La moderna teoria de l'elecció social: de la impossibilitat a la possibilitat

Jordi Massó*†

Universitat Autònoma de Barcelona i Barcelona GSE

Abril, 2012

Resum: Es presenten els dos teoremes d'impossibilitat més importants de la moderna teoria de l'elecció social: el teorema d'Arrow per a funcions de benestar social no dictatorials que satisfan el principi de Pareto i la propietat de la independència d'alternatives irrellevants, i el teorema de Gibbard-Satterthwaite per a funcions d'elecció social no trivials i no manipulables. Es descriuen set exemples de problemes concrets d'elecció social on l'estructura particular del conjunt d'alternatives socials permet restringir el domini de preferències individuals i dissenyar per a cada un d'ells funcions d'elecció social no manipulables en els corresponents dominis de preferències restringits.

Paraules clau: agregació de preferències, funció de benestar social, funció d'elecció social, teoremes d'impossibilitat, no manipulabilitat, preferència unimodal, votant medià, preferència separable, vot per comitès, problema de la divisió, regla uniforme, béns col·lectius, mètodes pivotals, subhasta de segon preu, assignació d'objectes indivisibles, nucli, algoritme d'intercanvi de millors, assignació bilateral estable, algoritme d'acceptació diferida.

Classificació MSC2010: 9102, 91A12, 91A80, 91B10, 91B12, 9114, 91B15 i 91B18.

*Departament d'Economia i d'Història Econòmica, Facultat d'Economia i Empresa, Edifici B, Universitat Autònoma de Barcelona. 08193 Bellaterra (Cerdanyola del Vallès). Correu electrònic: jordi.masso@uab.es

†L'autor agraeix els comentaris de Salvador Barberà, Dolors Berga, Núria Clos, Josep Maria Font i un avaluador, així com el suport rebut a través del premi "ICREA Acadèmia" per a l'excel·lència en la recerca, finançat per la Generalitat de Catalunya. També agraeix el suport de MOVE (on és un investigador afiliat), de la Generalitat de Catalunya, a través del projecte SGR2009-419, i del Ministerio de Ciencia e Innovación, a través dels projectes ECO2008-04756 (Grupo Consolidado-C) i CONSOLIDER-INGENIO 2010 (CDS2006-00016).

The Modern Social Choice Theory: From the Impossibility to the Possibility

Jordi Massó*

Universitat Autònoma de Barcelona and Barcelona GSE

April, 2012

Abstract: We present the two most important impossibility theorems of the modern social choice theory: Arrow's theorem for social welfare functions satisfying the Pareto principle and the independence of irrelevant alternatives property, and Gibbard-Satterthwaite's theorem for non-trivial and strategy-proof social choice functions. We describe seven examples of specific social choice problems where the particular structure of the set of social alternatives allows to restrict the domain of individual preferences and to design for each of them strategy-proof social choice functions on the corresponding restricted preference domains.

Keywords: preference aggregation, social welfare function, social choice function, impossibility theorems, strategy-proofness, single-peaked preference, median voter, separable preference, voting by committees, division problem, uniform rule, collective goods, pivotal methods, second price auction, assignment of indivisible objects, core, top trading cycle algorithm, stable matching, deferred acceptance algorithm.

MSC2010 Classification: 9102, 91A12, 91A80, 91B10, 91B12, 9114, 91B15 i 91B18.

*Departament d'Economia i d'Història Econòmica, Facultat d'Economia i Empresa, Edifici B, Universitat Autònoma de Barcelona. 08193 Bellaterra (Cerdanyola del Vallès). E-mail: jordi.masso@uab.es

1 Introducció

La teoria de l'elecció social estudia els procediments pels quals les societats prenen decisions col·lectives tenint en compte les preferències dels seus membres.[1] Es considera la publicació del llibre d'Arrow (1951) *Social Choice and Individual Values* com el naixement de la moderna teoria de l'elecció social.[2] Arrow es pregunta si és possible considerar una societat, composta per diversos agents racionals, com si fos un únic agent prenent decisions racionals. En particular, es pregunta si existeixen procediments amb propietats desitjables per agregar preferències individuals, potencialment diferents, en una única preferència social. El teorema d'impossibilitat d'Arrow respon a aquesta pregunta negativament. Considerem una societat amb n agents. Cada un d'ells té preferències que ordenen el conjunt A d'alternatives socials. Una *funció de benestar social* assigna, a cada llista de n preferències individuals en A (anomenada perfil de preferències o perfil), una preferència social en A. Arrow considera que una funció de benestar social hauria de satisfer les següents propietats.

- *Racionalitat individual i social.* Tant les preferències individuals com la preferència social són relacions binàries totals i transitives en A.
- *Domini universal.* Qualsevol preferència individual és legítima.
- *Principi de Pareto.* Si tots els agents consideren que una alternativa és millor que una altra, l'ordenació social de les dues alternatives ha de coincidir amb l'ordenació unànime.
- *Independència d'alternatives irrellevants.* L'ordenació social entre dues alternatives només depèn de les ordenacions individuals entre elles, i no de com les preferències individuals ordenen les altres alternatives.
- *No dictatorial.* No hi ha un agent la preferència del qual coincideix sempre amb la preferència social, independentment de les preferències dels altres agents.

El teorema d'Arrow diu que les cinc propietats són incompatibles: no existeix cap funció de benestar social que les satisfaci simultàniament. Hi ha una extensa literatura, ja insinuada per Black (1948) i continuada per Sen (1969),[3] que proposa propietats alternatives de les funcions de

[1] Els mètodes de votació són exemples d'aquests procediments. En les seccions 3 i 4 es presentaran diferents problems d'elecció social i procediments per resoldre'ls.

[2] Kenneth Arrow neix a New York el 1921. Actualment és professor emèrit del Departament d'Economia de la Stanford University. Arrow és un dels millors economistes del segle XX i rep el premi Nobel d'economia l'any 1972, juntament amb John Hicks, "per les seves contribucions pioneres a la teoria de l'equilibri general i a la teoria del benestar". La moderna teoria de l'elecció social té molts antecedents (el llibre de McLean i Urken (1995) en descriu els més importants). Entre ells, en destaquem tres. (1) Els tres escrits de Ramon Llull *Artificium electionis presonarum* (1274), *En qual manera Natana fo eleta a abadessa* (1283) i *De arte electionis* (1299). (2) L'escrit del filòsof, matemàtic i historiador francès Marie Jean Antoine Nicolas de Caritat, marquès de Condorcet ($1743-1794$) *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* (1785). (3) L'escrit del matemàtic i fisic francès Jean Charles de Borda ($1733-1799$) *Mémoire sur les élections au scrutin* (1781).

[3] Amartya Sen neix a Shantiniketan (India) el 1933. Actualment és professor d'economia i filosofia a la Harvard University. Rep el premi Nobel d'economia l'any 1998 "pel seu treball en el camp de l'economia del benestar". L'any 1997 fou guardonat amb el Premi Internacional Catalunya concedit per la Generalitat de Catalunya.

benestar social amb l'objectiu d'obtenir resultats de possibilitat. En aquest article no seguirem aquesta literatura sinó una altra, iniciada simultàniament i de forma independent per Gibbard (1973) i Satterthwaite (1975),[4] que modifica el context d'Arrow i situa el problema dels incentius estratègics en el centre de l'anàlisi. Sovint, per a resoldre el problema d'elecció social no és necessària una preferència social, n'hi ha prou amb seleccionar una alternativa social. En aquests casos ens interessa que el resultat de l'agregació de les n preferències individuals sigui una elecció, l'alternativa escollida per la societat en el perfil considerat, en comptes d'una ordenació total de totes les alternatives. Una *funció d'elecció social* assigna, a cada perfil de n preferències individuals, una única alternativa social. Una funció d'elecció social pot ser entesa com un mecanisme que sol·licita a cada un dels agents les seves preferències, i tenint-les en compte, escull l'alternativa social. Però, com que les preferències individuals són informació privada i l'elecció social pot dependre d'elles, quan un agent ha de revelar les seves preferències al mecanisme, es pot preguntar: quina preferència em convé revelar? En general, la resposta a aquesta pregunta depèn de la conjectura que l'agent faci sobre les preferències revelades pels altres agents. Una funció d'elecció social genera un problema d'incentius estratègics. Les funcions d'elecció social no manipulables són aquelles que eliminen el comportament estratègic dels agents. Una funció d'elecció social és *no manipulable* si els agents sempre volen declarar la seva verdadera preferència, ja que mai hi surten guanyant declarant unes preferències diferents a les pròpies. Per tant, una funció d'elecció social no manipulable indueix un problema d'optimització a cada agent (trobar la resposta a la pregunta: quina preferència em convé declarar?) que té solució (declarar la verdadera preferència és una solució òptima) independentment de les conjectures que l'agent pugui fer sobre les preferències declarades pels altres agents. A més, i des del punt de vista normatiu, la no manipulabilitat de la funció d'elecció social assegura que aquesta utilitza la informació correcta (el perfil de preferències verdaderes) i per tant, l'alternativa escollida satisfà altres propietats desitjables, relatives al perfil de preferències verdader.

El teorema de Gibbard-Satterthwaite ens diu que, excepte les trivials, no és possible dissenyar funcions d'elecció social no manipulables en el domini universal (quan totes les preferències individuals són legítimes).

L'objectiu d'aquest article és doble. Primer, presentar els dos resultats d'impossibilitat mencionats, que suggereixen les dificultats de dissenyar procediments de presa de decisió col·lectiva satisfactoris en el domini universal de preferències individuals. Segon, i a causa de l'interès en identificar funcions d'elecció social no manipulables en àmbits d'aplicació reduïts, donar set exemples de problemes d'elecció social on la hipòtesi de domini universal deixa de ser raonable i per als quals existeixen (sovint moltes) funcions d'elecció social no manipulables en el corresponent domini restringit de preferències individuals. Per aquests casos coneixem classes, o subclasses especialment interessants, de funcions d'elecció social no manipulables i les seves caracteritzacions a partir de conjunts de propietats desitjables. Els dos primers exemples tenen en comú que les alternatives socials no tenen components privats (i per tant, les restriccions del domini de preferències són anomenades de béns públics), mentre que en els altres cinc les alternatives socials tenen

[4] Allan Gibbard és professor de filosofia a la University of Michigan i Mark Satterthwaite és professor de direcció estratègica i economia de l'empresa a la Northwestern University.

components privats, aquells que només afecten a cada un dels agents (i per tant, les restriccions del domini de preferències són anomenades de béns privats).

L'estructura de l'article és la següent. En la secció 2 definim primer les peces fonamentals de tots els models estudiats: els agents, el conjunt d'alternatives socials i les preferències que els agents tenen sobre el conjunt d'alternatives. Després, presentem els dos resultats d'impossibilitat més importants de la moderna teoria de l'elecció social: el teorema d'Arrow per a funcions de benestar social i el teorema de Gibbard-Satterthwaite per a funcions d'elecció social no manipulables. També relacionem les propietats de la independència d'alternatives irrellevants de les funcions de benestar social monòtones i la no manipulabilitat de les funcions d'elecció social i, al final de la secció, suggerim la restricció del domini de preferències individuals com una via per a obtenir resultats de possibilitat. En les seccions 3 i 4, el nucli de l'article, presentem resultats de possibilitat indicant com es poden dissenyar funcions d'elecció social no manipulables en dominis de preferències restringits. La secció 3 conté les restriccions de domini de béns públics. El primer resultat és la caracterització de les extensions del "*votant medià*" com la classe de funcions d'elecció social no manipulables en el domini de preferències unimodals en un conjunt unànimement ordenat. El segon resultat és la caracterització del *vot per comitès* com la classe de funcions d'elecció no manipulables (i exhaustives) tant en el domini de preferències separables com en el de preferències additives pel problema d'elecció d'un subconjunt d'entre un conjunt donat de candidats. En la secció 4 presentem cinc restriccions de domini de bé privat. Per a la primera presentem dues caracteritzacions de la *regla uniforme* per resoldre el problema de la divisió d'un bé homogeni i perfectament divisible quan cada un dels agents té preferències unimodals en la part rebuda del bé (el component privat de les alternatives socials). Per a la segona describim totes les funcions d'elecció social no manipulables, que són extensions dels *mètodes pivotals*, per l'elecció d'un bé col·lectiu binari quan els agents tenen preferències quasi-lineals en el consum del bé i el preu pagat. Per a la tercera identifiquem la *subhasta de segon preu* com una funció d'elecció social no manipulable per assignar un objecte indivisible a un únic agent d'entre un conjunt d'agents, quan aquests estan interessats en rebre l'objecte i en pagar el preu més baix possible. Per a la quarta presentem la caracterització del *nucli* (computat a partir de l'algoritme d'intercanvi de millors de Gale) com l'única funció d'elecció social no manipulable per resoldre el problema d'assignar un conjunt d'objectes indivisibles a un conjunt d'agents a qui només interessa l'objecte rebut per ells, i no l'objecte assignat als altres agents. Per a la quinta i última presentem els dos *algoritmes d'acceptació diferida* de Gale i Shapley, els únics (parcialment) no manipulables per resoldre el problema d'assignació bilateral entre dos subconjunts disjunts d'agents quan aquests només estan interessats en l'agent a qui són assignats.

2 Preliminars, impossibilitats i restriccions de domini

2.1 Agents, alternatives i preferències

Considerem una societat formada per un conjunt d'*agents* $N = \{1, ..., n\}$, on $n \geq 2$, que ha de prendre col·lectivament una decisió que afecta a tots els seus membres. Sigui A el conjunt de les possibles decisions o *alternatives* socials. Depenent del problema de decisió social A pot ser un

conjunt finit $\{x, y, ..., z\}$, un subconjunt de $\mathbb{R}^k$, on $k \geq 1$, un espai mètric, o un espai més abstracte. Cada agent $i \in N$ té unes *preferències* (o gustos sobre el conjunt d'alternatives) representades per una relació binària R_i completa i transitiva en A; és a dir, i és capaç de comparar qualsevol parell d'alternatives amb un minim requeriment de racionalitat. Per un parell $x, y \in A$, diem que xR_iy si l'agent i considera que l'alternativa x és almenys tan preferida com l'alternativa y. Donades les preferències R_i, definim primer les preferències estrictes de l'agent i com la relació binària P_i en A, on per a cada parell $x, y \in A$, xP_iy si i només si, $\neg yR_ix$ (no és cert que yR_ix). Si xP_iy, interpretem que l'agent i considera que l'alternativa x és millor que (o és estrictament preferida a) l'alternativa y. És fàcil comprovar que P_i és irreflexiva i transitiva. També definim la indiferència com la relació binària I_i en A, on per a cada parell $x, y \in A$, xI_iy si i només si, xR_iy i yR_ix. Si xI_iy, interpretem que l'agent i és indiferent entre les alternatives x i y. És fàcil comprovar que I_i és reflexiva i transitiva.

Denotarem per $\mathcal{R}$ el conjunt de totes les relacions binàries totals i transitives en A; per tant, $\mathcal{R}$ és el conjunt de totes les possibles preferències de cada un dels agents. Sovint, serà convenient referir-nos a situacions on els agents no són indiferents entre cap parell d'alternatives diferents (si $x \neq y$ llavors xR_iy o yR_ix, però no ambdós). En aquests casos direm que l'agent i té preferències estrictes en A i denotarem per $\mathcal{P}$ al subconjunt de $\mathcal{R}$ de preferències estrictes en A. Un *perfil* (de preferències) $R = (R_1, ..., R_n) \in \mathcal{R}^n$ és una llista de n preferències, una per a cada agent. Un perfil descriu les preferències que tenen tots els agents de la societat sobre el conjunt d'alternatives A.

2.2 El teorema d'impossibilitat d'Arrow

Arrow (1951) es pregunta si és possible agregar de forma sistemàtica els perfils de preferències individuals en una preferència social. La formulació d'aquesta pregunta, i la seva resposta, es considera com el naixement de la moderna teoria de l'elecció social.

Una *funció de benestar social* $F : \mathcal{R}^n \to \mathcal{R}$ és un procediment per obtenir preferències socials a partir de les preferències individuals. Per a cada perfil de preferències $R = (R_1, ..., R_n) \in \mathcal{R}^n$, la preferència social $F(R)$ és una relació binària total i transitiva que ordena el conjunt A, potencialment tenint en compte el perfil de preferències individuals R.

La mateixa formulació d'una funció de benestar social imposa dos supòsits implícits sobre el procediment d'agregació. El primer, conegut com el de *domini universal*, és que totes les preferències individuals són possibles (el domini de F és $\mathcal{R}^n$).[5] El segon és que la preferència social és una relació binària en A amb les mateixes propietats de totalitat i transitivitat que les preferències individuals: per a tot $R \in \mathcal{R}^n$, $F(R) \in \mathcal{R}$. Del primer, en parlarem més endavant. Per adonar-nos que el segon supòsit implícit elimina procediments d'agregació potencialment interessants, considerem el següent exemple, conegut com el de la paradoxa del vot de Condorcet.

Exemple 1 Considerem una societat $N = \{1, 2, 3\}$ amb tres agents i un conjunt $A = \{x, y, z\}$ amb tres alternatives. Per simplificar, suposarem que les preferències són estrictes. El procediment d'agregació M de la *majoria simple* (en $\mathcal{P}$) consisteix en que, per a cada perfil $P = (P_1, P_2, P_3) \in \mathcal{P}^3$

[5] Donat un subconjunt de preferències $\mathcal{D} \subseteq \mathcal{R}$, definim la funció de benestar social *en* $\mathcal{D}$ com $F : \mathcal{D}^n \to \mathcal{R}$. Sovint ens referirem a $\mathcal{D}$ (o a $\mathcal{D}^n$) com el *domini* de preferències. En particular, considerarem funcions de benestar social $F : \mathcal{P}^n \to \mathcal{R}$ en $\mathcal{P}$.

i cada parell d'alternatives $x, y \in A$, diem que

$$xM(P)y \iff \#\{i \in N \mid xP_iy\} > \#\{i \in N \mid yP_ix\}.^6$$

Interpretem $xM(P)y$ com que socialment x és estrictament preferida a y en el perfil P. Considerem el perfil $P = (P_1, P_2, P_3) \in \mathcal{P}^3$ representat per

P_1	P_2	P_3
x	y	z
y	z	x
z	x	y

on, per exemple, la primera columna significa que xP_1y, yP_1z i xP_1z (que en general escriurem com xP_1yP_1z). Aplicant la majoria simple a cada parell d'alternatives obtenim:

- $xM(P)y$ ja que xP_1y, xP_3y i yP_2x.
- $yM(P)z$ ja que yP_1z, yP_2z i zP_3y.
- $zM(P)x$ ja que zP_2x, zP_3x i xP_1z.

La preferència social $M(P)$ no és transitiva, ja que té el cicle $xM(P)yM(P)zM(P)x$ (és a dir, $xM(P)y$, $yM(P)z$ i $\neg xM(P)z$), i per tant $M(P) \notin \mathcal{P}$. El procediment d'agregació M de la majoria simple no és una funció de benestar social en $\mathcal{P}$. □

Arrow proposa dues propietats desitjables de les funcions de benestar social. La primera considera que si tots els agents ordenen estrictament dues alternatives de la mateixa manera, aleshores l'ordenació social de les dues alternatives ha de respectar l'ordenació unànime dels agents. Aquesta propietat es coneix com el principi de Pareto.

Definició 1 Una funció de benestar social $F : \mathcal{R}^n \to \mathcal{R}$ satisfà el *principi de Pareto* quan per a tot perfil $R = (R_1, ..., R_n) \in \mathcal{R}^n$ i tot parell d'alternatives $x, y \in A$, si xP_iy per a tot $i \in N$ llavors $\neg yF(R)x$.

La segona propietat respon al principi que l'ordre social entre dues alternatives x i y només pot dependre de les ordenacions individuals entre x i y, i no de les ordenacions individuals entre, per exemple, x i z, y i z, i z i w. Es coneix com la propietat de la independència d'alternatives irrellevants.

Definició 2 Una funció de benestar social $F : \mathcal{R}^n \to \mathcal{R}$ satisfà la propietat de la *independència d'alternatives irrellevants* quan per a tot parell de perfils $R, R' \in \mathcal{R}^n$ i tot parell d'alternatives $x, y \in A$, si per a cada i, xR_iy si i només si xR'_iy, llavors $xF(R)y$ si i només si $xF(R')y$.

El principi de Pareto exclou, entre d'altres, les funcions constants. La independència d'alternatives irrellevants elimina funcions de benestar social potencialment interessants com el procediment de votació per puntuació, conegut com el *compte de Borda.* El següent exemple ens ho mostra.

[6] Denotem la cardinalitat d'un conjunt X per $\#X$.

Exemple 2 Considerem una societat $N = \{1, 2, 3, 4\}$ amb quatre agents i un conjunt $A = \{x, y, z, w\}$ amb quatre alternatives socials. Per simplificar, suposem que les preferències dels agents són estrictes. El compte de Borda $B : \mathcal{P}^4 \to \mathcal{R}$ és la funció de benestar social en $\mathcal{P}$ que consisteix en el següent procediment. Donat un perfil $P = (P_1, P_2, P_3, P_4) \in \mathcal{P}^4$, per a cada agent s'assignen 3 punts a la seva millor alternativa, 2 a la segona, 1 a la tercera i 0 a la quarta.[7] Es sumen els punt obtinguts per a cada alternativa i aquestes s'ordenen de forma descendent segons aquests punts. Aquest ordre és $B(P) \in \mathcal{R}$. Considerem els dos perfils $P, P' \in \mathcal{P}^4$ representats per

P_1	P_2	P_3	P_4	P'_1	P'_2	P'_3	P'_4
x	x	y	y	z	w	y	y
z	w	w	z	w	z	w	z
w	z	x	x	x	x	x	x
y	y	z	w	y	y	z	w

Aplicant el compte de Borda a les alternatives x i y, obtenim que $xB(P)y$ (socialment, x és estrictament preferida a y en el perfil P) i $yB(P')x$ (socialment, y és estrictament preferida a x en el perfil P'). Per comprovar-ho, veiem que en el perfil P, x obté 8 punts (3 dels agents 1 i 2 i 1 dels agents 3 i 4) i y obté 6 punts (0 dels agents 1 i 2 i 3 dels agents 3 i 4), mentre que en el perfil P', x obté 4 punts (1 dels quatre agents) i y obté 6 punts (0 dels agents 1 i 2 i 3 dels agents 3 i 4). Però observem que

$$\{i \in N \mid xP_iy\} = \{i \in N \mid xP'_iy\} = \{1, 2\} \text{ i } \{i \in N \mid yP_ix\} = \{i \in N \mid yP'_ix\} = \{3, 4\}.$$

Per tant, el compte de Borda no satisfà la propietat de la independència d'alternatives irrellevants. □

Les dificultats presentades en els exemples 1 i 2 es resolen en part en considerar funcions d'elecció social, les quals es limiten a escollir, per a cada perfil de preferències, una única alternativa social en comptes d'una preferència social. Les funcions d'elecció social passaran a ser l'objecte del nostre interès a partir de la propera subsecció.

Finalment, una propietat no desitjable d'una funció de benestar social és la de ser dictatorial, que consisteix en preseleccionar un agent (el dictador) i fer que, per a cada perfil, la funció assigni com a preferència social la preferència del dictador.

Definició 3 Una funció de benestar social $F : \mathcal{R}^n \to \mathcal{R}$ és *dictatorial* quan existeix un agent $i \in N$ tal que per a tot perfil $R = (R_1, ..., R_n) \in \mathcal{R}^n$, $F(R) = R_i$.

El teorema d'Arrow ens mostra la impossibilitat d'agregar les preferències individuals en una preferència social de manera satisfactòria.

Teorema d'Arrow (Arrow, 1951) *Suposem que $\#A \geq 3$. La funció de benestar social $F : \mathcal{R}^n \to \mathcal{R}$ satisfà el principi de Pareto i la propietat de la independència d'alternatives irrellevants si i només si F és dictatorial.*

[7] El número de punts assignats a l'alternativa x en la preferència d'un agent es pot interpretar com el número d'alternatives pitjors a x.

Abans de procedir amb les idees principals de la demostració del teorema quatre comentaris són pertinents.

