelig preferensordning. Med "preferensordning" menas även här en "svag ordning", dvs. att reflexivitet, transitivitet och fullständighet gäller för samhällets preferens över alternativen i A .

Intuitionen är att en SWF är en *generell beslutsregel* för kollektiva beslut.

7# Lite perspektiv

Även för en liten mängd alternativ och en liten mängd individer finns det ganska många möjliga preferensprofiler. Om det t.ex. finns tre alternativ i A så finns det 14 möjliga ordningar för varje individ...

8# Definitionsmängd för en SWF

... och om det finns tre individer så finns det $14 \cdot 14 \cdot 14 = 2.744$ möjliga sätt att kombinera dessa ordningar. Var och en av dessa kombinationer är en preferensprofil.

En SWF ska kunna definiera en (svag) samhällelig preferensordning för *alla* dessa profiler, för att kunna sägas vara en *generell beslutsregel*.

Som sagt finns det 2.744 möjliga preferensprofiler enbart för tre alternativ och tre individer. Definitionsmängden växer snabbt med antalet alternativ och individer.

9# Kriterier för en SWF

Vilka egenskaper bör en SWF ha? Vad krävs för att en SWF ska kunna kallas rationell eller moraliskt försvarbar?

Resnik definierar sex villkor för en SWF (Dessa villkor har lite olika ursprung, men förknippas med Arrows teorem; se nedan):

- Citizens' sovereignty (CS)
- Nondictatorship (D)
- Independence of irrelevant alternatives (I)
- Positive association between individual and social values (PA)
- Unrestricted domain (U)
- The Pareto condition (P)

Dessa kommer nu att beskrivas i tur och ordning.

10# Citizens' sovereignty (CS)

"For each pair of distinct alternatives x and y there is at least one preference profile for which the SWF yields a social ordering that ranks x above y ." (Resnik, sid. 181)

Intuitionen bakom detta kriterium är att individer i gruppen (medborgarna i samhället) ska kunna avgöra om ett alternativ är bättre än ett annat; det ska inte vara förutbestämt att x är bättre än y eller att y är bättre än x .

Exempel på brott mot detta villkor: "Vad än medborgarna tycker så föredrar samhället x framför y ."

11# Nondictatorship (D)

An SWF should not be dictatorial. "An SWF is dictatorial when it always identifies the social ordering with that of one of the citizens – the same citizen being used each time." (Resnik, sid. 180)

Mer precist skulle man säga: Det finns ingen individ i sådan att samhällets preferens för varje par av alternativ $\langle x, y \rangle$ är identisk med i 's preferens för samma par. Med predikatlogiskt språk skrivs detta $\neg(\exists i \in I)(\forall x, y \in A)(x \succsim_i y \rightarrow x \succsim y)$, där $\succsim_i$ betecknar i 's preferensrelation och $\succsim$ betecknar samhällets preferensrelation.

Exempel på brott mot detta villkor: "Samhällets preferenser är desamma som Holgers preferenser."

12# Independence of irrelevant alternatives (I)

"If each citizen ranks the alternatives x and y in the same order in the preference profiles Prof_1 and Prof_2 , x and y must be in the same order with respect to each other in the social orderings that the SWF yields for Prof_1 and Prof_2 ." (Resnik, sid. 184)

Intuitionen bakom detta är att om två preferensprofiler skiljer sig åt endast med avseende på ett tredje – irrelevant – alternativ så bör detta inte påverka den samhälleliga preferensen mellan två andra alternativ. En illustration, om än inte helt ekvivalent med ett brott mot I-villkoret, av hur bisarrt det är att ta hänsyn till irrelevanta alternativ kan ges med följande dialog:

KYPAREN: Vad får det lov att vara? Vi har soppa och fisk på menyn.

RESTAURANGGÄSTEN: Då väljer jag soppa.

KYPAREN: Dessutom har vi biff på menyn.

RESTAURANGGÄSTEN: Jaså, i så fall väljer jag fisk.

13# I-villkoret (exempel)

Låt oss återgå till vännerna som ska måla huset. Betty och Carl är lite obeslutsamma och ändrar sig över natten.

