

# 1 Introduction

- Nous avons distingué trois principaux domaines de la philosophie de l'économie :

(1) la méthodologie de l'économie

(2) les fondements de l'économie et la théorie du choix rationnel

(3) l'économie normative et, de manière plus générale, les relations entre économie et éthique sociale

Exemple : Quel barème adopter pour l'impôt sur le revenu ?

Pour répondre à cette question, on a besoin d'une théorie "positive" qui permette de prédire correctement les conséquences attendues (différents états socio-économiques) des différents barèmes envisageables. On a aussi besoin de *critères* pour *évaluer* les différents états pouvant résulter de différents barèmes. L'économie normative est l'ensemble des contributions qui ont pour vocation de construire, d'explorer et d'appliquer de tels critères d'évaluation des états socio-économiques.

- Cette portera sur la **théorie du choix social**. La théorie du choix social n'épuise certainement pas toute l'économie normative ni toute l'interface économie/éthique sociale. Je l'ai choisie parce que

1. c'est une des principales théories de l'économie normative ; son formalisme et ses principaux résultats sont absolument indispensables pour aller plus loin en économie normative
2. elle "part" de la théorie de la décision individuelle, que nous connaissons bien désormais
3. le principe de Pareto joue un rôle essentiel dans ses résultats séminaux
4. elle illustre parfaitement le rôle de la méthode axiomatique dans l'économie normative contemporaine (voir Fleurbaey (1996) pour une excellente discussion introductive à cette méthode)

- Nous procéderons de la manière suivante : nous allons introduire le cadre formel de la TCS et discuter certains de ses résultats fondamentaux, notamment le théorème d'Arrow.

## 2 La théorie du choix social

- La théorie du choix sociale est la théorie qui "traite" des **décisions et des évaluations collectives**.

▷ A. Sen (1986)

“Social choice problems arise in aggregating the interests, or preferences, or judgments, or views, of different persons (or groups) in a particular society.”

• Le cadre formel, introduit par Arrow (1951/1963) est le suivant :

### Définition 1

(i) On considère un ensemble  $N = \{1, \dots, n\}$  d'individus, chacun ayant une relation de préférence large  $R_i$  sur l'ensemble des alternatives  $X$ .

(ii) On note  $R_C$  (resp.  $P_C$ ) la **préférence collective ou sociale** large (resp. stricte) sur  $X$ .

(iii)  $X \times X$  est l'ensemble des relations binaires possibles sur  $X$ .  $\mathfrak{R}$  est l'ensemble des préférences rationnelles sur  $X$ , et  $\mathfrak{R}'$  un sous-ensemble non-vide de  $\mathfrak{R}$  que l'on peut concevoir comme des préférences rationnelles obéissant à certaines restrictions.

(iv) Une **fonction de préférence sociale**  $f : \mathfrak{R}'^n \rightarrow \mathfrak{R}$  associe à tout  $n$ -uplet de préférences individuelles  $(R_1, \dots, R_n)$  une préférence collective  $R_C = f(R_1, \dots, R_n)$

• Commentaires :

1.  $X$ , l'ensemble des options, est souvent conçu comme l'ensemble des états sociaux possibles :

▷ Arrow (1951/1963)

“In the present study the objects of choice are social states. The most precise definition of a social state would be a complete description of the amount of each type of commodity in the hands of each individual, the amount of labor to be supplied by each individual, the amount of each productive resource invested in each type of productive activity, and the amounts of various types of collective activity..”

Remarque : Dans une interprétation décisionnelle, on suppose en général qu'il est dans le pouvoir de la collectivité de faire advenir l'un ou l'autre de ces états sociaux possibles. Dans une interprétation évaluative en revanche, cette restriction ne s'impose plus.

2. quel est le rôle de la “fonction de préférence sociale” ? Hausman & McPherson (2006) distinguent deux interprétations ou fonctions : une interprétation décisionnelle et une interprétation évaluative (voir aussi Sen (1983)). .
  - décision : une fonction de préférence sociale guide la prise de décision collective  
 $xP_Cy$  signifie que  $y$  ne doit pas être choisi par la société quand  $x$  est disponible (Sen 1983)
  - évaluation : une fonction de préférence sociale évalue les états sociaux  
 $xP_Cy$  signifie que  $x$  est une meilleure configuration sociale que  $y$
3. les relations de préférences  $R_i$  sont souvent conçues comme les relations de préférence dont traite la théorie individuelle de la décision. Bien sûr, les relations binaires peuvent en principe faire l'objet d'autres interprétations.
4. si l'on considère l'interprétation des préférences comme expression comparative des **valeurs individuelles**, alors deux questions importantes du point de vue philosophique sont soulevées :

(Q1) quelle est la conception des valeurs individuelles sous-jacente ?

a) ordinalisme : seules les différences de position dans le classement des options sont pertinentes

b) pas de comparaison interpersonnelle d'utilité (CIU)

Qu'est-ce que cela peut signifier ou impliquer que  $u_i(x) \geq u_j(y)$  - que l'utilité de  $x$  pour  $i$  est supérieure à celle de  $y$  pour  $j$  ? Dans le cas individuel, l'utilité est censée représenter les préférences donc on peut au moins dire que  $u_i(x) \geq u_i(y)$  implique que  $i$  préfère strictement  $x$  à  $y$  - ce qui implique typiquement à son tour que  $y$  n'est pas choisi par  $i$  quand  $x$  est disponible.

