1 Les préférences

• Rappel : en toute généralité, la rationalité d'une décision dépend des opportunités, des désirs et des croyances. Question : Comment représenter les désirs de l'agent i.e. la façon dont il évalue les conséquences des différentes actions possibles et ces actions elles-mêmes ? La théorie de la décision retient une *relation comparative* (entre conséquences ou actions), les **préférences**.

Notation 1
• *préférences larges :* $a \succeq b$ *:*

(Pierre préfère "largement" a *à* b *; Pierre estime que* a *est au moins aussi bonne que* b*)*

• *préférences strictes :* $a \succ b$ *:*

(Pierre préfère "strictement" a *à* b *; Pierre estime que* a *est strictement meilleure que* b*)*

• *indifférence* $a \sim b$ *:*

(Pierre est indifférent entre a *et* b *; Pierre estime que* a *et* b *ont la même valeur)*

• Interprétation : Les préférences sont généralement conçues comme des **états mentaux** de l'agent. En tant que telles, elles ne sont pas directement observables par le modélisateur, qui doit les inférer à partir du comportement de l'agent. Elles sont néanmoins censées entretenir des relations extrêmement étroites avec le comportement observable : typiquement, on suppose que si un agent préfère x à y et qu'il a le choix entre les deux, alors il choisira x. Les préférences sont donc conçues comme des *déterminants immédiats de l'action.*

Les préférences sont censées intégrer toutes les considérations pertinentes pour l'évaluation (comparative) des actions (ou de leurs conséquences) ; on peut les considérer comme des *évaluations globales.*

• La théorie de la décision impose certaines **conditions de rationalité** sur les préférences. Par exemple,

si $a \succ b$, alors $\neg(b \succ a)$
si $a \succ b$, alors $\neg(a \sim b)$

Remarque : De telles situations peuvent ne pas être irrationnelles si l'agent *change* de désirs. L'hypothèse tacite sur laquelle repose la théorie des préférences est que la relation de préférence est une représentation *instantanée* des désirs de l'agent : ses désirs à un instant donné.

• Nous allons maintenant définir ce que l'on considère en général comme des *préférences rationnelles* - on dit parfois également des préférences cohérentes. On va le faire sur les préférences larges. Il y a 2 conditions de rationalité qui pèsent sur les préférences larges :

Définition 1 (Préférences rationnelles)
Une relation de préférence large $\succeq$ sur X est ***rationnelle*** *si*
(i) elle est ***transitive :*** $\forall x, y, z$ *si* $x \succeq y$ *et* $y \succeq z$ *alors* $x \succeq z$
(ii) elle est ***complète :*** $x \succeq y$ *ou* $y \succeq x$

Remarque : il suit de la Définition qu'une relation de préférence rationnelle est réflexive : $\forall x,\ x \succeq x$.

Définition 2
Soit $\succeq$ une relation de préférence large ;

- *la relation de préférence stricte associée $\succ$ est le résidu asymétrique de $\succeq$: $x \succ y$ ssi $x \succeq y$ et $\neg(y \succeq x)$*
- *la relation d'indifférence associée $\sim$ est le résidu symétrique de $\succeq$: $x \sim y$ ssi $x \succeq y$ et $y \succeq x$*

Proposition 1
Si $\succeq$ est une relation de préférence rationnelle, alors

(a) la relation de préférence stricte associée $\succ$ est (i) irréflexive ($\forall x,\ \neg(x \succ x)$), (ii) transitive ($\forall x, y, z$ si $x \succ y$ et $y \succ z$ alors $x \succ z$), (iii) asymétrique ($\forall x, y \in C$, si xRy, alors $\neg(yRx)$), (iv) négativement transitive ($\forall x, y, z \in X$, $\neg(xRy)$ et $\neg(yRz)$ $\Rightarrow \neg(xRz)$)

(b) la relation d'indifférence associée $\sim$ est (i) réflexive, (ii) transitive, (iii) symétrique ($\forall x, y$, si $x \sim y$ alors $y \sim x$)

Preuve (sketch). (a-iii) Asymétrie. Supposons que $x \succ y$; alors $\neg(y \succeq x)$ donc $\neg(y \succ x)$. (a-iv) Transitivité négative. Supposons que $\neg(x \succ y)$ et $\neg(y \succ z)$. Alors $\neg(x \succeq y)$ ou $y \succeq x$. Etant donné que $\succeq$ est totale, cela équivaut simplement à $y \succeq x$. De la même façon, on a $z \succeq y$. Par transitivité de $\succeq$, on obtient donc $z \succeq x$. Par définition de $\succ$, cela suffit à ce que l'on ait $\neg(x \succ z)$. ♠

2 Discussion des conditions de rationalité

• Nous avons parlé de *préférences rationnelles* pour désigner des préférences complètes et transitives. On considère en effet souvent que ce sont des conditions de rationalité (on parle aussi parfois de "cohérence") sur les préférences : on doit avoir des préférences transitives et complètes. Dans cette section, nous allons discuter la force normative de la transitivité : la rationalité requiert-elle vraiment qu'un agent ait des préférences transitives ?
Remarque : la transitivité est particulièrement attractive (et plausible descriptivement) quand les options ou conséquences sont unidimensionnelles et en particulier se laissent exprimer numériquement. Exemple : argent.