Primer, el supòsit que $\#A \geq 3$ és imprescindible. Altrament, si $A = \{x, y\}$ el sistema de majoria simple $F : \mathcal{R}^n \to \mathcal{R}$ (en $\mathcal{R}$) definit com, per a cada $R \in \mathcal{R}^n$,

$$xF(R)y \Longleftrightarrow \#\{i \in N \mid xR_iy\} \geq \#\{i \in N \mid yR_ix\},$$

és una funció de benestar social no dictatorial que satisfà el principi de Pareto i la propietat de la independència d'alternatives irrellevants (que quan $\#A = 2$ esdevé una propietat innòcua). May (1952) caracteritza pel cas $\#A = 2$ la majoria simple com l'única funció de benestar social anònima (tots els agents són tractats de la mateixa manera), neutral (totes les alternatives són tractades de la mateixa manera) i monòtona (si un agent canvia la seva preferència millorant en el seu ordre individual l'alternativa socialment millor, el resultat de la funció de benestar social ha de ser el mateix).

Segon, és sorprenent però no hi ha moltes funcions de benestar social que siguin no dictatorials i compleixin la propietat de la independència d'alternatives irrellevants simultàniament. Si prescindim del principi de Pareto en el teorema d'Arrow, només apareixen (i) les funcions constants que consisteixen en preseleccionar una preferència $\bar{R} \in \mathcal{R}$ i establir que $F(R) = \bar{R}$ per a tot $R \in \mathcal{R}^n$ i (ii) les funcions antidictatorials que consisteixen en preseleccionar un agent $i \in N$ i definir, per a tot $R \in \mathcal{R}^n$, $F(R) = R_i^c$, on R_i^c és l'ordre invers de R_i (si xP_iy llavors yP_i^cx i si xI_iy llavors xI_i^cy).[8]

Tercer, totes les variacions del compte de Borda (les regles de puntuació on els punts assignats a les alternatives són un vector $(s_1, ..., s_{\#A})$ decreixent) satisfan el principi de Pareto i no són dictatorials, però vulneren la propietat de la independència d'alternatives irrellevants.

Quart, i com ja hem dit abans, el teorema té dos supòsits implícits: el del domini universal i que la preferència social ha de ser una relació binària total i transitiva en A. En relació al primer supòsit implícit, si les preferències dels agents són estrictes, el teorema continua sent verdader; és a dir, si $\#A \geq 3$, una funció de benestar social $F : \mathcal{P}^n \to \mathcal{R}$ en $\mathcal{P}$ satisfà el principi de Pareto i la propietat de la independència d'alternatives irrellevants si i només si F és dictatorial. Hi ha una extensa literatura que es pregunta si és possible obtenir funcions d'agregació interessants debilitant els requeriments exigits a les preferències socials (en particular, el de transitivitat).[9]

L'estructura de la demostració del teorema d'Arrow és la següent. Primer, és immediat comprovar que qualsevol funció de benestar social dictatorial satisfà el principi de Pareto i la propietat de la independència d'alternatives irrellevants. Per veure que l'altre implicació és certa, suposem que $F : \mathcal{R}^n \to \mathcal{R}$ satisfà el principi de Pareto i la propietat de la independència d'alternatives irrellevants. S'ha de demostrar que F és dictatorial. L'obtenció del dictador es basa en l'anàlisi del poder que tenen alguns grups d'agents (*a priori*, amb més d'un agent) anomenats coalicions decisives. Els agents d'una coalició decisiva per un parell d'alternatives poden imposar l'ordenació social entre les dues alternatives declarant una ordenació unànime entre elles. La demostració procedeix en dos passos. Primer, demostrar que si una coalició és decisiva per un parell d'alternatives

[8] Veure Wilson (1972).

[9] Exemples de noves impossibilitats o de possibilitats (aquestes, no massa substancials) es poden trobar a Sen (1969), Mas-Colell i Sonnenschein (1972), Blau i Deb (1977), Deb (1981), Blair i Pollak (1982) i Pattanaik i Peleg (1986).

llavors és decisiva per a qualsevol parell. Segon, demostrar que qualsevol partició d'una coalició decisiva en dos subcoalicions té la propietat de que una, i només una, de les subcoalicions és decisiva. Aleshores, començant per N, que pel principi de Pareto és una coalició decisiva, i aplicant reiteradament el segon pas, obtenim que existeix un únic agent $i \in N$ tal que la coalició $\{i\}$ és decisiva per a qualsevol parell d'alternatives. Per tant, F és dictatorial.[10]

2.3 El teorema d'impossibilitat de Gibbard-Satterthwaite

La noció de funció de benestar social és molt ambiciosa ja que exigeix que l'agregació de les preferències dels agents generi una preferència social sobre tot el conjunt d'alternatives. Molt sovint però, observem que els procediments pels quals les societats prenen decisions col·lectives consisteixen en seleccionar una única alternativa social a partir de les opinions individuals dels agents.

A principis dels anys 70, i de forma independent, Gibbard (1973) i Satterthwaite (1975) reformulen el context d'Arrow i introdueixen en l'anàlisi el problema dels incentius estratègics que els agents han de resoldre en revelar les seves preferències: la possible manipulabilitat de la funció d'elecció social.[11]

Una *funció d'elecció social* $f : \mathcal{R}^n \to A$ representa un procediment per escollir, per a cada perfil de preferències, una alternativa social.

Ara, en el perfil $R \in \mathcal{R}^n$, la societat escull una única alternativa social $f(R) \in A$. Per tant, té sentit preguntar-se com cada agent i avalua les conseqüències de subministrar a la funció d'elecció social una preferència (la verdadera R_i) o una altra preferència ($R'_i \neq R_i$). Serem exigents i voldrem que aquesta pregunta tingui sentit independentment del grau d'informació que cada agent té sobre les preferències dels altres agents. La noció de *no manipulabilitat* recull aquesta exigència, però per a definir-la necessitem la següent notació. Donat el perfil $R = (R_1, ..., R_n) \in \mathcal{R}^n$ i la preferència $R'_i \in \mathcal{R}$ de l'agent i, denotem per $(R'_i, R_{-i}) \in \mathcal{R}^n$ el perfil de preferències obtingut a partir de substituir en R la preferència R_i per R'_i; per exemple, si $1 < i < n$, $(R'_i, R_{-i}) = (R_1, ..., R_{i-1}, R'_i, R_{i+1}, ..., R_n) \in \mathcal{R}^n$.

Definició 4 Un agent $i \in N$ *manipula* la funció d'elecció social $f : \mathcal{R}^n \to A$ quan existeix un perfil $R \in \mathcal{R}^n$ i una preferència $R'_i \in \mathcal{R}$ tals que

$$f(R'_i, R_{-i}) P_i f(R_i, R_{-i}).$$

En aquest cas diem que i *manipula* f *en el perfil* R *declarant* R'_i. Una funció d'elecció social $f : \mathcal{R}^n \to A$ és *no manipulable* quan cap agent pot manipular-la.

Abans de procedir, dues consideracions i un exemple són pertinents. La primera és fer notar que perquè un agent manipuli una funció d'elecció social $f : \mathcal{R}^n \to A$ només cal que hi hagi un perfil i

[10] Hi ha moltes demostracions alternatives del teorema d'Arrow. Per exemple Barberà (1980) es concentra directament en coalicions decisives amb un únic agent, anomenats votants pivotals (aquells que tenen poder per alterar l'ordre social). Veure també Barberà (1983b), Reny (2001), Ubeda (2003), Geanokoplos (2005) i Man i Takayama (2011).

[11] Arrow ja és conscient d'aquest problema en el context de les funcions de benestar social, però ho deixa al marge de la seva anàlisi.

una preferència alternativa per a l'agent, on aquest obtingui, canviant la seva preferència declarada, una alternativa estrictament preferida (d'acord amb la seva *verdadera* preferència). La segona és observar que una funció d'elecció social indueix un joc en forma normal on el conjunt de jugadors és el conjunt d'agents i el conjunt d'estratègies de cada jugador és el conjunt de totes les preferències individuals. Aleshores, la no manipulabilitat d'una funció d'elecció social és equivalent a que declarar la verdadera preferència sigui una estratègia (dèbilment) dominant en el joc induït; és a dir, independentment de les preferències declarades pels altres agents, dir la verdadera preferència és sempre una de les estratègies òptimes. La propietat de no manipulabilitat és molt exigent, però té l'enorme avantatge que els agents no necessiten conjecturar el comportament dels altres agents, i per tant no és necessari fer hipòtesis sobre el grau d'informació que tenen sobre les preferències dels altres agents. Revelar la verdadera preferència és el millor que poden fer, independentment de les creences que tinguin sobre les preferències dels altres agents.[12] Per il·lustrar el concepte de manipulabilitat, considerem el següent exemple.

Exemple 3 Considerem una societat $N = \{1, 2\}$ amb dos agents i un conjunt $A = \{x, y, z\}$ amb tres alternatives socials. Per simplificar, suposem que les preferències dels agents són estrictes.[13] Considerem la següent funció d'elecció social $f : \mathcal{P}^2 \to \{x, y, z\}$ en $\mathcal{P}$. Fixat un perfil, suposem que cada agent veta una alternativa (la pitjor en la seva preferència estricta). Si només queda una alternativa sense ser vetada, aquesta és l'escollida per f en el perfil; si en queden dues, s'escull la millor alternativa de l'agent 2 (qui té per tant un vot de qualitat). Considerem el següent perfil $P = (P_1, P_2) \in \mathcal{P}^2$ i la següent preferència $P_1' \in \mathcal{P}$ de l'agent 1 representats per

P_1	P_2	P_1'
x	y	x
y	x	z
z	z	y

En el perfil (P_1, P_2), els dos agents veten z i surt elegida l'alternativa y, la preferida per l'agent 2 entre les no vetades x i y. Per tant, $f(P_1, P_2) = y$. En el perfil (P_1', P_2), les alternatives vetades són y i z i per tant, $f(P_1', P_2) = x$. Efectivament, l'agent 1 manipula f en el perfil (P_1, P_2) declarant P_1' ja que $x = f(P_1', P_2) P_1 f(P_1, P_2) = y$. □

El teorema de Gibbard-Satterthwaite ens indica la impossibilitat de dissenyar funcions d'elecció social interessants no manipulables. Per enunciar el teorema necessitem dues peces addicionals de notació i una definició.

El *rang* d'una funció d'elecció social $f : \mathcal{R}^n \to A$ és el conjunt

$$r_f = \{x \in A \mid \text{existeix } R \in \mathcal{R}^n \text{ tal que } x = f(R)\}.$$

[12] Jackson (2001) dóna una visió panoràmica d'aquesta anàlisi estratègica quan s'exigeix a les funcions d'elecció social propietats més dèbils que la de no manipulabilitat.

[13] Tal i com hem fet per les funcions de benestar social, donat un subconjunt de preferències $\mathcal{D} \subseteq \mathcal{R}$, definim la funció d'elecció social *en* $\mathcal{D}$ com $f : \mathcal{D}^n \to A$. Sovint ens referirem a $\mathcal{D}$ (o a $\mathcal{D}^n$) com el *domini* de preferències. En particular, considerarem funcions d'elecció social $f : \mathcal{P}^n \to A$ en $\mathcal{P}$. Al principi de la secció 3 definirem formalment una funció d'elecció social en $\mathcal{D}$ no manipulable.

Donats un subconjunt no buit d'alternatives $B \subseteq A$ i una preferència individual $R_i \in \mathcal{R}$, denotem el conjunt d'*alternatives maximals* de B segons R_i per

$$m(R_i, B) = \{x \in B \mid xR_iy \text{ per a tot } y \in B\}.$$

Aquest conjunt pot ser buit, però si A és finit i $R_i \in \mathcal{P}$ és una preferència estricta, aleshores $\#m(P_i, B) = 1$ per a tot subconjunt no buit B; en aquest cas, també representem aquest únic element per $m(P_i, B)$.

Definició 5 Una funció d'elecció social $f : \mathcal{R}^n \to A$ és *dictatorial* quan existeix un agent $i \in N$ tal que per a tot perfil $R = (R_1, ..., R_n) \in \mathcal{R}^n$, $f(R) \in m(R_i, r_f)$.

Una funció d'elecció social dictatorial consisteix en preseleccionar un agent i un subconjunt d'alternatives (el que serà el rang de la funció), i a cada perfil assignar-li una de les millors alternatives de l'agent en el subconjunt. Per tant, qualsevol funció constant és (trivialment) dictatorial.

El teorema de Gibbard-Satterthwaite (Gibbard (1973) i Satterthwaite (1975)) ens diu que totes les funcions d'elecció social són manipulables, excepte les trivials (dictatorials, o amb dues alternatives en el rang). Per tant, no podem evitar de manera satisfactòria el comportament estratègic dels agents quan totes les preferències individuals són legítimes: declarar la verdadera preferència no és sempre un comportament òptim, independentment de les preferències declarades pels altres agents.

Teorema de Gibbard-Satterthwaite *Sigui $f : \mathcal{R}^n \to A$ una funció d'elecció social tal que $\#r_f \neq 2$. Llavors, f és no manipulable si i només si f és dictatorial.*

Abans de procedir amb les idees principals de la demostració del teorema, tres comentaris són pertinents.

Primer, el teorema de Gibbard-Satterthwaite també té el supòsit implícit de domini universal. Això vol dir que els agents poden declarar com a pròpies *totes* les preferències en A. Si les preferències dels agents són estrictes (suposició plausible quan A és finit) el resultat d'impossibilitat continua sent cert; en particular, les úniques funcions d'elecció social $f : \mathcal{P}^n \to A$ en $\mathcal{P}$ no manipulables amb un rang amb una cardinalitat diferent de 2 són dictatorials. En les seccions 3 i 4 veurem com, en contextos d'elecció social específics, quan té sentit fer la hipòtesi que no totes les preferències individuals són raonables, es poden dissenyar funcions d'elecció social no manipulables no trivials.

Segon, si $\#r_f = 1$ (f és constant i per tant no manipulable) aleshores f és trivialment dictatorial (tots els agents són dictadors) ja que per a tot perfil $R \in \mathcal{R}^n$, $m(R_i, r_f) = r_f$ per a tot $i \in N$.

Tercer, si $\#r_f = 2$ la conclusió del teorema no és certa ja que la majoria simple no és ni dictatorial ni manipulable. De fet, aquesta situació és gairebé equivalent a la d'Arrow (ho és quan les preferències socials són estrictes) ja que podem identificar de forma natural la funció d'elecció social amb una funció de benestar social, i viceversa. Efectivament, la funció d'elecció social que consisteix en preseleccionar dues alternatives del conjunt A, per exemple x i y, i per a cada perfil $R = (R_1, ..., R_n) \in \mathcal{R}^n$ definir

$$f^{x,y}(R) = \begin{cases} x & \text{si } \#\{i \in N \mid xR_iy\} \geq \frac{n}{2} \\ y & \text{si } \#\{i \in N \mid xR_iy\} < \frac{n}{2} \end{cases}$$

és no manipulable, no dictatorial i $r_{f^{x,y}} = \{x, y\}$. Recordeu que en l'exemple 1, a on la majoria simple generava un cicle, el conjunt A tenia tres alternatives.

Hi ha moltes demostracions diferents del teorema de Gibbard-Satterthwaite. Algunes es basen en el teorema d'Arrow (per exemple, la de Gibbard (1973)). És immediat comprovar que qualsevol funció d'elecció social dictatorial és no manipulable. Aquí indicarem l'estructura de la demostració de Barberà i Peleg (1990) de l'altra implicació, suposant que el conjunt d'alternatives és finit i que les preferències són estrictes. La demostració és per inducció en el número d'agents i fa servir els conjunts d'opcions.[14] Presentarem només la demostració del teorema pel cas $n = 2$. L'argument d'inducció està fora de l'abast d'aquest article.

Suposem que $n = 2$, $\#A$ és finit, $f : \mathcal{P}^2 \to A$ és no manipulable i $\#r_f \neq 2$. Si $\#r_f = 1$, f és trivialment dictatorial. Suposem doncs que $\#r_f \geq 3$. S'ha de demostrar que f és dictatorial. L'obtenció del dictador es basa en l'anàlisi del poder que tenen els dos agents per imposar una alternativa. Sigui $P_1 \in \mathcal{P}$. Definim el conjunt de les *opcions* (relatives a f) deixades per P_1 a l'agent 2 com

$$o_2(P_1) = \{x \in A \mid \text{existeix } P_2 \in \mathcal{P} \text{ tal que } f(P_1, P_2) = x\}.$$

El conjunt $o_1(P_2)$ es defineix similarment.

Com a conseqüència de la no manipulabilitat de la funció d'elecció social les següents sis afirmacions sobre els conjunts d'opcions són certes.[15]

Primera, la funció sempre selecciona la millor alternativa, d'acord amb les preferències d'un agent, d'entre les opcions deixades per l'altre agent.

Lema 1 *Per a tot* $(P_1, P_2) \in \mathcal{P}^2$, $f(P_1, P_2) = m(P_2, o_2(P_1))$.

Demostració Sigui $z = f(P_1, P_2)$ i $x = m(P_2, o_2(P_1))$. Com que $x \in o_2(P_1)$, existeix $P_2' \in \mathcal{P}$ tal que $x = f(P_1, P_2')$. Si $x \neq z$ llavors xP_2z ja que $z = f(P_1, P_2)$ implica que $z \in o_2(P_1)$, i $x = m(P_2, o_2(P_1))$. Per tant, $f(P_1, P_2')P_2 f(P_1, P_2)$; és a dir, l'agent 2 manipula f en el perfil (P_1, P_2) declarant P_2'. ■

Segona, la millor alternativa en el rang és sempre una de les opcions.

Lema 2 *Per a tot* $P_1 \in \mathcal{P}$, $m(P_1, r_f) \in o_2(P_1)$.

Demostració Sigui $x = m(P_1, r_f)$. Com que $x \in r_f$, existeix $\bar{P} = (\bar{P}_1, \bar{P}_2) \in \mathcal{P}^2$ tal que $f(\bar{P}_1, \bar{P}_2) = x$. Sigui $f(P_1, \bar{P}_2) = z$. Si $x = z$, $m(P_1, r_f) \in o_2(P_1)$. Si $x \neq z$, llavors $f(\bar{P}_1, \bar{P}_2)P_1 f(P_1, \bar{P}_2)$; és a dir l'agent 1 manipula f en el perfil $(P_1, \bar{P}_2)$ declarant $\bar{P}_1$. ■

Tercera, la funció respecta la unanimitat en el rang.

Lema 3 *Per a tot* $x \in r_f$, *si* $x = m(P_1, r_f) = m(P_2, r_f)$ *llavors* $f(P_1, P_2) = x$.

[14] Per demostracions alternatives, veure també Schmeidler i Sonnenschein (1978), Batteau, Blin i Montjardet (1981), Barberà (1983a), Benoit (2000), Sen (2001), Reny (2001), Ubeda (2003), Eliaz (2004) i Man i Takayama (2011). Alguns d'aquests articles presenten un marc general on els teoremes d'Arrow i de Gibbard-Satterthwaite són casos particulars d'un resultat més general.

[15] El raonament que segueix no avança simètricament per als dos agents (malgrat que la conclusió si que és simètrica); al final es veurà que és irrellevant en quin dels dos agents concentrem la nostra atenció. Aquesta és una característica interessant del raonament en conjunt.

Demostració Suposem que $x \in r_f$ i $x = m(P_1, r_f) = m(P_2, r_f)$. Pel Lema 2, $x \in o_1(P_2) \cap o_2(P_1)$. Per tant, $x = m(P_1, o_1(P_2)) = m(P_2, o_2(P_1))$. Pel Lema 1, $f(P_1, P_2) = x$. ■

Quarta, el conjunt d'opcions només depèn de la millor alternativa en el rang.

Lema 4 *Per a tot $x \in r_f$, si $x = m(P_1, r_f) = m(P_1', r_f)$ llavors $o_2(P_1) = o_2(P_1')$.*

Demostració Suposem que $x = m(P_1, r_f) = m(P_1', r_f) \in r_f$ i existeix $z \in o_2(P_1) \backslash o_2(P_1')$. Pel Lema 2, $x \in o_2(P_1) \cap o_2(P_1')$. Sigui $\bar{P}_2 \in \mathcal{P}$ qualsevol preferència tal que $z\bar{P}_2 x \bar{P}_2 y$ per a tot $y \in r_f \backslash \{x, z\}$ ($\bar{P}_2$ existeix per el supòsit de domini universal i $\# r_f \geq 3$). Aleshores, (i) $f(P_1, \bar{P}_2) = z$, ja que pel Lema 1 $z = m(\bar{P}_2, o_2(P_1)) = f(P_1, \bar{P}_2)$ i (ii) $f(P_1', \bar{P}_2) = x$, ja que $z \notin o_2(P_1')$, $x \in o_2(P_1')$ i, per la definició de $\bar{P}_2$ i el Lema 1, $x = m(\bar{P}_2, o_2(P_1'))$. Per tant, $f(P_1', \bar{P}_2) P_1 f(P_1, \bar{P}_2)$; és a dir, l'agent 1 manipula f en el perfil $(P_1, \bar{P}_2)$ declarant P_1'. ■

Quinta, el conjunt d'opcions deixades per *una* preferència o conté una única alternativa o coincideix amb el rang.

Lema 5 *Per a tot $P_1 \in \mathcal{P}$, o bé $\# o_2(P_1) = 1$ o bé $o_2(P_1) = r_f$.*

Demostració Suposem que existeixen $\bar{P}_1 \in \mathcal{P}$ i $x, y, z \in r_f$ tals que $x, y \in o_2(\bar{P}_1)$ i $z \notin o_2(\bar{P}_1)$. Sense pèrdua de generalitat podem suposar que $x = m(\bar{P}_1, A)$ i $z\bar{P}_1 y$ ja que si no, podem modificar $\bar{P}_1$ fins que això sigui cert degut a que, pel Lema 2, la millor alternativa en el rang pertany al conjunt d'opcions i, pel Lema 4, el conjunt d'opcions només depèn de la millor alternativa en el rang. Considerem qualsevol preferència $\bar{P}_2 \in \mathcal{P}$ tal que $z\bar{P}_2 y \bar{P}_2 w$ per a tot $w \in r_f \backslash \{y, z\}$ ($\bar{P}_2$ existeix pel supòsit de domini universal i $\# r_f \geq 3$). Aleshores, $f(\bar{P}_1, \bar{P}_2) = y$ ja que $z \notin o_2(\bar{P}_1)$ i, pel Lema 1 i $y \in o_2(\bar{P}_1)$, $f(\bar{P}_1, \bar{P}_2) = m(\bar{P}_2, o_2(\bar{P}_1))$. Considerem una altra preferència qualsevol $P_1' \in \mathcal{P}$ tal que $z = m(P_1', r_f)$. Pel Lema 2, $z \in o_2(P_1')$. Pel Lema 1, $f(P_1', \bar{P}_2) = z$. Per tant, $f(P_1', \bar{P}_2) \bar{P}_1 f(\bar{P}_1, \bar{P}_2)$; és a dir, l'agent 1 manipula f en el perfil $(\bar{P}_1, \bar{P}_2)$ declarant P_1'. ■

Sexta i última, el conjunt d'opcions deixades per *totes* les preferències o bé conté una única alternativa o bé coincideix sempre amb el rang.

Lema 6 *O bé $\# o_2(P_1) = 1$ per a tot $P_1 \in \mathcal{P}$, o bé $o_2(P_1) = r_f$ per a tot $P_1 \in \mathcal{P}$.*

Demostració Suposem que no. Pel Lema 5, existeixen $\hat{P}_1, \bar{P}_1 \in \mathcal{P}$ tals que $o_2(\hat{P}_1) = \{x\}$ i $o_2(\bar{P}_1) = r_f$. Com que $\# r_f \geq 3$, pel Lema 4 (el conjunt d'opcions només depèn de la millor alternativa en el rang) podem suposar que $x\bar{P}_1 z$ per alguna $z \in r_f$. Considerem qualsevol preferència $P_2 \in \mathcal{P}$ tal que $m(P_2, r_f) = z$. Aleshores, (i) $f(\bar{P}_1, P_2) = z$, pel Lema 1, i (ii) $f(\hat{P}_1, P_2) = x$ ja que $o_2(\hat{P}_1) = \{x\}$. Per tant, $f(\hat{P}_1, P_2) \bar{P}_1 f(\bar{P}_1, P_2)$; és a dir, l'agent 1 manipula f en el perfil $(\bar{P}_1, P_2)$ declarant $\hat{P}_1$. ■

Ara és fàcil demostrar que $f : \mathcal{P}^2 \to A$ és dictatorial. Pel Lema 6, o bé $\# o_2(P_1) = 1$ per a tot $P_1 \in \mathcal{P}$ o bé $o_2(P_1) = r_f$ per a tot $P_1 \in \mathcal{P}$. Si $\# o_2(P_1) = 1$ per a tot $P_1 \in \mathcal{P}$, pel Lema 2, $o_2(P_1) = m(P_1, r_f)$. Aleshores, $f(P_1, P_2) = m(P_1, r_f)$ per a tot $(P_1, P_2) \in \mathcal{P}^2$; és a dir, l'agent 1 és un dictador. Si $o_2(P_1) = r_f$ per a tot $P_1 \in \mathcal{P}$, pel Lema 1, $f(P_1, P_2) = m(P_2, o_2(P_1))$ per a tot $P_2 \in \mathcal{P}$. Per hipòtesi, $m(P_2, o_2(P_1)) = m(P_2, r_f)$ per a tot $P_2 \in \mathcal{P}$; és a dir, l'agent 2 és un dictador.

