



## Réflexivité et auto-référence dans les systèmes complexes

12<sup>èmes</sup> journées  
de Rochebrune

*Rencontres interdisciplinaires sur les  
systèmes complexes naturels et artificiels*



**du 24 au 28 janvier 2005**

Organisées avec le soutien

de l'European Conference on Artificial Life (ECAL)  
de la Délégation Générale pour l'Armement (DGA)

**ENST 2005 S 001**

École Nationale Supérieure des Télécommunications

## Coordination, cognition, auto - référence et réflexivité aux fondements des « croyances sociales »

Denis Phan \*

\* *Université de Rennes 1 – CREM – UMR CNRS 6211*

Communication aux 12<sup>es</sup> Journées de Rochebrune, Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels : «Réflexivité et auto-référence dans les systèmes complexes», Megève, 24-28 Janvier 2005

### Résumé

L'objectif de ce papier est d'explorer les aspects cognitifs de la réflexivité des représentations dans la résolution de certains problèmes de coordination formalisés au moyen de la théorie des jeux, dans le but de définir ultérieurement une ontologie pour la modélisation multi-agents des « croyances sociales » dans les jeux de coordination. Dans une première section, on utilise plusieurs définitions formelles relatives à une dimension « sociale » des croyances, proposées par André Orléan (2002, 2004). On cherche ensuite à en évaluer la signification dans le cadre d'une confrontation entre l'approche « standard » de la théorie des jeux et l'approche « comportementaliste » (ou « expérimentale »). Deux situations particulières de coordination sont passées en revues. Dans les deux cas, la réflexivité des croyances est en question.

Dans les jeux de « pure coordination », les joueurs doivent se coordonner sans communication préalable (jeu non-coopératif). Une caractéristique de ces jeux du point de vue de l'approche standard est la multiplicité des équilibres de Nash (de deux à une infinité, selon le jeu considéré). Du point de vue comportemental, le problème du « point focal » réside alors dans la nécessité de mobiliser des connaissances extérieures au cadre « standard » de la théorie afin de servir de référence « saillante » dans la sélection des équilibres. La notion de « croyances sociales » introduite par Orléan permet une interprétation féconde de ces résultats. Dans le « concours de beauté », les joueurs sont en concurrence pour évaluer un multiple ce que sera l'évaluation moyenne de l'ensemble des joueurs. Il n'y a, selon l'approche standard qu'un seul équilibre de Nash. Les résultats expérimentaux montrent que cet équilibre n'est jamais atteint dans le cadre d'un jeu unique. La réflexivité se manifeste d'abord à travers la question centrale de savoir ce que va produire en moyenne le collectif des joueurs. On est alors amené à poser la question « de second niveau » sur le rôle de la croyance des autres joueurs sur la réponse à cette question et ses incidences sur leurs choix. Ce bouclage récursif pourrait en théorie être infini, pourtant, l'économie expérimentale nous montre qu'il n'en est rien dans les raisonnements effectifs. Pour conceptualiser ces dimensions récursives du processus cognitif, on utilise les notions de « profondeur stratégique » et de « hiérarchie cognitive ». A un premier niveau, il s'agit d'évaluer le nombre d'étapes de raisonnement récursifs dont est capable un joueur. A un second niveau, il s'agit d'évaluer la capacité de raisonnement récursif qu'un joueur attribue aux autres joueurs, etc... Ici encore la notion de « croyances sociales » permet une réinterprétation intéressante du phénomène.

### Introduction

L'économie connaît un grand nombre de situations caractérisées par des références croisées, mais des problèmes particulièrement intéressants se posent lorsque l'on

---

\* L'auteur remercie Frédéric Amblard et Jean Louis Dessalles pour leurs commentaires constructifs sur une version préliminaire de ce papier.

s'intéresse aux *croyances* des acteurs en situation d'interactions "stratégiques". Par exemple, Keynes a utilisé la métaphore d'un "concours de beauté" pour mettre en évidence la nature auto-référentielle du comportement des spéculateurs :

« la technique du placement peut être comparée à ces concours organisés par les journaux où les participants ont à choisir les six plus jolis visages parmi une centaine de photographies, le prix étant attribué à celui dont les préférences s'approchent le plus de la sélection moyenne opérée par l'ensemble des concurrents. Chaque concurrent doit donc choisir non les visages qu'il juge lui-même les plus jolis, mais ceux qu'il estime les plus propres à obtenir les suffrages des autres concurrents, lesquels examinent tous le problème sous le même angle. Il ne s'agit pas pour chacun de choisir les visages qui, autant qu'il peut en juger, sont réellement les plus jolis, ni même ceux que l'opinion moyenne considérera réellement comme tels. Au troisième degré où nous sommes déjà rendus, on emploie ses facultés à *découvrir l'idée que l'opinion moyenne se fera à l'avance de son propre jugement*. Et il y a des personnes, croyons nous, qui vont jusqu'au quatrième ou au cinquième degré ou plus loin encore. » (Keynes, 1936, Chapitre 12 – V, souligné par nous)

Parmi les nombreuses questions soulevées par ce texte célèbre de Keynes, on notera l'idée des *niveaux* (les « degrés ») *dans la « profondeur » du raisonnement*, et celle de la focalisation sur un *objet intermédiaire*, « l'opinion moyenne » dont le statut ontologique (au sens philosophique) apparaît problématique par construction. C'est précisément cet « objet », système psycho-social complexe résultant directement des interactions entre les participants (et indirectement de la confrontation de leurs croyances dans leurs esprits) qui manifesterait des propriétés de « *réflexivité* ». En effet, les propriétés dynamiques de l'agrégat qui le caractérise fonctionnent « *comme si* » le système lui-même disposait *par lui-même* des capacités de s'auto-observer et de s'auto-modifier. Comme cet « objet » (qui est imparfaitement défini) n'a pas d'existence matérielle (et donc pas à proprement parler de « capacités »), la question de sa *nature* et de sa *manière d'être* doit être posée.