Måndag: Ahmed föredrar gult framför blått framför rött.
Betty föredrar blått framför rött framför gult.
Carl föredrar rött framför gult framför blått.
(De enas om att gult är lika bra som blått.)

Tisdag: Ahmed föredrar gult framför blått framför rött.
Betty föredrar blått framför gult framför rött.
Carl föredrar gult framför blått framför rött.
(De enas om att gult är bättre än blått.)

Det är något konstigt med det ovanstående! Enda skillnaderna är att Betty har kommit att föredra gult framför rött och att Carl har kommit att föredra blått framför rött. Men i båda fallen gäller ju att Ahmed och Carl föredrar gult framför rött och att Betty föredrar rött framför gult. *Det borde inte spela någon roll* för jämförelsen mellan gult och blått att folk har ändrat sig om vad de tycker om den blåa färgen.

14# Positive association between individual and social values (PA)

“If an SWF ranks an alternative x above an alternative y for a given profile, it must also rank x above y in any profile that is exactly like the original one except that one or more citizens have moved x up in their own rankings.” (Resnik, sid. 181)

Intuitionen är helt enkelt att samhället blir mer positivt inställt till ett alternativ ju mer positivt inställd individerna är till alternativet.

Exempel: Anta att gruppen föredrar x framför y , även ifall Holger föredrar y framför x . Om Holger ändrar sig, och börjar föredra x , så borde inte gruppens preferens ändras till att föredra y – så länge ingen annan ändrar sig.

15# Unrestricted domain (U)

An SWF is defined for a specific society and a specific set of alternatives. It “produces a social ordering for every preference profile (derived from those alternatives and those citizens).” (Resnik, sid. 180)

Intuitionen är att en SWF är en *generell* beslutsregel.

16# The Pareto condition (P)

“The SWF must rank x above y for a given profile if every citizen ranks x above y in that profile.” (Resnik, sid. 182)

Ett intuitivt ganska självklart villkor! Predikatlogiskt uttryckt: $(\forall i \in I)(x \succ_i y) \rightarrow x \succ y$, där $\succ_i$ betecknar i :s preferensrelation och $\succ$ betecknar samhällets preferensrelation.

Det kan bevisas att P-villkoret är uppfyllt i alla SWF som uppfyller PA-, CS- och I-villkoren (Paretos lemma; se Resnik, sid. 185)

17# Arrow teorem

Alla de ovanstående kriterierna verkar vara rationella och moraliskt försvarbara. Det verkar till och med vara irrationellt och moraliskt förkastligt att bryta mot dem. Men, som bevisades av Kenneth Arrow i boken *Social Choice and Individual Values*, de var för sig rimliga kriterierna är oförenliga.

ARROWS THEOREM: För tre eller fler alternativ och två eller fler individer finns det ingen SWF som uppfyller villkoren CS, D, I, PA och U.

Eller ekvivalent: Om en SWF (definierad för tre eller fler alternativ och två eller fler individer) uppfyller villkoren U, P och I så måste den vara diktatorisk.

Detta kan tyckas paradoxalt, eftersom villkoren var för sig verkar vara fullt rimliga.

18# 19# Bevis för Arrows teorem: Definitioner

Resik återger fullständigt beviset för arrows teorem. Vi ska inte följa beviset i detalj här, men som översikt kan det vara bra att hålla följande definitioner i minne:

1. A set of citizens is *decisive for x over y* if and only if x is socially preferred to y whenever each member of the set prefers x to y .
2. A citizen is a *dictator for x over y* if and only if the set consisting of him alone is decisive for x over y .
3. A citizen is a *dictator* if and only if he is decisive for every pair of distinct alternatives.
4. An SWF is *dictatorial* if and only if some citizen is a dictator under it.
5. A set of citizens is *almost decisive for x over y* if and only if the social ordering ranks x over y when (a) all members of the set do and (b) all members outside prefer y to x .
6. A citizen is *almost decisive for x over y* if and only if the set consisting of him alone is almost decisive for x over y .

20# Bevis för Arrows teorem: tre steg

På samma sätt kan det hjälpa att hålla följande minnesregler i åtanke när man läser Resniks återskildring av Arrows bevis:

LEMMA 1: For any set of citizens and any pair of distinct alternatives there is at least one decisive set.