2.4 Relació entre independència d'alternatives irrellevants i no manipulabilitat

En aquesta subsecció presentem, en el context de preferències estrictes, la relació entre la propietat de la independència d'alternatives irrellevants d'una funció de benestar social i la no manipulabilitat de la funció d'elecció social associada, obtinguda assignant a cada perfil de preferències l'únic element maximal de la preferència social. És a dir, donada una funció de benestar social $F : \mathcal{P}^n \to \mathcal{P}$ definim la funció d'elecció social associada $f_F : \mathcal{P}^n \to A$ com, per a tot $P \in \mathcal{P}^n$, $f_F(P) = m(F(P), A)$.

Una funció de benestar social $F : \mathcal{P}^n \to \mathcal{P}$ és *monòtona* quan té la següent propietat. Per a tot $x \in A$ i tot parell de perfils $P, P' \in \mathcal{P}^n$ que només difereixen en que la posició relativa de x millora en P', llavors la posició relativa de x en $F(P')$ no empitjora. Blair i Muller (1983) demostren la següent proposició.

Proposició 1 (Blair i Muller, 1983) *Les dues afirmacions següents són equivalents:*
(1) $F : \mathcal{P}^n \to \mathcal{P}$ *és monòtona i satisfà la independència d'alternatives irrellevants.*
(2) $f_F : \mathcal{P}^n \to A$ *és no manipulable.*

2.5 Restriccions de domini i possibilitats

En les dues properes seccions presentarem diferents problemes d'elecció social on l'estructura del conjunt d'alternatives obliga a restringir el conjunt de preferències individuals. L'eliminació de preferències individuals limita la capacitat dels agents de manipular les funcions d'elecció social. Això suggereix la possibilitat que existeixin funcions d'elecció social no manipulables (i no trivials) en el domini restringit. Aquest enfoc però, dependrà de l'estructura del conjunt d'alternatives i de la restricció de domini de preferències induïda. Per aquest motiu no disposem d'una teoria amb resultats de possibilitat generals, sinó que aquests són específics a cada una de les restriccions de domini.

Hi ha dues famílies de models que tractarem separadament. En els models de la primera família, que corresponen al model presentat en la secció 2, les alternatives no tenen cap component privat i per tant, les preferències individuals continuen estant definides en el conjunt d'alternatives A, el conjunt imatge de les funcions d'elecció social. Aquestes restriccions de domini les anomenarem de *bé públic*. En els models de la segona família les alternatives tenen components privats ja que per descriure una alternativa hem d'especificar aspectes, cada un dels quals només interessa a un dels agents. En aquests casos, la restricció de domini és conseqüència (sovint, conjuntament amb altres consideracions) del fet que cada agent, en no importar-li l'aspecte privat de l'alternativa corresponent als altres agents, és indiferent entre totes les alternatives que tracten diferent als altres agents però el tracten igual a ell. Per tant, serà convenient considerar funcions d'elecció social que escullin alternatives depenent exclusivament de les preferències individuals sobre els components privats de les alternatives socials. Aleshores, el domini d'aquestes funcions no serà el producte cartesià (n vegades) del mateix subconjunt de preferències individuals en A, sinó que serà el producte cartesià de subconjunts de preferències diferents, un per a cada agent. Per això, haurem de reformular una mica el nostre model. Aquestes restriccions de domini les anomenarem de *béns privats* i les presentarem en la secció 4.

3 Restriccions de domini de bé públic

Sigui $\mathcal{D} \subseteq \mathcal{R}$ un subconjunt arbitrari de preferències en el conjunt d'alternatives A. Una *funció d'elecció social* f (*en* $\mathcal{D}$) és una funció $f : \mathcal{D}^n \to A$ que selecciona una alternativa $f(R)$ per a cada perfil de preferències $R = (R_1, ..., R_n)$ en el conjunt $\mathcal{D}^n$. Ens referim al producte Cartesià $\mathcal{D}^n$ (o al mateix conjunt $\mathcal{D}$) com al *domini* de preferències.

Estem interessats en funcions d'elecció social sobre dominis de preferències que induexin als agents a declarar la verdadera preferència; és a dir, que siguin no manipulables en el seu domini. Per tant, reformulant la noció general, diem que una funció d'elecció social $f : \mathcal{D}^n \to A$ (en $\mathcal{D}$) és *no manipulable* (*en* $\mathcal{D}$) quan per a tot $R \in \mathcal{D}^n$, tot $i \in N$, i tot $R'_i \in \mathcal{D}$,

$$f(R_i, R_{-i}) R_i f(R'_i, R_{-i}).$$

També considerarem altres propietats de les funcions d'elecció social en un domini $\mathcal{D}$. Una funció d'elecció social $f : \mathcal{D}^n \to A$ és *anònima* quan no depèn dels noms del agents; és a dir, per a qualsevol bijecció $\sigma : N \to N$ i tot $R \in \mathcal{D}^n$, $f(R_1, ..., R_n) = f(R_{\sigma(1)}, ..., R_{\sigma(n)})$. Una funció d'elecció social $f : \mathcal{D}^n \to A$ és *eficient* quan per a tot $R \in \mathcal{D}^n$ no existeix $z \in A$ tal que, per a tot $i \in N$, $zR_i f(R)$ i $zP_j f(R)$ per algun $j \in N$. Una funció d'elecció social $f : \mathcal{D}^n \to A$ és *unànime* quan per a tot $R \in \mathcal{D}^n$ tal que $\bigcap_{i \in N} m(R_i, A) \neq \varnothing$, $f(R) \in \bigcap_{i \in N} m(R_i, A)$.

3.1 Bé públic amb preferències unimodals: el votant medià

Considerem problemes d'elecció social on les alternatives tenen unes propietats que indueixen un ordre "natural" en A, unànimament acceptat per tots els agents. Hi ha molts problemes d'elecció social on el conjunt d'alternatives està ordenat: la localització física d'un bé públic (un hospital, una escola, o una piscina municipal), els partits polítics ordenats en l'espectre esquerra-dreta, la temperatura en una habitació on han de conviure (o treballar) els agents, o el mínim exempt de l'impost de la renda sobre les persones físiques. És important assenyalar que l'ordre ha de ser unànime. Per exemple, tots els agents estan d'acord en que una habitació a 19 graus és més freda que a 21. Això no treu, però, que els agents discrepin sobre quina és la temperatura ideal de l'habitació; uns consideraran que és 19 graus i d'altres 21. En aquest cas, el problema d'elecció social és escollir una temperatura de l'habitació tenint en compte les preferències individuals. Aquesta estructura d'ordre sobre el conjunt d'alternatives imposa unes limitacions naturals sobre el conjunt de preferències individuals. Per exemple, si la temperatura ideal de l'habitació de l'agent i és 21 graus, és natural que i prefereixi 23 graus que 25, ja que és raonable suposar que les preferències R_i tenen la propietat que existeix una única *alternativa ideal* ($\#m(R_i, A) = 1$, que denotarem per $t(R_i)$), i que a mesura que considerem alternatives més allunyades de la ideal (segons l'ordre unànime), aquestes són considerades com a pitjors. Per simplicitat considerarem que el conjunt d'alternatives A és l'interval $[0, 1] \subset \mathbb{R}$ i que l'ordre "natural" sobre $[0, 1]$ és l'ordre $>$ dels números reals. Black (1948) és el primer en suggerir que, donat l'ordre del conjunt d'alternatives, les preferències dels agents han de ser unimodals.

Definició 6 Una preferència $R_i \in \mathcal{R}$ és *unimodal* quan

(1) existeix una única alternativa ideal $t(R_i)$: $t(R_i) P_i y$ per a tot $y \in [0, 1] \backslash \{t(R_i)\}$ i

(2) per a tot parell d'alternatives $x, y \in [0,1]$ tals que $y < x \leq t(R_i)$ o $t(R_i) \leq x < y$, xP_iy.

Sigui $\mathcal{UM} \subset \mathcal{R}$ el conjunt de preferències unimodals en $[0,1]$. Donada una preferència unimodal $R_i \in \mathcal{UM}$, pot passar que yP_ix fins i tot si $|t(R_i) - x| < |t(R_i) - y|$; però llavors, x i y estan necessàriament situades a diferents costats de l'alternativa ideal $t(R_i)$.

Una funció d'elecció social $f : \mathcal{UM}^n \to [0,1]$ és *un sistema de votació* quan per a tot parell de perfils $R, R' \in \mathcal{UM}^n$ tals que $t(R_i) = t(R'_i)$ per a tot $i \in N$, $f(R) = f(R')$; és a dir, els sistemes de votació escullen l'alternativa social tenint en consideració només el vector d'alternatives ideals.

Moulin (1980) caracteritza tots els sistemes de votació no manipulables en el domini de preferències unimodals. Aquesta família conté moltes funcions no dictatorials. Totes elles són extensions del *sistema de votació del votant medià*. Seguint el propi Moulin (1980), i abans de presentar el teorema general, presentarem dos dels seus corol·laris caracteritzant dues subfamílies enniuades (en anglès *nested*) que permeten desenvolupar intuicions útils per entendre la caracterització general, però que també són interessants per elles mateixes.

Considerem el cas d'un número n senar d'agents amb preferències unimodals en $[0,1]$ i la funció d'elecció social $f : \mathcal{UM}^n \to [0,1]$ que consisteix en escollir, per a cada perfil de preferències $R = (R_1, ..., R_n) \in \mathcal{UM}^n$, la mediana d'entre les alternatives ideals dels n agents; és a dir, $f(R) = med\{t(R_1), ..., t(R_n)\}$.[16] Aquesta funció d'elecció social és un sistema de votació (només depèn del vector d'alternatives ideals) anònim i eficient.[17] Per veure que a més és no manipulable, considerem un perfil R arbitrari de preferències unimodals. Un agent i per a qui l'alternativa ideal sigui la mediana no pot manipular f en R ja que $f(R) = t(R_i)$. Considerem ara un agent i amb una alternativa ideal diferent de la mediana. Suposem que $t(R_i) < f(R)$ (el cas contrari és simètric). L'agent i només pot modificar l'alternativa escollida declarant una preferència R'_i amb la propietat $f(R) < t(R'_i)$, però en aquest cas la nova alternativa escollida o no canvia o és més gran que $f(R)$, i per tant és igual o pitjor des del punt de vista de la preferència unimodal R_i. L'agent i tampoc pot manipular f en el perfil R, amb la qual cosa podem concloure que f és no manipulable.

Una funció d'elecció social potencialment interessant és la funció $f : \mathcal{UM}^n \to [0,1]$ que escull, per a cada perfil $R \in \mathcal{UM}^n$, la mitjana de les n alternatives ideals; és a dir, $f(R) = \frac{t(R_1)+...+t(R_n)}{n}$. Aquesta funció és manipulable ja que és massa sensible a les alternatives ideals declarades. Per veure-ho, considerem el cas $N = \{1,2\}$ i qualsevol perfil $(R_1, R_2) \in \mathcal{UM}^2$ on $t(R_1) = \frac{1}{4}$ i $t(R_2) = \frac{3}{4}$. Llavors, $f(R) = \frac{1}{2}$, però l'agent 1 manipula f en el perfil R declarant qualsevol preferència $R'_1 \in \mathcal{UM}$ amb la propietat que $0 \leq t(R'_1) < \frac{1}{4}$. Simètricament, l'agent 2 manipula f en R declarant qualsevol preferència $R'_2 \in \mathcal{UM}$ amb la propietat que $\frac{3}{4} < t(R'_2) \leq 1$. D'altra banda, la funció mediana és molt menys sensible respecte a l'alternativa ideal declarada per a cada agent: per a tot $R_{-i} \in \mathcal{UM}^{n-1}$, $t(R_i), t(R'_i) \leq f(R_i, R_{-i})$ implica $f(R_i, R_{-i}) = f(R'_i, R_{-i})$. Aquest és un principi bastant general. Una funció d'elecció social, per ser no manipulable, no pot ser massa sensible a les preferències individuals i per tant, ha de ser constant en bastants subconjunts de perfils. Aquest és

[16] Donat un conjunt senar de números reals $\{x_1, ..., x_K\}$ definim la seva *mediana* com $med\{x_1, ..., x_K\} = y$, on y és tal que $\#\{1 \leq k \leq K \mid x_k \leq y\} \geq \frac{K}{2}$ i $\#\{1 \leq k \leq K \mid x_k \geq y\} \geq \frac{K}{2}$. Si K és senar la mediana és única i pertany al conjunt $\{x_1, ..., x_K\}$.

[17] És fàcil comprovar que una funció d'elecció social $f : \mathcal{UM}^n \to [0,1]$ és eficient si i només si, per a tot $R \in \mathcal{UM}^n$, $\min\{t(R_i) \mid i \in N\} \leq f(R) \leq \max\{t(R_i) \mid i \in N\}$.

un dels motius més importants que fan difícil dissenyar funcions d'elecció social no manipulables i eficients. Aquest no és el cas, però, quan les preferències són unimodals.

Tornem ara a la funció mediana i suposem que afegim, a les alternatives ideals dels n agents, $n-1$ vots ficticis: $\frac{n-1}{2}$ vots per l'alternativa 0 i els altres $\frac{n-1}{2}$ vots per l'alternativa 1 (continuem suposant que n és senar). Aleshores, la mediana entre les alternatives ideals dels n agents i la mediana entre les n alternatives ideals i els $n-1$ vots ficticis coincideixen, ja que els $\frac{n-1}{2}$ zeros i els $\frac{n-1}{2}$ uns ficticis es cancel·len entre ells; és a dir, per a tot $R=(R_1,...,R_n)\in\mathcal{UM}^n$,

$$f(R)=med\{t(R_1),...,t(R_n),\ \underbrace{0,...,0}_{\frac{n-1}{2}\text{ vegades}}\ ,\ \underbrace{1,...,1}_{\frac{n-1}{2}\text{ vegades}}\ \}=med\{t(R_1),...,t(R_n)\}.$$

Podem fer un pas més, i en comptes de posar els $n-1$ vots ficticis en els extrems, i independentment de si n és parell o senar, els podem posar en qualsevol alternativa i considerar llavors la mediana entre les alternatives ideals dels n agents i els $n-1$ vots ficticis $v_{n-1},...,v_1\in[0,1]$. Per tant, donats $v_{n-1},...,v_1\in[0,1]$ definim la funció d'elecció social $f:\mathcal{UM}^n\to[0,1]$ (que depèn dels $n-1$ vots ficticis considerats) assignant, per a tot $R\in\mathcal{UM}^n$,

$$f(R)=med\{t(R_1),...,t(R_n),v_{n-1},...,v_1\}.$$

És menys obvi veure que aquesta família de funcions (una per a cada possible vector de $n-1$ vots ficticis) coincideix amb el conjunt de totes les funcions d'elecció social no manipulables, anònimes i eficients (és fàcil demostrar que tota funció d'elecció social no manipulable i eficient és un sistema de votació).

Corol·lari 1 (Moulin, 1980) *Una funció d'elecció social $f:\mathcal{UM}^n\to[0,1]$ és no manipulable, anònima i eficient si i només si existeixen $n-1$ vots ficticis $0\le v_{n-1}\le...\le v_1\le 1$ tals que, per a tot $R\in\mathcal{UM}^n$,*

$$f(R)=med\{t(R_1),...,t(R_n),v_{n-1},...,v_1\}.$$

Sorprenentment, si en comptes de prefixar $n-1$ vots ficticis en prefixem $n+1$, podem prescindir de la propietat d'eficiència en el resultat anterior (però llavors hem d'afegir la hipòtesi que la funció d'elecció social és un sistema de votació).

Corol·lari 2 (Moulin, 1980) *Una funció d'elecció social $f:\mathcal{UM}^n\to[0,1]$ és un sistema de votació no manipulable i anònim si i només si existeixen $n+1$ vots ficticis $0\le v_n\le...\le v_0\le 1$ tals que, per a tot $R\in\mathcal{UM}^n$,*

$$f(R)=med\{t(R_1),...,t(R_n),v_n,...,v_0\}.$$

Observem que les funcions constants, eliminades en el corol·lari 1 per la propietat d'eficiència, pertanyen a la classe identificada en el corol·lari 2, ja que la funció $f(R)=x$ per a tot $R\in\mathcal{UM}^n$ correspon a la funció mediana quan els $n+1$ vots ficticis són tots iguals a x. A més, cada una de les funcions identificades en el corol·lari 1 (amb els seus corresponents $n-1$ vots ficticis $v_{n-1},...,v_1$) pot ser representada també amb $n+1$ vots ficticis afegint als $n-1$ anteriors $v_n=0$ i $v_0=1$.

Per desenvolupar una intuició útil per entendre la caracterització sense anonimat, considerem el cas $n=2$. Donats $0\le v_{\{1,2\}}\le v_{\{1\}}\le v_{\{2\}}\le v_{\varnothing}\le 1$, un vot fictici per a cada coalició (subconjunt)

d'agents, definim la funció d'elecció social $f : \mathcal{UM}^2 \to [0,1]$ com, per a tot $R \in \mathcal{UM}^2$,

$$f(R) = \begin{cases} v_{\{1,2\}} & \text{si } t(R_1), t(R_2) \leq v_{\{1,2\}} \\ t(R_2) & \text{si } t(R_1) \leq v_{\{1,2\}} \leq t(R_2) \leq v_{\{1\}} \\ v_{\{1\}} & \text{si } t(R_1) \leq v_{\{1,2\}} \leq v_{\{1\}} \leq t(R_2) \\ med\{t(R_1), t(R_2), v_{\{1\}}\} & \text{si } v_{\{1,2\}} \leq t(R_1) \leq v_{\{1\}} \\ t(R_1) & \text{si } v_{\{1\}} \leq t(R_1) \leq v_{\{2\}} \\ med\{t(R_1), t(R_2), v_{\{2\}}\} & \text{si } v_{\{2\}} \leq t(R_1) \leq v_{\varnothing} \\ v_{\{2\}} & \text{si } v_{\varnothing} \leq t(R_1) \text{ i } t(R_2) \leq v_{\{2\}} \\ t(R_2) & \text{si } v_{\{2\}} \leq t(R_2) \leq v_{\varnothing} \leq t(R_1) \\ v_{\varnothing} & \text{si } v_{\varnothing} \leq t(R_1), t(R_2). \end{cases}$$

Notem que $r_f = [v_{\{1,2\}}, v_{\varnothing}]$. Podem interpretar aquesta funció com una manera d'assignar el poder als subconjunts d'agents (coalicions) per seleccionar l'alternativa en el conjunt $r_f = [v_{\{1,2\}}, v_{\varnothing}]$. És fàcil comprovar que aquesta funció es pot escriure com

$$f(R) = \min_{S \subseteq \{1,2\}} \max_{i \in S} \{t(R_i), v_S\}.$$

Per a presentar la caracterització de tots els sistemes de votació no manipulables en el domini de preferències unimodals, diem que una col·lecció $\{v_S\}_{S \in 2^N}$ és una *família monòtona de vots ficticis* si (i) $v_S \in [0,1]$ per a tot $S \in 2^N$ i (ii) $T \subset Q$ implica $v_Q \leq v_T$. El resultat general de possibilitat és el següent.

Teorema 1 (Moulin, 1980) *Una funció d'elecció social $f : \mathcal{UM}^n \to [0,1]$ és un sistema de votació no manipulable si i només si existeix una família monòtona de vots ficticis $\{v_S\}_{S \in 2^N}$ tal que, per a tot $R \in \mathcal{UM}^n$,*

$$f(R) = \min_{S \in 2^N} \max_{i \in S} \{t(R_i), v_S\}. \tag{1}$$

Les funcions d'elecció social identificades en el teorema 1 es coneixen com a *sistemes del votant medià generalitzat*. Una primera manera d'interpretar-los és la següent. Cada sistema del votant medià generalitzat (i la seva família monòtona de vots ficticis associada) pot ser entès com una manera particular de distribuir el poder entre les coalicions (els subconjunts) d'agents per influir en el resultat de l'elecció social. Per veureu-ho, considerem una coalició arbitrària S i el seu vot fictici associat v_S. Aleshores, la coalició S pot assegurar-se que, declarant tots els seus membres una alternativa ideal menor o igual a v_S, l'elecció social serà com a màxim v_S, independentment de les alternatives ideals declarades pels agents de la coalició complementària.[18] Una descripció alternativa de la distribució de poder entre les coalicions és la següent. Fixem una família monòtona de vots ficticis $\{v_S\}_{S \in 2^N}$ (i per tant, un sistema del votant medià generalitzat) i considerem un vector d'alternatives ideals $(t(R_1), ..., t(R_n))$. Comencem per l'extrem esquerre de l'interval i "pressionem" l'elecció social cap a la dreta, com si f volgués seleccionar l'alternativa més gran possible però, al mateix temps, les coalicions tinguessin el poder de parar aquesta tendència creixent amb el vot dels seus membres. Per tant, anem pujant des de l'alternativa 0 fins que arribem a

[18] Veure Barberà, Massó i Neme (1997) per a una interpretació similar pel cas en què el conjunt d'alternatives sigui un conjunt finit totalment ordenat.

una alternativa x, la *primera* en la que dues coses passen simultàniament: (i) existeix una coalició d'agents S tal que tots els seus membres han declarat una alternativa ideal igual o menor a x (*i.e.*, $t(R_i) \leq x$ per a tot $i \in S$) i (ii) el vot fictici v_S associat a S està situat per sota d'x (*i.e.*, $v_S \leq x$).

Els sistemes de votació identificats en el corol·lari 2 constitueixen la subclasse anònima dels sistemes del votant medià generalitzat. En aquest cas, els vots ficticis de dues coalicions amb la mateixa cardinalitat han de ser iguals. Podem identificar els $n+1$ vots ficticis $v_n \leq ... \leq v_0$ necessaris per descriure f com un sistema del votant medià generalitzat de la següent manera: per a cada $0 \leq s \leq n$, $v_s = v_S$ per a tot $S \in 2^N$ tal que $\#S = s$. A més, si n és senar, el (verdader) votant medià s'obté escollint $v_n = ... = v_{\frac{n+1}{2}} = 0$ i $v_{\frac{n+1}{2}-1} = ... = v_0 = 1$ ja que per a qualsevol $R = (R_1, ..., R_n) \in \mathcal{UM}^n$,

$$\begin{aligned} med\{t(R_1), ..., t(R_n), v_n, ..., v_0\} &= med\{t(R_1), ..., t(R_n), \underbrace{0, ..., 0}_{\frac{n+1}{2} \text{ vegades}}, \underbrace{1, ..., 1}_{\frac{n+1}{2} \text{ vegades}}\} \\ &= med\{t(R_1), ..., t(R_n)\}. \end{aligned}$$

Per acabar, la funció d'elecció social $f : \mathcal{UM}^n \to [0,1]$ on l'agent $j \in N$ és un dictador (és a dir, per a tot $R \in \mathcal{UM}^n$, $f(R) = t(R_j)$) es pot descriure com un sistema del votant medià generalitzat, seleccionant $v_S = 0$ per a tot $S \subset N$ tal que $j \in S$ i $v_S = 1$ per a tot $S \subset N$ tal que $j \notin S$. Aleshores, per a qualsevol $R \in \mathcal{UM}^n$, $\max_{i \in S}\{t(R_i), v_S\} = 1$ si $j \notin S$ i $\max_{i \in S}\{t(R_i), v_S\} = \max_{i \in S}\{t(R_i)\}$ si $j \in S$. Per tant, $\min_{S \in 2^N} \max_{i \in S}\{t(R_i), v_S\} = t(R_j)$.