Si l'on se déplace au niveau formel, la théorie des jeux utilise largement des formes variées d'auto-référence et de réflexivité, tant au niveau de la rationalité des joueurs que de celui des conditions dans lesquelles les choix sont effectués (i.e. l'hypothèse de « *connaissance commune* » qui sera discutée plus loin). Mais la théorie standard ne recourt pas à des croyances « sociales » au sens ou nous les entendons ici. C'est précisément pour cela que nous avons choisi d'aborder la question de la possible réflexivité des croyances portant des « objets sociaux » à travers le filtre de la théorie des jeux (évidemment réducteur, il permet cependant de clarifier la réflexion).

Plus spécifiquement, l'objectif de ce papier est d'explorer les aspects cognitifs et réflexifs des représentations « sociales » dans la résolution de certains problèmes de coordination formalisés au moyen de la théorie des jeux. L'objectif ultime, qui n'est pas traité ici, étant de proposer une « *ontologie* » (au sens informatique discuté plus loin, distinct du sens traditionnel de la philosophie) pour la modélisation multi-agents des « croyances sociales » dans les jeux de coordination (Ferber, Phan, 2005)

La notion informatique d'*ontologie* ne faisant pas l'objet d'une définition canonique ni même d'un consensus, nous retiendrons d'abord la définition large d'un « *modèle*

*d'organisation de la connaissance* ». Dans le cas qui nous intéresse la « connaissance » repose elle-même sur une famille de modèles, méthodes, expérimentations et discours « savants ». On pourra alors parler de « *modèle d'organisation et de conception de la connaissance* ». Comme notre objectif premier reste la modélisation multi-agents des « croyances sociales » dans les jeux de coordination, il est utile de restreindre cette définition à une représentation rapidement opérationnelle dans ce dernier domaine, liée aux *modèles de conception* « objet » et « agent ». On retiendra donc l'idée d'un *modèle de représentation formelle d'un système de concepts articulés*, qui fasse sens à la fois pour le théoricien des jeux, le « thématicien » ou le théoricien en sciences économiques ou sociales, et l'informaticien. Ce système de concepts articulés devrait permettre à la fois d'organiser les connaissances existantes et d'en concevoir de nouvelles : notre *ontologie* serait donc ainsi elle-même dotée de propriétés réflexives ! Plus spécifiquement, il serait formalisé par des « objets » structurés par des relations de *composition* et d'*héritage*. Les « objets » ainsi définis seraient encore caractérisés par des types d'*attributs* et types de *relations* entre objets, l'ensemble de cette construction étant interprétable à tous les niveaux dans une *sémantique spécifique*. Le présent travail ne prétend pas construire une telle ontologie, mais cherche plus modestement à « mettre à plat » certains éléments du débat, du point de vue du théoricien en sciences économiques ou sociales, pour initier le dialogue avec les informaticiens.

Dans une première section, on utilise plusieurs définitions formelles relative à une dimension « sociale » des croyances, introduites par Orléan (2002, 2004). On cherche ensuite à proposer une interprétation dans le cadre d'une confrontation entre l'approche « standard » de la théorie des jeux et l'approche « comportementaliste » - ou « expérimentale » (Camerer, 2003). Deux situations particulières de coordination sont passées en revue : les jeux de « pure coordination » et le « concours de beauté ». Dans les deux cas, la réflexivité des croyances est en question. En effet, l'analyse réflexive du problème sous la forme récursive associée à la théorie standard, conduit a des résultats qui ne correspondent pas a ceux que l'on obtient par expérimentation. Une première notion de « *hiérarchie cognitive* » permet alors de conceptualiser une forme stylisée de procédure réflexive limitée qui permet aux agents de sortir de l'aporie liée à la récursivité généralisée. L'interprétation de ces procédure à l'aide d'une notion spécifique de « *croyance sociale* » proposée par André Orléan permet de mettre a jour des perspectives nouvelles. Dans les deux jeux considérés, la connaissance qu'ont les joueurs du « *background culturel* » des autres joueurs, peut faciliter l'attitude cognitive qui consiste à « se mettre à la place de l'autre », sans empathie (la « *décentration* ») pour éliminer des solutions qui apparaissent alors comme peu (ou pas) pertinentes. Cette attitude peut être considérée comme un *réducteur d'incertitude*, qui aide à la coordination. Mais elle demande aussi un investissement cognitif important. C'est pourquoi le recours à une entité abstraite représentative : « le groupe », à qui l'on attribue les caractéristiques que l'on cherche à inférer chez les autres, peut *réduire la complexité du processus cognitif*, et donc le *coût computationnel* que l'on pourrait y associer. Lorsque cette démarche est partagée, le « groupe » peut prendre une forme spécifique de croyance collective, qui se traduit par un objet abstrait dotée d'une autonomie relative par rapport aux croyances individuelles qui en sont pourtant le vecteur.