• **Argument 1 : l'argument de la pompe à finance (*money pump*)** (Davidson & Suppes) :

Exemple 1 (La ruine de Pierre)
Soient les options :
a = *un poste très prestigieux à 50 000 euros/an*
b = *un poste assez prestigieux à 55 000 euros/an*
c = *un poste peu prestigieux à 60 000 euros/an*
Supposons que Pierre ait des préférences cycliques comme suit :

$$\boxed{\begin{array}{c} (i)\ a \succ b \\ (ii)\ b \succ c \\ (iii)\ c \succ a \end{array}}$$

Supposons en outre que Pierre possède a. Alors, en vertu de (iii), il devrait être prêt à payer ϵ pour obtenir c ; en vertu de (ii), il devrait ensuite être prêt à payer ϵ' pour obtenir b ; et en vertu de (i), il devrait être prêt à payer ϵ'' pour obtenir a. A la fin de cet échange, il se retrouve donc avec ce qu'il possédait au début, a, mais il a versé $\epsilon + \epsilon' + \epsilon''$.
Ce n'est pas fini : on peut itérer la suite d'échanges et ruiner Pierre.

L'argument de la pompe à finance est un argument *pragmatique* en faveur de la thèse d'irrationalité de l'intransitivité. En gros, l'argument est le suivant : intransitivité $\Rightarrow$ vulnérabilité $\Rightarrow$ irrationalité.

Objection :

▷ F. Schick (1986), "Dutch Bookies and Money Pumps", *The Journal of Philosophy*, vol.83, pp. 112-9

"Does a person with cyclical preferences have no grounds for declining offers ? Let him look back and see the arrangements he has already paid for. He may then come to see which way the wind is blowing, that if he accepts the current offer, he will then get another, and then another, and still another, every cycle bringing him back to where he was at the start, only poorer. Seeing what is in store for him, he may well reject the offer and thus stop the pump."

Idée générale : l'argument du *money pump* semble faire une hypothèse de myopie très forte sur le décideur ; sans cette hypothèse, il est peu plausible que le décideur aux préférences cycliques s'engage dans le processus. C'est un argument qui attaque donc "intransitivité $\Rightarrow$ vulnérabilité". Rabinowicz (2000) critique cette objection : si l'on formalise le problème rigoureusement sous forme séquentielle, on aboutit également à la vulnérabilité.

- **Argument 2 : les préférences sont analytiquement transitives** (Davidson)

▷ D. Davidson, (1980), *Essays on Actions and Events*

> "If length is not transitive, what does it mean to use of number of measure length at all ? We could find or invent an answer, but unless or until we do, we must strive to interpret 'longer than' so that it comes out transitive. Similarly for 'preferred to'."

Objection (Anand) : pourquoi accepter l'analogie avec les longueurs ? Pourquoi pas avec "l'équipe A bat l'équipe B" ? Dans ce cas, on aurait pas nécessairement transitivité.

Exemple 2 (Le paradoxe de la course de chevaux (Blyth 1972))
Trois chevaux a, b, c.

- *s'il pleut (s_1), $aBbBc$*
- *si le temps est humide (s_2), $bBcBa$*
- *si le temps est sec (s_3), $cBaBb$*

Supposons (i) que $x \succ y$ ssi $P(xBy) > 1/2$ (règle R) ; et (ii) que $P(s_1) = P(s_2) = P(s_3) = 1/3$. Alors,

$$b \succ c \succ a \succ b$$

Discussion : la règle R de détermination des préférences est sensée pour des *courses à deux chevaux*. Elle l'est beaucoup moins pour la course à trois chevaux puisque dans ce cas, x est préféré à y ssi x a une plus grande chance de terminer dans les deux premiers. Il faudrait plutôt adopter une règle comme la suivante :

$$x \succeq y \text{ ssi } P(xBy \wedge xBz) \geq P(yBx \wedge yBz) \text{ (R')}$$

...et dans ce cas, l'agent sera simplement indifférent entre les paris sur a, b et c.

- Sur la complétude.

▷ R. Aumann (1962) "Expected Utility without the Completeness Axiom"

> "Of all the axioms of the utility theory, the completeness axiom is perhaps the most questionable. Like others, it is inaccurate as a description of real life ; but unlike them we find it hard to accept even from a normative point of view."

• Digression sur l'idée selon laquelle la théorie de la décision ne dit pas quelles préférences il faut avoir mais impose seulement des contraintes de cohérence sur les préférences. Broome (1991, chap.5 et 1999, chap.5) distingue le **humeanisme extrême** du **humeanisme modéré**. Selon le humeanisme extrême, il n'y a aucune préférence irrationnelle. Selon le huméanisme modéré, il y a quand même des contraintes de cohérence qui pèsent sur les préférences. Comme par exemple la transitivité.

