Une science de l'action est-elle possible ?

Mikaël Cozic

IHPST (Paris I-ENS Ulm-CNRS)
GREGHEC (HEC-CNRS), Logiques de l'Agir (Besançon) & DEC (ENS Ulm)

Besançon - Séminaire SIPS
5/XII/2007

introduction

- objet : la **théorie de la décision** ou **théorie du choix rationnel**
- qu'est-ce que la théorie de la décision ? théorie mathématisée développée conjointement par des philosophes (Ramsey, Jeffrey) et des économistes (Von Neumann & Morgenstern, Savage)

introduction

- objet : la **théorie de la décision** ou **théorie du choix rationnel**
- qu'est-ce que la théorie de la décision ? théorie mathématisée développée conjointement par des philosophes (Ramsey, Jeffrey) et des économistes (Von Neumann & Morgenstern, Savage)
- la théorie de la décision est utilisée à la fois comme théorie normative et comme théorie positive de l'action
- la théorie de la décision constitue le socle fondamental de l'économie contemporaine ; elle trouve également de nombreuses applications en sciences politiques

introduction

- objet de l'exposé : la théorie de la décision comme science de l'action
- plusieurs modèles de scientificité :
  - sciences *a priori* (logique et mathématiques)
  - sciences normatives (sciences juridiques)
  - sciences de la nature (sciences physiques, sciences de la vie)

introduction

- objet de l'exposé : la théorie de la décision comme science de l'action
- plusieurs modèles de scientificité :
  - sciences *a priori* (logique et mathématiques)
  - sciences normatives (sciences juridiques)
  - sciences de la nature (sciences physiques, sciences de la vie)
- c'est le rapprochement avec les sciences de la nature qui m'intéresse et plus précisément la théorie de la décision comme science *positive et empirique* de l'action

introduction

- naturalisme = la théorie de la décision se laisse concevoir comme une science positive et empirique de l'action
- naturalisme $\neq$ la théorie de la décision se laisse *réduire* aux sciences naturelles (en particulier aux neurosciences)

introduction

- naturalisme = la théorie de la décision se laisse concevoir comme une science positive et empirique de l'action
- naturalisme $\neq$ la théorie de la décision se laisse *réduire* aux sciences naturelles (en particulier aux neurosciences)
- quels critères de scientificité ?
  - précision
  - contenu nomologique
  - pouvoir explicatif
  - **contenu empirique** (testabilité)

introduction

- thèse naturaliste : la théorie de la décision peut être envisagée de manière cohérente comme une théorie positive et empirique

(i) la théorie de la décision est pourvue de contenu empirique

(ii) la théorie de la décision classique souffre d'anomalies empiriques qui sont de mieux en mieux connues

(iii) rien n'interdit *a priori* que la théorie de la décision puisse être substantiellement améliorée

plan

- Plan de l'exposé :

(1) éléménts de théorie de la décision

(2) les causes et les raisons

(3) naturalisme en théorie de la décision

1. La théorie de la décision

l'action rationnelle

- le choix d'une action par un agent est rationnel si, étant donné ce qu'il **croit** et étant donné **ce qu'il peut choisir**, l'action qu'il choisit est celle dont les **conséquences** satisfont le mieux ses **désirs**

l'action rationnelle

- le choix d'une action par un agent est rationnel si, étant donné ce qu'il **croit** et étant donné **ce qu'il peut choisir**, l'action qu'il choisit est celle dont les **conséquences** satisfont le mieux ses **désirs**

(i) la rationalité dépend des **opportunités** offertes à l'agent: supposons que Pierre préfère le vin jaune au vin blanc et le vin blanc au vin rouge

- situation 1:

a_b	vin blanc
a_r	vin rouge

Pierre doit choisir a_b

l'action rationnelle

- situation 2:

a_b	vin blanc
a_r	vin rouge
a_j	vin jaune

Pierre doit cette fois choisir a_j (et non a_b)

l'action rationnelle

- situation 2:

a_b	vin blanc
a_r	vin rouge
a_j	vin jaune

Pierre doit cette fois choisir a_j (et non a_b)

(ii) la rationalité dépend des **désirs** de l'agent
Pierre préfère le vin blanc ou vin rouge ; Jean préfère le vin rouge au vin blanc

a_b	vin blanc
a_r	vin rouge

Pierre doit choisir a_b mais Jean doit choisir a_r

l'action rationnelle

(iii) la rationalité dépend des **croyances** de l'agent:

- Jean croit que s'il choisit l'action a_b, il obtiendra du vin rouge tandis que s'il choisit l'action a_r, il obtiendra du vin blanc
- les goûts de Jean sont les mêmes que ceux de Pierre

a_b	vin rouge
a_r	vin blanc

Jean doit choisir a_r (et non a_b)

l'action rationnelle

- la rationalité d'un choix est une relation entre (1) les opportunités, les désirs et les croyances de l'agent et (2) l'action choisie
- on dit parfois que la théorie de la décision exprime une conception **instrumentale** de la rationalité ou encore qu'elle traite de la rationalité des moyens et non de celle des fins
- *prima facie* la théorie de la décision prend les désirs de l'agent comme donnés, elle ne dit pas quels sont les "bons" désirs et quels sont les "mauvais"
- mais elle impose des conditions de *rationalité* ou de *cohérence* sur ces désirs

les désirs

- les désirs de l'agent portent sur les conséquences possibles de ses actions; l'ensemble des actions réalisables est noté A, l'ensemble des conséquences C
- comment la théorie de la décision représente-t-elle les désirs?

