

Décision, rationalité et interaction

Le pôle « Décision, rationalité et interaction » a pour objet les sciences de la décision. Ses activités de recherches se rassemblent en trois axes principaux :

1. La théorie de la décision individuelle
2. L'épistémologie des sciences de la décision
3. L'étude de la rationalité interactive

1 Théorie de la décision individuelle

Pour comprendre les êtres humains, aussi bien au niveau individuel qu'au niveau social, il est nécessaire de comprendre leurs actions. L'étude des décisions humaines, et des facteurs qui sont impliqués dans la décision, est non seulement importante du point de vue psychologique, ou dans la perspective de l'évaluation des actions humaines, mais elle occupe une position cruciale et fondatrice en sciences sociales. La théorie de la décision – discipline au croisement de l'économie, de la philosophie et de la psychologie cognitive – a pour but de décrire et de comprendre les rapports entre les décisions prises par les individus et les attitudes qui sont pertinentes pour ces décisions. Parmi les questions fondamentales de la théorie de la décision, on trouve les suivantes :

- comment devrait-on décider en fonction de ses attitudes ?
- comment décide-t-on effectivement sur la base de ses attitudes ?
- quelles attitudes sont (doivent être) pertinentes pour la décision ?
- qu'est-ce que c'est la croyance, le désir (ou éventuellement, certaines émotions, comme la tentation ou le regret) : peuvent-elles être définies par leur rôle dans la prise de décision ?
- à quel point est-il possible de révéler les attitudes d'un agent à partir des décisions prises ?
- quel est (devrait être) le rôle des facteurs comme le temps, la considération des situations contrefactuelles, la planification dans la décision ?

Les questions de ce genre se posent dans chacune des sortes de situations de décision que l'on distingue traditionnellement.

1.1 La décision en situation de choix certain

Dans ces situations, l'agent choisit entre un certain nombre d'options, avec la certitude d'obtenir l'option qu'il a choisit ; par exemple, il a le choix entre recevoir 100 EUR à coup sûr et recevoir 0 EUR à coup sûr. Dans de telles situations, on suppose traditionnellement que les utilités des objets de choix (les degrés auxquels il les désire) déterminent son choix ; les dites théories du choix propose de montrer les correspondances entre des propriétés des préférences entre les objets du choix et les utilités de l'agent.

1.2 La décision en situation de choix risqué

Dans ces situations, les différentes options offertes à l'agent peuvent conduire à différentes conséquences possibles selon des probabilités qui sont fixées objectivement (et supposées connues par l'agent). Un exemple d'une telle situation est le choix entre deux paris – un qui donne 100 EUR au cas où une pièce tombe sur face et 0 EUR si elle tombe sur pile, et l'autre qui donne 0 EUR si elle tombe sur face et 100 EUR si elle tombe sur pile – où la pièce n'est pas biaisée : la probabilité qu'elle tombe sur face et la probabilité qu'elle tombe sur pile valent un demi. Les théories de la décision pour le risque se proposent de révéler les liens entre les utilités des agents et leurs préférences pour les diverses options.

1.3 La décision en situation de choix incertain

Dans ces situations, les conséquences des choix sont incertaines (comme dans le cas de la décision risquée), mais à la différence de celle-ci, les probabilités d'atteindre telle conséquence par tel choix ne sont pas fixées objectivement. L'agent doit estimer ces probabilités sur la base de ses propres croyances ; d'ailleurs, on représente le plus souvent les croyances de l'agent dans une telle situation par une fonction de probabilité sur les états qui déterminent la conséquence qui résulte de chaque choix. Un exemple d'une telle situation est un pari sur une course de chevaux : il est encore possible de recevoir 100 EUR ou 0 EUR selon le résultat de la course, mais la probabilité qu'un cheval particulier gagne la course n'est pas donnée à l'avance – elle doit être estimée par l'agent qui l'utilise. Les théories de la décision en situation d'incertitude n'ont pas seulement affaire aux utilités de l'agent mais également à ses probabilités (croyances) : elles ont l'intention de comprendre les rapports entre les préférences de l'agent sur les options disponibles, ses utilités et ses probabilités.

2 Epistémologie des sciences de la décision

Le second axe de recherche du pôle « Décision, rationalité et interaction » porte sur l'épistémologie et la méthodologie des sciences de la décision et en particulier des sciences économiques.