## 1. Dans quelle mesure les croyances peuvent – elles être « sociales »

Dans des travaux récents Orléan (2002, 2004) propose une nouvelle définition des croyances « collectives » (*Collectives Belief*, in : Orlean, 2004), également appelées « croyances sociales » dans Orléan (2002). Cette initiative a entre autre deux intérêts pour notre démarche. Premièrement, elle propose une présentation formelle (logique) de la dimension « sociale » des croyances, qui a le mérite de clarifier les choses, en particulier dans la perspective de l'élaboration d'ontologies informatiques, au sens défini précédemment. Deuxièmement, elle soulève la problématique de l'existence « d'objets sociaux » (ici les croyances « sociales ») dotés d'une autonomie relative par rapport aux consciences individuelles qui en sont le support. L'intérêt de l'existence de tels « objets » dans la théorie des jeux (et plus largement en économie) est loin d'être unanimement établie. En particulier, la théorie standard n'y a généralement pas recours. L'objectif explicite des articles cités d'Orléan est d'ailleurs de promouvoir leur intérêt dans le cadre d'une économie « cognitive ». Le mode d'existence de telles entités est par ailleurs discutée aussi bien chez les philosophes (parmi ceux qui cherchent des fondements ontologiques aux objets sociaux on citera : Gilbert 1990, 2000, 2003 ; Pettit, 1993, 2000, 2004 ; Searle, 1995 ; sans oublier Descombes, 1996, sur un thème connexe) que chez les sociologues. Boudon, 2003 en conteste l'autonomie, mais il existe une solide tradition selon laquelle les « faits sociaux » doivent être traités « comme des choses » (Durkeim, 1895 ; pour une ré-interprétation: Douglas, 1986 ; pour un fondement cognitif : Kaufmann, Clement, 2003 et sur l'ensemble de la question : Livet, Ogien, 2000). Dans la mesure ou notre objectif ultime de modélisation multi-agents entraîne bien la réification d'un certain nombre de concepts, la recherche d'un positionnement dans cette discussion de l'ontologie (philosophique) et donc du mode d'existence de certains « faits sociaux » nous semble incontournable. La typologie d'Orléan nous offre précisément l'opportunité de clarifier ce débat dans le domaine de la dimension « collective » des croyances.

### 1.1 Croyances partagées et croyances communes.

Orléan nous propose quatre définitions. La « croyance partagée » permet de construire de manière récursive, la « croyance commune » (ré-interprétation dans l'ordre des croyances, de la notion de *common knowlegde*). Il introduit ensuite deux formes de « croyances sociales », elles-mêmes liées par une construction récursive. Dans ce but, il introduit la notation suivante :  $C_iQ$  signifie : « l'individu  $i$  croit que la proposition  $Q$  est vraie » que nous simplifierons en : « l'individu  $i$  croit  $Q$  ». On désigne ainsi les agents par leur indice  $i$ . La notation  $i \in G_e$  signifie qu'un individu  $i$  appartient à un groupe  $G_e$  ( $G_e$  est un ensemble d'agents). On notera  $N_e$  le nombre d'individus de ce groupe ( $N_e = \text{card} | G_e |$ )

Il y a « croyance partagée » dans un ensemble d'agents  $G_e$  lorsque l'on a :

$$(1) \quad C_iQ \text{ pour tout } i \in G_e$$

qui peut s'énoncer : « tous les individus de  $G_e$  croient  $Q$  ». La définition de la « croyance partagée » est homologue a celle de « connaissance mutuelle » en théorie des jeux (on l'a juste affaiblie en remplaçant « connaissance » par « croyance »)

Il y a « *croyance commune* » dans un ensemble d'agents  $G_e$  lorsque l'on a :

$$(2) \quad C_i Q ; C_j C_i Q ; C_k C_j C_i Q ; \dots \text{ ad infinitum pour tout } i, j, k, \dots \in G_e$$

La définition de la *croyance commune* est ainsi homologue à celle de *connaissance commune* en théorie des jeux. Le philosophe David Lewis a défini cette notion de la manière suivante : « quelque chose » est connaissance commune si (1) tout le monde le sait, (2) tout le monde sait que : tout le monde le sait, (3) tout le monde sait que : tout le monde sait que : tout le monde le sait, et ainsi de suite *ad infinitum*. Cette définition logique a ensuite été traduite en théorie des jeux par Rober Aumann.

Il apparaît ainsi que la croyance commune peut être construite de manière récursive à partir de la connaissance mutuelle. En décomposant la première en niveaux hiérarchiques imbriqués, la « sous-croyance commune » de niveau (2) n'est pas autre chose qu'une « croyance partagée » sur la sous-croyance commune de niveau (1) qui est elle-même simplement une croyance partagée. De même, la sous-croyance commune de niveau (3) est une croyance partagée sur la sous-croyance commune de niveau (2), et ainsi de suite... La caractéristique commune de ces deux définitions est qu'elles se fondent strictement sur des croyances individuelles. « Elles ne sont collectives qu'au sens où *tous* les d'individus [du groupe], d'une manière ou d'une autre, les ont fait leur » nous dit Orléan. De ce point de vue, elles sont conformes à la définition de l'individualisme méthodologique retenu par Boudon (2003).

## 1.2. Les croyances « sociales » de premier niveau : une tentative d'interprétation

Orléan nous propose d'introduire directement un « *objet collectif* » intitulé le « *groupe* », que l'on notera :  $G_s$  pour le distinguer de *l'ensemble d'agents*,  $G_e$ . Bien que cet « objet » ne soit pas un acteur comme les autres agents, au sens où il ne prend pas de décision par lui-même, les acteurs (les agents) lui prêtent des croyances qui sont formellement équivalentes à des croyances individuelles dans la formalisation introduite par Orléan :  $C_{G_s} Q$  signifie ainsi : « le groupe  $G_s$  croit  $Q$  ». C'est cette proposition qui est qualifiée par Orléan de « *croyance sociale* » : « dans ce cas, on attribue des croyances à une *entité abstraite*, à savoir le groupe lui-même » (souligné par nous).

La première justification qu'Orléan nous donne de l'introduction de cette entité abstraite est empirique et repose sur l'analyse du discours ordinaire. Il ne manque pas en effet de situation où les individus attribuent des caractéristiques anthropomorphiques à une entité abstraite représentant un collectif. C'est le cas par exemple avec cette phrase citée par Orléan : « *le marché croit que cette devise est sous évaluée* ». Bien entendu, on peut arguer qu'il s'agit là d'un « raccourci » et que personne ne croit que « le marché » soit doté de croyance. Quoiqu'il en soit, la simple utilisation de ce raccourci pourrait cependant introduire des *biais cognitifs* et il serait intéressant d'en explorer les conséquences. On sait par exemple que la dynamique d'un ensemble d'agents en interaction, en tant que *système complexe adaptatif* possède des propriétés spécifiques qui excluent généralement l'usage d'un agent représentatif (Anderson *et al.*, 1988 ; Weisbuch, 1989 ; Arthur *et al.*, 1997, Schuster 2001 ; Lesourne *et al.*, 2003, Phan, 2004). Nous ne discuterons pas ces questions ici, continuant à suivre le cheminement d'Orléan, avant d'en illustrer la portée à l'examen de quelques jeux significatifs, ce qui correspond également à la

démarche de cet auteur. Avec l'introduction des croyances attribuées au groupe, Orléan nous propose deux nouvelles catégories de croyances collectives : les « *croyances sociales* ».