La demostració del teorema 1 té dues parts. La primera, comprovar que qualsevol sistema del votant medià generalitzat és no manipulable. L'argument és molt semblant al ja presentat en el cas particular de la funció mediana: els agents només poden afectar l'elecció social allunyant-la de la seva alternativa ideal. La segona, donada una funció d'elecció social $f : \mathcal{UM}^n \to [0,1]$ que és un sistema de votació no manipulable, identificar la família monòtona de vots ficticis $\{v_S\}_{S \in 2^N}$, per la qual la condició (1) del teorema 1 es satisfà. Per identificar els vots ficticis, considerem qualsevol $S \in 2^N$ i perfil $R \in \mathcal{UM}^n$ tals que $t(R_i) = 0$ si $i \in S$ i $t(R_i) = 1$ si $i \notin S$. Definim $v_S = f(R)$. La comprovació que la f així definida satisfà (1) amb aquesta família monòtona de vots ficticis conclou la demostració, però la deixem per al lector.

Finalment, per veure que en les afirmacions del corol·lari 2 i el teorema 1 la hipòtesi que f és un sistema de votació és imprescindible (és a dir, no és una conseqüència de les propietats de no manipulabilitat i anonimat), considerem la següent funció d'elecció social $f : \mathcal{UM}^n \to [0,1]$ tal que per a tot $R \in \mathcal{UM}^n$,

$$f(R) = \begin{cases} 0 & \text{si } \#\{i \in N \mid 0R_i1\} \geq \#\{i \in N \mid 1P_i0\} \\ 1 & \text{si } \#\{i \in N \mid 0R_i1\} < \#\{i \in N \mid 1P_i0\}. \end{cases} \quad (2)$$

Aquesta funció f és no manipulable i anònima però no és un sistema de votació (la funció depèn de com els agents comparen les dues alternatives extremes, no de les seves alternatives ideals). A més, f tampoc és eficient, unànime, ni exhaustiva. No obstant, qualsevol funció d'elecció social no manipulable que satisfaci una d'aquestes tres propietats és de fet un sistema de votació.

Massó i Moreno de Barreda (2011) caracteritzen la classe de totes les funcions d'elecció social no manipulables en el domini de preferències unimodals simètriques, aquelles per a les quals la noció de distància a l'alternativa ideal determina les preferències entre els parells d'alternatives

en costats diferents de l'alternativa ideal. La classe de funcions d'elecció social no manipulables en aquest domini més petit és substancialment més gran que la dels sistemes del votant medià generalitzat. Per altra banda, Barberà i Peleg (1990) obtenen un resultat d'impossibilitat (les úniques funcions d'elecció social no manipulables són dictatorials) en el domini de preferències contínues quan el conjunt d'alternatives A és un espai mètric arbitrari.[19] Aquests dos resultats il·lustren la tensió entre considerar dominis més grans de preferències i la possibilitat de dissenyar-hi funcions d'elecció social no manipulables: com més restrictiu sigui el domini de preferències més possibilitats hi ha de poder-hi definir funcions d'elecció social no manipulables.

Border i Jordan (1983) és el primer d'una llarga llista d'articles que estenen els resultats de Moulin (1980) a contextos on el conjunt d'alternatives té una estructura més rica.[20] Per descriure una alternativa són necessàries més d'una característica, i cada una d'elles es pot donar en diferents intensitats; per exemple quan considerem una societat que ha d'escollir simultàniament entre diferents bens públics, o la seva localització física en un espai bidimensional, o la descripció dels partits polítics en dues dimensions, una per descriure els aspectes de les seves polítiques socials (esquerra-dreta) i l'altra el seu grau de nacionalisme (molt-poc). En aquests casos també té sentit considerar que les preferències dels agents en aquests conjunts d'alternatives són unimodals generalitzades. Però ara, hi ha més d'una noció natural de preferències unimodals en conjunts d'alternatives multi-dimensionals. Per cada possible extensió, tenim diferents resultats de possibilitat (o fins i tot d'impossibilitat).

3.2 Elecció de subconjunts de candidats: vot per comitès

Barberà, Sonnenschein i Zhou (1991) consideren problemes d'elecció social on les alternatives socials són els subconjunts d'un conjunt donat. Per exemple, l'elecció de nous membres d'una societat, les eleccions de representants en societats democràtiques, o les posicions públiques de partits polítics en diferents temes. Farem servir el primer d'aquests exemples com a referència bàsica.[21] Per tant, considerem una societat N de n agents que ha d'elegir un subconjunt de nous membres d'entre un conjunt donat K de k candidats. El conjunt d'alternatives és doncs 2^K, la família de subconjunts de K. En aquest cas és natural suposar que no totes les preferències individuals en 2^K són admissibles, ja que les preferències sobre els subconjunts de K vindran guiades, en part, per com els agents consideren els candidats individualment. Per simplicitat també suposarem que les preferències dels agents en 2^K són estrictes. En particular, direm que una preferència P_i és *separable* quan podem dividir el conjunt de candidats en dos subconjunts disjunts, el dels bons i el dels mals candidats, de manera que afegir un candidat a un conjunt donat millora el conjunt si i només si el candidat afegit és un bon candidat.

[19] Diem que la preferència R_i en un espai mètric A és *contínua* si, per a tot $x \in A$, els conjunts $\{y \in A \mid yR_ix\}$ i $\{y \in A \mid xR_iy\}$ són tancats.

[20] Per exemple, veure Barberà, Gul i Stachetti (1993), Barberà i Jackson (1994), Barberà, Massó i Neme (1997, 2005), Barberà, Massó i Serizawa (1998), Le Breton i Sen (1999), Nehring i Puppe (2007a, 2007b), Peters, van der Stel i Storcken (1992, 1993) i Zhou (1991).

[21] De fet, moltes societats escullen els seus nous membres fent servir el vot per comitès; per exemple, la *Econometric Society* escull cada any els seus nous membres d'honor amb un vot per comitès anònim i neutral.

Definició 7 Una preferència estricta P_i és *separable* en 2^K quan per a tot $x \in K$ i tot $S \in 2^K$ tals que $x \notin S$,

$$S \cup \{x\} P_i S \Longleftrightarrow \{x\} P_i \varnothing.$$

Una preferència estricta P_i és *additiva* en 2^K quan existeix una funció (d'utilitat) $u_i : K \cup \varnothing \to \mathbb{R}$ tal que $u_i(\varnothing) = 0$ i per a tot $S, T \in 2^K$,

$$SP_iT \Longleftrightarrow \sum_{x \in S} u_i(x) > \sum_{x \in T} u_i(x).$$

Les preferències additives són separables, i hi ha preferències separables que no són additives. Per exemple, amb $k = 3$, la preferència

$$\{x,y,z\} P_i \{y,z\} P_i \{x,z\} P_i \{x,y\} P_i \{x\} P_i \{y\} P_i \{z\} P_i \varnothing$$

és separable però no additiva ja que $\{x\} P_i \{y\} P_i \{z\}$ implicaria que $u_i(x) > u_i(y) > u_i(z)$, però llavors el subconjunt $\{x,y\}$ hauria de ser estrictament preferit a $\{y,z\}$. Per tant, per $k \geq 3$, $\mathcal{A} \subsetneq \mathcal{S}e \subsetneq \mathcal{P}$, on $\mathcal{A}$ i $\mathcal{S}e$ són els conjunts de preferències estrictes additives i separables en 2^K, respectivament. Donada $P_i \in \mathcal{P}$ denotarem per $m(P_i, 2^K)$ el millor subconjunt de 2^K segons P_i.

Seguint a Barberà, Sonnenschein i Zhou (1991), un sistema de vot per comitès es defineix a partir d'una col·lecció de famílies de coalicions guanyadores (comitès), una per a cada candidat. Els agents voten per un subconjunt de candidats (interpretat com el seu subconjunt de candidats ideal). Per ser escollit, un candidat ha d'obtenir els vots de tots els membres d'alguna coalició guanyadora per aquest candidat. Formalment, un *comitè per* $x \in K$, representat per $\mathcal{W}_x \subseteq 2^N$, és una família no buida de coalicions no buides de N amb la propietat de la monotonia coalicional: (i) $\mathcal{W}_x \neq \varnothing$, (ii) $\varnothing \notin \mathcal{W}_x$ i (iii) $S \in \mathcal{W}_x$ i $S \subset T$ implica $T \in \mathcal{W}_x$. Sigui $\mathcal{D} \subset \mathcal{P}$ un subconjunt arbitrari de preferències estrictes en 2^K. Diem que una funció $f : \mathcal{D}^n \to 2^K$ és *un sistema de vot per comitès* (*en* $\mathcal{D}$) quan existeix una família de comitès $\mathcal{W} = (\mathcal{W}_x)_{x \in K}$, un per a cada candidat, tal que per a tot perfil $P = (P_1, ..., P_n) \in \mathcal{D}^n$ i tot candidat $x \in K$,

$$x \in f(P) \Longleftrightarrow \{i \in N \mid x \in m(P_i, 2^K)\} \in \mathcal{W}_x.$$

Els sistemes de vot per comitès només depenen del vector $(m(P_1, 2^K), ..., m(P_n, 2^K))$ de subconjunts ideals. Noteu que si P_i és una preferència separable (additiva) aleshores $m(P_i, 2^K)$ coincideix amb el conjunt de candidats bons per i: $x \in m(P_i, 2^K)$ si i només si $\{x\} P_i \varnothing$ $(u_i(x) > 0)$. Per exemple, per $N = \{1,2,3\}$ i $K = \{x,y,z\}$ el sistema de vot per comitè $\mathcal{W} = (\mathcal{W}_x, \mathcal{W}_y, \mathcal{W}_z)$ on $\mathcal{W}_x = \{\{1\},\{2\},\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}$, $\mathcal{W}_y = \{\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$ i $\mathcal{W}_z = \{\{1,2,3\}\}$ té la propietat que perquè x pentanyi al subconjunt escollit ha de ser un bon candidat per a l'agent 1 o l'agent 2, perquè y pertanyi al subconjunt escollit ha de ser un bon candidat per almenys dos agents, i perquè z pertanyi al subconjunt escollit ha de ser un bon candidat per als tres agents.

Barberà, Sonnenschein i Zhou (1991) demostren que els sistemes de vot per comitès en el domini de preferències separables (i també de preferències additives) constitueixen la classe de funcions d'elecció social exhaustives i no manipulables en $\mathcal{S}e$ (i també en $\mathcal{A}$).

Teorema 2 (Barberà, Sonnenschein i Zhou, 1991)

► *Una funció d'elecció social exhaustiva* $f : \mathcal{S}e^n \to 2^K$ *és no manipulable en* $\mathcal{S}e$ *si i només si és un sistema de vot per comitès en* $\mathcal{S}e$.

► *Una funció d'elecció social exhaustiva $f : \mathcal{A}^n \to 2^K$ és no manipulable en $\mathcal{A}$ si i només si és un sistema de vot per comitès en $\mathcal{A}$.*

Les dues afirmacions del teorema 2 només difereixen en els dominis de les funcions d'elecció social. En aquest cas l'augment del domini (i per tant, del conjunt de preferències que els agents poden declarar), en passar d'$\mathcal{A}$ a $\mathcal{S}e$, no redueix el conjunt de funcions d'elecció social exhaustives i no manipulables. Barberà, Sonnenschein i Zhou (1991) també caracteritzen les subclasses anònimes (els comitès només depenen de la cardinalitat de les coalicions) i neutrals (tots els comitès dels diferents candidats són iguals). Finalment, la hipòtesi que f és exhaustiva és restrictiva.[22]

La demostració que qualsevol funció d'elecció social exhaustiva i no manipulable en $\mathcal{S}e$ o en $\mathcal{A}$ és un sistema de vot per comitès queda fora de l'abast d'aquest article;[23] és per inducció en el número d'agents i utilitza intensament el conjunt de les opcions deixades per la preferència d'un agent als altres agents. No obstant, donada una funció d'elecció social en $\mathcal{S}e$ o en $\mathcal{A}$, és senzill veure com s'obtenen els comitès. Per a cada candidat $x \in K$ contruïm el seu comité $\mathcal{W}_x$ com: per a tot $\varnothing \neq T \subseteq N$, $T \in \mathcal{W}_x$ si i només si $f(P) = \{x\}$ per a qualsevol perfil P (en $\mathcal{S}e$ o en $\mathcal{A}$) tal que $m(P_i, 2^K) = \{x\}$ per a tot $i \in T$ i $m(P_i, 2^K) = \varnothing$ per a tot $i \notin T$. Per veure que un sistema de vot per comitès és una funció d'elecció social exhaustiva considerem, per a cada subconjunt de candidats $S \in 2^K$, qualsevol perfil P (en $\mathcal{S}e$ o en $\mathcal{A}$) tal que $m(P_i, 2^K) = S$ per a tot $i \in N$. Aleshores, $f(P) = S$ ja que per a tot $x \in S$, $N \in \mathcal{W}_x$, i per a tot $y \notin S$, $\{i \in N \mid y \in m(P_i, 2^K)\} = \varnothing \notin \mathcal{W}_y$. Per tant, f és exhaustiva en $\mathcal{S}e$ o en $\mathcal{A}$. Finalment, per veure que un sistema de vot per comitès és no manipulable, fixem un perfil P en $\mathcal{S}e$ o en $\mathcal{A}$ i considerem, per a qualsevol agent $i \in N$, les quatre possibles relacions entre els conjunts $m(P_i, 2^K)$ i $f(P)$. Si $x \in m(P_i, 2^K) \cap f(P)$, i potser podria excloure x de l'elecció, però no vol. Si $x \notin m(P_i, 2^K) \cup f(P)$, i potser podria incloure x en l'elecció, però no vol. Si $x \in m(P_i, 2^K) \backslash f(P)$, i ja ha donat el seu suport per incloure x en l'elecció, però no ha estat suficient. Finalment, si $x \in f(P) \backslash m(P_i, 2^K)$, i no ha donat suport a x, però x ha tingut suficient suport per a ser inclòs en l'elecció. Donada la separabilitat o additivitat de P_i és fàcil concloure que i no pot manipular f en el perfil P; per tant, f és no manipulable en $\mathcal{S}e$ i en $\mathcal{A}$.

4 Restriccions de domini de bé privat

En aquesta secció considerarem problemes d'elecció social on les alternatives tenen components privats, aquells que només interessen a cada un dels agents, i potser també components públics. Per a cada agent i, podrem fer una partició del conjunt d'alternatives A en classes d'equivalència de manera que, des del punt de vista de i, les diferències entre les alternatives en la mateixa classe siguin irrellevants. Sigui A_i el conjunt quocient de totes les classes d'equivalència. La interpretació

[22] Noteu però, que qualsevol funció d'elecció social unànime és exhaustiva. Barberà, Massó i Neme (1997, 2005) obtenen caracteritzacions de totes les funcions d'elecció social no manipulables. L'interès d'aquesta extensió és que les funcions d'elecció social no exhaustives són necessàries per resoldre problemes on existeixen restriccions que fan que no tots els subconjunts de candidats puguin ser escollits (degut a restriccions pressupostàries o a polítiques de promoció de paritat de sexes entre els candidats escollits, per exemple).

[23] La part més complicada és la que les funcions d'elecció social exhaustives i no manipulables en $\mathcal{S}e$ on en $\mathcal{A}$ només depenen del vector de subconjunts ideals.

d'aquestes classes d'equivalència és que, tot i que l'alternativa y és diferent a l'alternativa x, que x i y estiguin a la mateixa classe vol dir que les dues alternatives difereixen en aspectes que no afecten a i; en particular, qualsevol preferència d'i en A les ha de considerar com indiferents. Aquesta serà una de les raons (a vegades l'única) per restringir el domini de preferències. Sovint, el conjunt A tindrà una estructura de producte cartesià on cada alternativa $x \in A$ pot ser descrita com $x = (x_1, ..., x_n) \in A_1 \times ... \times A_n$.

Sigui $\mathcal{R}_i$ el conjunt de preferències de l'agent i en A_i induïdes pel conjunt de preferències $\mathcal{R}$ en A. Abusant de la notació, farem servir R_i per denotar tant una preferència del conjunt $\mathcal{R}_i$ definida en A_i com del conjunt $\mathcal{R}$ definida en A.

En aquests casos serà convenient considerar funcions d'elecció social $f : \mathcal{D}_1 \times ... \times \mathcal{D}_n \to A$ que siguin mecanismes de revelació directa en el sentit que cada agent i ha de revelar preferències d'un subconjunt $\mathcal{D}_i \subset \mathcal{R}_i$ de preferències en A_i. Donat $R = (R_1, ..., R_n) \in \mathcal{D}_1 \times ... \times \mathcal{D}_n$ i $i \in N$, denotarem per $f_i(R)$ la classe d'equivalencia d'A_i que conté $f(R)$. Aleshores, direm que $f : \mathcal{D}_1 \times ... \times \mathcal{D}_n \to A$ és *no manipulable* (en el domini $\mathcal{D}_1 \times ... \times \mathcal{D}_n$) quan per a tot $R = (R_1, ..., R_n) \in \mathcal{D}_1 \times ... \times \mathcal{D}_n$, tot $i \in N$ i tot $R'_i \in \mathcal{D}_i$,

$$f_i(R) R_i f_i(R'_i, R_{-i}).$$

En cada una de les cinc restriccions que presentem a continuació, serem més explícits sobre aquesta construcció.

4.1 El problema de la divisió: la regla uniforme

Considerem el problema d'un conjunt d'agents que han de repartir-se una quantitat d'un bé homogeni i perfectament divisible.[24] Per exemple, un grup d'agents participa en una activitat que requereix una quantitat fixa de treball M (mesurat en unitats de temps). Cada un dels agents ha de contribuir amb una quantitat de treball, la suma de les quals ha de ser igual a M, i per la qual rebrà un salari de w unitats monetàries per unitat de temps. Els agents consideren el treball com a no desitjable. Per tant, volen disfrutar del màxim número d'hores de lleure (hores no treballades) i de diners, però per tenir més diners han de treballar més, i això vol dir disminuir les hores de lleure. Donat un salari $w > 0$, preferències monòtones i quasi-còncaves en el conjunt de parells (lleure,diners)[25] generen preferències unimodals en el conjunt $[0, M]$ d'hores que l'agent pot treballar, on la quantitat de treball ideal és aquella associada al parell (lleure,diners) òptim. La figura 1 ilustra aquesta construcció, comú en la teoria microeconòmica, on (x_i^*, d_i^*) és l'elecció òptima d'i, i per tant, $M - x_i^*$ és la part ideal d'M que i voldria rebre.

[24] Sprumont (1991) és el primer en formular aquest problema com un problema d'incentius en el context de la teoria de l'elecció social.

[25] Unes preferències en $\mathbb{R}^2_+$ són *monòtones* si per a tot $x, y \in \mathbb{R}^2_+$, $x \neq y$, $x_1 \geq y_1$ i $x_2 \geq y_2$ implica xP_iy, i són *quasi-còncaves* si per a tot $x \in \mathbb{R}^2_+$, el conjunt $\{y \in \mathbb{R}^2_+ \mid yR_ix\}$ és convex.

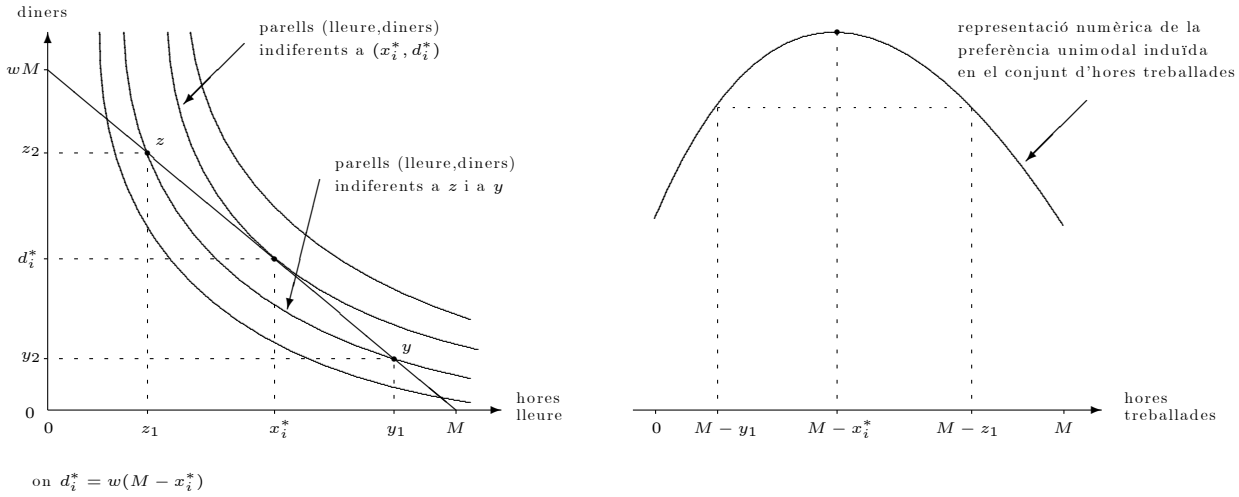

Figura 1

Un altre exemple és el d'un grup d'agents que volen invertir en un projecte (o en un actiu financer amb un valor donat) que requereix una quantitat determinada de diners (ni més ni menys). El rendiment de l'actiu, les actituds dels agents davant del risc i les seves riqueses indueixen preferències unimodals en la quantitat a invertir. Finalment, un grup d'empreses de diferents grandàries han d'empendre conjuntament un projecte d'una determinada dimensió. Com que les empreses poden estar involucrades en altres projectes, les seves preferències són unimodals en les seves respectives parts del projecte.

En general, el *problema de la divisió* consisteix en un conjunt d'agents $N = \{1, ..., n\}$ i una quantitat fixada d'un bé perfectament divisible $M > 0$, que ha de dividir-se en n parts, una per a cada agent. Sense pèrdua de generalitat, suposarem que $M = 1$. El conjunt d'alternatives socials A és doncs

$$A = \{x = (x_1, ..., x_n) \in \mathbb{R}_+^n \mid \textstyle\sum_{i \in N} x_i = 1\}.$$

En aquest cas, de l'alternativa $x = (x_1, ..., x_n)$, a cada agent i només li interessa el component $i-$èssim que especifica la seva part (és indiferent entre tots els parells $x, y \in A$ tals que $x_i = y_i$). Per a qualsevol $x \in A$, la classe d'equivalència de i que conté x és $[x]_i = \{y \in A \mid x_i = y_i\}$ i podem identificar A_i amb l'interval $[0, 1]$. A més, suposarem que aquestes preferències induïdes en $[0, 1]$ són unimodals. Un *perfil* de preferències unimodals $R = (R_1, ..., R_n) \in \mathcal{UM}^n$ descriu ara la preferència de cada agent en l'interval $[0, 1]$ de possibles parts rebudes i $(t(R_1), ..., t(R_n))$ és el vector de parts ideals dels n agents. La restricció de domini que ens allunya de la hipòtesi de domini universal té un doble component: (i) a cada agent només li preocupa la seva part rebuda i (ii) la unimodalitat de les preferències en el conjunt $[0, 1]$ de possibles parts rebudes.

Una funció d'elecció social $f : \mathcal{UM}^n \to A$ assigna a cada perfil de preferències unimodals en $[0, 1]$ un vector de n parts. En aquest cas, $f_i(R)$ és la part rebuda per l'agent i en el perfil R. Com ja hem assenyalat, estem lluny de la hipòtesi de domini universal del teorema de Gibbard-Satterthwaite. De fet, aquesta restricció de domini en el problema de la divisió permet una família

molt gran de funcions d'elecció social no manipulables.[26]

En general, la suma de les parts ideals $\sum_{i\in N} t(R_i)$ serà més gran o més petita que 1. En el primer cas parlem d'un problema de racionament positiu i en el segon cas negatiu. Les funcions d'elecció social difereixen entre elles en com resolen aquest problema de racionament en termes dels incentius que generen, de la seva eficiència, equitat, consistència, monotonicitat, etc.