▷ Arrow (1951/1963)

“The viewpoint will be taken here that interpersonal comparison of utilities has no meaning and, in fact, that there is no meaning relevant to welfare comparisons in the measurability of individual utility.”

“If we look away from the mathematical aspects of the matter, it seems to make no sense to add the utility of one individual, a psychic magnitude in his mind, with the utility of another individual.”

▷ Robbins (1938)

“Every mind is inscrutable to every other mind and no common denominator of feelings is possible.”

(i) *old welfare economics* : la OWE s'inspire l'idée de la conception **utilitariste** de maximisation de la somme des utilités individuelles, que l'on peut exprimer formellement de la manière suivante :

$$xR_Cy \Leftrightarrow \sum_i u_i(x) \geq \sum_i u_i(y)$$

Cette conception suppose que l'on puisse comparer les utilités individuelles.

(ii) *new welfare economics* : la NWE se base essentiellement sur le principe de Pareto. Bergson (1939), Samuelson (1947)

On parle parfois de *base informationnelle* pour désigner le type de données sur les individus dont on part. La TCS se fonde initialement sur une base informationnelle étroite ou pauvre (seulement les préférences).

(Q2) quelle est la conception des valeurs collectives ou sociales sous-jacente ?

On appelle généralement **welfarisme** l'idée selon laquelle les seules informations pertinentes pour l'évaluation/la décision collective sont les informations relatives aux préférences individuelles. Il y a deux hypothèses que l'on peut distinguer, qui correspondent à deux inspirations de la TCS (l'économie du bien-être d'une part, la théorie du vote d'autre part) :

1/ les préférences de  $i$  reflètent le bien-être de  $i$  voire le bien de  $i$ , et une fonction de préférence sociale évalue un état social possible en fonction du sort qu'il réserve aux différents individus.

L'assimilation du bien-être individuelle à la satisfaction des préférences individuelles a par exemple été critiqué par A. Sen :

▷ Sen (1985, cité par Mongin & d'Aspremont 1998)

“...a person who is illfed, undernourished, unsheltered and ill can still be high up in the scale of happiness or desire-fulfilment if he or she has learned to have ‘realistic’ desires and to take pleasure in small mercies. The physical conditions of a person do not enter the view of well-being seen entirely in terms of happiness or desire-fulfilment, except insofar as they are indirectly covered by the mental attitudes of happiness or desire.”

Les préférences ne coïncident que par accident avec le bien-être individuel. Le bien-être individuel inclut des choses objectives qui peuvent échapper aux préférences individuelles. Et par ailleurs le bien individuel inclut d'autres choses que le bien-être (ex : la liberté). MA98 défendent que des préférences *rationnelles et bien informées* coïncident essentiellement avec le bien-être individuel.

2/ les préférences individuelles de  $i$  reflètent l'évaluation par l'individu  $i$  des états sociaux possibles et la fonction de préférence sociale synthétise ou agrège les différentes évaluations en une évaluation globale. Le paradigme est ici plutôt le vote, et l'hypothèse conceptuelle implicite est, si l'on suit Arrow, une hypothèse anti-Platoniste :

▷ Arrow (1951/1963)

“The relative ranking of a fixed pair of alternative social states will vary, in general, with changes in the values of at least some individuals; to assume that the ranking does not change with any changes in individual values is to assume, with traditional social philosophy of the Platonic realist variety, that there exists an objective social good defined independently of individual desires.”

La discussion qui précède repose sur deux interprétations des préférences individuelles : les préférences comme bien-être individuel et les préférences comme jugements individuels. Si l'on combine ces deux interprétations avec la distinction décision sociale/évaluation sociale, on obtient 4 interprétations possibles ( analogue à Hausman & McPherson 2006).

Il faut avoir en tête ces distinctions tant pour interpréter les résultats de la TCS et les conceptions philosophiques que l'on peut chercher à exprimer dans la TCS.

### 3 Le théorème d'impossibilité d'Arrow

- L'acte de naissance de la TCS contemporaine est le livre d'Arrow qui contient le célèbre “théorème d'impossibilité”. Nous allons présenter désormais ce théorème. On peut concevoir le théorème comme affirmant qu'un certain ensemble restreint de propriétés normatives portant sur la fonction de préférence sociale et intuitivement très attractives ne sont pas logiquement compatibles. Nous allons détailler chacune de ces conditions.

#### Définition 2

**Condition U** (universalité) : une fonction de préférence sociale satisfait la Condition U ssi elle peut avoir pour argument  $n$  importe quel  $n$ -uplet de préférences rationnelles :  $\mathfrak{R}^n = \mathfrak{R}^n$ .