Selon une première définition, l'individu  $i$  croit que « le groupe croit » que la proposition  $Q$  est vraie s'il croit qu'une « grande partie » du groupe croit que la proposition  $Q$  est vraie. Orléan ne propose pas de définition formelle, mais nous pouvons certainement traduire une des interprétations possibles de sa pensée en écrivant :

$$(3) \quad C_i C_{G_s} Q \stackrel{i}{\Leftrightarrow} C_i C_{G_{pi}} Q \equiv (C_i C_j Q \text{ pour tout } j \in G_{pi} / N_e \geq \text{card} | G_{pi} | > N_{pi})$$

C'est à dire : «  $i$  croit que *le groupe croit*  $Q$  » si, *de manière équivalente pour lui* ( $\stackrel{i}{\Leftrightarrow}$ ) il croit qu'un nombre de membres du groupe supérieur à un certain seuil  $N_{pi}$  le croit aussi, ce que l'on notera :  $C_i C_{G_{pi}} Q$  ( $i$  croit qu'une partie « suffisamment grande » du groupe,  $G_{pi}$  croit  $Q$ ). *La référence de base renvoie ici aux croyances individuelles* sur  $Q$  ( $C_i C_j Q$  pour tout  $j \in G_{pi}$ ) *et non au groupe en tant qu'objet social identifié*. Les conditions d'équivalence avec la *croyance sociale* de l'agent  $i$  :  $C_i C_{G_s} Q$  doivent encore être spécifiées. Notre formalisation va plus loin que l'article initial informel d'Orléan ne le suggère, puisque notre notation entraîne que les croyances de  $i$  intègrent explicitement un périmètre d'individus  $G_{pi}$  qui est une croyance spécifique à  $i$  et permet de renvoyer à des croyances sur les croyances individuelles des membres de ce groupe. Ce sous-groupe  $G_{pi}$  « existe » donc a ce niveau en tant qu'*entité ontologiquement subjective* propre à chaque agent, mais il est de fait distinct d'une référence commune au groupe dans son ensemble (noté  $G_s$ ) qui est supposée renvoyer pour tous les agents au même objet dont ils partagent la représentation. André Orléan fait remarquer que cette première notion de croyance collective « sociale » est proche de la croyance partagée. Si l'on ne considère que la partie droite de l'équivalence  $C_i C_{G_{pi}} Q$ , on pourrait même considérer que cette forme de croyance collective n'est qu'une forme affaiblie de cette croyance partagée. En effet, *l'étant* de cette entité  $G_{pi}$  se réduit à la liste des agents qui le composent. Mais la partie droite  $C_i C_{G_s} Q$  est d'*une autre nature* qu'il faudrait préciser, tout comme le mécanisme qui permet de rendre en pratique ces deux visions des choses équivalentes, pour les agents.

Ce n'est que qu'à un second niveau, par un « *raccourci de pensée* » qu'il peut y avoir ce que nous définissons comme une *subsumption* de la croyance sur en ensemble de croyances individuelles ( $C_i C_{G_{pi}} Q$ ) par une forme de croyance spécifique sur un « groupe » de référence commune ( $C_i C_{G_s} Q$ ). Cette opération substitue ainsi une entité abstraite à un ensemble d'agents, et la notation adoptée présente l'avantage de mettre en avant la logique « floue » sous-jacente à l'opération, qui traduit selon nous une première manifestation du « raccourci de pensée » qui a été évoqué précédemment. Plus spécifiquement, le flou se manifeste en particulier dans le passage des sous-groupe individualisés  $C_{G_{pi}}$  qui sont une construction subjective spécifique aux agents, et qui ont donc le statut d'une croyance personnelle sur des entités concrètes à la notion abstraite du « groupe dans son ensemble » :  $G_s$  qui est une « entité collective » en principe reconnue de tous, même si par définition, il n'est pas nécessaire que cette référence partagée repose sur une même évaluation des conditions d'équivalence pour les agents (i.e. les ensembles  $G_{pi}$  et les seuils  $N_{pi}$  peuvent différer selon les agents). Autrement dit, lorsque l'on parle de « groupe » (ou de « marché ») en tant qu'entité collective, tout le monde

fait *comme si* ce mot désignait bien sans ambiguïté la même chose pour tout le monde, bien que ces termes puissent renvoyer à des représentations différentes selon les agents.

Notons que selon Orléan (2002), la notion vague de « grande partie » devrait être « précisée par des travaux ultérieurs ». Dans Orléan (2004), cette difficulté est contournée en remplaçant « grande partie » par « majorité ». Les exemples étudiés plus loin suggèrent que la notion (à préciser) de « grande partie » aurait l'avantage sur la notion de majorité de pouvoir capturer l'idée qu'il existe d'autres *seuils critiques* dans les groupes que la majorité, mais l'inconvénient de rendre ce critère très dépendant du contexte.