En aquest context, una funció d'elecció social $f : \mathcal{UM}^n \to A$ és *no manipulable* quan per a tot perfil $R = (R_1, ..., R_n) \in \mathcal{UM}^n$, tot agent $i \in N$, i tota preferència $R'_i \in \mathcal{UM}$,

$$f_i(R)\, R_i f_i\left(R'_i, R_{-i}\right),$$

i és *eficient* quan per a cada perfil $R = (R_1, ..., R_n) \in \mathcal{UM}^n$ no existeix cap altra alternativa $y \in A$ amb la propietat que $y_i R_i f_i(R)$ per a tot $i \in N$ i $y_j P_j f_j(R)$ per algun $j \in N$. És fàcil comprovar que una funció d'elecció social és eficient si i només si tots els agents són racionats en la mateixa direcció; és a dir, si per a tot perfil $R = (R_1, ..., R_n) \in \mathcal{UM}^n$,

(1) si $\sum_{i\in N} t(R_i) \geq 1$ llavors $f_i(R) \leq t(R_i)$ per a tot $i \in N$,

(2) si $\sum_{i\in N} t(R_i) < 1$ llavors $f_i(R) \geq t(R_i)$ per a tot $i \in N$.

Finalment una funció d'elecció social $f : \mathcal{UM}^n \to A$ *no genera enveja* si per a tot perfil $R = (R_1, ..., R_n) \in \mathcal{UM}^n$ i tot parell d'agents $i, j \in N$,

$$f_i(R) R_i f_j(R);$$

és a dir, els agents prefereixen la seva part assignada per f (en el perfil R) que la part assignada a qualsevol altre agent.

Sprumont (1991) dóna dues caracteritzacions d'una funció d'elecció social anomenada *regla uniforme*. Aquesta regla, que ha jugat un paper central en l'estudi del problema de la divisió, intenta assignar les n parts de la forma més igualitària possible sense violar l'eficiència.

La *regla uniforme* $U : \mathcal{UM}^n \to A$ assigna, per a cada perfil $R = (R_1, ..., R_n) \in \mathcal{UM}^n$ i a cada agent $i \in N$, la part

$$U_i(R) = \begin{cases} \min\{\beta, t(R_i)\} & \text{si } \sum_{j\in N} t(R_j) \geq 1 \\ \max\{\beta, t(R_i)\} & \text{si } \sum_{j\in N} t(R_j) < 1, \end{cases}$$

on β és l'única solució a l'equació $\sum_{j\in N} \min\{\beta, t(R_j)\} = 1$ si $\sum_{j\in N} t(R_j) \geq 1$ i és l'única solució a l'equació $\sum_{j\in N} \max\{\beta, t(R_j)\} = 1$ si $\sum_{j\in N} t(R_j) < 1$. Noteu que β depèn del perfil R.

Abans de presentar el resultat de Sprumont (1991) considerem el següent exemple que ilustra la regla uniforme.

Exemple 4 Considerem el problema de la divisió on $N = \{1, 2, 3\}$. Sigui $R \in \mathcal{UM}^3$ qualsevol perfil tal que $t(R_1) = 0'1$, $t(R_2) = 0'2$ i $t(R_3) = 0'5$. Llavors $f(R) = (0'25, 0'25, 0'5)$ i $\beta = 0'25$. Sigui $R' \in \mathcal{UM}^3$ qualsevol altre perfil tal que $t(R'_1) = 0'3$, $t(R'_2) = 0'4$ i $t(R'_3) = 0'7$. Llavors $f(R') = (0'3, 0'35, 0'35)$ i $\beta' = 0'35$. En cap dels dos perfils l'alternativa igualitària $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ és eficient, però $f(R)$ i $f(R')$ són, per ambdós casos, les alternatives més igualitàries d'entre totes les eficients. □

[26] El conjunt de totes les funcions d'elecció social no manipulables és extraordinàriament gran; fins ara només se n'han caracteritzat algunes de les seves subclasses. Veure per exemple Barberà, Jackson i Neme (1997).

Una possible metàfora per descriure la regla uniforme és la següent. Suposem que hem de dividir un litre d'un líquid entre n agents. Cada agent es presenta amb un recipient d'una determinada capacitat (la seva part ideal), i suposem que la suma de les capacitats dels n recipients és més gran que 1 (l'altre cas es pot descriure amb una metàfora simètrica). La regla uniforme uneix tots els recipients connectant-los amb vasos comunicants, i comença a omplir el recipient més gran de manera que tots s'omplen uniformement. Dues coses són possibles: (i) El litre s'acaba abans de que cap recipient s'hagi omplert completament; llavors cada recipient conté $\frac{1}{n}$ i $(\frac{1}{n}, ..., \frac{1}{n})$ és l'assignació proposada per la regla uniforme ($\beta = \frac{1}{n} \leq t(R_i)$ per a tot $i \in N$). (ii) Abans de que s'acabi el litre, un dels recipients (el més petit) s'omple completament, i en aquest cas l'agent corresponent rep la seva part ideal i és desconnectat del sistema de vasos comunicants; la regla uniforme procedeix similarment amb els $n-1$ agents restants fins que o bé s'acaba el litre (la quantitat de líquid dels $n-1$ recipients és igual a β) o un altre recipient (el segon més petit) s'omple completament, l'agent corresponent rep la seva part ideal i és desconnectat del sistema de vasos comunicants, etc.; com que la suma de les capacitats és més gran que 1 aquest procés s'acaba i la quantitat de líquid que conté cada un dels recipients encara connectats és la part (β) que reben els corresponents agents, els que són finalment racionats.

Teorema 3 (Sprumont, 1991)

▶ *Una funció d'elecció social $f : \mathcal{UM}^n \to A$ és no manipulable, eficient i anònima si i només si f és la regla uniforme.*

▶ *Una funció d'elecció social $f : \mathcal{UM}^n \to A$ és no manipulable, eficient i no genera enveja si i només si f és la regla uniforme.*

La demostració del teorema consisteix primer en verificar que la regla uniforme satisfà les quatre propietats. Les altres dues implicacions es demostren recursivament en el número d'agents però són complicades, poc intuitives i fora de l'abast d'aquest article. Hi ha una llarga llista d'articles caracteritzant la regla uniforme amb altres conjunts de propietats.

4.2 Béns col·lectius amb preferències quasi-lineals: mètodes pivotals

Una societat formada per un grup d'agents ha de decidir conjuntament sobre la possible provisió d'un bé col·lectiu. Per exemple, la construcció d'un pont per travessar un riu en un poble, la instal·lació d'un ascensor o d'una antena de televisió col·lectiva en una escala de veïns, o la construcció d'una piscina en un jardí comunitari. La característica fonamental del bé col·lectiu és que és binari (es proveeix o no) i que si es proveeix, pot ser utilitzat simultàniament per diferents agents, tot i que algun d'ells pot ser exclòs del seu ús. Cada agent té una valoració monetària del bé col·lectiu, o la màxima disposició a pagar per ser-ne usuari. Tres decisions col·lectives s'han de prendre: (i) si el bé col·lectiu es proveeix o no, (ii) el conjunt dels seus usuaris, si es proveeix, i (iii) les contribucions monetàries de (o preus a pagar per) cada un dels agents. Una alternativa social és una terna especificant les tres decisions. Seria desitjable que aquestes depenguessin de les valoracions que els agents tenen del bé col·lectiu; però com que són informació privada han de ser sol·licitades als agents, donant lloc a un problema d'incentius. Voldríem doncs que la funció d'elecció social, que assigna a cada perfil de valoracions una alternativa especificant les tres

decisions, fos no manipulable.

Per poder ser més precisos, sigui com sempre $N = \{1, ..., n\}$ el conjunt d'agents que ha de decidir sobre la provisió del bé col·lectiu. Sigui $X = \{0, 1\}$ el conjunt de les dues decisions, on 1 significa que el bé col·lectiu es proveeix i 0 que no es proveeix, i sigui $x \in X$ una decisió genèrica. Es diu que el bé col·lectiu és *excloïble* quan un subconjunt d'agents poden ser exclosos del seu ús (els no usuaris), fins i tot quan $x = 1$. El conjunt d'agents que no són exclosos són els *usuaris*, que denotarem per S. El bé col·lectiu és *pur* si no és possible excloure cap agent del seu ús ($x = 1$ implica que $S = N$). De moment posposem les consideracions relacionades amb el cost de proveir el bé col·lectiu, ja que el resultat de possibilitat que presentarem cobreix tots els problemes de béns col·lectius binaris, independentment de les especificacions dels costos de provisió.

Per a cada agent $i \in N$, sigui $\alpha_i \in \mathbb{R}_+$ la *valoració* monetària que i assigna al bé col·lectiu si aquest és proveït i l'agent i n'és un usuari. En suposar que α_i és independent del conjunt d'usuaris S estem implícitament suposant que no hi ha efectes externs en l'ús del bé col·lectiu, com les aglomeracions. Un *perfil* $\alpha = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n_+$ és ara un vector de valoracions, una per a cada agent. Per a cada subconjunt d'agents $S \subseteq N$, sigui $\mathbf{1}_S : N \to \{0, 1\}$ la funció indicatriu, on per a cada $i \in N$,

$$\mathbf{1}_S(i) = \begin{cases} 1 & \text{si } i \in S \\ 0 & \text{si } i \notin S. \end{cases}$$

Sigui $p = (p_1, ..., p_n) \in \mathbb{R}^n$ un vector de *preus* (o contributions monetàries).[27]

El conjunt d'agents ha de triar col·lectivament una *alternativa*, que és una terna $(x, S, p) \in X \times 2^N \times \mathbb{R}^n$ amb la propietat que $x = 0$ implica $S = \emptyset$.[28] En aquest cas doncs, el conjunt d'alternatives socials és

$$A = \{(x, S, p) \in X \times 2^N \times \mathbb{R}^n \mid x = 0 \text{ implica } S = \varnothing\}.$$

Les preferències de l'agent i sobre el conjunt A depenen de la seva valoració α_i, per això les denotem per $R_i^{\alpha_i}$, i suposem que poden ser representades per la funció d'utilitat $u_i : A \times \mathbb{R}_+ \to \mathbb{R}$, on per a cada $(x, S, p, \alpha_i) \in A \times \mathbb{R}_+$,

$$u_i(x, S, p, \alpha_i) = \mathbf{1}_S(i) \cdot x \cdot \alpha_i - p_i;$$

és a dir, per a tot $\alpha_i \in \mathbb{R}_+$ i tot parell d'alternatives socials $(x, S, p), (x', S', p') \in A$,

$$u_i(x, S, p, \alpha_i) = \mathbf{1}_S(i) \cdot x \cdot \alpha_i - p_i \geq \mathbf{1}_{S'}(i) \cdot x' \cdot \alpha_i - p'_i = u_i(x', S', p', \alpha_i) \iff (x, S, p) R_i^{\alpha_i} (x', S', p').$$

Noteu que a l'agent i, d'una alternativa (x, S, p), només l'interessa si ell és o no usuari ($\mathbf{1}_S(i) \cdot x$) i el preu que ell paga (p_i). Per tant, el conjunt quocient A_i de les classes d'equivalència de i del conjunt A ve donat per $[(x, S, p)]_i = \{(x', S', p') \in A \mid \mathbf{1}_S(i) \cdot x = \mathbf{1}_{S'}(i) \cdot x' \text{ i } p_i = p'_i\}$. La valoració $\alpha_i > 0$ ordena el conjunt A de manera que totes les alternatives d'una mateixa classe d'equivalència són indiferents per i. Noteu que en aquest cas és possible que i consideri també com indiferents dues alternatives (x, S, p) i (x', S', p') en diferents classes d'equivalència, però llavors, $\mathbf{1}_S(i) \cdot x \cdot \alpha_i - p_i = \mathbf{1}_{S'}(i) \cdot x' \cdot \alpha_i - p'_i$.

[27] Admetem la possibilitat de que els preus siguin negatius; és a dir, que els agents puguin ser compensats.

[28] De moment no imposem cap condició en el vector de preus p ni excloem la possibilitat de que $x = 1$ i $S = \emptyset$.

Com que la societat N es mantindrà fixa, un perfil $\alpha = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n_+$ descriu completament un *problema*. Escriurem (α_i, α_{-i}) per emfatitzar el paper de l'agent i en el perfil α.

En aquest context, una *funció d'elecció social* $f : \mathbb{R}^n_+ \to A$ selecciona, per a cada perfil de valoracions $\alpha \in \mathbb{R}^n_+$, una alternativa $f(\alpha) \in A$. Per tant, una funció d'elecció social f pot ser identificada pels seus tres components $f \equiv (x^f, S^f, p^f)$, on $x^f : \mathbb{R}^n_+ \to \{0,1\}$, $S^f : \mathbb{R}^n_+ \to 2^N$ i $p^f : \mathbb{R}^n_+ \to \mathbb{R}^n$. És a dir, per a cada $\alpha \in \mathbb{R}^n_+$, $f(\alpha) = (x^f(\alpha), S^f(\alpha), p^f(\alpha))$. Donada una funció d'elecció social $f : \mathbb{R}^n_+ \to A$ i un agent $i \in N$, definim la funció $\beta^f_i : \mathbb{R}^n_+ \to \{0,1\}$ com: per a tot $\alpha \in \mathbb{R}^n_+$, $\beta^f_i(\alpha) = \mathbf{1}_{S^f(\alpha)}(i) \cdot x^f(\alpha)$. Per tant, $\beta^f_i(\alpha) = 1$ si i és un usuari a α i $\beta^f_i(\alpha) = 0$ si i no és un usuari a α.

Diem que un agent $i \in N$ *manipula* $f : \mathbb{R}^n_+ \to A$ *en el perfil* $\alpha \in \mathbb{R}^n_+$ *declarant* $\alpha'_i \in \mathbb{R}_+$ quan

$$\begin{aligned} u_i(x^f(\alpha'_i, \alpha_{-i}), S^f(\alpha'_i, \alpha_{-i}), p^f(\alpha'_i, \alpha_{-i}), \alpha_i) &= \beta^f_i(\alpha'_i, \alpha_{-i}) \cdot \alpha_i - p^f_i(\alpha'_i, \alpha_{-i}) \\ &> \beta^f_i(\alpha_i, \alpha_{-i}) \cdot \alpha_i - p^f_i(\alpha_i, \alpha_{-i}) \\ &= u_i(x^f(\alpha), S^f(\alpha), p^f(\alpha), \alpha_i). \end{aligned}$$

Una funció d'elecció social $f : \mathbb{R}^n_+ \to A$ és no manipulable quan cap agent pot manipular f en cap perfil.

El següent teorema caracteritza *totes* les funcions d'elecció social no manipulables per a béns col·lectius binaris. El teorema és conseqüència d'un resultat fonamental degut a Myerson (1981).[29]

Teorema 4 *Una funció d'elecció social $f : \mathbb{R}^n_+ \to A$ és no manipulable si i només si per a tot $i \in N$ existeixen dues funcions $\phi^f_i : \mathbb{R}^{n-1}_+ \to \mathbb{R}_+ \cup \{+\infty\}$ i $h^f_i : \mathbb{R}^{n-1}_+ \to \mathbb{R}$ tals que*

► *si $\alpha_i > \phi^f_i(\alpha_{-i})$ llavors $\beta^f_i(\alpha) = 1$ i $p^f_i(\alpha) = \phi^f_i(\alpha_{-i}) - h^f_i(\alpha_{-i})$;*

► *si $\alpha_i < \phi^f_i(\alpha_{-i})$ llavors $\beta^f_i(\alpha) = 0$ i $p^f_i(\alpha) = -h^f_i(\alpha_{-i})$;*

► *si $\alpha_i = \phi^f_i(\alpha_{-i})$ llavors o bé $[\beta^f_i(\alpha) = 1$ i $p^f_i(\alpha) = \phi^f_i(\alpha_{-i}) - h^f_i(\alpha_{-i})]$ o bé $[\beta^f_i(\alpha) = 0$ i $p^f_i(\alpha) = -h^f_i(\alpha_{-i})]$.*

Les funcions d'elecció social no manipulables identificades en el teorema 4 tenen dues propietats bàsiques. La primera, la decisió sobre si l'agent i és o no un usuari es pren a partir d'una funció simple monòtona creixent de la valoració α_i amb un punt de tall $\phi^f_i(\alpha_{-i})$ que només depèn de les valoracions dels altres agents: si $\alpha_i < \phi^f_i(\alpha_{-i})$ llavors i no és usuari, i si $\alpha_i > \phi^f_i(\alpha_{-i})$ sí que ho és (si $\alpha_i = \phi^f_i(\alpha_{-i})$ pot ser-ho o no, depenent de la funció f concreta). La segona és que, per a tots els perfils de valoracions α en els quals i no canvia la seva condició o no d'usuari, el preu que i ha de pagar és independent de la seva pròpia valoració; a més, la funció $h^f_i(\alpha_{-i})$ és la constant d'integració que apareix en aplicar el Teorema Fonamental del Càlcul. Per il·lustrar l'origen d'aquesta constant d'integració, considerem una part de la demostració del teorema 4. Sigui $f = (x^f, S^f, p^f)$ no manipulable. Fixem $i \in N$ i siguin $\alpha_i, \alpha'_i \in \mathbb{R}_+$ i $\alpha_{-i} \in \mathbb{R}^{n-1}_+$ arbitràries. Per la no manipulabilitat de f,

$$\beta^f_i(\alpha_i, \alpha_{-i}) \cdot \alpha_i - p^f_i(\alpha_i, \alpha_{-i}) \geq \beta^f_i(\alpha'_i, \alpha_{-i}) \cdot \alpha_i - p^f_i(\alpha'_i, \alpha_{-i}) \tag{3}$$

[29] Roger Myerson neix a Boston el 1951. Actualment és professor d'economia a la University of Chicago. Rep el premi Nobel d'economia l'any 2007, juntament amb Leonid Hurwicz i Eric Maskin, "per les seves contribucions a la teoria del disseny de mecanismes". Una demostració completa del resultat pot trobar-se a Massó, Nicolò i Sen (2010). Hi ha molts resultats similars a la literatura; per exemple, veure Dobzinski, Mehta, Roughgarden i Sundararajan (2008), Mehta, Roughgarden i Sundararajan (2007) i Nisan (2008).

i

$$\beta_i^f(\alpha_i', \alpha_{-i}) \cdot \alpha_i' - p_i^f(\alpha_i', \alpha_{-i}) \geq \beta_i^f(\alpha_i, \alpha_{-i}) \cdot \alpha_i' - p_i^f(\alpha_i, \alpha_{-i}). \tag{4}$$

Sumant (3) i (4),

$$(\beta_i^f(\alpha_i, \alpha_{-i}) - \beta_i^f(\alpha_i', \alpha_{-i})) \cdot (\alpha_i - \alpha_i') \geq 0. \tag{5}$$

Suposem, sense pèrdua de generalitat, que $\alpha_i > \alpha_i'$. Per (5),

$$\beta_i^f(\alpha_i, \alpha_{-i}) \geq \beta_i^f(\alpha_i', \alpha_{-i}).$$

És a dir, per a tot $\alpha_{-i} \in \mathbb{R}_+^{n-1}$, $\beta_i^f(\alpha_i, \alpha_{-i})$ és una funció creixent de α_i, i, com que només pren valors en el conjunt $\{0, 1\}$, té com a màxim un únic punt de discontinuïtat. Sigui $\phi_i^f(\alpha_{-i}) \in \mathbb{R}_+ \cup \{+\infty\}$ aquest punt de discontinuïtat (si $\phi_i^f(\alpha_{-i}) = +\infty$ la funció és de fet contínua i constant a 0 i l'agent i no és un usuari independentment de la seva valoració).

Per obtenir la forma funcional dels preus del teorema 4, per (3), i per a tot $\alpha_i, \alpha_i' \in \mathbb{R}_+$ i tot $\alpha_{-i} \in \mathbb{R}_+^{n-1}$,

$$u_i(x^f(\alpha_i, \alpha_{-i}), S^f(\alpha_i, \alpha_{-i}), p^f(\alpha_i, \alpha_{-i}), \alpha_i) \geq \beta_i^f(\alpha_i', \alpha_{-i}) \cdot \alpha_i - p_i^f(\alpha_i', \alpha_{-i}). \tag{6}$$

Com que

$$\begin{aligned} \beta_i^f(\alpha_i', \alpha_{-i}) \cdot \alpha_i - p_i^f(\alpha_i', \alpha_{-i}) &= u_i(x^f(\alpha_i', \alpha_{-i}), S^f(\alpha_i', \alpha_{-i}), p^f(\alpha_i', \alpha_{-i}), \alpha_i') \\ &\quad + \beta_i^f(\alpha_i', \alpha_{-i}) \cdot (\alpha_i - \alpha_i'), \end{aligned}$$

podem escriure (6) com

$$u_i(x^f(\alpha), S^f(\alpha), p^f(\alpha), \alpha_i) \geq u_i(x^f(\alpha_i', \alpha_{-i}), S^f(\alpha_i', \alpha_{-i}), p^f(\alpha_i', \alpha_{-i}), \alpha_i') + \beta_i^f(\alpha_i', \alpha_{-i}) \cdot (\alpha_i - \alpha_i'). \tag{7}$$

Similarment, per (4), i per a tot $\alpha_i, \alpha_i' \in \mathbb{R}_+$ i tot $\alpha_{-i} \in \mathbb{R}_+^{n-1}$,

$$u_i(x^f(\alpha_i', \alpha_{-i}), S^f(\alpha_i', \alpha_{-i}), p^f(\alpha_i', \alpha_{-i}), \alpha_i') \geq u_i(x^f(\alpha), S^f(\alpha), p^f(\alpha), \alpha_i) + \beta_i^f(\alpha) \cdot (\alpha_i' - \alpha_i). \tag{8}$$

Fixem $\alpha_{-i} \in \mathbb{R}_+^{n-1}$ i suposem, sense pèrdua de generalitat, que $\alpha_i' > \alpha_i$. Aleshores, per (7) i (8),

$$\begin{aligned} \beta_i^f(\alpha_i', \alpha_{-i}) &\geq \frac{u_i(x^f(\alpha_i', \alpha_{-i}), S^f(\alpha_i', \alpha_{-i}), p^f(\alpha_i', \alpha_{-i}), \alpha_i') - u_i(x^f(\alpha), S^f(\alpha), p^f(\alpha), \alpha_i)}{(\alpha_i' - \alpha_i)} \\ &\geq \beta_i^f(\alpha_i, \alpha_{-i}). \end{aligned} \tag{9}$$

Com acabem de veure, la funció (de t) $\beta_i^f(t, \alpha_{-i})$ només és discontínua en el punt $\phi_i^f(\alpha_{-i}) \in \mathbb{R}_+ \cup \{+\infty\}$. Fixem $\hat{\alpha}_i \in \mathbb{R}_+$. Aquí només considerarem el cas que $\beta_i^f(t, \alpha_{-i})$ és contínua a $\hat{\alpha}_i$ (veure Massó, Nicolò i Sen (2010) pel cas discontinu); *i.e.*, $\hat{\alpha}_i \neq \phi_i^f(\alpha_{-i})$. Sigui $\{\alpha_i'^k\}_{k=1}^{\infty} \to \hat{\alpha}_i$ una successió tal que per a tot $k \geq 1$, $\alpha_i'^k > \hat{\alpha}_i$. Com que $\beta_i^f(t, \alpha_{-i})$ és contínua en $\hat{\alpha}_i$, $\{\beta_i^f(\alpha_i'^k, \alpha_{-i})\}_{k=1}^{\infty} \to \beta_i^f(\hat{\alpha}_i, \alpha_{-i})$. Per (9),

$$\lim_{k \to \infty} \frac{u_i(x^f(\alpha_i'^k, \alpha_{-i}), S^f(\alpha_i'^k, \alpha_{-i}), p^f(\alpha_i'^k, \alpha_{-i}), \hat{\alpha}_i) - u_i(x^f(\hat{\alpha}_i, \alpha_{-i}), S^f(\hat{\alpha}_i, \alpha_{-i}), p^f(\hat{\alpha}_i, \alpha_{-i}), \hat{\alpha}_i)}{\alpha_i'^k - \hat{\alpha}_i}$$

existeix i és igual a $\beta_i^f(\hat{\alpha}_i, \alpha_{-i})$. Per tant,

$$\frac{\partial u_i(x^f(\hat{\alpha}_i, \alpha_{-i}), S^f(\hat{\alpha}_i, \alpha_{-i}), p^f(\hat{\alpha}_i, \alpha_{-i}), \hat{\alpha}_i)}{\partial \alpha_i} = \beta_i^f(\hat{\alpha}_i, \alpha_{-i})$$

per a tot $\hat{\alpha}_i \in \mathbb{R}_+$, on $\beta_i^f(t, \alpha_{-i})$ és contínua. Pel Teorema Fonamental del Càlcul,

$$u_i(x^f(\hat{\alpha}_i, \alpha_{-i}), S^f(\hat{\alpha}_i, \alpha_{-i}), p^f(\hat{\alpha}_i, \alpha_{-i}), \hat{\alpha}_i) = \int_0^{\hat{\alpha}_i} \beta_i^f(t, \alpha_{-i})dt + h_i^f(\alpha_{-i}),$$

on $h_i^f(\alpha_{-i})$ és una constant (*i.e.*, no depèn de α_i). Com que

$$u_i(x^f(\hat{\alpha}_i, \alpha_{-i}), S^f(\hat{\alpha}_i, \alpha_{-i}), p^f(\hat{\alpha}_i, \alpha_{-i}), \hat{\alpha}_i) = \beta_i^f(\hat{\alpha}_i, \alpha_{-i}) \cdot \hat{\alpha}_i - p_i^f(\hat{\alpha}_i, \alpha_{-i}),$$

$$p_i^f(\hat{\alpha}_i, \alpha_{-i}) = \beta_i^f(\hat{\alpha}_i, \alpha_{-i}) \cdot \hat{\alpha}_i - \int_0^{\hat{\alpha}_i} \beta_i^f(t, \alpha_{-i})dt - h_i^f(\alpha_{-i}). \tag{10}$$

Per tant, si $\hat{\alpha}_i < \phi_i^f(\alpha_{-i})$ llavors $\beta_i^f(\hat{\alpha}_i, \alpha_{-i}) = \beta_i^f(t, \alpha_{-i}) = 0$ per a tot $t \in [0, \hat{\alpha}_i]$. Per (10), $p_i^f(\hat{\alpha}_i, \alpha_{-i}) = -h_i^f(\alpha_{-i})$. Si $\hat{\alpha}_i > \phi_i^f(\alpha_{-i})$ llavors $\beta_i^f(\hat{\alpha}_i, \alpha_{-i}) = 1$, $\beta_i^f(t, \alpha_{-i}) = 0$ per a tot $t \in [0, \phi_i^f(\alpha_{-i}))$, i $\beta_i^f(t, \alpha_{-i}) = 1$ per a tot $t \in (\phi_i^f(\alpha_{-i}), \hat{\alpha}_i]$. Per (10),

$$\begin{aligned} p_i^f(\hat{\alpha}_i, \alpha_{-i}) &= \hat{\alpha}_i - \int_0^{\phi_i^f(\alpha_{-i})} \beta_i^f(t, \alpha_{-i})dt - \int_{\phi_i^f(\alpha_{-i})}^{\hat{\alpha}_i} \beta_i^f(t, \alpha_{-i})dt - h_i^f(\alpha_{-i}) \\ &= \hat{\alpha}_i - \hat{\alpha}_i + \phi_i^f(\alpha_{-i}) - h_i^f(\alpha_{-i}) \\ &= \phi_i^f(\alpha_{-i}) - h_i^f(\alpha_{-i}), \end{aligned}$$

que és el que volíem comprovar.