**Condition R** (*rationalité*) : une fonction de préférence sociale satisfait la Condition R ssi les préférences collectives  $\succeq_C$  sont rationnelles (réflexives, transitives et complètes)

**Condition P** (*Pareto faible*) : une fonction de préférence sociale satisfait la Condition P ssi pour toutes options  $x, y \in X$ , si pour tout  $i \in N$ ,  $xP_iy$ , alors  $xP_Cy$

Commentaire sur la Condition P : la Condition P dit que s'il y a unanimité des préférences individuelles, alors la préférence sociale doit s'aligner sur les préférences individuelles.

Dans l'interprétation où les préférences individuelles sont les jugements des individus et où la fonction de préférence sociale évalue les états sociaux possibles, la Condition P n'est pas indiscutable : il se peut que tout le monde se trompe ! Par contraste, si les préférences individuelles reflètent le bien être des individus, la Condition P a plus de force : si le bien être de tous les individus est supérieur dans l'état social  $x$  que dans l'état social  $y$ , alors l'état  $x$  est socialement préférable.

### Définition 3

**Condition I** (*Indépendance des options non-pertinentes*) : une fonction de préférence sociale  $f$  satisfait la Condition I ssi pour toute paire d'options  $x, y \in X$ , la préférence collective  $\succeq_C$  sur  $\{a, b\}$  ne dépend que des préférences individuelles sur  $\{x, y\}$ . Formellement : soient deux profils de préférences  $(R_1, \dots, R_n)$  et  $(R'_1, \dots, R'_n)$  tels que pour tout individu  $i$

$$xR_iy \text{ ssi } xR'_iy$$

alors  $xf(R_1, \dots, R_n)y \text{ ssi } xf(R'_1, \dots, R'_n)y$

Commentaire :

1/ la règle majoritaire satisfait manifestement la Condition I.

### Définition 4

La **règle majoritaire** est la fonction de préférence sociale telle que

$$xR_Cy \text{ ssi } N(xP_iy) \geq N(yP_ix)$$

où  $N(xP_iy)$  désigne le nombre d'individus  $i$  tels que  $xP_iy$ .

2/ la *règle de Borda* consiste à attribuer un score à chaque option  $x \in X$  : ce score est la somme de points obtenus pour chaque agent  $i \in N$  : l'option préférée par  $i$  reçoit le maximum de points, la seconde meilleure option un peu moins, la troisième un peu moins encore, etc.

**Définition 5**

Le **score** de l'option  $x$  relativement à  $R_i$  est le nombre d'options  $y$  telles que  $i$  préfère  $x$  à  $y$ . Le **score de Borda** d'une option  $x$  est la somme des scores individuels soit  $b(x, \bar{R}) = \sum_i b(x, R_i)$ .

La **règle de Borda** est la suivante :

$$xR_Cy \text{ ssi } b(x, \bar{R}) \geq b(y, \bar{R})$$

**Fait 1**

La règle de Borda ne satisfait pas la Condition I.

Le contre-exemple suivant suffit à démontrer le Fait.

**Exemple 1 (Salles 2006)**

$wP_1xP_1yP_1z$	$wP'_1yP'_1zP'_1x$
$xP_2wP_2zP_2y$	$xP'_2wP'_2zP'_2y$
$yP_3xP_3wP_3z$	$yP'_3xP'_3wP'_3z$

Calculons les scores de Borda pour  $w$  et  $x$  :

$b(w) = 3 + 2 + 1 = 6$	$b'(w) = 6$
$b(x) = 2 + 3 + 2 = 7$	$b'(x) = 0 + 3 + 2 = 5$

La Condition I est violée puisque  $wP_Cx$  alors que  $xP'_Cw$ .

- La Condition I est cruciale pour le Théorème d'Arrow, et elle a été beaucoup discutée. Passons sur les justifications pratiques (cela permet facilement de déterminer les préférences sociales). Du point de vue normatif, l'argument principal est essentiellement un argument d'évidence : les préférences (sociales) entre  $x$  et  $y$  ne doivent pas dépendre des préférences individuelles qui mettent en jeu d'autres options que  $x$  et  $y$ . L'argument est parfois renforcé en faisant valoir que si la Condition I n'est pas satisfaite, alors *le choix des autres options* peut avoir un effet sur les préférences entre  $x$  et  $y$ . On pourrait donc manipuler la préférence sociale en choisissant de manière appropriée l'agenda i.e. l'ensemble des options  $X$ .

**Exemple 2**

$wP_1x$	$wP'_1x$
$xP_2w$	$xP'_2w$
$xP_3w$	$xP'_3w$

Calculons les scores de Borda pour  $w$  et  $x$  :

$b(w) = 1 + 0 + 0 = 1$	$b'(w) = 1$
$b(x) = 0 + 1 + 1 = 2$	$b'(x) = 2$

Avec  $X = \{w, x, y, z\}$ , on avait  $wP_Cx$ . En passant à l'agenda  $X = \{w, x\}$ , avec la même méthode de choix, on reverse les préférences puisque désormais  $xP_Cw$ .