### 1.3. Subsumer un collectif d'agents par une entité abstraite : quelle signification ?

L'intérêt de l'introduction de la notion de « groupe » pour la formalisation multi-agents des croyances résiderait selon nous dans l'utilisation que l'on peut faire dans le temps de l'opération qui consiste à *subsumer*<sup>1</sup> la croyance individuelle  $C_i C_{G_{pi}} Q$  sur le sous groupe  $G_{pi}$ , dont « l'existence » est une construction ontologiquement subjective de l'agent  $i$  par la croyance  $C_i C_{G_s} Q$  sur le groupe  $G_s$  dont la construction est d'emblée « sociale » reste cependant à définir. En particulier, on remarquera que la condition d'équivalence individuelle (3) est une condition nécessaire de subsumption, mais elle n'est pas suffisante : encore faut-il préciser (modéliser) les conditions sous lesquelles cette subsumption peut se réaliser et se maintenir (ou être remise en cause). Autrement dit, à quelle(s) condition(s) les individus remplacent-ils des croyances sur d'autres individus par des croyances sur le « comportement » d'une entité collective représentative de ces derniers, le « groupe », dont la nature est tout aussi ontologiquement subjective, mais en principe partagée par tous ? On montrera dans la dernière section que cette formalisation du « raccourci de pensée » peut avoir des effets non neutres sur la dynamique des modèles (et corrélativement sur celle des comportements).

Le mouvement qui permet de considérer comme équivalent pour l'agent  $i$  le groupe vu comme un ensemble d'agents ( $G_e$ ) au groupe en tant qu'objet social ( $G_s$ ) est problématique. Selon la perspective de l'individualisme méthodologique orthodoxe, le groupe n'existe qu'en tant que représentation commode pour organiser la pensée. Comme la monnaie « voile » de la théorie standard, le groupe ( $G_e$ ) est une *abstraction neutre* au sens où il peut à tout moment être remplacé par la liste des agents qui le compose. Descombes (1996) oppose à cette vision logique de totalité abstraite celle de *totalité concrète*, irréductible à l'ensemble des éléments du collectif considéré, car les prédicats qui peuvent lui être attribués ne sont pas distribuables aux composantes. En effet, les propriétés relationnelles qui structurent le système ne sont ni de l'ordre de tout ni de l'ordre des parties, tout en étant constitutifs des deux (Descombes parle de *holisme structural*). La *dimension diachronique* permet alors de révéler la fonction particulière de la relation dans la forme d'existence de cet objet collectif : « *du point de vue historique et social, les collectivités ne sont justement pas des collections d'individus,*

---

<sup>1</sup> du latin sub, sous, et sumere, rendre : - Lalande « penser un fait comme constituant l'application d'une loi » (notion de philosophie kantienne qui signifie aussi penser une entité comme comprise dans une entité plus large)

*puisque leur identité n'est pas fixée par une liste d'individu* » (Descombes, 1996, p.134). Commentant Wittgenstein, Descombes montre qu'une « bibliothèque » est plus qu'un ensemble de livres (totalité logique), car c'est un tout concrètement organisé, décomposable selon d'autres niveaux d'individuation que le livre lui-même, et dont le sens est lié dynamiquement à cette structure organisationnelle. L'existence de la bibliothèque et sa signification dépasse les éventuels mouvements qui portent sur les ouvrages qui la composent.

De la même manière, les marchés financiers d'Orléan sont également des formes durables, organisés pour coordonner les échanges. Ils ne peuvent être réduits à l'ensemble des agents qui y participent à un moment donné indépendamment du rapport qu'ils entretiennent avec cette structure (Orléan 2000). Cette *structure sociale complexe* est ainsi *porteuse d'un sens* qui se maintient dans le temps indépendamment des agents qui y participent et qui dépasse leur dimension individuelle. C'est peut-être précisément pour cela que les agents lui attribuent spontanément dans le langage quotidien des qualités anthropomorphiques qui ramènent cette structure à une dimension plus interprétable.

Une conjecture intéressante sur une possible interprétation *cognitive* de la subsumption anthropomorphique (du « *raccourci de pensée* ») pourrait être dérivée de l'explication cognitiviste des croyances collectives proposée par Boudon (1995, 1997). Pour ce dernier, « les sujets tentent de se débattre dans une question complexe en émettant des conjectures raisonnables. C'est parce ces conjectures sont raisonnables qu'une grande majorité les adopte » (1995, p.28). Boudon appelle « raisons transsubjectives » les raisons perçues comme *suffisantes* (au sens de la rationalité limitée de Simon) par cette « grande majorité » pour être adoptées en lieu et place d'une représentation plus complète du phénomène complexe en question. On ne discutera pas ici des causes (éventuellement infra individuelles) pour lesquelles ce principe de raison suffisante est mis en œuvre par cette grande majorité des individus, et on se concentrera sur l'objet du processus cognitif lui-même : le *raccourci de pensée*. On sait que Boudon considère que la connaissance « ordinaire » et la connaissance « scientifique » renvoient à des mécanismes cognitifs de même nature et qu'il n'y a pas lieu d'établir « une ligne de démarcation franche entre la pensée scientifique et la pensée ordinaire » (op.cit. p. 48). Selon sa vision de l'individualisme méthodologique, Boudon considère qu'une « croyance collective ne s'explique que dans la mesure où elle fait sens pour un *sujet idéaltypique*, pour un individu *quelconque* » (op.cit. p.38). On est tenté d'appliquer ces principes de manière réflexive au-delà de l'usage qu'en fait habituellement leur auteur.

On pourrait ainsi dire que le *fait social* désigné par Orléan sous le nom de « croyance sociale », qui consiste à substituer une entité abstraite (le groupe) à un collectif d'agents lié par des relations complexes, est une opération cognitive de même nature que celle qui consiste à substituer à la diversité des individus une entité conceptuelle *idéaltypique* (au sens de Weber) pour *comprendre les raisons* qui motivent le comportement de ces individus. En d'autres termes, la subsumption cognitive d'un collectif d'individus par une entité sociale chez une « grande majorité » d'individus pourrait être fondée sur des « raisons transsubjectives ». Celles – ci se traduiraient par la croyance (en fait fautive) qu'il

est plus aisé de *comprendre* la dynamique (complexe) d'un groupe en la « personnalisant », c'est-à-dire en réduisant l'ensemble des agents et des relations qui caractérisent ce collectif à une entité abstraite mais *porteuse de sens* pour les individus, grâce à l'attribution de caractéristiques anthropomorphiques.