Considerem ara tres propietats desitjables (a més de la de no ser manipulables) que les funcions d'elecció social poden satisfer. La primera és la d'*individualitat racional*: tots els agents consideren que l'alternativa seleccionada és millor o igual a no participar (no sent usuari i pagant un preu igual a 0). La classe de funcions d'elecció social individualment racionals i no manipulables és la descrita al teorema 4 amb la condició addicional que per a tot $\alpha \in \mathbb{R}_+^n$ i tot $i \in N$, $h_i^f(\alpha_{-i}) \geq 0$.

Suposem ara que la provisió del bé col·lectiu requereix un cost de c unitats monetàries. Aleshores, una funció d'elecció social $f : \mathbb{R}_+^n \to A$ és *equilibrada pressupostàriament* quan per a tot perfil $\alpha \in \mathbb{R}_+^n$, $x^f(\alpha) = 0$ implica $\sum_{i \in N} p_i^f(\alpha) = 0$ i $x^f(\alpha) = 1$ implica $\sum_{i \in N} p_i^f(\alpha) = c$. Una funció d'elecció social $f : \mathbb{R}_+^n \to A$ és *eficient* quan per a tot $\alpha \in \mathbb{R}_+^n$, $\sum_{i \in N} \alpha_i > c$ implica $x^f(\alpha) = 1$ i $S^f(\alpha) = N$, i $\sum_{i \in N} \alpha_i < c$ implica $x^f(\alpha) = 0$ (i per tant, $S^f(\alpha) = \varnothing$).

Acabem aquesta subsecció amb un conegut resultat d'impossibilitat: no existeix cap funció d'elecció social no manipulable, individualment racional, efficient i equilibrada pressupostàriament. Per veureu-ho, considerem el cas on $N = \{1, 2, 3\}$ i $c = 9$ i suposem que f satisfà les tres primeres propietats. Veurem que f no és equilibrada pressupostàriament. Sigui $\alpha = (4, 4, 2)$ un vector de valoracions. Pel teorema 4 i l'eficiència de f, $\phi_1^f(4, 2) = \phi_2^f(4, 2) = 3$ i $\phi_3^f(4, 4) = 1$. Pel teorema 4 i l'individualitat racional de f, $p_i^f(4, 4, 2) \leq \phi_i^f(\alpha_{-i})$ per a tot $i \in N$ ja que $h_i^f(\alpha_{-i}) \geq 0$. Per tant, $\sum_{i \in N} p_i^f(4, 4, 2) \leq \sum_{i \in N} \phi_i^f(\alpha_{-i}) = 7 < 9 = c$; és a dir, f no és equilibrada pressupostàriament. Una part molt important de la literatura s'ha concentrat en estudiar funcions d'elecció social no manipulables i eficients quan el bé col·lectiu és pur (quan no és possible excloure als agents del seu ús: $x = 1$ implica $S = N$). Aquesta subclasse de funcions d'elecció social no manipulables i eficients és coneguda com el mètode pivotal, que Clarke (1971) i Groves (1973) varen proposar i estudiar independentment.

4.3 Assignació d'un objecte indivisible: la subhasta de segon preu de Vickrey

S'ha d'assignar un únic objecte indivisible a un agent d'entre un conjunt $N = \{1, ..., n\}$ d'agents. Per a cada agent $j \in N$, sigui $e^j \in \mathbb{R}^n$ el vector

$$e_i^j = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

que indica que l'agent j rep l'objecte. Sigui $X = \{e^1, ..., e^n\}$ el conjunt de totes les possibles assignacions de l'objecte als n agents. Per a cada agent $i \in N$, sigui $\alpha_i \in \mathbb{R}_+$ la *valoració* monetària que l'agent i atribueix a l'objecte. Un *perfil* $\alpha = (\alpha_1, ..., \alpha_n) \in \mathbb{R}_+^n$ és també ara un vector de valoracions, una per a cada agent. Sigui $p = (p_1, ..., p_n) \in \mathbb{R}^n$ un vector de preus, les contribucions monetàries de cada un dels agents. Una alternativa és un parell $(e^j, p) \in X \times \mathbb{R}^n = A$ que indica que l'agent j rep l'objecte i que cada agent i paga p_i. Suposem que les preferències de l'agent i sobre el conjunt d'alternatives A poden ser representades per la funció d'utilitat $u_i : A \times \mathbb{R}_+ \to \mathbb{R}$, on per a cada $(e^j, p, \alpha_i) \in X \times \mathbb{R}^n \times \mathbb{R}_+$,

$$u_i(e^j, p, \alpha_i) = e_i^j \cdot \alpha_i - p_i.$$

En aquest cas també, de cada alternativa $(e^j, p) \in A$, a l'agent i només li interessa si ell rep o no l'objecte (si $e_i^j = 1$ o $e_i^j = 0$) i el preu que paga (p_i). Per tant, el conjunt quocient A_i de les classes d'equivalència de i del conjunt A ve donat per $\left[(e^j, p)\right]_i = \{(\bar{e}^k, \bar{p}) \in A \mid \bar{e}_i^k = e_i^j \text{ i } \bar{p}_i = p_i\}$.

Una *funció d'elecció social* $f : \mathbb{R}_+^n \to A$ selecciona per a cada perfil $\alpha \in \mathbb{R}_+^n$ una alternativa $f(\alpha) \in A$. Qualsevol funció d'elecció social pot ser interpretada com una subhasta on els agents són postors i la valoració α_i és l'oferta del postor i. Vickrey (1961) proposa la subhasta de segon preu.[30] Per definir-la com una funció d'elecció social, reordenem qualsevol vector d'ofertes $\alpha \in \mathbb{R}_+^n$ com $\alpha^{\geq} = (\alpha_{i_1}^{\geq}, ..., \alpha_{i_n}^{\geq})$, on $\alpha_{i_1}^{\geq} \geq ... \geq \alpha_{i_n}^{\geq}$ i si $\alpha_j = \alpha_k$, $j < k$ implica $i_j < i_k$. Una funció d'elecció social $f : \mathbb{R}_+^n \to A$ és una *subhasta de segon preu* quan el postor amb l'oferta més gran guanya la subhasta, rep l'objecte i paga la segona oferta més gran mentre que els altres postors perden la subhasta i no paguen cap quantitat; és a dir, per a tot $\alpha \in \mathbb{R}_+^n$, $f(\alpha) = (e^j, p)$ on $j = i_1$ i per a tot $k \in N$,

$$p_k = \begin{cases} \alpha_{i_2} & \text{si } k = i_1 \\ 0 & \text{si } k \neq i_1. \end{cases}$$

El següent teorema estableix que la subhasta de segon preu és no manipulable; és a dir, per a tot perfil de valoracions $\alpha \in \mathbb{R}_+^n$, tot vector d'ofertes $b = (b_1, ..., b_n) \in \mathbb{R}_+^n$ i tot $i \in N$,

$$u_i(f(\alpha_i, b_{-i}), \alpha_i) \geq u_i(f(b), \alpha_i).$$

Teorema 5 (Vickrey, 1961) *La subhasta de segon preu és no manipulable.*

Demostració Per comprovar-ho, considerem qualsevol perfil de valoracions $\alpha \in \mathbb{R}_+^n$ i qualsevol vector d'ofertes $b \in \mathbb{R}_+^n$. Sense pèrdua de generalitat, suposem que $b_1 \geq ... \geq b_n$. El postor 1 guanya la subhasta, rep l'objecte i paga b_2; és a dir, $f(b) = (e^1, (b_2, 0, ..., 0))$.

[30] William Vickrey neix a Victoria (Canadà) el 1914 i mor l'octubre del 1996, pocs dies després de rebre el premi Nobel d'economia, juntament amb James Wirrlees, "per les seves contribucions a la teoria de la informació asimètrica".

Considerem primer el postor 1. Dos casos són possibles. Primer, $\alpha_1 \geq b_2$. Si $b'_1 \geq b_2$, el postor 1 guanya, paga b_2 i $u_1(f(b'_1, b_{-1}), \alpha_1) = \alpha_1 - b_2 \geq 0$; en particular, $u_1(f(\alpha_1, b_{-1}), \alpha_1) = \alpha_1 - b_2 \geq 0$. Si $b'_1 < b_2$, el postor 1 perd i $u_1(f(b'_1, b_{-1}), \alpha_1) = 0$. Per tant, $u_1(f(\alpha_1, b_{-1}), \alpha_1) \geq u_1(f(b'_1, b_{-1}), \alpha_1)$ per a tot $b'_1 \in \mathbb{R}_+$. Segon, $\alpha_1 < b_2$. Si $b'_1 \geq b_2$, el postor 1 guanya, paga b_2 i $u_1(f(b'_1, b_{-1}), \alpha_1) = \alpha_1 - b_2 < 0$. Si $b'_1 < b_2$, el postor 1 perd i $u_1(f(b'_1, b_{-1}), \alpha_1) = 0$; en particular, $u_1(f(\alpha_1, b_{-1}), \alpha_1) = 0$. Per tant, $u_1(f(\alpha_1, b_{-1}), \alpha_1) \geq u_1(f(b'_1, b_{-1}), \alpha_1)$ per a tot $b'_1 \in \mathbb{R}_+$.

Considerem ara qualsevol postor $i \neq 1$. Dos casos són possibles. Primer, $\alpha_i > b_1$. Si $b'_i > b_1$, el postor i guanya, paga b_1 i $u_i(f(b'_i, b_{-i}), \alpha_i) = \alpha_i - b_1 > 0$; en particular, $u_i(f(\alpha_i, b_{-i}), \alpha_i) = \alpha_i - b_1 > 0$. Si $b'_i \leq b_1$, el postor i perd i $u_i(f(b'_i, b_{-i}), \alpha_i) = 0$. Per tant, $u_i(f(\alpha_i, b_{-i}), \alpha_i) \geq u_i(f(b'_i, b_{-i}), \alpha_i)$ per a tot $b'_i \in \mathbb{R}_+$. Segon, $\alpha_i \leq b_1$. Si $b'_i > b_1$, el postor i guanya, paga b_1 i $u_i(f(b'_i, b_{-i}), \alpha_i) = \alpha_i - b_1 \leq 0$. Si $b'_i \leq b_1$, el postor i perd i $u_i(f(b'_i, b_{-i}), \alpha_i) = 0$; en particular, $u_i(f(\alpha_i, b_{-i}), \alpha_i) = 0$. Per tant, $u_i(f(\alpha_i, b_{-i}), \alpha_i) \geq u_i(f(b'_i, b_{-i}), \alpha_i)$ per a tot $b'_i \in \mathbb{R}_+$. ■

És fàcil comprovar que la subhasta de primer preu, en la qual el postor amb l'oferta més gran guanya, rep l'objecte i paga la seva pròpia oferta, és manipulable. Per exemple, considerem $n = 2$, $\alpha_1 = 2$, $\alpha_2 = 1$ i $b_2 = 1$. L'oferta $b_1 = 2$ és estrictament pitjor pel postor 1 que l'oferta $b'_1 = 1'5$.

La subhasta de segon preu (o algunes de les seves variants) és extensament utilitzada. Per exemple, en els mercats de segells, en el sistema d'eBay de licitació automàtica, en els programes de Google i Yahoo de publicitat en línia i en el mercat de bons del Tresor de molts països (per exemple, l'Índia).

4.4 Assignació d'objectes indivisibles: l'algoritme d'intercanvi de millors

Considerem el problema d'assignar n objectes indivisibles a n *agents*, de manera que cada agent rebi un objecte. Suposem que les compensacions monetàries no són possibles. Una *assignació* $\alpha : N \to O$ és una funció bijectiva que assigna a cada agent del conjunt $N = \{1, ..., n\}$ un objecte del conjunt $O = \{o_1, ..., o_n\}$. Suposem que existeix una *assignació inicial* d'agents a objectes $\mu : N \to O$ i que sense pèrdua de generalitat, $\mu(i) = o_i$ per a tot $i = 1, ..., n$.[31] Cada agent $i \in N$ té una *preferència estricta* P_i sobre el conjunt d'objectes O.[32]

Shapley i Scarf (1974) són els primers que proposen i estudien aquest model d'assignació d'objectes indivisibles. Els exemples de problemes amb aquestes característiques van des del mercat de cases, originalment suggerit per Shapley i Scarf (1974), o el de l'assignació d'estudiants a habitacions d'una residència en un campus universitari, fins al problema dels trasplantaments creuats de ronyons de donants vius, estudiat per primera vegada des del punt de vista de l'elecció social per Roth, Sönmez i Ünver (2004). En aquest últim cas, els agents són pacients amb insuficiència renal que necessiten un trasplantament de ronyó i que, per alguna característica específica

[31] El model i els resultats que presentem a continuació es poden modificar sense grans complicacions per admetre situacions on el número d'agents i objectes sigui diferent, o on els agents puguin rebre més d'un objecte.

[32] El supòsit de preferències estrictes és natural si tots els objectes són diferents, però no ho és quan alguns dels objectes són idèntics ja que aleshores haurien de ser considerats com a indiferents per a tots els agents. Aquesta extensió no és trivial; les solucions proposades per incloure les indiferències són molt més complicades. El lector interessat pot trobar dues solucions diferents en els articles de Alcalde-Unzu i Molis (2011) i Jaramillo i Manjunath (2012).

de la seva malaltia, tenen dificultats per a rebre ronyons provinents de donants cadàvers. Cada pacient té un donant viu, normalment un familiar pròxim, i formen una parella (pacient,donant). Si són compatibles, el trasplantament es realitza entre ells i la parella (pacient,donant) surt del problema d'assignació.[33] Si són incompatibles, potser per incompatibilitat del grup sanguini, el trasplantament no és possible. El ronyó del donant del pacient i és l'objecte $\mu(i)$. L'assignació μ enumera totes les n parelles (pacient,donant) incompatibles $(1, o_1), ..., (n, o_n)$ presents en el problema d'assignació. Donades les caracterísitiques dels ronyons, la preferència estricta P_i en O reflexa el grau de desitjable que els ronyons de tots els donants vius tenen pel pacient i. És el nefròleg del pacient qui determina aquest ordre sobre el conjunt de ronyons disponibles depenent de la compatibilitat genètica, l'edat, el pes, etc. En particular, $o_j P_i o_i$ significa que el ronyó del donant j és compatible amb el pacient i mentre que $o_i P_i o_j$ significa que el pacient i i el donant j són incompatibles. A més, $o_j P_i o_{j'}$ significa que el ronyó del donant j és *a priori* millor que el del donant j' per al pacient i; per exemple, per ser el donant j significativament més jove que el j', tot i tenir tots dos característiques genètiques similars (un HLA semblant).[34]

Un *perfil* $P = (P_1, ..., P_n)$ és una llista de preferències estrictes en O, una per a cada agent. Recordem que donada la preferència estricta P_i de l'agent i, definim la preferència dèbil R_i en O com: per a tot $j, j' = 1, ..., n$, $o_j R_i o_{j'}$ si i només si o bé $o_j = o_{j'}$ o bé $o_j P_i o_{j'}$.

Fixem N, O, μ i P, i anomenem al quàdruple (N, O, μ, P) un *problema* (*d'assignació*). Una solució del problema (N, O, μ, P) és una assignació $\alpha : N \to O$. En l'exemple de la donació creuada de ronyons de donants vius, una solució α assigna a cada pacient un ronyó (si $\alpha(i) = \mu(i) = o_i$ interpretem que el pacient i no rep cap ronyó ja que l'agent i és incompatible amb el ronyó o_i del seu donant). Noteu que en aquest model ordinal no es permeten compensacions monetàries; aquest fet és rellevant per a l'aplicació del moldel al problema dels trasplantaments creuats de ronyons de donants vius ja que en la majoria de països les compensacions monetàries entre pacients i donants estan prohibides explícitament. Perquè una assignació $\alpha : N \to O$ pugui ser considerada una solució del problema d'assignació, ha de satisfer algunes propietats bàsiques. La primera és indispensable si la participació dels agents en el problema d'assignació és voluntària. Una assignació $\alpha : N \to O$ és *individualment racional* en el problema (N, O, μ, P) quan cada agent rep un objecte almenys tant bo com el seu objecte inicial; és a dir, si per a cada $i \in N$, $\alpha(i) R_i \mu(i)$. En el cas contrari, $\mu(i) P_i \alpha(i)$, l'agent i podria bloquejar l'assignació α (en el cas del nostre exemple, estariem proposant al pacient i un trasplantament d'un ronyó incompatible, al qual podria negar-s'hi). La segona propietat exigeix que l'assignació faci un bon ús dels objectes disponibles. Una asignació $\alpha : N \to O$ és *eficient* en el problema (N, O, P, μ) quan no existeix cap altre assignació $\nu : N \to O$ tal que per a tot $i \in N$, $\nu(i) R_i \alpha(i)$ i $\nu(j) P_j \alpha(j)$ per algún $j \in N$. La tercera, més exigent que les dues anteriors ja que les implica, requereix que l'assignació sigui immune a secessions en el sentit que no existeixi cap subconjunt d'agents que puguin millorar, en relació a l'assignació proposada, reassignant-se entre ells els objectes inicialment assignats per μ; és a dir, que l'assignació no pugui ser bloquejada per cap subconjunt d'agents.

[33] Estem simplificant el problema quan suposem que no hi ha pacients amb dos o més donants incompatibles. El model es pot modificar sense grans complicacions per incloure aquest cas.

[34] Aquest subapartat es basa en Massó (2010), on es considera més detalladament el problema de donació creuada de ronyons de donants vius.

Definició 8 Una assignació $\alpha : N \to O$ pertany al *Nucli* del problema (N, O, P, μ) quan no existeix cap subconjunt (bloquejador) d'agents $S \subseteq N$ ni cap assignació $\nu : N \to O$ tals que:

(1) $\nu(i) \in \mu(S)$ per a tot $i \in S$,

(2) $\nu(i) R_i \alpha(i)$ per a tot $i \in S$, i

(3) $\nu(i) P_i \alpha(i)$ per algun $i \in S$.

Tota assignació del Nucli és individualment racional i eficient. Per veure-ho només cal considerar a la definició 8 els subconjunts amb un únic agent ($S = \{i\}$) i el conjunt de tots els agents ($S = N$). En l'exemple dels trasplantaments creuats de ronyons de donants vius, la pertinença al Nucli assegura que cap hospital o comunitat autònoma vulgui separar-se de l'organització nacional i resoldre separadament el subproblema amb el seu conjunt de parelles (pacient,donant). En aquest cas, la grandària del problema és important ja que com més parelles incompatibles hi hagi en el problema més fàcil serà que un pacient pugui rebre un ronyó d'un altre donant viu. Shapley i Scarf (1974) demostren que el Nucli de qualsevol problema és no buit.

Proposició 2 (Shapley i Scarf, 1974) *Tots els problemes d'assignació tenen un Nucli no buit.*

L'article de Shapley i Scarf (1974) conté dues demostracions alternatives de la proposició 2. La primera, la seva original, és indirecta i no constructiva. La segona, atribuïda per Shapley i Scarf a una suggerència de David Gale, consisteix en definir un algoritme, actualment conegut com l'algoritme d'intercanvi de millors i que anomenarem l'algoritme TTC de Gale (per *Top Trading Cycles* de l'anglès), que produeix per a cada problema una assignació en el Nucli.

L'algoritme TTC de Gale resol el problema d'assignació per etapes. En cada etapa (i) es construeix un graf on els vèrtexs són les parelles (agent,objecte) l'agent de les quals encara no ha estat assignat en les etapes anteriors; (ii) es dirigeix el graf (de cada vèrtex surt una fletxa assenyalant un altre vèrtex) fent que cada agent assenyali el seu millor objecte entre els que encara són presents en l'etapa; (iii) s'identifiquen els vèrtexs dels cicles del graf dirigit i (iv) es satisfan els cicles, assignant a cada agent dels vèrtexs dels cicles l'objecte que assenyala. L'algoritme TTC de Gale va identificant i satisfent successivament els cicles. Observeu que en cada etapa sempre existeix almenys un cicle, si hi ha diversos cicles aquests no s'intersecten entre ells i que un cicle pot tenir un únic vèrtex l'agent del qual assenyala al seu propi objecte (determinat per μ). Una mica més formalment, l'algoritme TTC de Gale pot descriure's com segueix.

- **Input:** Un problema d'assignació (N, O, μ, P).
- **Etapa 1:**
  - Cada agent "assenyala" el seu millor objecte. Com que hi ha un nombre finit n d'agents i objectes, el graf dirigit té almenys un cicle.
  - Cada agent d'un cicle és assignat a l'objecte que assenyala i es treu del problema d'assignació amb l'objecte assignat (és a dir, es satisfan els cicles).
  - Si queda almenys un agent, es passa a la següent etapa. En cas contrari, el resultat de l'algoritme és l'assignació definida en satisfer els cicles.

- **Etapa k:**
  - Cadascun dels agents que encara no ha estat assignat a les etapes anteriors "assenyala" el seu millor objecte d'entre els que encara queden per assignar.
  - Cada agent d'un cicle és assignat a l'objecte que assenyala i es treu del problema d'assignació amb l'objecte assignat (és a dir, es satisfan els cicles).
  - Si queda almenys un agent, es passa a la següent etapa. En cas contrari, el resultat de l'algoritme és l'assignació definida en satisfer els cicles en totes les etapes anteriors.

Denotarem per $\eta : N \to O$ l'assignació obtinguda en aplicar l'algoritme TTC de Gale a un problema (N, O, μ, P) i per K l'última etapa de l'algoritme. És fàcil comprovar que la complexitat del algoritme TTC de Gale és polinòmica i per tant, els problemes amb un número gran d'agents i objectes poder ser resolts amb temps d'ordinador raonables. L'exemple 5 il·lustra el funcionament de l'algoritme TTC de Gale.