2.1 Méthodologie de l'économie

Parmi les sciences sociales, l'économie est celle dont la construction théorique se fonde le plus sur les modèles formels de décision, modèles qui constituent l'objet du premier axe de recherches (modèles en situations certaines, risquées et incertaines). La méthodologie de l'économie est la branche de la philosophie des sciences dite « spéciales » dont l'objet est la science économique. Voici un échantillon de certaines des questions les plus discutées de la méthodologie de l'économie :

- Y a-t-il des lois économiques ?
- Quel est le statut des hypothèses fondamentales de l'économie ?
- Quel rôle joue l'hypothèse de rationalité dans la stratégie explicative de l'économie ?
- Quelle est la relation entre la macroéconomie et la microéconomie ?
- Quel est le rapport de la science économique à l'expérience ?

Voir <http://plato.stanford.edu/entries/economics/>

2.2 Rapports entre les sciences de la décision

Les sciences de la décision sont passablement hétérogènes : de la théorie de la décision « abstraite » à ses applications économiques, de la psychologie cognitive à la philosophie de l'action en passant par l'informatique théorique ou les neurosciences.

L'une des questions épistémologiques privilégiées du pôle « Décision, rationalité et interaction » est celle de savoir comment s'articulent ces différentes disciplines. Une attention toute particulière est accordée au *rapport entre psychologie et économie* et, de manière plus générale, entre sciences cognitives et sciences sociales : faut-il concevoir ces deux familles de sciences, quand elles abordent la décision, comme des approches séparées d'un même objet ? Si oui, sont-elles complémentaires ? Ou faut-il tenter de fonder, par exemple, les secondes sur les premières ? Peut-on attendre, par exemple, des neurosciences de la décision qu'elles réaménagent la théorie du choix rationnel toute entière et qu'elles améliorent sensiblement notre compréhension des phénomènes sociaux ?

2.3 Normativité et valeurs dans les sciences de la décision

Les sciences de la décision entretiennent un rapport étroit aux normes et aux valeurs. Ce rapport, toutefois, est loin d'être transparent. Le troisième thème de recherches en épistémologie des sciences de la décision est la clarification du rôle que la normativité et les valeurs jouent dans ce champ. La question est d'autant plus pressante que, dans les sciences économiques, il existe un domaine entier, l'*économie normative*, qui est consacré à l'évaluation des états et des institutions économiques. Les principales questions qui commandent ce thème de recherches sont les suivantes :

- Les sciences de la décision sont-elles neutres du point de vue axiologique ? Doivent-elles intégrer des conceptions éthiques ?
- Une discipline peut-elle intégrer des conceptions éthiques et rester scientifique ?
- Peut-il (et doit-il) exister une économie normative ?

Voir <http://plato.stanford.edu/entries/economic-justice/>

3. Rationalité interactive

Le troisième axe de recherches du pôle « Décision, rationalité, interaction » porte sur la caractérisation et les implications de la rationalité dans les groupes ou les sociétés.

3.1. Théorie des jeux

La théorie des jeux est l'étude des interactions stratégiques entre agents : elle s'intéresse aux situations de choix où les conséquences des actions des agents dépendent des actions entreprises par les autres agents. Elle constitue donc une extension de la théorie de la décision, dans laquelle la façon dont l'agent fait ou devrait faire son choix n'affecte pas la nature du problème. En théorie des jeux au contraire, l'anticipation par chacun des choix d'autrui fait partie intégrante de l'ensemble des paramètres pertinents pour son propre choix.

La théorie des jeux fournit ainsi un langage pour la description des interactions stratégiques, fondé avant tout sur des concepts d'équilibre, qui constituent l'« issue » du jeu selon la théorie. Dans un équilibre, les choix des différents agents impliqués forment ainsi un ensemble stable : ils constituent chacun simultanément la meilleure réponse possible à tous les autres (en fonction de l'information disponible dans une situation donnée) et se justifient donc mutuellement. A partir de ce principe fondamental, les divers types de jeux existants ont mené à autant de concepts d'équilibre : équilibre de Nash pour les jeux sous forme normale, équilibre en sous-jeu parfait pour les jeux sous forme extensive, équilibres bayésiens pour les jeux à information incomplète, etc.

Outre son succès au niveau des modèles, la notion d'équilibre a l'avantage de fixer précisément le niveau de rationalité nécessaire pour faire les « bons » choix en situation d'interaction. Elle ne résout pas pour autant tous les problèmes. D'abord, lorsqu'un jeu contient plusieurs équilibres de même type, la théorie ne peut en général pas en conseiller l'un plutôt que l'autre (problème de sélection des équilibres). Ensuite, le niveau de rationalité nécessaire pour toujours jouer selon un équilibre est considérable. Le principe fondamental de maximisation de l'utilité espérée doit être complété, dans un cadre dynamique, par des règles de détermination de l'information pertinente, où se mêlent estimation des coups futurs d'autrui et prise en compte signaux envoyés par les coups passés. Entrent de plus en jeu les connaissances et croyances des agents sur la rationalité d'autrui.