#### 1.4. Croyances sociales auto-référentielles de second niveau

La dernière définition proposée par Orléan est d'emblée *auto-référentielle* : « l'individu  $i$  croit que le groupe croit à  $Q$  pour autant qu'il croit qu'une grande partie du groupe croit également que le groupe croit à  $Q$  ». La nature auto-référentielle de cette définition ressort de l'indétermination associée à la référence à des *croyances du second degré* : puisque les membres du groupe ne se limitent plus « croire  $Q$  », mais « croient que le groupe croit  $Q$  »

$$(4) \quad C_i C_{G_s} C_{G_s} Q \Leftrightarrow C_i C_{G_{pi}} C_{G_s} Q \equiv (C_i C_j C_{G_s} Q \text{ pour tout } j \in G_{pi} / N_e \geq \text{card} | G_{pi} | > N_{pi})$$

Remarquons que cette notation induit des croyances de type (3) de second degré sur la relation d'équivalence de l'agent  $i$  elle-même, qui traduit une subsumption au premier degré. Ce qui signifie qu'il y a déjà eu remplacement au premier degré des croyances de l'ensemble d'individus  $G_{pi}$  par une « croyance » attribuée à l'entité sociale  $G_s$ . Le principal avantage de l'introduction directe de cette croyance attribuée à une entité collective  $C_{G_s}$  peut être trouvé au niveau cognitif, puisqu'il permet d'économiser des capacités computationnelles, en particulier d'éviter une récursivité d'une profondeur infinie, comme dans le cas de la croyance commune. Par contre, le prix à payer est celui de *l'indétermination liée à l'auto-référence du second degré* : « le groupe croit que le groupe croit ». Mais cette indétermination est aussi une des sources de l'autonomie relative des croyances « sociales ». Autonomie car le pouvoir d'action des individus sur cet objet collectif (pourtant par nature ontologiquement subjectif) est d'autant plus faible que le groupe est grand, comme on peut le montrer avec des modèles formels d'interdépendance sociale ou l'ensemble formé par les agents en interaction peut être considéré comme un système complexe adaptatif (Lesourne *et al.*, 2002, Phan *et al.* 2004). Mais cette autonomie n'est que relative, car cet objet collectif émane bien des subjectivités individuelles qui en sont le support. Il n'existerait pas sans elles et sa dynamique dépend en dernier ressort à la fois des interactions « externes » entre les agents du groupe considéré et des raisonnements « internes » (en particulier la subsumption) de ces derniers. C'est ce que nous allons chercher à montrer à partir d'une étude critique de certains jeux de population à la lumière de l'économie expérimentale.

## 2. Jeux de coordination et « saillance à la Schelling »

Dans un jeu de « *pure coordination* », les joueurs ne sont pas en compétition, mais cherchent simplement à se coordonner sur un « fait saillant » (ou « point focal ») *sans communication préalable* (jeu *non-coopératif*). Une caractéristique de ces jeux du point de vue de l'approche « standard » est la multiplicité des équilibres de Nash (de deux à une infinité, selon le jeu considéré). Le tableau 1 représente le jeu de pure coordination le plus simple sous la forme dite « normale » (ou « stratégique »). Cette forme ne nous permet de connaître ni l'ordre des interventions et les enchaînements possibles, ni

l'information dont disposent les joueurs au cours de ces enchaînements. Mais elle nous permet d'identifier les équilibres non coopératifs.

## 2.1. Le cadre du jeu

Sur le tableau 1, les stratégies du joueur  $P_1$  figurent en ligne, et celles du joueur  $P_2$  en colonne. Ceux-ci doivent choisir entre deux stratégies dans l'ensemble :  $S = \{A, B\}$ . On désigne par  $S_1$  la stratégie du joueur  $P_1$  et par  $S_2$  celle du joueur  $P_2$ . Les gains associés à chaque couple de stratégies sont désignés par :  $\pi_i(S_i, S_j)$ . Ils sont représentés entre parenthèse dans les cases de la matrice de jeu, le gain du joueur  $P_1$  :  $\pi_1$  d'abord, puis celui du joueur  $P_2$  :  $\pi_2$  ensuite, soit :  $(\pi_1, \pi_2)$ . *L'équilibre de Nash est une situation où aucun joueur n'a intérêt à dévier unilatéralement.* (A, A) est un équilibre de Nash car si le joueur  $P_1$  dévie unilatéralement et joue B son gain passe de  $\pi_1(A, A) = 1$  à :  $\pi_1(B, A) = 0$ . Il en est de même pour le joueur  $P_2$ . Le lecteur vérifiera facilement que : (B, B) est également un équilibre de Nash.

Tableau 1 : jeu de pure coordination

	$P_2$ joue $S_2 = A$	$P_2$ joue $S_2 = B$
$P_1$ joue $S_1 = A$	(1, 1)	(0, 0)
$P_1$ joue $S_2 = B$	(0, 0)	(1, 1)

Formellement, dans le cas à 2 joueurs, pour le joueur  $P_i$ , la stratégie  $S_i^* \in S$  est un équilibre de Nash si la condition suivante est vérifiée :

$$(5) \quad \pi_i(S_i^*, S_j^*) \geq \pi_i(S_i, S_j^*) \quad \text{pour tout } S_i \in S \text{ et tout } i, j \in \{1, 2\}$$

Pour une stratégie donnée (quelconque) de l'autre joueur, on peut définir la *meilleure réponse* d'un joueur comme la stratégie qui maximise son gain, conditionnellement à cette stratégie donnée de l'autre joueur :

$$(6) \quad \forall i, j \in \{1, 2\}, \tilde{S}_i | S_j \in \arg \max_{S_i \in S} \{ \pi_i(S_i, S_j) \}$$

On peut alors également définir un équilibre de Nash comme un ensemble de choix tel que *la stratégie de chaque acteur est la meilleure réponse aux stratégies des autres acteurs* :

$$(7) \quad \forall i, j \in \{1, 2\}, S_i^* \in \arg \max_{S_i \in S} \{ \pi_i(S_i^*, S_j^*) \}$$

Par exemple, si mon adversaire joue A, ma meilleure réponse consiste à jouer A également. Si mon adversaire joue B, ma meilleure réponse est B. Dans ces deux cas c'est en effet la stratégie qui maximise mon gain. Ces trois définitions s'étendent facilement à n joueurs en remplaçant la stratégie de l'adversaire  $S_j$  par le vecteur  $S_{-i}$  des stratégies des adversaires du joueur  $P_i$ . Dans ce qui suit, on supposera que les joueurs jouent toujours leur meilleure réponse, conditionnellement à leurs croyances sur le comportement des autres joueurs.