> "The moderate Humean view, then, comes down to this. You may, rationnally, have any preferences, provided only that they are consistent with each other. And what consistency requires is spelt out in decision theory."

Broome propose l'exemple suivant pour montrer que le humeanisme modéré n'est pas une position tenable.

Exemple 3 (L'intransitivité de Maurice)
Maurice préfère
visiter Rome plutôt qu'aller à la montagne : $R \succ M$
rester à la maison plutôt que visiter Rome : $H \succ R$
aller à la montagne plutôt que rester à la maison : $M \succ H$
Intransitivité ? Maurice répond que ses préférences ne sont pas intransitives : il ne faut pas confondre l'option de rester à la maison quand elle est comparée avec celle de visiter Rome H_R *et l'option de rester à la maison quand elle est comparée avec celle d'aller à la montagne* H_M*. Si l'on opère cette distinction, on n'a plus d'intransitivité :* $R \succ M$, $H_R \succ R$ *et* $M \succ H_M$.

Inteprétation de Broome : la transitivité implique que $H_R \succ M$, mais il ne s'agit pas de *préférences pratiques*, c'est-à-dire de préférences auxquelles peut correspondre un choix. La leçon de l'exemple de Maurice est donc que les préférences pratiques ne sont pas contraintes par la seule transitivité. Si l'on juge Maurice irrationnel, ce ne peut être que parce que l'on juge irrationnel qu'il ait une préférence entre H_R et H_M. Il faut donc que les contraintes de rationalité portent sur des préférences/indifférences entre deux options : "Rational principles of indifference are needed to give consistency a grip on practical preferences"

3 Le modèle de choix certain

• Pour le moment, nous n'avons abordé que les déterminants mentaux du choix, pas le choix à proprement parler. Dans le cas du choix certain, néanmoins, l'essentiel est fait : le modèle de choix certain (MCC) ne fait qu'affirmer la relation attendue entre préférences et choix (relation que nous appelons (MEC RAT) et que Sugden (1985) appelle la *revealed preference condition*) :

<u>Modèle de choix certain</u> :

(MEC A) l'agent peut choisir entre les éléments d'un ensemble d'opportunités ou d'actions réalisables $X \subseteq A$

(MEC P) les préférences de l'agent sur A sont rationnelles

(MEC RAT) l'agent choisit, s'il en existe au moins une, une des actions dont la conséquence est $\succeq$-maximale relativement à X. On note $c(.)$ la fonction de choix qui s'écrit donc

$$c(X) = \{x : \forall y \in X, x \succeq y\}$$

• le MCC impose des conditions sur la relation entre $\succ$ et $c(.)$. Une telle connexion permet d'inférer certains des choix de l'agent à partir de certaines de ses préférences, ou inversement certaines de ses préférences à partir de certains de ses choix :

(i) *inférence des choix à partir des préférences.* Supposons que l'agent doive au moins choisir une action parmi celles qui sont réalisables (condition de non-vacuité). Le critère de rationalité des modèles élémentaires de choix a pour conséquence directe que si $x \succ y$, alors $c(\{x, y\}) = \{x\}$[1].

(ii) *inférence des préférences à partir des choix.* Inversement, on peut se baser sur $c(.)$ pour attribuer des préférences à l'agent. Si $c(\{x, y\}) = \{x\}$, alors on peut inférer que $x \succeq y$ et que $\neg(y \succ x)$. Ce que l'on va pouvoir inférer des décisions de l'agent va dépendre de la notion exacte de décision en jeu.
Supposons en effet que, par construction, l'agent ne puisse choisir qu'une action par problème de choix (condition d'exclusivité) ; cela revient à poser que $c(.)$ est monovalente (pour tout argument X, $c(X)$ est un singleton). Dans ce cas, on ne peut pas inférer de $c(\{x, y\}) = \{x\}$ que $x \succ y$: il se peut très bien que l'agent soit indifférent entre x et y mais choisisse x parce qu'il doit nécessairement faire un choix . En revanche, si l'on autorise que $c(.)$ soit plurivalent, c'est-à-dire si l'on autorise que $c(.)$ recouvre les choix que l'agent considère comme acceptables, alors ce que l'on peut inférer de $c(\{x, y\})$ devient plus précis : si $c(\{x, y\}) = \{x\}$, alors $x \succ y$ et si $c(\{x, y\}) = \{x, y\}$, alors $x \sim y$.

De manière plus générale, les relations entre $c(.)$ et $\succ$ déterminent **le contenu empirique comportemental** du MCC, et elles ont fait l'objet d'investigations très approfon-

[1]Bien sûr, ce n'est pas le cas si l'on suppose simplement que $x \succeq y$ puisqu'il se peut très bien que l'agent soit *indifférent* entre x et y.

dies. Les relations entre $\succ$ et $c(.)$ ont été systématiquement étudiées sous l'impulsion de la *théorie de la préférence révélée*. Voici quelques-uns des concepts et résultats fondamentaux.