3.2. Théorie du choix social

3.2.1. Le cadre classique

A l'origine, la théorie du choix social ou théorie du choix collectif étudie l'agrégation des préférences individuelles en une relation de préférence collective. C'est l'une des branches principales de l'économie normative contemporaine. Son formalisme ainsi que son résultat fondateur remontent aux travaux pionniers d'Arrow (1951). Le théorème d'impossibilité d'Arrow affirme qu'il est impossible de concilier un petit groupe de réquisits particulièrement naturels sur les propriétés que doivent satisfaire l'agrégation des préférences.

3.2.2. Agrégation des jugements

Le problème de l'agrégation des jugements concerne la question de savoir selon quels principes il est possible d'agréger les jugements de différents individus d'un groupe en un jugement « collectif ». C'est une question qui, depuis le début de la décennie, intéresse l'économie (en particulier les spécialistes de théorie du choix social), la philosophie, les sciences politiques et l'informatique théorique.

Les difficultés théoriques de l'agrégation des jugements sont illustrées par le « paradoxe doctrinal » (appelé également « dilemme discursif ») : supposons qu'un groupe d'experts ait à se prononcer sur un projet qui est évalué selon les critères c_1 et c_2 ; que tous les experts sont d'accord pour considérer que le projet est acceptable (A) si et seulement si il satisfait les critères c_1 et c_2 mais qu'ils divergent sur la question de savoir si le projet satisfait ces critères. On pourrait résumer ainsi leurs jugements sur les différentes propositions en jeu :

	C_1	C_2	$A \leftrightarrow (C_1 \wedge C_2)$	A
Expert 1	Oui	Oui	Oui	Oui
Expert 2	Oui	Non	Oui	Non

Expert 3	Non	Oui	Oui	Non
Majorité	Oui	Oui	Oui	Non

Supposons maintenant que l'on engendre les jugements collectifs de la manière suivante : une proposition est acceptée par le groupe si et seulement si elle est acceptée par la majorité. On voit immédiatement qu'on aboutit à des jugements de groupe logiquement incohérents.

Comme en théorie du choix social, dont leurs démonstrations s'inspirent d'ailleurs à l'occasion, les principaux travaux portant sur l'agrégation des jugements établissent des résultats d'impossibilité du type : toute procédure d'agrégation des jugements possédant certaines propriétés naturelles est pourtant susceptible mener à des ensembles de jugements incomplets ou inconsistants.

Voir <http://personal.lse.ac.uk/LIST/doctrinalparadox.htm>

3.3. Epistémologie interactive

L'épistémologie interactive porte sur les questions de *rationalité théorique* (croyances et connaissances) dans les groupes.

3.3.1. Logique épistémique multi-agent

La logique épistémique est la branche de la logique modale qui est consacrée à la modélisation de la croyance et de la connaissance. La logique épistémique multi-agent s'intéresse aux relations épistémiques qui peuvent exister dans un groupe d'agents, comme par exemple le fait qu'une connaissance soit *partagée* par tous dans le groupe ou qu'une proposition soit de *connaissance commune* dans un groupe (chacun sait la proposition, chacun sait que chacun sait la proposition, chacun sait que chacun sait que chacun sait que chacun sait la proposition, etc.)

Voir <http://plato.stanford.edu/entries/common-knowledge/>

3.3.2. Logique probabiliste multi-agent

La logique probabiliste multi-agent est l'analogue probabiliste de la logique épistémique multi-agent. Elle comporte une famille d'opérateur probabiliste du type $L_a^n \phi$ qui signifie : l'agent n croit au moins au degré a que ϕ . La logique probabiliste multi-agent est la version logique des *espaces de types* qui sont des hiérarchies de croyances probabilistes dont on se sert en théorie des jeux.

Au niveau de la modélisation, passer de la logique épistémique à cette logique probabiliste permet par exemple d'introduire des considérations de réalisme cognitif. Certains résultats de théorie des jeux sont ainsi conservés lorsqu'on remplace le réquisit de connaissance commune par celui de croyance commune, fondé sur les opérateurs probabilistes et considéré comme cognitivement moins exigeant.

3.3.3. Caractérisation épistémique des équilibres

Depuis un résultat fondamental montrant l'équivalence entre la représentation de la connaissance par une partition d'information et par un opérateur logique satisfaisant à certains

axiomes, il est notoire que logique épistémique et théorie des jeux entretiennent un rapport étroit. Cette relation trouve sa meilleure illustration dans les travaux de recherche des *fondements épistémiques* des équilibres de théorie des jeux, c'est-à-dire de la détermination de conditions sur les connaissances et croyances des agents qui, lorsqu'elles sont remplies, les conduit nécessairement à jouer selon un équilibre. Ces conditions s'expriment généralement en termes de connaissance partagée ou commune.