Un autre problème est lié au "vote stratégique", c'est-à-dire le fait que les individus peuvent avoir intérêt à mentir sur les préférences pour faire que l'option qu'ils souhaitent faire gagner gagne effectivement. Regardons le premier des deux exemples, et le profil  $(P'_1, P'_2, P'_3)$ . Le résultat était  $wP'_Cx$ . Cela n'arrange pas l'individu 2 qui préfère  $x$  à  $w$ . Celui-ci classe  $w$  juste après  $x$ , ce qui fait que  $w$  remporte 2 points. S'il plaçait  $w$  en dernier,  $w$  aurait 0 point et au total  $b'(w) = 4$ . Donc  $x$  serait collectivement préféré à  $w$ .

- La Condition I a cependant été critiquée, en voici un exemple :
- ▷ Hausman & McPherson (2006)

"Condition I may be attractive to welfare economists because they so often take welfare to be the satisfaction of ordinal preferences, but it is hard to believe that nothing should be relevant to the social evaluation of  $A$  versus  $B$  except individual rankings of  $A$  and  $B$ . In actual political systems the choice between  $A$  and  $B$  typically depends on many other factors, such as precedent, constitutional provision, the extent to which  $A$  and  $B$  respect rights, and the intensity of preferences. All of these factors are ruled out by I...Like many other social scientists, we believe that I is unacceptable."

### Définition 6

**condition D** (absence de dictature) : il n'existe pas d'individu  $i$  tel que pour tout  $a, b \in X$ , si  $aP_ib$  alors  $aP_Cb$

Commentaire sur la Condition D : un dictateur est un individu qui impose ses préférences (strictes) au groupe dans son ensemble.

### Théorème 1 (Théorème d'impossibilité d'Arrow (1951))

Si  $|N| \geq 2$  et  $|X| \geq 3$ , il n'existe pas de fonction de préférence sociale qui satisfasse simultanément les Conditions R, U, P, I et D.

### Exemple 3 (Le paradoxe de Condorcet)

La règle majoritaire satisfait les Conditions I, U, et P. Il faut donc conclure du Théorème d'Impossibilité qu'elle ne satisfait pas la Condition R. C'est Condorcet qui le premier a mis ce phénomène en évidence.

$$\begin{array}{c} xP_1yP_1z \\ yP_2zP_2x \\ zP_3xP_3y \end{array}$$

Si l'on suit la règle majoritaire, on obtient un cycle :  $xP_CyP_CzP_Cx$ .

*Preuve.* La preuve la plus répandue du Théorème repose sur les deux définitions suivantes :

### Définition 7

Un ensemble d'individus  $V \subseteq N$  est **presque décisif** pour  $x$  contre  $y$  si, si  $xP_iy \forall i \in V$  (ce que l'on note  $xP^V y$ ) et  $yP_jx \forall j \in N - V$  (ce que l'on note  $yP^{V^c} x$ ), alors  $xP_Cy$ . On le note  $D_V(x, y)$ .

Un ensemble d'individus  $V \subseteq N$  est **décisif** pour  $x$  contre  $y$  si, si  $xP_i \forall i \in V$ , alors  $xP_Cy$ . On le note  $D * _V(x, y)$ .

Remarquons que  $D * _V(x, y)$  implique  $D_V(x, y)$ . Dans tout ce qui suit, on suppose que  $f$  est une fonction de préférence sociale satisfaisant les Conditions P, U et I.

### Lemme 1 (Lemme de contagion)

Si  $V$  est presque décisif pour  $x$  contre  $y$ , alors pour tous  $a, b$ ,  $V$  est décisif pour  $a$  contre  $b$ .

*Preuve.*

Supposons que  $D_V(x, y)$  et soit  $a, b \in X$

Cas 1 :  $a = x$  et  $b \neq x, y$ . Supposons que

$$xP_iyP_ib \forall i \in V$$

$$yP_jx \text{ et } yP_jb \forall j \in N - V$$

En vertu de la Condition P,  $yP_Cb$  et puisque  $D_V(x, y)$ ,  $xP_Cy$  donc puisque  $R_C$  est rationnelle,  $xP_Cb$ . En vertu de la Condition I, seules les préférences des individus sur  $x$  et  $b$  déterminent  $xP_Cb$  donc on a  $xP_Cb$  quelles que soient les préférences des membres de  $N - V$  sur  $x$  et  $b$ . Donc  $D * _V(a, b)$ .

Cas 2 :  $b = y$  et  $a \neq x, y$ . Supposons que

$$aP_ixP_iy \forall i \in V$$

$$yP_jx \text{ et } aP_jx \forall j \in N - V$$

En vertu de la Condition P,  $aP_Cx$  et puisque  $D_V(x, y)$ ,  $xP_Cy$  donc  $aP_Cy$ . En vertu de la Condition I,  $D * _V(a, b)$ .

Cas 3 :  $(a, b) = (x, y)$ . Soit  $z \neq x, y$ .  $D_V(x, y)$  implique  $D_V^*(x, z)$  (Cas 1) donc  $D_V(x, z)$  donc  $D^*_{*V}(x, y)$  (derechef par le Cas 1).