Nous avons donc un jeu à deux équilibres de Nash. La question de savoir lequel des deux sera choisi est qualifiée dans la littérature de problème de « sélection » d'un

équilibre. Dans le cas à deux joueurs, *en l'absence d'une référence extérieure*, il est impossible de savoir lequel des deux équilibres sera sélectionné. Si, par contre, on introduit une référence extérieure, par exemple le fait que A vienne avant B dans l'alphabet, on peut imaginer, comme on le verra plus loin, que cet équilibre a plus de chance d'être sélectionné, en particulier si chaque joueur a des croyances (de premier degré ou au-delà) qui viendraient renforcer ce biais « culturel ».

## 2.2. Le « jeu de population »

Considérons maintenant le cas d'un « jeu de population » (Blume, 1997) où chaque joueur est apparié avec un autre joueur choisi aléatoirement (sans redondance) dans la population. Les joueurs doivent jouer en « *stratégie pure* », c'est à dire choisir A ou B. Que se passe-t-il si les joueurs ont des croyances *a priori* sur la distribution des stratégies dans la population ? Supposons que le joueur  $i$  croit qu'il y a dans la population une proportion  $\alpha_i$  de joueurs qui jouent A (de manière complémentaire :  $(1-\alpha_i)$  qui jouent B). L'espérance de gains de sa stratégie  $S_i$  dans un appariement aléatoire (*random pairwise*) sera donc égal à :

$$(8) \quad \pi_i(S_i, \alpha_i) = \alpha_i \cdot \pi_i(S_i, A) + (1-\alpha_i) \cdot \pi_i(S_i, B)$$

Le lecteur vérifiera que si le joueur  $i$  choisit sa meilleure réponse avec des gains définis par l'équation (8), alors il choisira A si  $\alpha_i > 0,5$  et B si  $\alpha_i < 0,5$ . On dit qu'il choisit ainsi sa meilleure réponse contre la « *stratégie mixte* »  $\alpha_i$ . Il importe de remarquer que chaque agent  $i$  a une croyance spécifique :  $\alpha_i$  et qu'il n'y a aucune raison que cette croyance soit la même pour tous les agents. Dans le cadre qui nous intéresse ici, les joueurs doivent choisir une stratégie pure dans un jeu de population avec appariement aléatoire et tout se passe comme s'ils faisaient face à un seul joueur qui joue A et (respectivement : B) avec une probabilité  $\alpha_i$  et (respectivement :  $1-\alpha_i$ ). Dans le cas d'un *jeu de population itéré*, les joueurs peuvent « apprendre » à partir de leur expérience. Si les agents révisent ainsi leurs croyances à partir des observations qu'il font à chaque nouvelle rencontre, ces croyances évolueront en fonction de l'histoire des rencontres et si l'environnement est stable, on pourra converger vers l'un des équilibres de Nash (A, A) ou (B, B) selon un processus qui dépendra à la fois de la distribution initiale des croyances et l'histoire des rencontres.

Si l'on attribue les croyances initiales à la « culture » et la révision des croyances à « l'histoire », la sélection d'un équilibre résultera donc de facteurs historico-culturels, dans une proportion qui dépendra de la part accordée à chacun de ses facteurs dans le comportement. C'est ainsi par exemple que les anglais se sont « coordonnés » sur la conduite à gauche, alors que les continentaux l'ont fait sur la conduite à droite.

Remarquons que les croyances décrites ci-dessus peuvent se formaliser à l'aide de notation utilisée en (3) pour *la croyance sociale*. Le lecteur vérifiera que si le joueur  $i$  croit qu'une « grande partie du groupe » croit que sa meilleure réponse est de jouer A ( $\text{card}G_{pi}/N_e \equiv \alpha_i > 0,5$  dans le cas où l'on adopte le seul majoritaire :  $N_{pi} = 0,5$ ), alors sa meilleure réponse sera de jouer A. En notant  $Q \equiv (\tilde{S} = A)$  et  $C_j(\tilde{S} = A) \equiv$  (le joueur  $j$  croit que sa meilleure réponse est de jouer A) on a :

$$(9) \quad \tilde{S}_i = A \text{ si : } C_i C_{G_{pi}}(\tilde{S} = A) \equiv (C_i C_j(\tilde{S} = A)) \text{ pour tout } j \in G_{pi} / 1 \geq \text{card}|G_{pi}|/N_e > 0,5)$$

Mais aussi, en cas de *subsumption* :

$$(9 \text{ bis}) \quad \tilde{S}_i = A \text{ si : } C_i C_{G_s}(\tilde{S} = A) \stackrel{i}{\Leftrightarrow} C_i C_{G_{pi}}(\tilde{S} = A)$$

Mais d'où proviennent ces croyances « de premier niveau » ? En reprenant nos hypothèses, et en supposant que les croyances de *i* sur les croyances de *j* sont exactes, si le joueur *j* croit qu'il faut jouer *A* :  $C_j(\tilde{S} = A)$ , c'est parce qu'il croit lui-même, qu'il y a un groupe majoritaire  $G_{pj}$  qui joue ainsi ( $\alpha_j > 0,5$ ) donc :  $C_j C_{G_j}(\tilde{S} = A)$  et sous les conditions de la *subsumption* :  $C_j C_G(\tilde{S} = A)$ . Si nous réintégrons ce raisonnement dans les croyances de l'agent *i* nous *pourrions* re-formuler nos hypothèses sur les croyances initiales sous la forme de second degré des croyances sociales selon (4) :

$$(10) \quad \tilde{S} = A \text{ si : } \\ C_i C_{G_{pi}} C_{G_s}(\tilde{S} = A) \equiv (C_i C_j C_{G_s}(\tilde{S} = A)) \text{ pour tout } j \in G_{pi} / 1 \geq \text{card}|G_{pi}|/N_e > 0,5)$$