On peut définir, pour tout sous-ensemble $X \subseteq A$ l'ensemble des actions qui peuvent être choisies par l'agent selon le MCC :

$c_{\succ}(X) = \{x \in X : \forall y \in X, \neg(y \succ x)\}$ si l'on part d'une relation de préférence stricte,

$c_{\succeq}(X) = \{x \in X : \forall y \in X, x \succeq y\}$ si l'on part d'une relation de préférence large.

Ces définitions sont équivalentes pour des préférences dotées des propriétés usuelles en vertu de la totalité de la préférence large. On appelera $c_{\succ}$ (resp. $c_{\succeq}$) la **fonction de choix engendrée** par la relation de préférence $\succ$ (resp.$\succeq$). De manière générale, on peut définir ainsi une fonction de choix, c'est-à-dire une fonction qui décrit les choix de l'agent pour des ensembles possibles d'actions réalisables :

Définition 3
*Soit A un ensemble d'actions réalisables et $\mathbb{F} \subseteq \wp(A)$; une **fonction de choix** pour $\mathbb{F}$ est une fonction $c : \mathbb{F} \to \wp(A)$ t.q. $\forall X \in \mathbb{F}$, $c(X) \subseteq X$. On appelle $(\mathbb{F}, c)$ une **structure de choix**.*

La fonction de choix indique, pour un certain nombre de problèmes de choix, quels sont les choix de l'agent. On remarquera que l'on n'exige pas que les fonctions de choix soient monovalentes : $c(X)$ peut ne pas être réduit à un singleton. Pourtant, on conserve la même interprétation des actions réalisables que dans les modèles préférentiels : les actions réalisables sont mutuellement exclusives et par conséquent un agent ne peut pas choisir *effectivement* plusieurs actions pour le même problème de choix. Une fonction de choix (non monovalente) peut recevoir plusieurs interprétations possibles : en général, on considère qu'elle identifie les actions que l'agent *accepterait* de choisir. Dans ce qui suit, on supposera que la fonction de choix est non-vide pour tout argument de son domaine :

Condition de non-vacuité : pour tout $X \in \mathbb{F}$, si $X \neq \emptyset$, alors $c(X) \neq \emptyset$

On supposera en particulier la condition de non-vacuité pour les fonctions de choix engendrées par des relations de préférence constituant un ordre faible ; ceci équivaut logiquement à supposer que la relation de préférence stricte $\succ$ n'a pas de chaîne infinie ascendante. Les fonctions de choix permettent de clarifier systématiquement les rapports entre les préférences de l'agent et ses choix.

- 2 questions théoriques fondamentales :

1. *propriétés nécessaires* : quelles conditions les décisions de l'agent doivent-elles nécessairement satisfaire s'il se conforme au MCC ?
2. *propriétés suffisantes* : quelles conditions suffisent pour que les décisions de l'agent puissent être expliquées par le MCC ?

Remarques :

- si une propriété nécessaire n'est pas satisfaite, alors, par contraposition, il n'est pas possible de rendre compte du comportement décisionnel en question. Les violations des propriétés nécessaires constituent donc des réfutations du modèle.

- Si une propriété suffisante est satisfaite, alors le comportement décisionnel de l'agent est *compatible* avec le modèle de choix rationnel. Autrement dit, le modèle ne peut pas être réfuté par le comportement de l'agent ; les propriétés suffisantes circonscrivent les réfutations possibles du modèle.

Définition 4
Soit $(\mathbb{F}, c)$ *une structure de choix ;*

(i) $(\mathbb{F}, c)$ *est* ***rationalisée*** *par une relation de préférence* $\succ$ *si* $(\mathbb{F}, c) = (\mathbb{F}, c_{\succ})$

(ii) $(\mathbb{F}, c)$ *est* ***rationalisable*** *s'il existe une relation de préférence qui la rationalise.*

On peut prendre comme exemple de propriété décisionnelle impliquée par le modèle élémentaire de choix la Propriété α :

Définition 5
Une structure de choix $(\mathbb{F}, c)$ *satisfait la Propriété* α[2] *si pour tout* $X, Y \in \mathbb{F}$,

$$\text{si } X \subseteq Y \text{ et si } x \in c(Y) \cap X, \text{ alors } x \in c(X) \text{ (Propriété } \alpha)$$

"Ìf the world champion in a particular discipline is a Pakistani, he must be a Pakistani champion" (Sen)

Exemple 4 (contre-exemple)
$X = \{a, b\}$, $Y = \{a, b, c\}$, $c(Y) = \{b\}$ *mais* $c(X) = \{a\}$.

On peut facilement constater qu'une fonction de choix engendrée par une relation de préférence dotée des propriétés usuelles satisfait la Propriété α. En revanche, toute fonction de choix qui satisfait la Propriété α n'est pas rationalisable : la Propriété α est une propriété nécessaire mais pas suffisante pour qu'une fonction de choix soit compatible avec le modèle élémentaire de choix. On peut considérer une propriété plus forte, l'axiome faible de congruence.

Définition 6
La structure de choix $(\mathbb{F}, c)$ *satisfait l'axiome faible de congruence (WAC) si :*

> *si pour un* $X \in \mathbb{F}$ *avec* $x, y \in X$, $x \in c(X)$, *alors pour tout* $X' \in \mathbb{F}$ *avec* $x, y \in X'$, $y \in c(X')$, $x \in c(X')$ *(WAC)*[3]

[2] Aussi appelée "basic contraction consistency" ou "independence of irrelevant alternatives".