(i) préférences: relation comparative entre conséquences

- $c_i \succeq c_j$:

Pierre préfère "largement" c_i à c_j
Pierre estime que c_i est au moins aussi bonne que c_j

- $c_i \succ c_j$:

Pierre préfère "strictement" c_i à c_j
Pierre estime que c_i est strictement meilleure que c_j

- $c_i \sim c_j$:

Pierre est indifférent entre c_i et c_j
Pierre estime que c_i et c_j ont la même valeur

les désirs

(ii) utilité : fonction numérique $u : C \longrightarrow \mathbb{R}$

$$u(\text{vin jaune}) = 5$$
$$u(\text{vin blanc}) = 3$$
$$u(\text{vin rouge}) = 1$$

les désirs

(ii) utilité : fonction numérique $u : C \rightarrow \mathbb{R}$

$$u(\text{vin jaune}) = 5$$
$$u(\text{vin blanc}) = 3$$
$$u(\text{vin rouge}) = 1$$

- conditions de *rationalité* sur les désirs:

(i) **transitivité:** si Pierre préfère le vin jaune au vin blanc et s'il préfère le vin blanc au vin rouge, alors il préfère le vin jaune au vin rouge

$$\text{si } x \succeq y \text{ et } y \succeq z \text{ alors } x \succeq z$$

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

$$c_i \succ c_j$$
$$c_j \succ c_k$$
$$c_k \succ c_i$$

les désirs

(ii) **complétude:** Pierre est capable de comparer toute paire de conséquences possibles de ses actions :

$$x \succeq y \text{ ou } y \succeq x$$

les désirs

(ii) **complétude:** Pierre est capable de comparer toute paire de conséquences possibles de ses actions :

$$x \succeq y \text{ ou } y \succeq x$$

- la représentation par une fonction d'utilité induit automatiquement transitivité et complétude
- si l'ensemble de conséquences est fini, toute relation de préférence rationnelle peut être représentée par une fonction d'utilité

choix certain

- situations de certitude: Pierre sait précisément pour chaque action a quelle est la conséquence c qui est produite par a
- on note $a(C)$ la conséquence de a

choix certain

- situations de certitude: Pierre sait précisément pour chaque action a quelle est la conséquence c qui est produite par a
- on note $a(C)$ la conséquence de a
- règle de décision [préférences]: choisir l'action $a \in A$ dont la conséquence est préférée aux conséquences des autres actions (s'il en existe une!)

$$\text{pour tout } a', a(C) \succeq a'(C)$$

- règle de décision [utilité]: choisir l'action a dont l'utilité de la conséquence est supérieure à celle des autres actions

$$\text{pour tout } a', u(a(C)) \geq u(a'(C))$$

choix incertain

- situations d'**incertitude**: Pierre ne sait pas quelle est la conséquence d'une action réalisable $a \in A$
- situations de **risque**: Pierre sait pour toute action a et pour toute conséquence c avec quelle probabilité l'action a produit la conséquence c
- exemple: choix entre deux **paris** ou **loteries**

Comment choisir entre a_1 et a_2? La règle de choix en certitude ne s'applique plus: $5 \prec 10$ mais $3 \succ 2$

- comment déterminer la valeur d'une action risquée ?
  comment choisir entre différentes actions risquées ?

- comment déterminer la valeur d'une action risquée ? comment choisir entre différentes actions risquées ?
- idée fondamentale: pondérer les différents gains possibles par la probabilité qu'ils arrivent

  - *Logique de Port-Royal*:

    "...pour juger de ce que l'on doit faire pour obtenir un bien ou pour éviter un mal, il ne faut pas seulement considérer le bien ou le mal en soi, mais aussi la probabilité qu'il arrive ou n'arrive pas, et regarder géométriquement la proportion que toutes ces choses ont ensembles"

  cf. Pascal et le problèmes des points du Chevalier de Méré

l'espérance de gain

- 

$$EG(a_1) = (0.6 \times 5) + (0.4 \times 3) = 4.2 \text{ ;}$$
$$EG(a_2) = (0.1 \times 10) + (0.9 \times 2) = 2.8$$

l'espérance de gain

- 

$EG(a_1) = (0.6 \times 5) + (0.4 \times 3) = 4.2$;
$EG(a_2) = (0.1 \times 10) + (0.9 \times 2) = 2.8$

- formule générale
  supposons que l'action a ait pour conséquences possibles $c_1, ..., c_m$, chaque conséquence c_j survenant avec une probabilité p_j^a

$$EG(a) = \sum_{j=1}^{m} p_j^a(c_j)$$

- règle de décision: choisir l'action dont l'espérance de gain est maximum

St-Petersbourg

- une pièce non-biaisée est lancée de manière répétée jusqu'à ce qu'elle tombe sur face (F)

série	F	PF	PPF	...	P...PF	...
proba.	1/2	1/4	1/8	...	$1/2^n$	...
gain	2	4	8	...	2^n	...