LEMMA 2: If there is a decisive set, there is a citizen who is almost decisive for some pair of alternatives.

LEMMA 3: Any citizen who is almost decisive for a single pair of alternatives is decisive for every pair of alternatives.

21# Relaxering av villkor U

Vissa beslutsregler, t.ex. majoritetsregeln, bryter endast mot villkor U. Om man begränsar sig till vissa preferensprofiler så kan man ändå formulera en SWF som uppfyller de rimliga villkoren, även om man bryter mot U.

22# Entoppiga preferenser

En sådan begränsning är att preferenser ska vara "entoppiga"; de ska kunna jämföras längs en viss "dimension" och varje individs preferens ska ha en maxpunkt längs denna dimension.

Resniks exempel; politiska kandidater:

- Individ a vill ha en kandidat som är så liberal som möjligt.
- Individ b vill ha en kandidat som är så nära mitten som möjligt.
- Individ c vill ha en kandidat som är så konservativ som möjligt.

23# Icke-entoppiga preferenser

Individuella preferenser har mer än en "maxpunkt" längs dimensionen.

Resniks exempel: Individ d är antimoderat och vill ha en kandidat som är så långt från den politiska mitten som möjligt.

Är inte d:s preferens konstig?

Det är naturligtvis inte alltid lätt att definiera preferenstoppar på detta sätt; t.ex. i fallet med husmålningen är det svårt att jämföra de tre färgerna längs en och samma "dimension".

24# Interpersonella nyttojämförelser

Observera att det ramverk vi hittills har tittat på endast formulerar preferenser som binära relationer. Man pratar inte om "nyttointervall" eller "nyttoenheter" som kan jämföras mellan olika personer.

Utilitarismen är en moralfilosofisk tradition där nytta bör maximeras och antas vara interpersonellt jämförbar.

25# Sens liberala paradox

Amartya Sen (1970) inleder paradoxen med att formulera ett villkor som han kallar "minimal liberalism" (L); det finns åtminstone två individer k och j och två olika par av alternativ $\{x, y\}$ icg $\{z, w\}$ som är sådana att k och j är bestämmande över $\{x, y\}$ respektive $\{z, w\}$.

Intuitionen bakom L-villkoret är att det finns beslut som enbart bör fattas av enskilda individer; beslutet om jag ska sova på mage eller på rygg, t.ex., är det bara jag som bestämmer och det spelar ingen roll vad andra säger om saken.

Alltså: $x \succ_k y \rightarrow x \succ y$, $y \succ_k x \rightarrow y \succ x$, $z \succ_j w \rightarrow z \succ w$ och $w \succ_j z \rightarrow w \succ z$.

Det finns ingen generell beslutsregel som uppfyller villkoren U, P och L.

26# Den liberala paradoxen: Sens exempel

En pryd man (Mr. A), en liderlig man (Mr. B) och en erotisk roman. Tre alternativ:

- (a) Mr. A läser boken.
- (b) Mr. B läser boken.
- (c) Ingen läser boken.

Mr. A vill helst att ingen läser boken, men om någon absolut måste läsa boken så föredrar han att läsa den själv än att den snuskige Mr. B får möjlighet att läsa den. Mr. B vill däremot gärna ge boken till den stele Mr. A, men om han inte kan få honom att läsa den så föredrar Mr. A att själv läsa den framför att ingen läser den. Antag att inga andra individer är inblandade i beslutssituationen.

Alltså: $c \succ_A a$, $a \succ_A b$, $a \succ_B b$ och $b \succ_B c$.

Notera att båda föredrar a framför b , vilket enligt paretovillkoret skulle innebära att "samhället" (dvs. de båda personerna) föredrar a framför b . Alltså: $a \succ b$.

MEN: Detta är oförenligt med det "liberala" villkoret, ty om vi (rimligen) antar att $c \succ_A a \rightarrow c \succ a$ och att $b \succ_{BC} \rightarrow b \succ c$ (dvs. att var och en själv bestämmer om han ska läsa boken eller inte) så får vi en motsägelse; $c \succ a$ och $b \succ c$ skulle nämligen tillsammans innebära $b \succ a$, enligt transitivitetsvillkoret (en del av U-villkoret).