[3] (WAC) est logiquement équivalent à ce que Sen (1971) appelle l'axiome de la préférence révélée : (WARP) S'il existe un $X \in \mathbb{F}$ tel que $x \in c(X)$ et $y \in X - c(X)$, alors il n'existe pas $Y \in \mathbb{F}$ tel que $x \in Y$ et $y \in c(Y)$.

Idée : s'il arrive que x soit choisi alors que y est disponible, x est toujours choisi quand il est disponible et que y est choisi.

Dans le cas particulier où $c(.)$ est monovalente, la Propriété α est équivalente à l'axiome (WAC)[4]. Dans le cas général, pour passer de la Propriété α à l'axiome (WAC), il faut ajouter la Propriété β :

Définition 7
Une structure de choix $(\mathbb{F}, c)$ satisfait la Propriété β si pour tout $X, Y \in \mathbb{F}$,

si $X \subseteq Y$ $x, y \in c(X)$, alors $y \in c(Y)$ ssi $x \in c(Y)$ (Propriété β)

"If a Pakistani is world champion, then all the Pakistani champions are world champions"

On vérifie aisément, là encore, que si $\succ$ est une préférence (stricte) rationnelle alors la structure de choix $(2^A, c_\succ)$ satisfait l'axiome faible de congruence (WAC). La réciproque, en revanche, n'est toujours pas vraie : il existe des structures de choix qui satisfont (WAC) mais qui ne sont pas rationalisables. Il suffit cependant de postuler que $\mathbb{F}$ contient au moins les sous-ensembles finis de A[5] - on dira alors que $(\mathbb{F}, c)$ est *régulière* - pour obtenir la réciproque : si une fonction de choix est régulière et satisfait (WAC), alors elle est rationalisable (Sen 1971, (T.3)). Attention : ce n'est pas toujours une condition bénigne - voir par exemple en théorie du consommateur.

4 L'utilité

• On a introduit un premier concept, celui de préférences ; nous allons désormais passer à un second concept central de la théorie de la décision, celui d'**utilité**. Autant la notion technique de préférence est relativement proche de la notion préthéorique, autant la notion d'utilité reçoit un sens très spécifique en théorie de la décision.

• L'utilité est *numérique* : c'est un nombre que l'on attache aux mêmes objets que ceux qui figurent dans la relation de préférence - les actions réalisables ou les conséquences de ces actions. Une **fonction d'utilité** est une fonction qui associe un nombre à chaque objet de son domaine. Le terme "utilité" a un certain nombre de connotations préthéoriques. On peut par exemple penser que l'utilité d'une action a pour Pierre est la mesure du *bien* qu'elle peut constituer pour Pierre ; ou encore la mesure du *plaisir* qu'elle peut procurer à Pierre. En théorie de la décision, et plus précisément en situation de certitude, l'utilité n'a pas cette signification : **l'utilité est une représentation numérique des préférences**. Voici ce que cela signifie exactement.

[4]Sen 1971, (T.6)

[5]En réalité, il suffit de supposer que $\mathbb{F}$ contient au moins les sous-ensembles qui contiennent trois éléments ou moins.

Définition 8
Une fonction d'utilité $u : X \to \mathbb{R}$ ***représente*** *une relation de préférence* $\succeq$ *(resp.* $\succ$*) sur* X *ssi pour tout* $x, y \in X$, $x \succeq y$ *ssi* $u(x) \geq u(y)$ *(resp.* $x \succ y$ *ssi* $u(x) > u(y)$*)*

En d'autres termes, une fonction d'utilité représente une relation de préférence si l'ordre naturel sur les nombres associés aux objets correspond à l'ordre décrit par la relation de préférence. Les fonctions numériques ne sont pas les seules à pouvoir représenter des relations de préférence. Penser par exemple aux lettres dans la notation des copies : l'échelle des lettres $\{A, B, C, D, E\}$ peut (dans certaines circonstances favorables) représenter la relation "x est une meilleure copie que y". Par ailleurs, dans d'autres domaines on a ce genre de correspondance entre fonction numérique sur des objets et relations comparatives entre paires d'objets. Penser par exemple à la température et à la relation "x est plus chaud que y".