- quelle est la valeur du pari de St-Petersbourg?
  $EG(StP) = 1 + 1 + 1 + ... = \infty$

espérance d'utilité

- proposition de Daniel Bernouilli (1738): découpler le gain monétaire et la satisfaction que l'agent en retire ou l'utilité
- on note $u(c_j)$ l'utilité que l'agent retire du gain monétaire c_j

$$EU(a) = \sum_{j=1}^{m} p_j^a \times u(c_j)$$

espérance d'utilité

- proposition de Daniel Bernouilli (1738): découpler le gain monétaire et la satisfaction que l'agent en retire ou l'utilité
- on note $u(c_j)$ l'utilité que l'agent retire du gain monétaire c_j

$$EU(a) = \sum_{j=1}^{m} p_j^a \times u(c_j)$$

- règle de décision: choisir l'action dont l'espérance d'utilité est maximum

espérance d'utilité

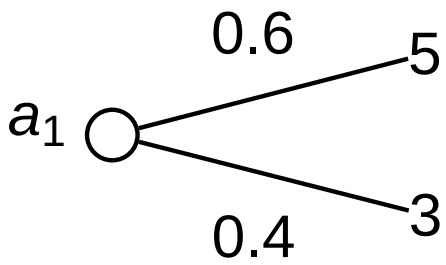
- proposition de Daniel Bernouilli (1738): découpler le gain monétaire et la satisfaction que l'agent en retire ou l'utilité
- on note $u(c_j)$ l'utilité que l'agent retire du gain monétaire c_j

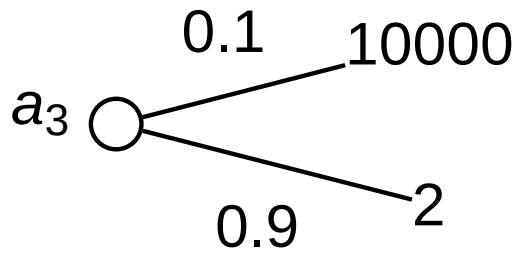
$$EU(a) = \sum_{j=1}^{m} p_j^a \times u(c_j)$$

- règle de décision: choisir l'action dont l'espérance d'utilité est maximum
- l'argent a une utilité marginale décroissante
- $u(x) = log(x)$
  Dans ce cas, l'espérance d'utilité $EU(StP) \simeq 0.6$ et la valeur monétaire du Pari est (environ) 4 euros

remarques

- rem 1: on peut construire un super-paradoxe de St-Petersburg pour l'utilité (subjective): si la pièce tombe sur F au n-ème lancer, l'agent reçoit l'équivalent de 2^n "utiles"
- rem 2 : la notion d'utilité change de signification par rapport à celle qui servait simplement à représenter les préférences; elle reflète désormais l'*intensité* des désirs





• supposons que les conséquences soient évaluées en "utiles"; alors cette fois Pierre doit choisir a_3 et non a_1

théorème de représentation

- on connaît un ensemble d'axiomes AX sur les préférences entre actions (loteries) tels que
  • **si** les préférences d'un agent satisfont AX,
  • **alors** une fonction d'utilité $u(.)$ (dite utilité von Neumann -Morgenstern) sur les conséquences telle que

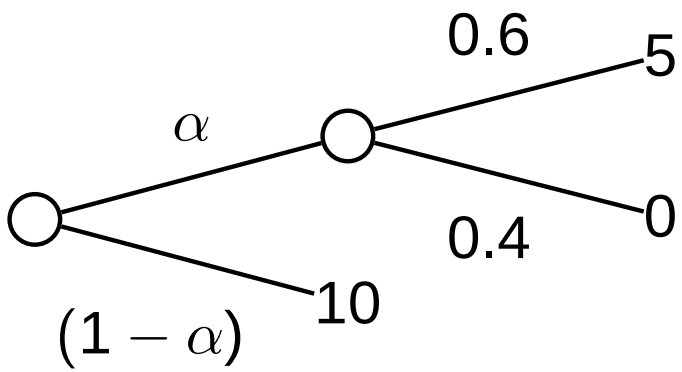
$$a_i \succ a_j \text{ ssi } EU(a_i) > EU(a_j)$$

l'axiome d'indépendance

- loterie composée : une loterie faite de loteries :

$$\alpha L \oplus (1-\alpha)L'$$

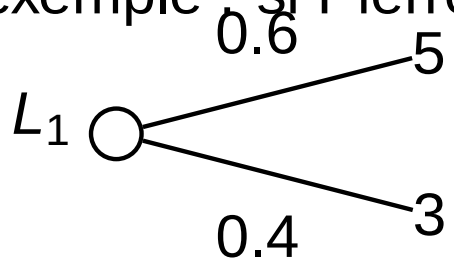
signifie : avec probabilité α, l'agent joue la loterie L et avec probabilité $(1-\alpha)$, L'.

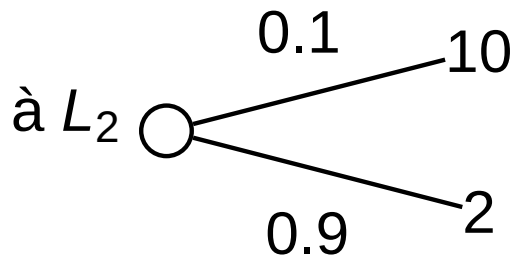


- axiome d'indépendance : les préférences entre deux loteries ne changent pas quand on compose chaque loterie avec une même troisième dans les mêmes proportions