Cas 4 :  $(a, b) = (y, x)$ . Soit  $z \neq x, y$ .  $D_V(x, y)$  implique  $D_V^*(x, z)$  (Cas 1) donc  $D_V(x, z)$  donc  $D^*_{*V}(y, z)$  (Cas 2) donc  $D_V(y, z)$  donc  $D^*_{*V}(y, x)$  (Cas 1).

Cas 5 :  $b = x$  et  $a \neq x, y$ .  $D_V(x, y)$  implique  $D^*_{*V}(x, a)$  (Cas 1) donc  $D_V(b, a)$  donc  $D^*_{*V}(a, b)$  (Cas 4).

Cas 6 :  $a = y$  et  $b \neq x, y$ .  $D_V(x, y)$  implique  $D^*_{*V}(b, y)$  (Cas 2) donc  $D_V(b, a)$  donc  $D^*_{*V}(a, b)$  (Cas 4).

Cas 7 :  $a, b \neq x, y$ .  $D_V(x, y)$  implique  $D^*_{*V}(x, b)$  (Cas 1) donc  $D_V(x, b)$  donc  $D^*_{*V}(a, b)$  (Cas 2). ♠.

### Lemme 2 (Lemme de contraction)

Si  $|V| \geq 2$  et si  $V$  est décisif sur toute paire de  $X$ , alors il existe un sous-ensemble strict de  $V$  qui est décisif pour toute paire de  $X$ .

*Preuve.* Soient  $V_1$  et  $V_2$  deux sous-ensembles stricts disjoints et exhaustifs de  $V$ .

$xP_i y P_i z \forall i \in V_1$

$yP_i z P_i x \forall i \in V_2$

$zP_i x P_i y \forall i \in N - V$

Puisque  $V$  est décisif,  $yP_C z$ . Deux cas possibles : (a)  $xP_C z$  et alors  $D_{V_1}(x, z)$ ; (b)  $zR_C x$  et alors par transitivité  $yP_C x$  donc  $D_{V_2}(x, y)$ . Dans les deux cas le Lemme d'expansion garantit que l'un de  $V_1$  ou  $V_2$  est décisif sur toute paire de  $X$ . ♠.

### Corollaire 1

Il existe un individu  $d \in N$  qui est décisif sur toute paire de  $X$  i.e. un dictateur.

*Preuve.* En vertu de la Condition P, l'ensemble  $N$  lui-même est décisif sur toute paire de  $X$ . En appliquant itérativement le Lemme précédent, on obtient un singleton qui contient un dictateur.

• La leçon immédiate du Théorème est que toute fonction de préférence sociale présente au moins l'une des caractéristiques suivantes :

- son domaine n'est pas universel (violation de la Condition U), ou
  - elle engendre une préférence collective non rationnelle (violation de la Condition R),
- ou
- elle engendre une préférence qui ne s'aligne pas toujours sur les préférences individuelles quand celles-ci sont identiques (violation de la Condition P), ou
  - elle engendre une préférence entre  $x$  et  $y$  qui ne dépend pas que des préférences individuelles entre  $x$  et  $y$  (violation de la Condition I), ou

- elle engendre une préférence stricte qui s'aligne systématiquement sur les préférences strictes de l'un des agents (violation de la Condition D).

## 4 La “possibilité” du choix social : stratégies de type I

- Le Théorème d'Arrow est en général considéré comme délivrant un message pessimiste. Une bonne partie de l'immense littérature de la TCS consiste à tenter de contourner le Théorème d'Arrow. Les stratégies de type I sont les stratégies qui consistent à affaiblir certaines des Conditions. Nous aborderons les stratégies de type II au prochain cours : il s'agit, en gros, de remettre en question non plus telle ou telle condition mais le *cadre* de la TCS et en particulier l'hypothèse welfariste.

### 4.1 La condition U

- La condition U exige que le domaine d'une fonction de préférence sociale  $f$  soit l'ensemble des  $n$ -uplets de préférences rationnelles. Une fonction de préférence sociale satisfaisant la Condition U doit donc être capable de traiter une grande variété de profil de préférences individuelles. Une restriction peu intéressante est celle où les profils sont unanimes.

- La restriction la plus célèbre de la Condition U consiste à restreindre le domaine aux **préférences unimodales**. L'idée intuitive est la suivante : on suppose qu'il existe un ordre sur les options  $S$  qui est commun à tous les agents. Attention : cet ordre ne reflète pas nécessairement les préférences des différents agents. Mais il est tel qu'il existe un seul “pic”, c'est-à-dire une seule option  $x$  qui est préférée à son prédécesseur (au sens de  $S$ ) et à son successeur (toujours au sens de  $S$ ).

#### Exemple 4

(a) On peut supposer que les partis politiques sont ordonnés selon un axe. Chaque citoyen a un parti préféré et ses préférences décroissent ensuite au fur et à mesure qu'on s'éloigne du point de l'axe qui correspond à son parti, que l'on s'éloigne vers la droite ou vers la gauche.