• Il y a un point qui est mathématiquement trivial mais conceptuellement très important, c'est le fait que si l'on n'exige seulement d'une fonction d'utilité qu'elle représente des préférences, alors, s'il existe au moins une fonction qui peut le faire, "beaucoup" d'autres fonctions peuvent le faire également :

Proposition 2
Si $u : X \to \mathbb{R}$ *représente la relation de préférence* $\succ$*, alors pour toute fonction strictement croissante* $f : \mathbb{R} \to \mathbb{R}$, $v(x) = f(u(x))$ *représente* $\succ$ *également.*

La représentation d'une relation de préférence est unique à une transformation strictement croissante près. Pour cette raison, on dit que le concept d'utilité que nous venons d'introduire est ***ordinal*** : il ne fait qu'encoder des différences de position dans la hiérarchie évaluative de l'agent, pas des différences d'*intensité*. Pour une fonction d'utilité ordinale, cela ne fait pas sens de dire que

(1) l'utilité de b est deux fois celle de a

(2) l'utilité de c est juste à mi-chemin entre celle de a et celle de b

En effet, si la fonction d'utilité u_1 est telle que $u_1(b) = 4$ et $u_1(a) = 2$, il se peut parfaitement que la fonction d'utilité u_2 représente aussi bien les préférences et soit telle que $u_2(b) = 4$ et $u_2(a) = 3$.

• Il est aisé de voir que si une relation de préférence stricte est représentée par une fonction d'utilité, alors elle est rationnelle. La réciproque est vraie si X est fini ou dénombrable, mais pas dans le cas général (Kreps 1988, chap.3).

Théorème 1
Supposons que X soit dénombrable. $\succ$ une relation de préférence rationnelle sur X ssi il existe une fonction $u : X \rightarrow \mathbb{R}$ qui représente $\succ$.

Preuve (pour X fini).
($\Leftarrow$). Si $u(.)$ représente $\succ$, alors $\succ$ est asymétrique et négativement transitive : (i) $x \succ y$ ssi $u(x) > u(y)$ ssi $\neg(u(y) \geq u(x))$ ssi $\neg(y \succ x)$. (ii) Si $\neg(x \succ y)$ et $\neg(y \succ z)$ alors $\neg(u(x) > u(y))$ et $\neg(u(y) > u(z))$ ssi $u(x) \leq u(y) \leq u(z)$. Donc $u(x) \leq u(z)$ ssi $\neg(u(x) > u(z)$ donc $\neg(x \succ z)$.

($\Rightarrow$). Kreps (1988) donne une preuve par induction sur la taille de X. Le départ de la preuve est immédiat ; si $X = \{x\}$, on pose $u(x) = 1/2$. On suppose qu'on a montré que cela vaut pour tout ensemble de cardinalité $n-1$ et on considère un ensemble X de cardinalité n. On choisit un élément $x* \in X$ et on note $X' = X - x*$. La restriction de $\succ$ sur X' est une une relation de préférence rationnelle et par conséquent, par hypothèse d'induction, il existe une fonction u' qui représente $\succ$ sur X'.

Cas 1 : il existe $x' \in X'$ t.q. $x' \sim x*$. Dans ce cas, $u(x*) = u'(x')$ et on ne modifie pas $u'(x)$ pour $x \in X'$.

Cas 2 : pour tout $x \in X'$, $x* \succ x$. Dans ce cas, $u(x*) = [\max_{x \in X'} u'(x) + 1]/2$ et on ne modifie pas $u'(x)$ pour $x \in X'$. Noter que l'on pourrait prendre tout réel strictement supérieur à $\max_{x \in X'} u'(x)$. Le choix présent permet de conserver le co-domaine de $u(.)$ dans $(0, 1)$.

Cas 3 : pour tout $x \in X'$, $x \succ x*$. Dans ce cas, $u(x*) = [\min_{x \in X'} u'(x)]/2$ et on ne modifie pas $u'(x)$ pour $x \in X'$.

Cas 4 : il n'existe pas $x' \in X'$ t.q. $x' \sim x*$ et $x*$ n'est ni strictement préféré ni strictement préférable aux éléments de X'. L'idée est alors la suivante : on va considérer l'option $\underline{x}$ qui est la meilleure des options à qui $x*$ est préférée et $\bar{x}$, la moins bonne des options qui sont préférées à $x*$. Dans ce cas, $u(x*) = [u'(\bar{x}) + u'(\underline{x}]/2$. ♠

L'une des vertus de cette preuve est qu'elle permet de voir comment construire pas à pas la fonction d'utilité d'un agent en lui demandant, pour chaque nouvelle option $x \in X$, quelles relations préférentielles elle entretient avec les options auxquelles une utilité a déjà été affectée.

Exemple 5
On considère d'abord $x_1 \in X$; on lui attribue

$u(x_1) = 1/2.$

On considère ensuite x_2. Supposons que $x_2 \succ x_1$. On se trouve alors dans le Cas 3. Aussi

$u(x_2) = [u(x_1) + 1]/2 = [1/2 + 1]/2 = 3/4$

On considère ensuite x_3 que le décideur classe entre x_1 et x_2 : $x_2 \succ x_3 \succ x_1$. On se trouve alors dans le Cas 4. Aussi

$u(x_3) = [u(x_1) + u(x_2)]/2 = [1/2 + 3/4]/2 = 5/8$

Etc.

• Le théorème ne s'étend pas tel quel pour un X de cardinalité non-dénombrable. Les préférences lexicographiques en sont l'une des illustrations canoniques.