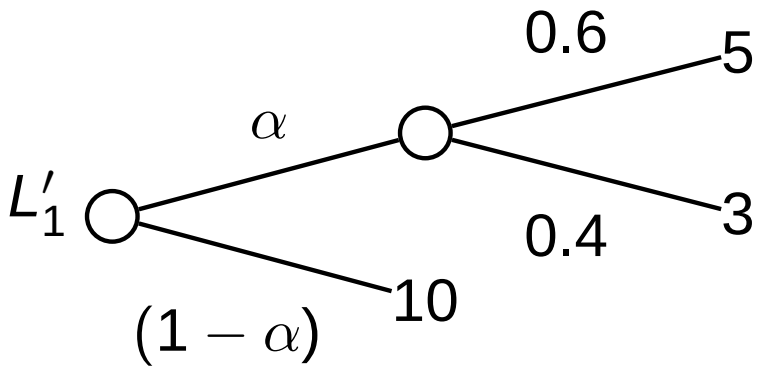
axiome d'indépendance, suite

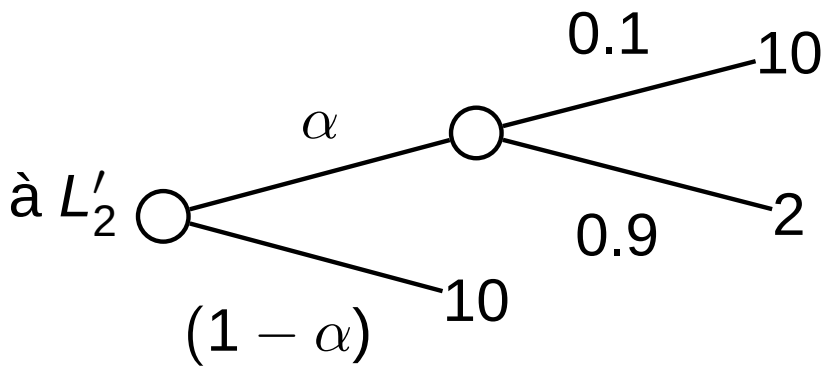
- exemple : si Pierre préfère





alors il préfère également





mesure de l'utilité VNM

- le théorème de représentation offre la possibilité de *mesurer* l'utilité vis-à-vis des différentes conséquences:
  • soient 3 conséquences : j, r, b classées ainsi :
  $j \succ b \succ r$
  • fixons conventionnellement $u(j) = 1$ et $u(r) = 0$.
  • comment mesurer $u(b)$? On utilise les loteries :
  $(j, 3/4; r, 1/4) \succ b$
  $(j, 1/4; r, 3/4) \prec b$
  dans ce cas, $3/4 \succ u(b) \succ 1/4$
  $u(b) = \alpha$ où $b \sim (j, \alpha; r, (1 - \alpha))$
  l'utilité obtenue est cardinale = unique à une transformation affine près - cf température

l'incertitude *s.s.*

- situations d'incertitude *stricto sensu*: une probabilité "objective" n'est pas donnée aux agents
- **bayésianisme:** les degrés de croyance d'un agent obéissent aux lois des probabilités
- chaque agent dispose d'une distribution de probabilité sur l'ensemble des paramètres qui peuvent affecter sa décision
- exemple : Pierre est invité, il hésite entre acheter une bouteille de vin rouge et une bouteille de vin blanc. La conséquence de son choix dépend de ce que son hôte a choisi de faire (viande ou poisson)
- on parle de **probabilité subjective** ; elles peuvent varier d'un individu à l'autre

l'ESU

- le cadre général du modèle d'espérance subjective d'utilité:

(i) ontologie:
  - un ensemble d'états de la nature S
  - un ensemble d'actions réalisables A
  - un ensemble de conséquence C ; chaque action a a une conséquence déterminée c étant donné un état de la nature s

(ii) déterminants subjectifs:
  - une fonction de probabilité P sur S (les croyances partielles de l'agent)
  - une fonction d'utilité u sur C (les désirs de l'agent)

- L. Savage, *The Foundations of Statistics* (1954/1972)

l'ESU

- un problème de décision:

	s_1	s_2	...	s_m
a_1	c_{11}	c_{12}	...	c_{1m}
a_2	c_{21}	c_{21}	...	c_{2m}
...	...	...	...	....
a_n	c_{n1}	c_{n2}	...	c_{nm}

- règle de décision: choisir l'action a dont l'espérance subjective d'utilité est maximale

$$ESU(a_i) = P(s_1) \times u(c_{i1}) + ... + P(s_m) \times u(c_{im})$$

$$ESU(a_i) = \sum_j P(s_j) \times u(c_{ij})$$

	poisson 1/4	viande 3/4
vin rouge (VR)	0	4
vin blanc (VB)	4	2

$$ESU(VR) = (1/4 \times 0) + (3/4 \times 4) = 3$$
$$ESU(VB) = (1/4 \times 4) + (3/4 \times 2) = 5/2$$

théorème de représentation

- Savage (1954) propose un ensemble d'axiomes AX' sur les préférences entre actes tels que
  • **si** les préférences d'un agent satisfont AX',
  • **alors** il existe une unique distribution de probabilité P sur les états de la nature et une fonction d'utilité $u(.)$ sur les conséquences tels que

$$a_i \succ a_j \text{ ssi } ESU(a_i) > ESU(a_j)$$

l'axiome de la chose sûre

- axiome de la chose sûre : une modification commune de la partie commune de deux actes ne change pas les préférences entre ces actes
- exemple :
  a_1 = [1000 euros si Prince de Bretagne gagne la 5è, une bière sinon]
  a_2 = [1000 euros si Prince de Bretagne gagne la 5è, un pastis sinon]
  a'_1 = [100 euros si Prince de Bretagne gagne la 5è, une bière sinon]
  a'_2 = [100 euros si Prince de Bretagne gagne la 5è, un pastis sinon]

la mesure des états mentaux

- le modèle d'ESU autorise une variation explicite des croyances (proba. subj.) *et* des désirs (utilité)
- le problème de la mesure devient beaucoup plus complexe : dans le cas du modèle d'EU, les croyances étaient fixées (les préférences aussi) et c'est ce qui permettait la mesure des utilités
- comment faire quand les deux familles d'états mentaux sont inconnues ? comment faire pour séparer proba. et utilités ?