(b) exemple analogue avec les voitures chez Gaertner (2006) :

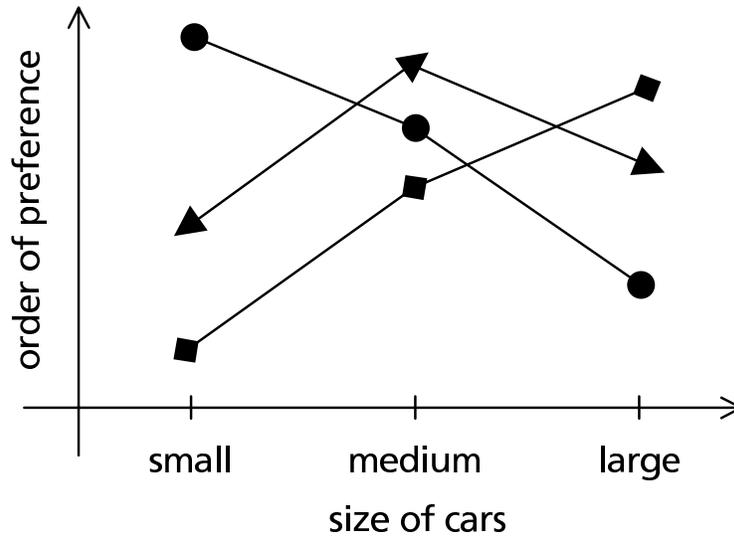


FIG. 1 -

(c) Si l'on considère le cas problématique pour le vote majoritaire décrit ci-dessus, on se rend compte que les préférences ne sont pas unimodales : quel que soit l'ordre  $S$  que l'on choisisse, on aura toujours l'une des relations  $P_i$  qui aura plus d'un pic. Voici par exemple l'ordre  $S = x > y > z$ .

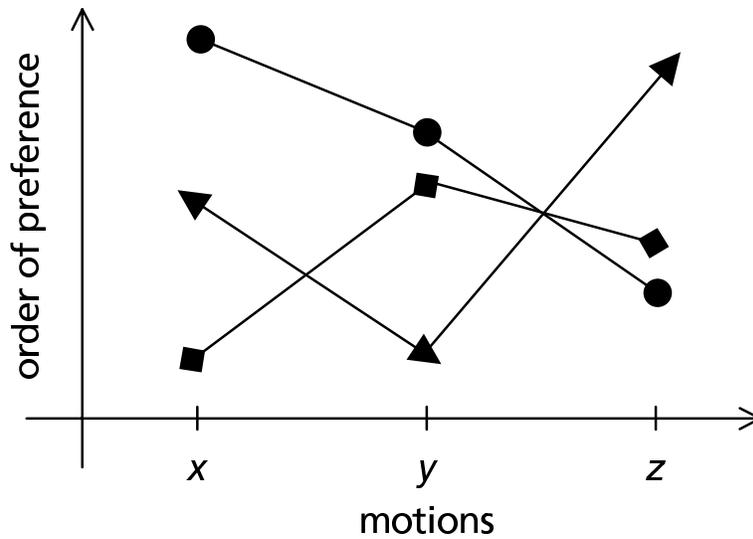


FIG. 2 -

**Définition 8**

Un profil de préférences  $(R_1, \dots, R_n)$  est **unimodal** (single-peak) ssi il existe un ordre  $S$  tel que pour tout  $i$ ,  $x$  et  $y$ ,  $xR_iy$  et  $B(x, y, z)$  implique  $yP_iz$  où  $B(., ., .)$  est la relation d'intermédiarité induite par  $S$ .

**Théorème 2 (Théorème de possibilité pour les préférences unimodales)**

Si  $|N|$  est impair et si le domaine ne contient que des profils de préférences unimodaux, alors la règle majoritaire induit des préférences collectives rationnelles.

Remarque 1 : on sait par ailleurs que la règle majoritaire satisfait les Conditions I, P et D.

Remarque 2 : on appelle **vainqueur de Condorcet** une option  $x$  telle que pour tout  $y \in X$ ,  $|i : xP_iy| \geq |j : xP_jy|$ . Le Paradoxe de Condorcet montre qu'il n'existe pas nécessairement de vainqueur de Condorcet. La condition d'unimodalité implique l'existence d'une telle option - si le nombre d'agents est impair, le vainqueur de Condorcet est unique.

**4.2 Les autres conditions**

- La Condition U n'est pas la seule que l'on ait cherché à affaiblir. On a exploré également les affaiblissements de la Condition R qui exige que les préférences collectives soient rationnelles. Par exemple : on peut exiger que la préférence collective  $\succeq_C$  soit seulement **quasi-transitive** :  $P_C$  est transitive mais  $I_C$  peut ne pas l'être.