Exemple 5 Sigui (N, O, μ, P) un problema d'assignació amb $\#N = \#O = 8$, $\mu(i) = o_i$ per a cada $i = 1, ..., 8$, i el perfil P representat a la taula 1, on l'objecte dintre d'un quadrat indica l'assignació inicial μ de cada agent.

Taula 1

P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
o_2	o_3	o_1	o_8	o_4	o_8	o_4	o_6
o_3	o_1	$\boxed{o_3}$	o_7	o_7	o_1	o_8	$\boxed{o_8}$
o_5	$\boxed{o_2}$	o_7	$\boxed{o_4}$	o_3	$\boxed{o_6}$	o_3	o_1
o_6	o_8	o_2	o_1	o_6	o_5	o_6	o_2
o_8	o_6	o_5	o_2	o_1	o_4	o_1	o_3
$\boxed{o_1}$	o_4	o_8	o_3	o_8	o_3	o_5	o_7
o_7	o_7	o_6	o_5	o_2	o_2	o_2	o_5
o_4	o_5	o_4	o_6	$\boxed{o_5}$	o_7	$\boxed{o_7}$	o_4

En la figura 2 representem les 3 etapes de l'algoritme TTC de Gale aplicat al perfil P per obtenir l'assignació η. □

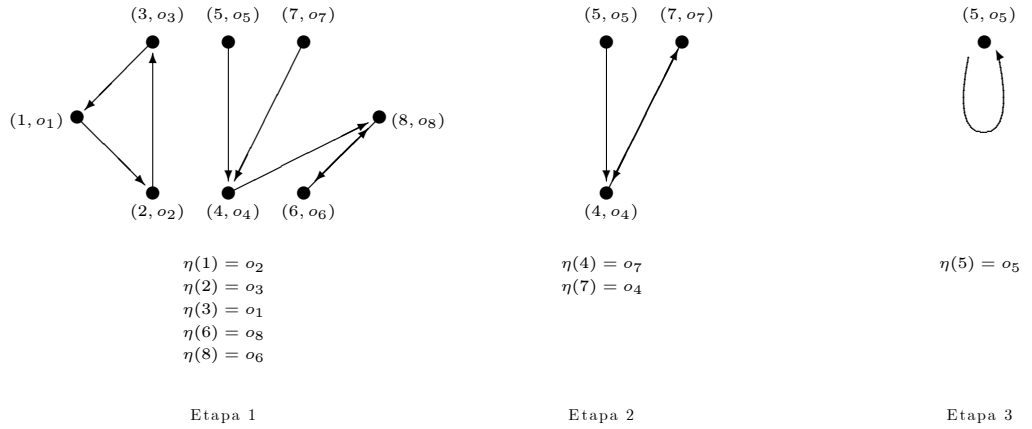

Figura 2

Passem ara a demostrar la proposició 2 comprovant que l'algoritme TTC de Gale selecciona una assignació en el Nucli.

Demostració de la proposició 2 Sigui η l'assignació obtinguda per l'algoritme TTC de Gale en el problema (N, O, μ, P). Per comprovar que η està en el Nucli de (N, O, μ, P), siguin $S_1, ..., S_K$ els conjunts d'agents que formen part dels cicles i que són assignats i trets del problema d'assignació en les etapes $1, ..., K$ de l'algoritme, respectivament. Noteu que $S_1, ..., S_K$ és una partició de N i que en S_k (el conjunt d'agents que són assignats a objectes i trets del problema d'assignació en l'etapa k) hi pot haver-hi diversos cicles. Observem que cap agent de S_1 pot ser agent d'un subconjunt bloquejador de l'assignació η que la prefereixi estrictament, ja que cada un d'ells ha estat assignat al seu millor objecte. Donat això, cap agent de S_2 pot ser agent d'un subconjunt bloquejador de η que la prefereixi estrictament, ja que cada un d'ells ha estat assignat al seu millor objecte entre el conjunt d'objectes restants $O \backslash \eta(S_1)$. Procedint iterativament, obtenim que η és una assignació del Nucli del problema (N, O, μ, P) ja que no pot ser bloquejada per cap subconjunt d'agents. ■

Roth i Postlewaite (1977) es van preguntar si existien altres assignacions en el Nucli diferents de la seleccionada per l'algoritme TTC de Gale. La resposta és negativa: el Nucli només conté l'assignació obtinguda per mitjà de l'algoritme TTC de Gale.

Proposició 3 (Roth i Postlewaite, 1977) *El Nucli de cada problema d'assignació només conté una única assignació.*

Demostració Sigui $\eta : N \to O$ l'assignació obtinguda per l'algoritme TTC de Gale en el problema (N, O, μ, P), i considerem qualsevol altra assignació diferent $\nu \neq \eta$. Volem demostrar que ν no està en el Nucli del problema (N, O, μ, P). Sigui k la primera etapa de l'algoritme TTC de Gale en la que hi ha un agent i a S_k (el conjunt d'agents que, en pertanyer a algun cicle en l'etapa k, són assignats a algun objecte i trets del problema en l'etapa k) amb la propietat que $\nu(i) \neq \eta(i)$; si hi ha varis agents amb aquesta propietat, escollim i de forma arbitrària. És a dir, i pertany a S_k i, per a tot agent j assignat per η abans dels cicles de S_k (és a dir, per a tot $j \in S_1 \cup ... \cup S_{k-1}$), es compleix que $\nu(j) = \eta(j)$. Per tant, per a tot $j \in S_k$, $\eta(j) R_j \nu(j)$ i $\eta(j) \in \mu(S_k)$. A més, per la

definició de η, $\eta(i)P_i\nu(i)$ ja que $\nu(i) \neq \eta(i)$. Això vol dir que el subconjunt S_k bloqueja l'assignació ν. Per tant, ν no està en el Nucli del problema (N, O, μ, P). ■

L'assignació η seleccionada per l'algoritme TTC de Gale en el problema (N, O, μ, P) depèn del perfil P i, en particular, l'objecte rebut per l'agent $i \in N$ a η depèn de la seva preferència P_i. Per tant, és natural preguntar-se si l'algoritme TTC de Gale, entès com a funció d'elecció social, incentiva els agents a revelar les seves verdaderes preferències.

Fixats N, O i μ sigui A el conjunt d'alternatives socials; és a dir, $A = \{\alpha : N \to O \mid \alpha$ és bijectiva$\}$. En aquest cas, a cada agent i només li interessa l'objecte que ell rep. Les seves preferències estrictes estan definides en el conjunt d'objectes O (i no en A). Per tant, el conjunt quocient A_i de les classes d'equivalència d'alternatives socials, que descriu les característiques de les alternatives d'A que tenen interés per a i, s'obté definint, per a tot $\alpha \in A$, $[\alpha]_i = \{\beta \in A \mid \beta(i) = \alpha(i)\}$, a on $[\alpha]_i$ representa la classe d'equivalència que conté l'assignació α. Donada la preferència estricta P_i en O podem definir, abusant de la notació, la preferència dèbil R_i en el conjunt d'alternatives socials A de la següent manera: per a tot parell $\alpha, \alpha' \in A$, $\alpha R_i \alpha'$ si i només si, o bé $\alpha(i) = \alpha'(i)$ o bé $\alpha(i)P_i\alpha'(i)$; efectivament, a l'agent i només li interessa, de les alternatives d'A, l'objecte que ell rep. La preferència dèbil R_i en A té moltes indiferències ja que i és indiferent entre totes les assignacions en les que i rep el mateix objecte (el fet que R_i representi la preferència dèbil en el conjunt d'objectes O així com la preferència dèbil en el conjunt d'alternatives socials A no ha de confondre al lector). Una vegada més, aquest problema concret d'assignació fa que el conjunt de preferències individuals en el conjunt d'alternatives socials A sigui un subconjunt de totes les possibles preferències a A, i que per tant estiguem lluny de la hipòtesi de domini universal de preferències implícita en el teorema d'impossibilitat de Gibbard-Satterthwaite. Per tant, podem esperar que existeixin funcions d'elecció social no manipulables.

Donat el conjunt d'objectes O, recordem que $\mathcal{P}$ és ara el conjunt de preferències estrictes en O. Una funció d'elecció social $f : \mathcal{P}^n \to A$ és doncs un mètode sistemàtic que proposa, fixats N, O i μ, per a cada perfil de preferències $P \in \mathcal{P}^n$ una assignació $\alpha : N \to O$. Donat $i \in N$, denotarem per $f_i(P) = \alpha(i)$, on $f(P) = \alpha$.

Com sempre, diem que una funció d'elecció social $f : \mathcal{P}^n \to A$ és *manipulable* quan existeix un perfil $P = (P_1, ..., P_n) \in \mathcal{P}^n$, un agent $i \in N$ i una preferència $P_i' \in \mathcal{P}$ tal que

$$f_i(P_i', P_{-i})P_i f_i(P_i, P_{-i});$$

és a dir, l'agent i obté un millor objecte (segons P_i) declarant P_i' en lloc de P_i. Roth (1982a) estableix que la funció d'elecció social que escull per a cada perfil P el Nucli del problema (N, O, μ, P) (és a dir, l'assignació seleccionada per l'algoritme TTC de Gale) és no manipulable.

Teorema 6 (Roth, 1982a) *El Nucli com a funció d'elecció social és no manipulable.*

Demostració Fixem N, O i μ. Sigui $\varphi : \mathcal{P}^n \to A$ la funció d'elecció social que selecciona per a cada problema (N, O, μ, P) l'única assignació del Nucli, l'obtinguda per l'algoritme TTC de Gale aplicat al problema (N, O, μ, P). Sigui $P \in \mathcal{P}^n$ un perfil arbitrari. Siguin $S_1, ..., S_K$ els conjunts d'agents que formen part dels cicles, els quals, en aplicar l'algoritme TTC de Gale per obtenir $\eta = \varphi(P)$, són assignats a objectes i trets del problema d'assignació en les etapes $1, ..., K$; és a dir,

$i \in S_k$ significa que l'agent i pertany a un dels cicles de l'etapa k. La demostració és per iteració en els cicles:

$k = 1$: Cada un dels agents de S_1 rep a l'assignació η el seu millor objecte segons P. Per tant, cap d'ells pot beneficiar-se declarant una preferència diferent. Observem, que els cicles de S_1 a l'etapa 1 són els mateixos, independentment que els agents de $N \backslash S_1$ declarin unes preferències diferents.

$k \geq 2$: Cada un dels agents de S_k (per a $k \geq 2$) rep a l'assignació η el seu millor objecte, entre el conjunt d'objectes $O \backslash (\eta(S_1) \cup ... \cup \eta(S_{k-1}))$, segons P. Com que els cicles anteriors en $S_1 \cup ... \cup S_{k-1}$ no es veuen afectats si els agents de S_k subministren unes altres preferències, si un agent de S_k canvia la seva preferència els agents de $S_1 \cup ... \cup S_{k-1}$ continuaran rebent els mateixos objectes que rebien. Per tant, en aplicar l'algoritme TTC de Gale, cap agent de S_k es pot beneficiar declarant unes preferències diferents de les verdaderes. ■

En algunes aplicacions, la noció de Nucli pot ser massa exigent ja que potser és difícil per als agents d'un subconjunt bloquejador identificar-se mútuament com a tals, i reassignar-se entre ells els objectes inicials. No obstant, sí que pot ser raonable exigir que la funció d'elecció social sigui individualment racional i eficient (per a cada perfil $P \in \mathcal{P}^n$, $f(P)$ és una assignació individualment racional i eficient a P), i que incentivi els agents a declarar les seves verdaderes preferències. Ma (1994) demostra que els subconjunts intermedis d'agents ($S \neq \{i\}$ i $S \neq N$) no tenen poder de bloqueig addicional.

Proposició 4 (Ma, 1994) *La funció d'elecció social $f : \mathcal{P}^n \to A$ és individualment racional, eficient i no manipulable si i només si f és la funció d'elecció social del Nucli (la que selecciona l'assignació d'acord amb l'algoritme TTC de Gale).*

La demostració de la proposició 4 està fora de l'abast d'aquest article. No obstant, és interessant assenyalar que les tres propietats (individualitat racional, eficiència i no manipulabilitat) són mútuament independents en el següent sentit. Primer, qualsevol funció d'elecció social dictatorial en sèrie[35] és eficient i no manipulable, però no és individualment racional. Segon, la funció d'elecció social que sempre selecciona l'assignació inicial ($f(P) = \mu$ per a tot P) és individualment racional i no manipulable, però no és eficient. Tercer, per veure que existeixen funcions d'elecció social individualment racionals i eficients però a la vegada manipulables, considerem el problema (N, O, μ, P) on $N = \{1, 2, 3\}$, $O = \{o_1, o_2, o_3\}$, $\mu(i) = o_i$ per a tot $i = 1, 2, 3$, i el perfil de preferències P és

P_1	P_2	P_3
o_2	o_1	o_1
o_3	o_3	o_3
o_1	o_2	o_2

[35] Una funció d'elecció social dictatorial en sèrie ordena els agents, i permet que, seguint aquest ordre, els agents es quedin successivament el millor dels objectes d'entre aquells que no han estat escollits pels seus predecessors. Les funcions d'elecció social dictatorials en sèrie no són individualment racionals ja que en general no garantitzen que els agents es puguin retirar del problema d'assignació amb un objecte millor o igual als seus objectes inicials.

Sigui $\varphi : \mathcal{P}^n \to A$ la funció d'elecció social del Nucli associada a l'algoritme TTC de Gale. Observem que $\varphi_1(P) = o_2$, $\varphi_2(P) = o_1$ i $\varphi_3(P) = o_3$ i definim una altra funció d'elecció social $\phi : \mathcal{P}^n \to A$ amb la propietat que ϕ coincideix amb φ en tots els perfils excepte en el perfil P, on l'assignació seleccionada és $\phi_1(P) = o_2$, $\phi_2(P) = o_3$ i $\phi_3(P) = o_1$. Com que la funció d'elecció social φ selecciona assignacions individualment racionals i eficients en tots els perfils, i com que l'assignació $\phi(P)$ és individualment racional i eficient en el perfil P, la funció ϕ és individualment racional i eficient. Per comprovar que ϕ és manipulable considerem el perfil $P' = (P_1, P_2, P_3')$ on $o_2 P_3' o_1 P_3' o_3$. Llavors, $\phi_1(P') = \varphi_1(P') = o_2$, $\phi_2(P') = \varphi_2(P') = o_1$ i $\phi_3(P') = \varphi_3(P') = o_3$. Per tant, $\phi_3(P) = o_1 P_3' o_3 = \phi_3(P')$. L'agent 3 manipula la funció d'elecció social ϕ en el perfil P' declarant P_3, obtenint així un millor objecte. La funció d'elecció social ϕ és individualment racional, eficient i manipulable.

Finalment, una altra propietat desitjable de la funció d'elecció social del Nucli associada a l'algoritme TTC de Gale és que la qualitat de l'objecte rebut per un agent depèn positivament de la qualitat del seu objecte. Una manera indirecta de veure aquesta propietat és que l'assignació del Nucli correspon a la que s'obtindria descentralitzadament a través d'un mercat, on les compensacions monetàries són possibles. Suposem que a cada objecte o_i se li assigna un preu $p_{o_i} \geq 0$. Diem que un objecte o_j és *accessible* per a l'agent i en el vector de preus $p = (p_{o_1}, ..., p_{o_n}) \in \mathbb{R}^n_+$ quan $p_{o_j} \leq p_{\mu(i)}$; és a dir, si i pot comprar l'objecte o_j al preu p_{o_j} després de vendre el seu objecte $\mu(i)$ al preu p_{o_i}. Una assignació $\nu : N \to O$ és un *equilibri* del problema (N, O, μ, P) quan existeix un vector de preus $p = (p_{o_1}, ..., p_{o_n})$ tal que per a cada agent i, $\nu(i)$ és el millor objecte d'entre el conjunt d'objectes accessibles per a ell en $p = (p_{o_1}, ..., p_{o_n})$. Roth i Postlewaite (1977) demostren que tot problema d'assignació té un únic equilibri, i aquest coincideix amb l'assignació del Nucli (la seleccionada per l'algoritme TTC de Gale).

Proposició 5 (Roth i Postlewaite, 1977) *Per a cada problema d'assignació només hi ha una assignació que sigui un equilibri, i aquesta és l'assignació del Nucli.*

La demostració de la proposició 5 també queda fora de l'àmbit d'aquest article.

En resum, l'assignació escollida per l'algoritme TTC de Gale té molt bones propietats. No només selecciona l'única assignació del Nucli, sinó que a més, com a funció d'elecció social, és no manipulable (de fet, és l'única d'entre totes les individualment racionals i eficients que ho és) i és la que s'obtindria a partir d'un mercat descentralitzat, i per tant, agents amb objectes amb més qualitat reben millors objectes. De fet, en la majoria de països o regions amb programes centralitzats de donacions de ronyons creuades fan servir l'algoritme TTC de Gale, o algunes de les seves variants.

4.5 Assignacions bilaterals estables: l'algoritme d'acceptació diferida

Considerem problemes d'elecció social que consisteixen en assignar dos subconjunts disjunts d'agents, de manera que cada agent d'un dels subconjunts estigui o bé assignat a un agent de l'altre subconjunt, o bé sense assignar. Aquesta classe de problemes va ser proposada i estudiada per Gale i Shapley (1962). Exemples típics d'assignació bilaterals són els d'un conjunt d'homes a un con-

junt de dones, d'estudiants a titulacions universitàries, de nens de 6 anys a les escoles públiques de primària d'una ciutat, de metges interns residents a places hospitalàries, o la de treballadors novells a empreses en mercats de treball professionals.[36] Farem servir com a referència l'exemple de treballadors i empreses. Per tant, el conjunt d'*agents* del nostre problema d'assignació bilateral està format per dos conjunts disjunts, el conjunt d'*empreses* E i el conjunt de *treballadors* T. Suposarem que $\#E \geq 2$ i $\#T \geq 2$. Un agent genèric serà denotat per $i \in N = E \cup T$, mentre que empreses i treballadors genèrics seran denotats per e i t, respectivament. Cada treballador $t \in T$ té una preferència estricta P_t en el conjunt $E \cup \{t\}$, el conjunt d'empreses E a les quals t pot ser assignat o no. La possibilitat de quedar-se sense empresa l'identifiquem amb ser assignat a ell mateix. Cada empresa $e \in E$ té una preferència estricta P_e en el conjunt $T \cup \{e\}$, el conjunt de treballadors T als quals e pot ser assignat o no. La possibilitat de quedar-se sense cap treballador la identifiquem amb ser assignada a ella mateixa. Sigui $\mathcal{P}_i$ el conjunt de totes les preferències estrictes de l'agent i. Sigui $\mathcal{P} = \prod_{i \in N} \mathcal{P}_i$ el conjunt de *perfils* de preferències, una per a cada agent.

Un *problema* (*d'assignació bilateral*) és una terna (E, T, P), on E és un conjunt d'empreses, T és un conjunt de treballadors i P és un perfil de preferències. El problema d'assignació consisteix en assignar treballadors a empreses, mantenint la naturalesa bilateral de les seves relacions i permetent la possibilitat que tant empreses com treballadors quedin sense ser assignats. És a dir, una *assignació bilateral* és una funció $\mu : N \to N$ amb les següents propietats: (i) per a tot $t \in T$, $\mu(t) \in E \cup \{t\}$; (ii) per a tota $e \in E$, $\mu(e) \in T \cup \{e\}$; i (iii) per a tot $i \in N$, $\mu(\mu(i)) = i$. Fixats E i T, el conjunt d'alternatives socials A és el conjunt de totes les assignacions bilaterals. En aquest cas, d'una assignació bilateral μ a cada agent $i \in N$ només li interessa l'agent de l'altre subconjunt assignat a i, $\mu(i)$, i li és igual com estan assignats els altres agents. Ara, el conjunt quocient A_i de classes d'equivalència d'alternatives socials es defineix com: per a tot $\mu \in A$, $[\mu]_i = \{\nu \in A \mid v(i) = \mu(i)\}$. Podem estendre la preferència estricta P_i sobre el conjunt original d'assignats potencials ($E \cup \{i\}$ si $i \in T$ i $T \cup \{i\}$ si $i \in E$) al conjunt d'assignacions bilaterales A (d'alternatives socials) identificant una assignació bilateral μ amb $\mu(i)$. Per tant, i abusant de la notació, dir que l'empresa e prefereix estrictament μ' a μ ($\mu' P_e \mu$) significa que $\mu'(e) P_e \mu(e)$, i prefereix dèbilment μ' a μ ($\mu' R_e \mu$) significa que o bé $\mu'(e) = \mu(e)$ o bé $\mu'(e) P_e \mu(e)$.

Una assignació bilateral és *estable* quan totes les empreses i tots els treballadors estan assignats a agents almenys tan preferits com a ells mateixos (*racionalitat individual*) i cap parella empresa-treballador no assignats es prefereixen mútuament més que als seus agents assignats (*estabilitat a parelles*). És a dir, donat un perfil $P \in \mathcal{P}$ una assignació bilateral $\mu \in A$ és *estable a* P quan (i) per a tot $i \in N$, $\mu(i) R_i i$, i (ii) no existeix cap parell $(e, t) \in E \times T$ tal que $e P_t \mu(t)$ i $t P_e \mu(e)$. La noció d'estabilitat juga un paper central en l'estudi dels problemes d'assignació bilateral. Si l'assignació bilateral és voluntària, la seva estabilitat és una condició indispensable perquè l'assignació proposada perduri, ja que si no ho és, o bé un agent trencant unilateralment amb la seva parella proposada o bé una parella (empresa,treballador), identificant-se com a parella

[36] Aquí presentarem resultats del model en què els agents només poden ser assignats a un altre agent. Hi ha una extensa literatura que estudia models en què, com en alguns d'aquests exemples, els agents d'un dels subconjunts (el de les institucions) poden ser assignats a més d'un agent de l'altre subconjunt (el de les persones). Roth i Sotomator (1990) és una monografia magistral sobre aquests models bilaterals d'assignació.

bloquejadora (enviant-se un missatge electrònic o fent-se una trucada de telèfon), poden trencar amb els respectius agents assignats i formar una parella entre ells, fent inviable l'assignació proposada. Gale i Shapley (1962) demostren que el conjunt d'assignacions estables a P és no buit i coincideix amb el Nucli del problema (E, T, P); és a dir, no hi ha pèrdua de generalitat si suposem que tot el poder de bloqueig recau només en els agents individuals i en les parelles (empresa,treballador), ja que les coalicions d'agents més grans no tenen cap poder addicional de bloqueig. Denotem per $S(P)$ el conjunt d'assignacions estables a P. Donat un perfil $P \in \mathcal{P}$ una assignació bilateral $\mu \in A$ és *eficient* quan no existeix una altra assignació bilateral $\mu' \in A$ tal que $\mu'(i) R_i \mu(i)$ per a tot $i \in N$ i $\mu'(j) P_j \mu(j)$ per algun agent $j \in N$.

El conjunt d'assignacions estables $S(P)$ té una estructura reticular completa i dual.[37] Per tant, el conjunt d'assignacions estables $S(P)$ conté dues assignacions estables, μ_E i μ_T, els màxims dels dos ordres dels dos reticle complets (anomenades assignació estable òptima de les empreses i assignació estable òptima dels treballadors, respectivament) que tenen la propietat que les empreses (els treballadors) unànimament estan d'acord que μ_E (μ_T) és la millor assignació entre totes les estables; a més, l'assignació estable òptima per a uns és l'assignació estable subòptima per als altres. És a dir, per a tot $\mu \in S(P)$, (i) $\mu_E(e) R_e \mu(e) R_e \mu_T(e)$ per a tot $e \in E$ i (ii) $\mu_T(t) R_t \mu(t) R_t \mu_E(t)$ per a tot $t \in T$.

Que una assignació sigui o no estable depèn de les preferències dels agents i com que aquestes preferències són informació privada, els agents han de ser preguntats sobre elles. Una funció d'elecció social en aquest context és una funció $f : \mathcal{P} \to A$ que assigna a cada perfil de preferències $P \in \mathcal{P}$ una assignació bilateral $f(P) \in A$. Escriurem $f_i(P)$ per indicar l'agent assignat a i al perfil P per la funció d'elecció social f. Una funció d'elecció social f és *estable* si per a tot $P \in \mathcal{P}$, $f(P) \in S(P)$. Molts dels exemples mencionats a la introducció d'aquesta subsecció utilitzen mecanismes centralitzats per proposar una solució al problema d'assignació, que són de fet funcions d'elecció social estables.