la méthode de Ramsey

- Ramsey, “Truth and Probability” (1926) ; part de la notion de *proposition éthiquement neutre* (deux mondes possibles ne différent que par une telle proposition sont d’égales valeurs)
- E proposition éthiquement neutre de probabilité 1/2 : pour toute paire de conséquences c_i, c_j, l’agent est indifférent entre les actes suivants :
  $[c_i \text{ si } E, c_j \text{ sinon}] \sim [c_i \text{ si } E, c_j \text{ sinon}]$

(1) mesure des utilités
$EU([c_1 \text{ si } E, c_2 \text{ sinon}]) = P(E) \times u(c_1) + P(E) \times u(c_2)$
$EU([c_3 \text{ si } E, c_4 \text{ sinon}]) = P(E) \times u(c_3) + P(E) \times u(c_4)$
$[c_1 \text{ si } E, c_2 \text{ sinon}] \sim [c_3 \text{ si } E, c_4 \text{ sinon}]$
ssi $u(c_1) - u(c_3) = u(c_4) - u(c_2)$
la différence d’utilité entre c_1 et c_3 est égale à celle entre c_4 et c_3.

la méthode de Ramsey

(2) mesure des croyances

- Si l'on dispose de $u(c_1)$, $u(c_2)$, et $u(c_3)$, si l'agent maximise sont espérance d'utilité, alors
  $EU([c_1 \text{ avec certitude}]) = u(c_1)$
  $EU([c_2 \text{ si } E, c_3 \text{ sinon}]) = P(E) \times u(c_2) + 1 - P(E) \times u(c_3)$
- Si Pierre est indifférent entre [c_1 avec certitude] et [c_2 si E, c_3 sinon], alors
  $u(c_1) = P(E) \times u(c_2) + 1 - P(E) \times u(c_3)$
  $P(E) = (u(c_1) - u(c_2))/(u(c_1) - u(c_3))$
- on peut vérifier que $P(.)$ ainsi défini obéit aux lois des probabilités

remarques

- le terme “bayésianisme” est très souvent utilisé dans les théories formelles de la rationalité; il y a en fait trois sens distincts:

(i) les degrés de croyances d’un agent rationnel se laissent représenter par une fonction de probabilité
• justifications: *Dutch Book* (Ramsey, de Finetti) ou axiomes sur les préférences (Savage)

(ii) quand un agent rationnel apprend une information I, il modifie ses croyances en suivant la règle de conditionalisation

$$P(E/I) = P(E \wedge I)/P(E)$$

(iii) un agent rationnel choisit l’action qui maximise son espérance subjective d’utilité

2. Les causes et les raisons

la philosophie de l'action

- les philosophes contemporains se sont parfois invertis dans le développement normatif de la théorie de la décision
  exemple : débat théorie causale vs. théorie évidentielle de la décision
- mais les rôles explicatif et descriptif de la théorie de la décision ont reçu assez peu d'attention

la philosophie de l'action

- les philosophes contemporains se sont parfois invertis dans le développement normatif de la théorie de la décision
  exemple : débat théorie causale vs. théorie évidentielle de la décision
- mais les rôles explicatif et descriptif de la théorie de la décision ont reçu assez peu d'attention
- ce qui a en revanche reçu une attention considérable, c'est la clarification du statut de nos **explications communes** ("naïves") **de l'action**
- la clarification de l'explication de l'action est l'un des objectifs fondamentaux de la **philosophie de l'action**

philosophie de l'action

- l'une des figures centrales de la philosophie de l'action contemporaine est Donald Davidson
- Davidson a été l'un des pionniers de la théorie expérimentale de la décision (Davidson, Suppes & Siegel, 1957) - travaux directement inspirés de ceux de Ramsey
- Davidson est l'un des seuls philosophes contemporains à prendre des positions fortes sur le statut de la théorie de la décision, positions qui prolongent celles qu'ils développent en philosophie de l'action

l'explication commune de l'action

- Q : "Pourquoi Pierre entre dans la cuisine ?"
  - "Parce qu'il a soif."
  - "Parce qu'il a l'intention de prendre une bière dans le réfrigirateur."
  - "Parce qu'il aime boire une bière lorsqu'il a soif."
  - "Parce qu'il croit qu'il y a des bières dans le réfrigirateur."
  - "Parce qu'il a soif, qu'il aime boire une bière lorsqu'il a soif et qu'il croit qu'il y a des bières dans le réfrigirateur."
- de manière générale, une action est typiquement expliquée par un ou plusieurs états mentaux (désirs, croyances, intentions...) = complexe d'états mentaux

explication commune de l'action

- Davidson ("Actions, Reasons and Causes", 1963) parle de **raison primaire** :
  "*R* n'est une raison primaire pour laquelle un agent a accompli l'action *A* sous la description *d* que si *R* consiste en une pro-attitude de l'agent à l'égard d'actions qui ont une certaine propriété et en la croyance de l'agent que *A*, sous la description *d*, a cette propriété"
- expliquer une action, c'est en donner une (ébauche de) raison primaire
- la raison primaire d'une action fournit une justification pour cette action, une raison d'entreprendre cette action
- une explication de l'action qui invoque une raison primaire ou un complexe mental de ce type est une **explication intentionnelle**

les causes et les raisons

- question : est-ce qu'affirmer

  "Pierre entre dans la cuisine *parce qu'*il croit qu'il y a des bières dans le réfrigirateur "

  est analogue à affirmer, par ex.,

  "La fenêtre s'est brisée *parce qu'*elle a été frappée par une pierre" ?