Le résultat n'est cependant pas très satisfaisant : on peut montrer qu'une fonction de préférence sociale qui est quasi-transitive et qui satisfait les conditions U, P et I est nécessairement **oligarchique** : il existe un groupe d'individus  $O \subseteq N$  tel que

1.  $xR_Cy$  ssi il existe  $i \in O$  tel que  $xP_iy$
2.  $xP_Cy$  ssi pour tout  $i \in O$ ,  $xP_iy$

**5 Le théorème de Gibbard-Satterthwaite**

- Jusqu'à présent, nous avons supposé un accès aux préférences individuelles. Bien souvent, ces préférences sont *exprimées* (ex : vote). Il ne va pas de soi que les préférences exprimées correspondent aux préférences authentiques de l'individu. L'une des raisons qui peut expliquer que les deux divergent est tout simplement que l'individu exprime délibérément de fausses préférences. Pourquoi ? Parce qu'il se pourrait que la fonction de préférence sociale (de choix social en fait ci-après) aboutisse à un résultat qu'il préfère quand il exprime de fausses préférences !

On parle de **manipulabilité** d'une procédure d'agrégation quand certains individus peuvent avoir intérêt à *ne pas* rapporter leurs préférences fidèlement.

**Définition 9**

Une **fonction de choix social**  $h : \mathfrak{R}^n \rightarrow X$  associe à tout  $n$ -uplet de préférences individuelles  $(P_1, \dots, P_n)$  une option de l'ensemble  $X$ .

**Définition 10**

La fonction de choix social  $h$  est manipulable par  $i$  en  $(P_1, \dots, P_i, \dots, P_n)$  s'il existe une  $i$ -variante  $(P_1, \dots, P'_i, \dots, P_n)$  de  $(P_1, \dots, P_i, \dots, P_n)$  telle que

$$h(P_1, \dots, P'_i, \dots, P_n) P_i h(P_1, \dots, P_i, \dots, P_n)$$

• Pour le théorème de Gibbard-Satterthwaite, on s'intéresse à une classe plus restreinte de relations de préférence : elles ne doivent pas seulement être rationnelles ( $\mathfrak{R}$ ), mais être en outre linéaire. On note ces relations  $\mathfrak{L}$ .

**Définition 11**

**Condition U** (Universalité) :  $\mathfrak{L}^n = \mathfrak{L}^n$ .

**Condition S** (Surjectivité) :  $h(\mathfrak{L}^n) = X$

**Condition NM** (Non Manipulabilité) : il n'existe pas de profil  $(P_1, \dots, P_n)$  et d'individu  $i$  tels que  $h$  est manipulable par  $i$  en  $(P_1, \dots, P_n)$ .

**Condition D** (Absence de dictature) : il n'existe pas de dictateur pour  $h$  c'est-à-dire d'individu  $i$  tel que pour tout  $(P_1, \dots, P_n)$ , pour tout  $x \in X - \{h(P_1, \dots, P_n)\}$ ,  $h(P_1, \dots, P_n) P_i x$ .

Il existe bien sûr des fonctions de choix social non manipulables.

**Théorème 3 (Théorème de Gibbard-Satterthwaite)**

Si  $n \geq 2$  et  $|X| \geq 3$ , aucune fonction de choix social ne satisfait les Conditions U, S, NM et D.

• La propriété d'unimodalité des préférences individuelles permet également de contourner le Théorème de Gibbard-Satterthwaite.

Sous les hypothèses du Théorème de possibilité, on sait qu'il existe un unique vainqueur de Condorcet. La fonction de choix social majoritaire  $h$  associe à chaque profil le vainqueur de Condorcet correspondant.

## 6 Le théorème d'impossibilité du Parétien Libéral

• Sen (1970) introduit dans le cadre de la théorie du choix social la notion de *liberté individuelle* via une condition de liberté minimale. Supposons que les états sociaux  $a$  et  $b$  ne diffèrent qu'en ceci :  $a$  l'individu porte du rouge et dans  $b$  d'ivoire.  $i$  doit avoir la liberté de choisir entre  $a$  et  $b$  - même si les autres individus ont des préférences qui contredisent celle de  $i$ . La condition est la suivante : il y a au moins deux options pour un individu  $i$  sur lesquelles l'individu est *décisif* : la préférence sociale doit alors s'aligner sur les préférences de  $i$ . On parle parfois de **sphère privée** pour désigner les paires entre lesquelles un individu devrait être décisif.

### Définition 12

**Condition LM** (*Liberté Minimale*) : il existe deux individus  $i$  et  $j$  tels que pour chacun il existe deux options sur lesquelles ils sont décisifs

### Théorème 4 (Théorème d'impossibilité du Parétien Libéral (Sen, 1970))

S'il y a au moins deux individus et deux options, il n'existe pas de fonction de choix social qui satisfasse simultanément les conditions  $R^*$ ,  $U$ ,  $P$  et  $LM$ .

*Preuve* : supposons que

- $i$  est décisif sur  $(x, y)$ ,
- $j$  est décisif sur  $(z, w)$ ,
- $xP_i y$  et  $zP_j w$
- $wP_k x$  et  $yP_k z \forall k$

Par hypothèse,  $xP_C y$  et  $zP_C w$ . Par la Condition  $P$ ,  $wP_C x$  et  $yP_C z$ . On obtient donc un cycle. ♠.