Exemple 6 (Préférences lexicographiques)
Soit $X = \mathbb{R}^2_+$ et pour $x = (x_1, x_2)$ et $y = (y_1, y_2)$, $x \succ y$ ssi $x_1 > y_1$ ou $x_1 = y_1$ et $x_2 > y_2$. Intuitivement, de telles préférences signifient que l'agent accorde une priorité totale à la première des deux composantes des options ; et qu'en cas d'égalité seulement, il juge en fonction de la seconde composante. Une telle relation de préférence est rationnelle.

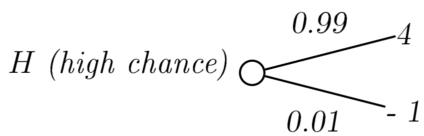
Il faut donc imposer certaines conditions "en plus" de la rationalité sur $\succ$ pour garantir la représentabilité par une fonction d'utilité. Existe-t-il une propriété P telle que $\succ$ est rationnelle et satisfait la propriété P ssi elle est représentable par une fonction d'utilité ? Oui. Nous ne developerons pas ce point, mais on peut mentionner cette propriété, la **séparabilité** : il existe un ensemble dénombrable $Z \subset X$ tel que pour tous $x, y \in X - Z$, si $x \succ y$, alors il existe $z \in Z$ tq $x \succ z \succ y$. Il est aisé de voir que les préférences lexicographiques de l'exemple précédent ne sont pas séparables. Intuitivement, un ensemble dénombrable Z est trop petit pour qu'un de ses éléments s'intercale entre chaque $x, y \in X - Z$.

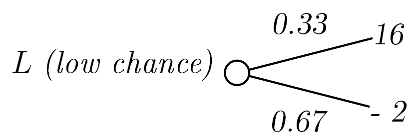
En général, en micro-économie du consommateur (où X est un sous-ensemble de l'espace euclidien à n dimensions), on fait une hypothèse suffisante mais non nécessaire sur les préférences rationnelles : on suppose qu'elles sont en outre *continues*, ce qui garantit la représentation par une fonction numérique (continue qui plus est, Debreu 1954). En gros, $\succeq$ est continue si, pour toute suite de paires $\{(x_n, y_n)\}$ telle que (i) pour tout n, $x_n \succeq y_n$ et (ii) x (resp. y) est la limite de $\{x_n\}$ (resp. $\{y_n\}$), $x \succeq y$.

5 Le renversement de préférences

• Pour terminer cette introduction au modèle de choix certain, nous allons revenir à la notion de préférence. La notion de préférence a été remise en question par le *renversement* *de préférences*, un phénomène découvert par Lichtenstein & Slovic (1971) et qui a fait l'objet de nombreuses explorations empiriques et théoriques depuis - voir l'anthologie Lichtenstein & Slovic (2006). Le coeur du phénomène est qu'un changement de la méthode de révélation aboutit à un renversement des préférences : typiquement, on considère deux méthodes, et alors que selon la méthode 1, $H \succ_1 L$, selon la méthode 2, $L \succ_2 H$.

Exemple 7
Soient les deux options suivantes, H et L :





Comportement modal :

- la plupart des sujets choisissent H plutôt que L

- pour la plupart des sujets, le prix minimal de vente (PMin) de H est inférieur à celui de L

• Selon Lichtenstein & Slovic (2006), les renversements des préférences se produisent typiquement dans des situations de choix (ou d'évaluation) qui (1) ne sont pas familières, (2) mettent en jeu des options où les préférences connues sont conflictuelles (ex., la certitude d'un gain vs. la magnitude d'un gain) et (3) il est difficile de synthétiser numériquement ces préférences conflictuelles.

• D.Grether & C.Plott (1979) ont mené des recherches empiriques pour tenter de "discréditer" des phénomènes potentiellement catastrophiques pour la théorie économique. Ils mettent à l'épreuve 12 tentatives de "sauvetage" de la théorie de la décision standard, comme par exemple : (i)les sujets ne sont pas motivés financièrement à jouer le jeu ; (ii) il peut y avoir un "effet revenu" ; (iii) les sujets peuvent en fait être indifférents entre les options ; (iv) les sujets peuvent être méfiants vis-à-vis d'expérimentateurs psychologues, etc. Mais à la grande satisfaction des psychologues, l'analyse des résultats de Grether & Plott montre que tous ces paramètres affectent pas ou peu les comportements. Le renversement des préférences semble "réel" et robuste.

• Différentes possibilités interprétatives (Tversky & Thaler 1991) :

(1) violation de l'invariance procédurale : les préférences changent en fonction de la méthode de révélation. L'heuristique qui a guidé Slovic & Lichtenstein (1971) est l'idée que quand on demande les préférences "simples" entre loteries, c'est la probabilité de gagner quelque chose qui gouverne les réponses ; tandis que lorsqu'on demande un prix, c'est la magnitude du gain qui est déterminante.