- autrement dit : est-ce que les explications communes de l'action sont des explications causales ordinaires ?
  Les **causalistes** répondent positivement à cette question, les **anti-causalistes** négativement.

anti-causalisme

- la psychologie du sens commun nous fait sans doute naturellement pencher vers le causalisme
- en s'inspirant de certaines remarques de Wittgenstein, un courant de pensée anti-causaliste a eu une influence très importante dans la philosophie anglo-saxonne des années 1950-60
- ces anti-causalistes ont ravivé des débats plus anciens sur la méthodologie des sciences humaines et sociales et en particulier sur la comparaison entre "la" méthode des sciences naturelles et "la" méthode des sciences humaines et sociales
- l'idée centrale est qu'une explication intentionnelle permet d'interpréter une action et non pas de lui assigner une cause

interprétativisme

- une explication de l'action en tant qu'action est intentionnelle
- une explication intentionnelle ne peut être une explication causale
- une explication intentionnelle fournit une inteprétation de l'action dont l'objectif est de faire comprendre l'action
- il ne faut pas chercher à ériger une théorie causale "scientifique" de l'action à partir de nos explications communes

l'argument de la connexion logique

- l'argument principal des anti-causalistes contemporains est appelé dans la littérature l'**argument de la connexion logique** (ACL)
- en réalité, il existe de nombreuses variantes de cet argument (sémantique, modale, épistémique)

l'argument de la connexion logique

- l'argument principal des anti-causalistes contemporains est appelé dans la littérature l'**argument de la connexion logique** (ACL)
- en réalité, il existe de nombreuses variantes de cet argument (sémantique, modale, épistémique)
- l'idée centrale est la suivante : la relation entre les raisons R d'une action a et l'action a est conceptuelle ou logique qui fait que R et a ne sont pas concevables indépendamment l'un de l'autre :
  - si a avait pour raison R' et non R, ce ne pourrait être a
  - si R expliquait a et non a', ce ne pourrait être R
- l'hypothèse auxiliaire fondamentale de l'argument est l'hypothèse humienne selon laquelle la relation de causalité est contingente et *a posteriori*

l'argument de la connexion logique

- exemple :
  R = Pierre a l'intention de mettre *Kind Of Blue* dans le lecteur CD
  a = Pierre met *Kind Of Blue* dans le lecteur CD
  Application : si la suite de mouvements que l'on considère comme étant *a* n'était pas supposée procéder de *C* ie de l'intention de mettre *Kind Of Blue* dans le lecteur CD, alors on ne la considèrerait plus comme l'action : Pierre met *Kind of Blue* dans le lecteur CD.
- conséquences :
  • *a* ne se conçoit pas réellement sans *R* : s'il s'était agi de C' plutôt que de *C*, alors on aurait eu a' et non pas *a*
  • *R* n'est pas réellement la cause de *a*
  • *R* et *a* (l'action et l'antécédent mental) entretiennent une relation non pas causale mais conceptuelle

réponses à ACL

(1) réponse 1 : la conception interprétativiste ne permet pas de rendre compte de la différence entre :
- donner **un** complexe de raisons vs. donner **le** complexe de raisons
- avoir des raisons vs. agir pour certaines raisons

- "...une personne peut avoir une raison de faire une action, et accomplir cette action bien que la raison ne soit pas la raison pour laquelle elle a accompli l'action. Il y a une idée qui est indissociable de la relation entre une action et la raison qui l'explique : c'est l'idée que l'agent a accompli l'action *parce qu'*il avait une certaine raison."

réponses à l'ACL

(2) réponse 2
- distinguer les événéments et leurs descriptions
- on peut de manière générale redécrire les événements en termes de leurs causes

exemple 1 : un coup de soleil est causé par le soleil

- or, "décrire un événement en termes de sa cause, ce n'est pas confondre l'événement avec sa cause, pas plus qu'expliquer un événement en le redécrivant n'exclut qu'on fournisse par là une explication causale." (Davidson, 1963)

causalisme

- réponses (1) + (2) $\Rightarrow$ il ne faut pas conlure du fait que l'explication par les raisons donne une justification qu'elle ne peut pas être causale
- "une rationalisation est une forme d'explication causale ordinaire"

3. Naturalisme en théorie de la décision

théorie de la décision et explication commune

- la théorie de la décision est la “psychologie du sens commun formalisée” (Rosenberg, 1988)
- D. Lewis : “La théorie de la décision est une exposition systématique des conséquences de certaines platitudes bien choisies concernant la croyance, le désir, la préférence et le choix. C’est le noyau de notre théorie commune de la personne, disséquée, et systématisée avec élégance.”

théorie de la décision et explication commune

- la théorie de la décision est la “psychologie du sens commun formalisée” (Rosenberg, 1988)
- D. Lewis : “La théorie de la décision est une exposition systématique des conséquences de certaines platitudes bien choisies concernant la croyance, le désir, la préférence et le choix. C’est le noyau de notre théorie commune de la personne, disséquée, et systématisée avec élégance.”
- *prima facie* si l’on transpose à la théorie de la décision l’analyse causaliste des explications communes de l’action, il semble qu’on fait un pas vers une conception naturaliste de la théorie de la décision
- dans cette perspective, la théorie de la décision semble expliciter (et éventuellement améliorer) les principes qui font passer des raisons aux actions

exemple

- exemple :

(1) antécédents mentaux R :
- Pierre voulait v
- Pierre croyait que faire a est un moyen de faire en sorte qu'il advienne v
- Pierre n'avait pas conscience d'actions a' préférables à a et qui permettaient également de faire en sorte que v advienne
- Pierre n'avait aucun autre désir que celui que v advienne
- Pierre savait comment faire a

exemple, suite

(2) principe d'action :
[P] Si quelqu'un veut v, croit que a est un moyen de faire que v advienne, n'a pas conscience d'actions a' préférables à a et qui permettent également que v advienne, n'a aucun autre désir que celui que v advienne et sait comment faire a, alors il fait a

mais...