Les dues versions de l'*algoritme d'acceptació diferida* (AAD) presentat per Gale i Shapley (1962) defineixen dues funcions d'elecció social estables que produeixen, per a cada $P \in \mathcal{P}$, o bé μ_E o bé μ_T, depenent de quin subconjunt d'agents faci les ofertes. En qualsevol etapa de l'algoritme on les empreses fan les ofertes, cada empresa e es proposa al treballador més preferit d'entre tots els que encara no han rebutjat e durant les etapes prèvies, mentre que cada treballador t accepta (i és assignat provisionalment a) l'empresa més preferida entre el conjunt format per les ofertes rebudes en aquesta etapa més l'empresa provisionalment assignada a t en l'etapa anterior (en el cas que aquesta existeixi). L'algoritme es para a l'etapa en què o bé totes les ofertes són acceptades o bé les empreses no tenen treballadors acceptables a qui fer una oferta; llavors

[37] Veure Roth i Sotomayor (1990) per una demostració d'aquesta afirmació que Knuth (1976) atribueix a John Conway. Definim dues relacions d'ordre $\succeq_E$ i $\succeq_T$ en el conjunt $S(P)$ de la següent manera. Donats $\mu, \mu' \in S(P)$ diem que $\mu \succeq_E \mu'$ si $\mu(e) R_e \mu'(e)$ per a tot $e \in E$ i diem que $\mu \succeq_T \mu'$ si $\mu(t) R_t \mu'(t)$ per a tot $t \in T$. Llavors, per a tot perfil $P \in \mathcal{P}$, $(S(P), \succeq_E)$ i $(S(P), \succeq_T)$ són dos reticles complets (tot subconjunt de $S(P)$ té màxim i mínim tant segons $\succeq_E$ com segons $\succeq_T$) i a més, un ordre és dual o invers de l'altre, és a dir, per a tot parell $\mu, \mu' \in S(P)$, $\mu \succeq_E \mu'$ si i només si $\mu' \succeq_T \mu$. Això vol dir que en el conjunt d'assignacions estables a P, els agents de cada subconjunt E i T comparteixen parcialment les seves opinions i hi ha una oposició parcial d'opinions entre els agents dels dos subconjunts.

l'assignació bilateral provisional passa a ser definitiva. Gale i Shapley (1962) demostren que aquesta assignació és estable i coincideix amb l'assignació μ_E. L'algoritme d'acceptació diferida on els treballadors fan les ofertes es defineix simètricament i té com a resultat l'assignació μ_T. Aquests algoritmes, o senzilles adaptacions, són extensament utilitzats per resoldre problemes concrets d'assignació bilateral. Per exemple, per l'assignació d'estudiants a places universitàries en molts països (com Espanya, Turquia o Hongria), de nens de 6 anys a escoles públiques en moltes ciutats (com Barcelona, Boston o New York) i de metges interns residents a places hospitalàries (com als Estats Units des de l'any 1956). L'exemple 6 il·lustra l'algoritme d'aceptació diferida.

Exemple 6 Considerem el problema d'assignació (E, T, P) on $E = \{e_1, e_2, e_3, e_4, e_5\}$, $T = \{t_1, t_2, t_3, t_4\}$, i P és

P_{e_1}	P_{e_2}	P_{e_3}	P_{e_4}	P_{e_5}	P_{t_1}	P_{t_2}	P_{t_3}	P_{t_4}
t_1	t_4	t_4	t_1	t_1	e_2	e_3	e_4	e_1
t_2	t_2	t_3	t_4	t_2	e_3	e_1	e_5	e_4
t_3	t_3	t_1	t_3	t_4	e_1	e_2	e_1	e_5
t_4	t_1	t_2	t_2	e_5	e_4	e_4	e_2	e_2
e_1	e_2	e_3	e_4	t_3	e_5	e_5	e_3	e_3
					t_1	t_2	t_3	t_4

- **L'AAD on les empreses fan les ofertes:**

Etapa 1		Etapa 2		Etapa 3		Etapa 4		Final
$e_1 \to t_1$	Si	$e_1 \to t_1$	Si	$e_1 \to t_1$	Si	$e_1 \to t_1$	Si	$\mu_E(e_1) = t_1$
$e_2 \to t_4$	Si	$e_2 \to t_4$	No	$e_2 \to t_2$	Si	$e_2 \to t_2$	Si	$\mu_E(e_2) = t_2$
$e_3 \to t_4$	No	$e_3 \to t_3$	Si	$e_3 \to t_3$	Si	$e_3 \to t_3$	Si	$\mu_E(e_3) = t_3$
$e_4 \to t_1$	No	$e_4 \to t_4$	Si	$e_4 \to t_4$	Si	$e_4 \to t_4$	Si	$\mu_E(e_4) = t_4$
$e_5 \to t_1$	No	$e_5 \to t_2$	Si	$e_5 \to t_2$	No	$e_5 \to t_4$	No	$\mu_E(e_5) = e_5$

- **L'AAD on els treballadors fan les ofertes:**

Etapa 1		Final
$t_1 \to e_2$	Si	$\mu_T(t_1) = e_2$
$t_2 \to e_3$	Si	$\mu_T(t_2) = e_3$
$t_3 \to e_4$	Si	$\mu_T(t_3) = e_4$
$t_4 \to e_1$	Si	$\mu_T(t_4) = e_1$
		$\mu_T(e_5) = e_5$

Observem que $\mu_E \neq \mu_T$ i que $\mu_E(e_5) = \mu_T(e_5) = e_5$.[38] □

L'algoritme d'acceptació diferida on les empreses fan les ofertes genera una funció d'elecció social $f^E : \mathcal{P} \to A$, on $f^E(P) = \mu_E$ per a tot $P \in \mathcal{P}$. Simètricament, l'algoritme d'acceptació diferida on els treballadors fan les ofertes genera una funció d'elecció social $f^T : \mathcal{P} \to A$, on $f^T(P) =$

[38] Sempre es verifica que si $\mu \in S(P)$ i $\mu(i) = i$ llavors, $\mu'(i) = i$ per a tot $\mu' \in S(P)$. No estar assignat a un agent de l'altre subconjunt és una propietat global del conjunt d'assignacions estables.

μ_T per a tot $P \in \mathcal{P}$. És fàcil comprovar que l'algoritme d'acceptació diferida té complexitat polinòmica i per tant, pot ser utilitzat en problemes amb molts agents.

Roth (1982a) demostra un primer resultat d'impossibilitat: no existeix cap funció d'elecció social estable i no manipulable. Noteu que les funcions dictatorials en sèrie[39] no són estables però són no manipulables.

Proposició 6 (Roth, 1982a) *Sigui $f : \mathcal{P} \to A$ una funció d'elecció social estable. Llavors, f és manipulable.*

Demostració Sorprenentment, per a demostrar el resultat d'impossibilitat és suficient mostrar un exemple per al qual no existeix cap funció d'elecció social estable i no manipulable, com mostrarem al final. Considerem un exemple amb dues empreses $E = \{e_1, e_2\}$, dos treballadors $T = \{t_1, t_2\}$ i el següent perfil de preferències P :

P_{e_1}	P_{e_2}	P_{t_1}	P_{t_2}
t_2	t_1	e_1	e_2
t_1	t_2	e_2	e_1
e_1	e_2	t_1	t_2

És fàcil comprovar que $S(P) = \{\mu_E, \mu_T\}$, on $\mu_E(e_1) = t_2$, $\mu_E(e_2) = t_1, \mu_T(e_1) = t_1$ i $\mu_T(e_2) = t_2$. Sigui $f : \mathcal{P} \to A$ qualsevol funció d'elecció social estable. Aleshores, $f(P) \in \{\mu_E, \mu_T\}$.

- Si $f(P) = \mu_T$ llavors e_1 pot manipular f en el perfil P declarant $t_2 P'_{e_1} e_1 P'_{e_1} t_1$ ja que $S(P'_{e_1}, P_{-e_1}) = \{\mu_E\}$ i per tant, per l'estabilitat de f, $f(P'_{e_1}, P_{-e_1}) = \mu_E$ i

$$f_{e_1}(P'_{e_1}, P_{-e_1}) = t_2 P_{e_1} t_1 = f_{e_1}(P).$$

- Si $f(P) = \mu_E$ llavors t_1 pot manipular f en el perfil P declarant $e_1 P'_{t_1} t_1 P'_{t_1} e_2$ ja que $S(P'_{t_1}, P_{-t_1}) = \{\mu_T\}$ i per tant, per l'estabilitat de f, $f(P'_{t_1}, P_{-t_1}) = \mu_T$ i

$$f_{t_1}(P'_{t_1}, P_{-t_1}) = e_1 P_{t_1} e_2 = f_{t_1}(P).$$

Ara bé, noteu que independentment del nombre d'empreses i treballadors, sempre podrem seleccionar un perfil on el perfil P de l'exemple descrigui la part rellevant des del punt de vista de l'estabilitat d'aquest perfil més general, on tots els treballadors diferents de t_1 i t_2 són no acceptables per a les empreses e_1 i e_2, i totes les empreses diferents de e_1 i e_2 són no acceptables per als treballadors t_1 i t_2. A més, la restricció a $\{e_1, e_2\}$ i $\{t_1, t_2\}$ de qualsevol assignació estable del problema general es redueix al conjunt $\{\mu_E, \mu_T\}$. Per tant, quasevol funció d'elecció social és manipulable. ■

Alcalde i Barberà (1994) donen el següent resultat d'impossibilitat que és més fort que el de Roth (1982a) ja que tota funció d'elecció social estable és individualment racional i eficient. La demostració és senzilla i similar a la de Roth (1982a).

[39] Les funcions dictatorials en sèrie per assignar agents a objectes indivisibles han estat definides en la subsecció anterior, i es poden adaptar fàcilment a aquest context d'assignació d'agents a agents.

Proposició 7 (Alcalde i Barberà, 1994) *Sigui $f : \mathcal{P} \to A$ una funció d'elecció social individualment racional i eficient. Llavors, f és manipulable.*

Per a cada agent $i \in N$, sigui $\mathcal{D}_i \subseteq \mathcal{P}_i$ un subconjunt de preferències estrictes en $T \cup \{i\}$ si $i \in E$ o en $E \cup \{i\}$ si $i \in T$. Sigui $n = \#E + \#T$. Diem que una funció d'elecció social $f : \mathcal{D}_1 \times \cdots \times \mathcal{D}_n \to A$ *és no manipulable per* $S \subseteq N$ quan per a tot perfil $P = (P_1, ..., P_n) \in \mathcal{D}_1 \times \cdots \times \mathcal{D}_n$, tot $i \in S$ i tot $P_i' \in \mathcal{D}_i$,

$$f_i(P) R_i f_i(P_i', P_{-i}).$$

No és difícil verificar que els dos algoritmes d'acceptació diferida entesos com a funcions d'elecció social són no manipulables per al subconjunt d'agents que fan les ofertes.

Teorema 7 (Dubins i Freedman, 1981; Roth, 1982b) *La funció d'elecció social $f^E : \mathcal{P} \to A$ és no manipulable per a les empreses i la funció d'elecció social $f^T : \mathcal{P} \to A$ és no manipulable per als treballadors.*

Alcalde i Barberà (1994) també identifiquen (i justifiquen amb molts exemples) una restricció de domini de preferències natural amb la que es pot obtenir un resultat de possibilitat: existeix una única funció d'elecció social estable i no manipulable en aquest domini. El que queda de l'anàlisi ho farem des del punt de vista dels treballadors. Es pot fer l'anàlisi simètrica des del punt de vista de les empreses.

Definició 8 Sigui $t \in T$ un treballador. Un conjunt de preferències $\widetilde{\mathcal{P}}_t \subseteq \mathcal{P}_t$ en $E \cup \{t\}$ satisfà la condició de *dominància en el millor per a t* quan per a tot parell $P_t, P_t' \in \widetilde{\mathcal{P}}_t$ i per a qualsevol parell $x, y \in E \cup \{t\}$ tals que (i) $xR_t t$, (ii) $yR_t' t$, (iii) $xP_t y$, i (iv) $yP_t' x$ llavors, no existeix cap agent $z \in E \cup \{t\}$ per al qual $zP_t x$ i $zP_t' y$.

Un conjunt de perfils de preferències estrictes $\widetilde{\mathcal{P}} = \prod_{i \in E \cup T} \widetilde{\mathcal{P}}_i \subseteq \mathcal{P}$ satisfà la condició de *dominància en el millor* si per a tot $t \in T$, $\widetilde{\mathcal{P}}_t$ satisfà la condició de dominància en el millor per a t, i per a tot $e \in E$, $\widetilde{\mathcal{P}}_e = \mathcal{P}_e$. És important notar que la condició de dominància en el millor només s'imposa en les preferències dels agents d'un dels dos subconjunts del problema d'assignació bilateral (els treballadors), mentre que les preferències dels agents de l'altre subconjunt (les empreses) es deixa sense restriccions.

Proposició 8 (Alcalde i Barberà, 1994) *La funció d'elecció social $f^E : \widetilde{\mathcal{P}} \to \mathcal{M}$ és l'única funció no manipulable en el conjunt de perfils de preferències que satisfan la condició de dominància en el millor.*

Sigui $\widehat{\mathcal{P}} \subset \mathcal{P}$ el subconjunt de perfils de preferències estrictes on els agents prefereixen qualsevol agent de l'altre subconjunt abans que quedar-se sense assignar; és a dir, $P \in \widehat{\mathcal{P}}$ si i només si per a tot $e \in E$ i tot $t \in T$, $tP_e e$ i $eP_t t$.

Proposició 9 (Alcalde i Barberà, 1994) *Hi ha funcions d'elecció social eficients, individualment racionals i no manipulables en el domini $\widehat{\mathcal{P}}$.*

La proposició és certa ja que totes les funcions d'elecció social dictatorials en sèrie són eficients, individualment racionals i no manipulables en $\widehat{\mathcal{P}}$. En aquest domini de preferències la condició d'individualitat racional no té cap poder restrictiu. És fàcil comprovar que qualsevol funció d'elecció social dictatorial en sèrie és eficient i no manipulable. Noteu però que no és estable.

5 Referències

1. J. Alcalde i S. Barberà. "Top dominance and the possibility of strategy-proof stable solutions to matching problems", *Economic Theory* 4, 417-435 (1994).

2. J. Alcalde-Unzu i E. Molis. "Exchange of indivisible goods and indifferences: the top trading absorbing sets mechanisms", *Games and Economic Behavior* 73, 1-16 (2011).

3. K. Arrow. *Social Choice and Individual Values*, Cowles Comission Monograph No.: 12, New York: J. Weley & Sons (1951). Segona edició (1963).

4. S. Barberà. "Pivotal voters: a new proof of Arrow's theorem", *Economics Letters* 6, 13-16 (1980).

5. S. Barberà. "Strategy-proofness and pivotal voters: a direct proof of the Gibbard-Satterthwaite theorem", *International Economic Review* 24, 413-418 (1983a).

6. S. Barberà. "Pivotal voters: a simple proof of Arrow's theorem", en *Social Choice and Welfare*, editat per P. Pattanaik i M. Salles. North Holland (1983b).

7. S. Barberà, F. Gul i E. Stachetti. "Generalized median voter schemes and committees", *Journal of Economic Theory* 61, 262-289 (1993).

8. S. Barberà i M. Jackson. "A characterization of strategy-proof social choice functions for economies with pure public goods", *Social Choice and Welfare* 11, 241-252 (1994).

9. S. Barberà, M. Jackson i A. Neme. "Strategy-proof allotment rules", *Games and Economic Behavior* 18, 1-21 (1997).

10. S. Barberà, J. Massó i A. Neme. "Voting under constraints", *Journal of Economic Theory* 76, 298-321 (1997).

11. S. Barberà, J. Massó i A. Neme. "Voting by committees under constraints", *Journal of Economic Theory* 122, 185-205 (2005).

12. S. Barberà, J. Massó i S. Serizawa. "Strategy-proof voting on compact ranges", *Games and Economic Behavior* 25, 272–291 (1998).

13. S. Barberà i B. Peleg. "Strategy-proof voting schemes with continuous preferences", *Social Choice and Welfare* 7, 31-38 (1990).

14. S. Barberà, H. Sonnenschein i L. Zhou. "Voting by committees", *Econometrica* 59, 595–609 (1991).

15. P. Batteau, J. Blin i B. Montjardet. "Stability of aggregation procedures, ultrafilters and simple games", *Econometrica* 40, 527-534 (1981).

16. J. P. Benoit. "The Gibbard-Satterthwaite theorem: a simple proof", *Economics Letters* 69, 319-322 (2000).

17. D. Black. "On the rationale of group decision making", *Journal of Political Economy* 56, 23-34 (1948).

18. D. Blair i E. Muller. "Essential aggregation procedures and restricted domain of preferences", *Journal of Economic Theory* 30, 34-53 (1983).

19. D. Blair i R. Pollak. "Acyclic collective choice rules", *Econometrica* 50, 931–943 (1982).

20. J. Blau i R. Deb. "Social decision functions and the veto", *Econometrica* 45, 871-879 (1977).

21. K. Border i J. Jordan. "Straightforward elections, unanimity and phantom agents", *Review of Economic Studies* 50, 153-170 (1983).

22. E. Clark. "Multipart pricing of public goods", *Public Choice* 11, 17-33 (1971).

23. R. Deb. "A monotone social decision functions and the veto", *Econometrica* 49, 899–910 (1981).

24. S. Dobzinski, A. Mehta, T. Roughgarden i M. Sundararajan. "Is Shapley cost sharing optimal?", en *SAGT '08 Proceedings of the 1st International Symposium on Algorithmic Game Theory* (2008).

25. L. Dubins i D. Freedman. "Machiavelli and the Gale-Shapley algorithm", *American Mathematical Monthly* 88, 485-494 (1981).

26. K. Eliaz. "Social aggregators", *Social Choice and Welfare* 22, 317-330 (2004).

27. D. Gale i L. Shapley. "College admissions and the stability of marriage", *American Mathematical Monthly* 69, 9-15 (1962).

28. J. Geanakoplos. "Three brief proofs of Arrow's impossibility theorem", *Economic Theory* 26, 211-215 (2005).

29. A. Gibbard. "Manipulation of voting schemes: a general result", *Econometrica* 41, 587–601 (1973).

30. T. Groves. "Incentives in teams", *Econometrica* 41, 617-631 (1973).

31. M. Jackson. "A crash course in implementation theory", *Social Choice and Welfare* 18, 655-708 (2001).

32. P. Jaramillo i V. Manjunath. "The difference indifference makes in strategy-proof allocation of objects", de pròxima aparició en *Journal of Economic Theory* (2012).

33. D. Knuth. *Marriages Stables*, Les Presses de l'Université de Montréal, Montréal, Canada (1976). Versió anglesa: *Stable marriages and its relation to other combinatorial problems*, CRM Proceedings and Lecture Notes, number 10, American Mathematical Society, Providence (Rhode island), USA (1991).

34. M. Le Breton i A. Sen. "Separable preferences, strategy-proofness and decomposability", *Econometrica* 67, 605-628 (1999).

35. J. Ma. "Strategy-proofness and the strict core in a market with indivisibilities", *International Journal of Game Theory* 23, 75-83 (1994).

36. P. Man i S. Takayama. "A unifying impossibility theorem", sense publicar (2011).

37. A. Mas-Colell i H. Sonnenschein. "General possibility theorems for group decision", *Review of Economic Studies* 39, 185-192 (1972).

38. J. Massó. "El intercambio de riñones y la matemática discreta", *Paseo por la Geometría*, curso 2009-2010. Facultad de Ciencias de la Universidad del País Vasco, 89-114 (2010).

39. J. Massó i I. Moreno de Barreda. "On Strategy-proofness and symmetric single-peakedness", *Games and Economic Behavior* 72, 467-484 (2011).

40. J. Massó, A. Nicolò i A. Sen. "Strategy-proofness and equal-cost sharing for a binary excludable public good with fixed cost", sense publicar (2010).

41. K. May. "A set of independent necessary and sufficient conditions for simple majority decisions", *Econometrica* 20, 680–684 (1952).

42. I. McLean i A. Urken. *Classics of Social Choice*, The University of Michigan Press (1995).

43. A. Mehta, T. Roughgarden i M. Sundararajan. "Beyond Moulin mechanisms", en *Proceedings of the 8th ACM Conference on Electronic Commerce (EC)*, 1-10 (2007).

44. H. Moulin. "On strategy-proofness and single peakedness", *Public Choice* 35, 437-455 (1980).

45. R. Myerson. "Optimal auction design", *Mathematics of Operations Research* 6, 58-73 (1981).

46. K. Nehring i C. Puppe. "The structure of strategy-proof social choice— Part I: General characterization and possibility results on median spaces", *Journal of Economic Theory* 135, 269-305 (2007a).

47. K. Nehring i C. Puppe. "Efficient and strategy-proof voting rules: a characterization", *Games and Economic Behavior* 59, 132-153 (2007b).

48. N. Nisan. "Introduction to mechanism design (for computer scientists)", en *Algorithmic Game Theory*, editat per N. Nisan, T. Roughgarden, É. Tardos i V. Vazirani. Cambridge University Press, New York (2008).

49. P. Pattanaik i B. Peleg. "Distribution of power under stochastic social choice rules", *Econometrica* 54, 909–921 (1986).

50. H. Peters, H. van der Stel i T. Storcken. "Pareto optimality, anonymity, and strategy-proofness in location problems", *International Journal of Game Theory* 21, 221-235 (1992).

51. H. Peters, H. van der Stel i T. Storcken. "Generalized median solutions, strategy-proofness and strictly convex norms", *ZOR—Methods Models of Operations Research* 38, 19-53 (1993).

52. P. Reny. "Arrow's theorem and the Gibbard-Satterthwaite theorem: a unified approach", *Economics Letters* 70, 99-105 (2001).

53. A. Roth. "Incentive compatibility in a market with indivisible goods", *Economics Letters* 9, 127-132 (1982a).

54. A. Roth. "The economics of matching: stability and incentives", *Mathematics of Operations Research* 7, 617-628 (1982b).

55. A. Roth i A. Postlewaite. "Weak versus strong domination in a market with indivisible goods", *Journal of Mathematical Economics* 4, 131-137 (1977).

56. A. Roth, T. Sönmez i U. Ünver. "Kidney exchange", *Quarterly Journal of Economics* 119, 457-488 (2004).

57. A. Roth i M. Sotomayor. *Two-sided Matching: A Study in Game-theoretic Modelling and Analysis*, Cambridge University Press i Econometric Society Monographs No. 18 (1990).

58. M. Satterthwaite. "Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions", *Journal of Economic Theory* 10, 187–217 (1975).

59. D. Schmeidler i H. Sonnenschein. "Two proofs of the Gibbard-Satterthwaite theorem on the possibility of strategy-proof social choice functions", en *Decision Theory and Social Ethics, Issues in Social Choice*, editat per H. Gottinger i W. Leinfellner, 227-234. D. Reidel Publishing Company, Dordrecht, Holanda (1978).

60. A. Sen. "Quasi-transitivity, rational choice and collective decisions", *Review of Economic Studies* 36, 381-393 (1969).

61. A. Sen. "Another direct proof of the Gibbard-Satterthwaite theorem", *Economics Letters* 70, 381-385 (2001).

62. L. Shapley i H. Scarf. "On Cores and indivisibilities", *Journal of Mathematical Economics* 1, 23-28 (1974).

63. Y. Sprumont. "The division problem with single-peaked preferences: a characterization of the uniform allocation rule", *Econometrica* 59, 509-519 (1991).

64. L. Ubeda. "Neutrality in Arrow and in other impossibility theorems", *Economic Theory* 23, 195-204 (2003).

65. W. Vickrey. "Counterspeculation, auctions and competitive sealed tenders", *Journal of Finance* 16, 8-37 (1961).

66. R. Wilson. "Social choice theory without the Pareto principle", *Journal of Economic Theory* 5, 478 486 (1972).

67. L. Zhou. "Impossibility of strategy-proof mechanisms for economies with pure public goods", *Review of Economic Studies* 58, 107-119 (1991).