(2) violation de la transitivité : si c'est la même relation de préférence qui est révélée par les deux méthodes et si l'on suppose qu'elle est monotone par rapport au gain, alors on obtient

$$PMinV(H) \sim H \succ L \sim PMinV(L) \succ PMinV(H)$$

(3) violation d'un axiome du modèle d'espérance d'utilité (indépendance) : procédure d'élicitation par PMinV repose sur un mauvais schéma d'incitation : la procédure standard (BDM) ne reçoit de garantie théorique que si le sujet est un maximisateur d'espérance d'utilité. Si le décideur viole l'axiome d'indépendance (et on a de bonnes raisons de penser que dans bien des situations c'est le cas, voir le Paradoxe d'Allais), alors on a plus cette garantie.
La procédure BDM : soit une loterie x. Après que le sujet donne son $PMin(x)$, on tire au sort un montant. Si le montant excède $Pmin(x)$, on donne le montant au sujet ; sinon, on joue la loterie x. Cette procédure reçoit une justification théorique : si l'agent maximise son espérance d'utilité, alors le $PMin(x)$ est l'équivalent certain de x.

Tversky, Slovic & Kahneman (1990) ont mis en place un protocole dont l'objectif est de discriminer entre les interprétations par violation de l'invariance procédurale et celles par violation de la transitivité. Ils introduisent un gain certain X tel que $C_L > X > C_H$ et proposent des choix binaires ($H - L$, $H - X$, $L - X$). Il y a quatre cas possibles :

(1) si $L \succ X \succ H$, alors on a intransitivité puisque $L \succ X \succ H \succ L$
(2) si $X \succ H$ et $X \succ L$, alors on a sur-évaluation de L puisque $C_L \succ X \succ L$
(3) si $H \succ X$ et $L \succ X$, alors on a sous-évaluation de H puisque $H \succ C \succ C_H$
(4) si $H \succ X$ et $X \succ L$, alors on a sur-évaluation de L et sous-évaluation de H

Ils estiment avoir des résultats particulièrement univoques : selon eux, 90 % des renversements de préférences sont imputables à une violation de l'invariance procédurale (cas (2)-(3) avec une grande prépondérance du cas (2)) tandis que 10 % seulement le seraient à une violation de la transitivité (cas (1)).

• Si les préférences changent avec la méthode d'élicitation, (1) comment connaît-on les "vraies" préférences d'un agent ?, et (2) existe-t-il quelque chose comme les "vraies" préférences de l'agent ?

▷ D.Grether & C.Plott (1979), "Economic Theory of Choice and the Preference Reversal Phenomenon", *American Economic Review*

"[The preference reversal phenomenon] suggests that no optimization principles of any sort lie behind even the simplest of human choices..."

▷ A. Tversky & R. Thaler (1991), "Anomalies : Preference Reversals", *The Journal of Economic Perspectives*, vol.4, n°2, p.210

" The discussion of the meaning of preference and the status of value may be illuminated by the well-known exchange among three baseball umpires. "I call them as I see them" said the first. "I call them as they are" said the second. The third disagreed, "They ain't nothing till I call them". Analogously, we can describe three different views regarding the nature of values. First, values exist - like body temperature - and people perceive and report them as best they can, possibly with bias (I call them as I see them). Second, people know their values and preferences directly - as they know the multiplication table (I call them as they are). Third values or preferences are commonly constructed in the process of elicitation (they ain't nothing till I call them). The research reviewed in this article is most compatible with the third view of preference as constructive, context-dependent process."

6 Références

6.1 sur la théorie des préférences et de l'utilité

• D. Kreps (1988), *Notes on the Theory of Choice*, Westview Presschap.3

• Resnik, M. *Choices. An Introduction to Decision Theory*, University of Minnesota Press [chap.2]

6.2 sur la justification des conditions de rationalité

• P. Anand (1987), "Are the Preference Axioms Really Rational ?", *Theory and Decision*, 23 :2, p.189

• P. Anand (1993), "The Philosophy of Intransitive Preferences", *The Economic Journal*, 103 :417, pp. 337-46

J. Broome (1991), *Wheighing Goods*, Blackwell [Chap. 5 et en particulier section 5.4]

• Davidson, D., McKinsey, J.C.C. & Suppes, P. (1955) "Outlines of a Formal Theory of Value", *Philosophy of Science*, vol. 22 (2), pp. 140-60 [l'article qui introduit l'argument de la "pompe à finance"]

• P. Fishburn (1991), "Nontransitive Preferences in Decision Theory", *Journal of Risk and Uncertainty*, 4, pp. 113-34

• Ph. Mongin (2000), "Does Optimization Implies Rationality ?", *Synthese*, 124, pp. 73-111

• F. Schick (1986), "Dutch Bookies and Money Pumps", *The Journal of Philosophy*, vol.83, pp. 112-9

6.3 Sur le renversement des préférences

• D.Grether & C.Plott (1979), "Economic Theory of Choice and the Preference Reversal Phenomenon", *American Economic Review*

• Lichtenstein & Slovic (1971), "Reversals of Preference Between Bids and Choices in Gambling Decisions", *Journal of Experimental Psychology*, vol.89, pp. 46-55.

• Lichtenstein & Slovic (2006), *Contruction of Preferences*, Cambridge UP

• A. Tversky & R. Thaler (1991), "Anomalies : Preference Reversals", *The Journal of Economic Perspectives*, vol.4, n°2, pp. 201-11