- mais Davidson impose deux limites sévères à une conception anturaliste de la théorie de la décision :

(1) l'**anomalisme**

(2) l'**a priorisme*** ou scepticisme naturaliste

anomalisme

- anomalisme ou irréductibilité nomologique du psychologique : pas de lois psychologiques strictes - en particulier pas de lois strictes de l'action
- conséquence : les sciences sociales ne peuvent se développer comme les sciences physiques :
  - "...nous ne pouvons nous attendre à expliquer et prédire le comportement humain avec une précision du genre de celle qui est en principe possible dans le cas des phénomènes physiques..."
  - "...nous ne pouvons pas transformer ce mode d'explication en quelque chose qui aurait des allures plus scientifiques"

anomalisme

- on ne peut pas comparer
  “si un homme veut manger une omelette aux glands, alors en général il le fera si l’opportunité s’en présente et si aucun désir concurrent ne l’emporte sur celui-ci”
  ”un corps dans le vide tombe à la vitesse *x*”
- comment conciliser causalisme et anomalisme ?
  • les relations causales entre évéément sont indépendantes de leurs descriptions
  • les relations nomologiques elles dépendent de leur description
  • s’il y a relation causale, il existe une description des événements avec relation nomologique entre les événements ainsi décrits
- “Supposons qu’un ouragan, rapporté à la page 5 du *Times* de mardi, cause une catastrophe, rapportée page 13 de la

- pourquoi anomalisme ? En raison de la seconde “limitation”, l’**a priorisme**
- Davidson, “Representation and Interpretation”, 1990
  ”Si quelqu’un croit que Tahiti est à l’est d’Honolulu, alors il devrait croire qu’Honolulu est à l’ouest de Tahiti. C’est pourquoi, si nous sommes certains que la personne croit qu’Honolulu est à l’ouest de Tahiti, c’est probablement une erreur d’interpréter ce qu’elle dit comment exprimant qu’elle croit également que Tahiti est à l’ouest d’Honolulu. S’il s’agit probablement d’une erreur, *ce n’est pas parce que ce serait un fait empirique que les gens ont rarement des conceptions contradictoires*, mais parce que les croyances (et les autres attitudes) sont largement identifiées par les relations, notamment logiques, qu’elles entretiennent entre elles ; changer ces relations, c’est changer l’identité de la pensée.”

a priorisme

- le respect du principe de non-contradiction n'est pas une régularité associée à l'attribution d'une croyance mais une *contrainte conceptuelle* qui pèse sur l'attribution de croyances
  si Pierre croit que ϕ et que $\neg\phi$, alors en réalité il ne croit pas vraiment que ϕ
- les principes de rationalité ne sont pas empiriquement évaluables mais sont des contraintes *a priori* sur l'explication de l'action et l'attribution d'attitudes
- "la rationalité n'est pas un trait empirique que nous pourrions découvrir chez les agents et inclure à titre de prémisses dans nos explications, mais une norme fondamentale de l'intelligibilité de leurs comportements présupposés par toute tentative d'explication." (Engel, 1993)

a priorisme et theorie de la décision

- Davidson reconnaît que, par rapport à la la psychologie de sens commune, la théorie de la décision “fait un pas important en direction de la respectabilité scientifique”
- question : “Pouvons-nous accepter une telle théorie de la décision comme théorie scientifique du comportement et la placer sur le même plan qu’une théorie physique ? ”
- réponse :
  “the laws, so called, of decision theory...are not empirical generalizations about all agents. What they do is to define what is meant...by being rational” (“Problems in the Explanation of Action”, 1987)
  “que l’on soit ou non scientifique, quand on attribue des pensées à autrui, on emploie nécessairement nos propres normes dans l’attribution”

scepticisme et théorie de la décision

- quand il discute de la théorie de la décision, Davidson avance souvent des affirmations moins fortes qui tirent vers une forme de scepticisme par rapport à une conception naturaliste de la théorie de la décision
- l'idée semble être en substance que l'on peut toujours sauver les principes de la théorie de la décision en modifiant l'interprétation des axiomes
  • exemple : choix intransitif $a \succ b$, $b \succ c$ et $c \succ a$
  on peut considérer que l'option "a en présence de b" n'est pas la même que "a en présence de c"

argument de Rosenberg

- argument de Rosenberg
  supposons que l'on ait un principe du genre
  (P') Si croyance c et désir d, alors action a
- **explication** de l'action selon (P') :
  • Pierre entreprend l'action a
  • explication : Pierre a la croyance c et le désir d - d'après (P'), si c et d, alors a
  Question : comment peut-on savoir que c'est bien le complexe mental (c, d) et non pas (c', d') qui implique d'après (P") la même action ?
  Réponse : à partir des autres comportements de Pierre, et de ses comportements verbaux. On utilisera toujours (P') pour inférer les croyances et désirs des comportements.

argument de Rosenberg

- **prédiction** de l'action selon (P') :
  • on fait l'hypothèse que Pierre a la croyance *c* et le désir *d*
  • conformément à (P'), on déduit de l'hypothèse que Pierre entreprendra l'action *a*
  Question : comment peut-on savoir que Pierre a bien la croyance *c* et le désir *d* ?
  Réponse : en utilisant (P'), on infère *c* et *d* des comportements passés de Pierre, notamment de ses comportements verbaux

argument de Rosenberg

- **test** de (P') :
  • on fait l'hypothèse que Pierre a la croyance *c* et le désir *d*
  • conformément à (P'), on déduit de l'hypothèse que Pierre entreprendra l'action *a*
  • le test est en faveur de (P') si Pierre entreprend effectivement *a*, en sa défaveur sinon
  Question : comment peut-on savoir que Pierre a bien la croyance *c* et le désir *d* ?
  Réponse : en utilisant (P'), on infère *c* et *d* des comportements passés de Pierre, notamment de ses comportements verbaux