### Exemple 5 (L'Amant de Lady Chatterley)

Un livre, *l'Amant de Lady Chatterley*. 2 individus : l'un prude ( $pr$ ), l'autre libertin ( $lib$ ).

Trois états sociaux possibles :

- $a$  :  $pr$  lit le roman
- $b$  :  $lib$  lit le roman
- $c$  : personne ne le lit

Les préférences du prude et du libertin sont les suivantes :

$$\begin{array}{l} cP_{pr} a P_{pr} b \\ aP_{lib} b P_{lib} c \end{array}$$

Par hypothèse,  $pr$  est décisif sur  $\{a, c\}$  et  $lib$  sur  $\{b, c\}$ .

On a donc :

$$cP_C a, bP_C c$$

en vertu de la Condition  $P$ , on a :

$$aP_Cb$$

et on obtient donc une préférence collective cyclique :

$$cP_CaP_CbP_Cc$$

- Interprétation de Sen :

- ▷ Sen (1995)

“The “impossibility of the Paretian liberal” captures the conflict between (i) the special importance of a person’s preferences over her own personal sphere, and (ii) the general importance of people’s preferences over any choice, irrespective of field.”

- ▷ “...in very basic sense liberal values conflict with the Pareto Principle”

- “the ultimate guarantee for individual liberty may rest not on rules for social choice but on developing individual values that respect each other’s personal choices”

- Critique de Sugden :

- ▷ Sugden, cité par Sen (1983)

“La faille de cet ingénieux argument réside, à mon avis, dans la formulation par Sen du principe de liberté. Bien qu’il prétende faire appel à la même conception de la liberté que Mill, il existe une différence capitale dans la conception de la liberté chez ces deux auteurs. Mill aurait certes admis qu’“il existe certaines questions personnelles à propos desquelles chaque personne devrait être libre de décider ce qui va se passer” ; mais aurait-il convenu que “dans le choix concernant ces questions, toute solution jugée meilleure par l’individu doit être considérée comme meilleure pour l’ensemble de la société” ? La première de ces deux propositions constitue un jugement de valeur sur les procédures ; elle indique que certaines questions devraient être déléguées, ou réservées, à la décision individuelle. La seconde proposition, en revanche, est un jugement de valeur sur les états finaux : elle affirme en effet que la procédure consistant à réserver ces questions à la décision individuelle conduit invariablement à la sélection des meilleurs états finaux possibles. Mais pourquoi un libéral devrait-il affirmer cela ?”

- ▷ réponse de Sen :

“L’argumentation est convaincante, mais elle se fonde sur une vision excessivement étroite du contenu possible de la “préférence sociale”, et du statut du “meilleure pour la société”...

Lorsque la préférence sociale, au sens de l’évaluation du résultat, est située dans le contexte d’une conception purement procédurale de la liberté, un résultat considéré comme “meilleure pour la société” du point de vue de la liberté est considéré comme tel précisément parce que c’est ce qui serait choisi par la personne en question.”

## 7 Références

### 7.1 sur la théorie du choix social et l'économie normative

- Fleurbaey, Marc, "Economics and Economic Justice", The Stanford Encyclopedia of Philosophy (Spring 2006 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2006/entries/economic-justice/>.

- M. Salles (2006), "La théorie du choix social : une introduction à quelques résultats fondamentaux", *Leçons de philosophie économique*, t.II

- Picavet, E. (1996), *Choix rationnel et vie publique*, PUF [Partie III ; exposition et interprétation philosophique des théorèmes d'impossibilité]

- Sen (1995), "Rationality and Social Choice", *The American Economic Review*, 85 :1, pp. 1-24

- Sen (1999), "The possibility of social choice", *AER*, 89, 3, pp. 349-78

### 7.2 sur le théorème d'impossibilité d'Arrow

- Arrow, K.J. (1951) *Social Choice and Invididual Values*, Wiley, NY [le fameux "théorème d'impossibilité d'Arrow"]

- Gaertner (2006), chap. 2

### 7.3 sur les droits individuels

- Gaertner (2006), chap. 6

- Sen, A. (1970) "The Impossibility of a Paretian Liberal", *Journal of Political Economy*, vol. 78, pp. 152-7

- Sen, A. (1983), "Liberty and Social Choice", *The Journal of Philosophy*, LXXX, 1, pp. 5-28 ; trad.fr. "Liberté et choix social" dans *Ethique et économie*, PUF, 1993

### 7.4 sur les comparaisons interpersonnelles

- P. Hammond (1991), "Interpersonal Comparisons of Utility : Why and how they are and should be made" in J. Elster and J.E. Roemer (eds.), *Interpersonal Comparisons of Well-Being* (Cambridge : Cambridge University Press, 1991), ch. 7, pp. 200-254 ; reprinted

in A.P. Hamlin (ed.) *Ethics and Economics*, Vol. I (Edward Elgar, 1996), ch. 22, pp. 410-464.

- Ph. Mongin & C. d'Aspremont (1998), "Utility Theory and Ethics", in P. Hammond (ed.), *Utility Theory*