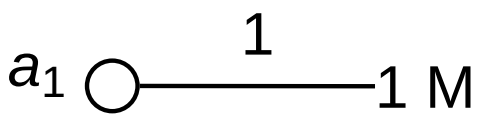
- ce que les remarques qui précèdent suggèrent, c'est qu'il y a une forme de circularité et/ou de régression dans l'usage que l'on fait d'un principe comme (P')
- cette circularité et/ou régression est le signe que (P') n'a que les apparences d'une généralisation empirique
- Rosenberg semble en tirer notamment la conclusion qu'un principe comme (P') n'est pas testable, dans la mesure où pour le tester on le présuppose

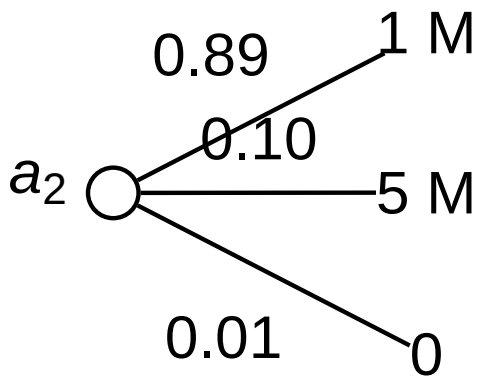
la théorie de la décision expérimentale

- les modèles d'espérance d'utilité sont testés !
- tests "horizontaux" (sur les préférences) vs. tests "verticaux" (états mentaux/préférences)
- non seulement les modèles d'espérance d'utilité sont testés, mais on a identifié des situations de choix où les sujets dévient de manière systématique des prédictions de la théorie de la décision

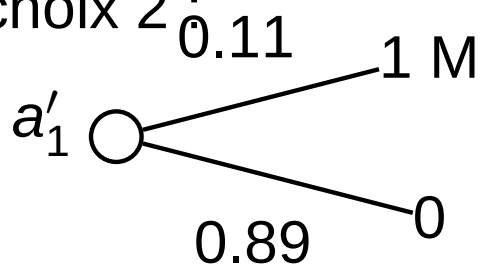
le paradoxe d'Allais

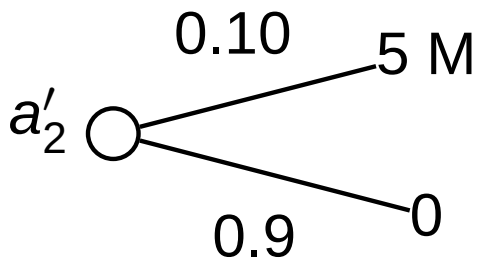
- choix 1 :

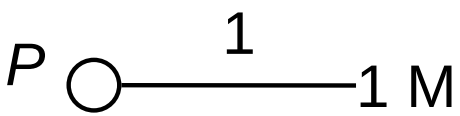


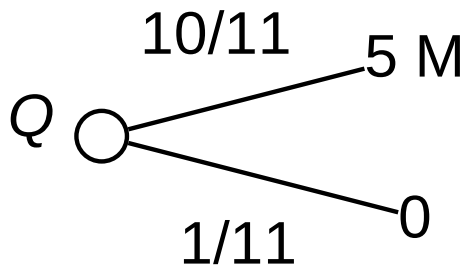


- choix 2 :









soit Φ_1 = gagner $1M$ avec probabilité 1 et Φ_0 = gagner 0 avec probabilité 1

- $a_1 = 0.11P \oplus 0.89\Phi_1$
- $a_2 = 0.11Q \oplus 0.89\Phi_1$
- $a'_1 = 0.11P \oplus 0.89\Phi_0$
- $a'_2 = 0.11Q \oplus 0.89\Phi_0$

le paradoxe d'Ellsberg

- une urne contient 90 boules, 30 rouges (R) et 60 bleues (B) ou jaunes (J)
- choix 1 : parier 100 E sur (R) ou parier 100 E sur (B)
- choix 2 : parier 100 E sur (R $\cup$ J) ou parier 100 E sur (B $\cup$ J)

naturalisme fort

- la ligne naturaliste “forte” soutient que les particularités épistémologiques de la théorie de la décision (et de l’explication commune de l’action) se laissent subsumer sous des facteurs parfaitement communs d’épistémologie des sciences empiriques :
  • l’inexactitude des lois de l’action reflète la complexité du sujet (cf. Mill et les lois *inexactes*)
  • le fait que l’attribution de désirs et de croyances “présuppose” que l’agent est rationnel reflète le fait que désirs et croyances sont des entités non observables introduites par la théorie*
  • le fait que la vérification du complexe mental invoqué pour expliquer une action nécessite l’usage de la théorie reflète le fait que nos explications de l’action sont marquées par une forte asymétrie entre les concepts

naturalisme modéré

- la ligne naturaliste modérée accepte l'idée qu'une certaine dose de rationalité est présupposée dans les attributions d'attitudes et l'explication de l'action, mais soutient (i) que cette dose est faible, ou (ii) que cette est en gros celle que l'on s'attribuerait volontiers (principe d'humanité)
- l'explication intentionnelle ne présuppose pas l'attribution d'une rationalité parfaite et en tous cas pas des principes qui sont codifiées par la théorie de la décision
  exemple : jouer aux échecs
- idée cruciale : il existe un espace important pour l'investigation empirique entre les principes de rationalité minimale et la théorie de la décision