Mais aussi, en cas de *subsumption* (croyances auto-référentielles) :

$$(10 \text{ bis}) \quad S_i = A \text{ si : } C_i C_{G_s} C_{G_s}(\tilde{S} = A) \stackrel{i}{\Leftrightarrow} C_i C_{G_{pi}} C_{G_s}(\tilde{S} = A)$$

Comme ces trois manières de formuler le problème des croyances différent, peut-on dire qu'il y en a une qui soit plus pertinente qu'une autre ?

### 1.3. Distinguer les croyances partagées et les croyances sociales : le « point focal »

Pour savoir quelle est l'interprétation la plus adéquate à une situation de coordination, on peut considérer un jeu de coordination un peu plus sophistiqué, où il ne s'agit plus de choisir entre deux stratégies indifférenciées, mais où il faut choisir un nombre entier entre zéro et cent. Comme dans le jeu précédent, le joueur gagne un s'il trouve le même nombre que son adversaire, zéro sinon. Dans ce cas, il y a potentiellement 101 équilibres de Nash, puisque tous les choix sont des candidats possibles pour un équilibre, pourvu que l'autre joueur joue de manière semblable (la matrice stratégique de ce jeu serait une matrice « identité » avec des couples (1,1) sur la première diagonale et des couples (0,0) partout ailleurs). Pourtant, lorsque l'on expérimente ce jeu avec des joueurs « réels », ceux-ci arrivent assez bien à se coordonner avec des stratégies « saillantes », telles que choisir : 0, 1, 7, 50 ou 100. Pour bien comprendre le mécanisme cognitif qui permet de passer aux croyances sociales, il est commode, suivant Orléan, d'utiliser les résultats de Metha, Starmer et Sugden (1994). Ceux-ci proposent une variante du jeu précédent qui permet une meilleure focalisation (en évacuant les deux extrêmes : la borne inférieure 0 et la supérieure 100). Ils demandent ainsi aux sujets de l'expérience de *choisir un nombre entier supérieur à 1*. Ils réalisent cette expérience sur deux groupes de référence. Dans le premier groupe, ils demandent juste de choisir un nombre en dehors de toute situation de coordination. Dans le second groupe ils proposent le jeu de coordination (gain de 1 s'il y a coordination, 0 sinon). Leurs résultats sont reproduits dans le tableau 2.

Tableau 2 : le jeu de coordination de Metha, Starmer et Sugden (1994) p. 667.

Groupe test « P » (N = 90)		Groupe jeu coordination « G » (N = 90)	
Réponses	Proportion	Réponses	Proportion
7	11,4 %	1	40,0 %
2	10,2 %	7	14,4 %
10	5,7 %	10	13,3 %
1	4,5 %	2	11,1 %
r = 28	c = 5,2 %	r = 17	c = 20,6 %

r = nombre de réponses différentes c = probabilité que deux individus donnent la même réponse

On peut interpréter les résultats du groupe test « P » comme une traduction des « préférences personnelles » des expérimentés, qui peuvent révéler des références « culturelles sous-jacentes. C'est par exemple le cas du nombre « magique » 7 qui arrive en tête. Lorsqu'il s'agit de se coordonner, 40% des joueurs choisissent le chiffre 1 qui sert ici de « point focal ». On remarque que le nombre de réponses différentes s'est réduit, et que la concentration des joueurs sur les deux premiers résultats les plus saillants (44,4% de la population) se traduit par une augmentation significative de la probabilité que deux joueurs tirés au hasard donnent la même réponse (de 5,2% à 20,6%). On remarquera enfin que l'inversion spectaculaire dans le classement du nombre 1 ne se fait pas au détriment du nombre « magique » 7, qui a été choisi par 3 % de sujets de plus dans la population (de 10,2% à 14,4%).

Hors de toute situation de coordination, les individus du groupe de contrôle « P » choisissent leur chiffre préféré. Dans la situation de coordination (le groupe « G »), si les individus avaient cherché à « projeter » sur les autres leurs propres croyances individuelles, où s'ils s'étaient demandé quelles pouvaient être les croyances individuelles des autres, cela aurait conduit, selon Orléan, au choix du chiffre 7, qui est intrinsèquement préféré par le plus grand nombre. Au lieu de cela, nous dit-il, dans cette situation : « *les individus réfléchissent ... en se plaçant à un niveau plus général d'abstraction, de façon à déterminer un principe capable de faire émerger, aux yeux de tous, un équilibre unique* » Ainsi, « *l'activité cognitive est tournée vers le groupe en tant qu'entité séparée* ». Ainsi donc, pour Orléan, la modélisation adéquate du comportement des joueurs correspond bien à la forme auto-référentielle (4) des croyances sociales. Une forme (3) des croyances sociales, basée sur un ensemble d'opinions individuelles aurait conduit selon lui à faire émerger le chiffre 7. Au lieu de cela « la force de la définition (4) tient au fait qu'elle fait totalement abstraction de la variabilité des préférences intrinsèques des individus [contrairement à (3)] pour s'attacher à définir une croyance propre au groupe  $G_s$ , en tant que groupe ». Si cette démarche cognitive est problématique, c'est parce que la dynamique du groupe, en tant que système complexe adaptatif, peut s'écarter durablement de ce que serait celle d'un agent représentatif de la majorité. Le groupe dans son ensemble pourrait ainsi se trouver durablement dans un état contraire à celui souhaité par la majorité de ses membres (Galam, 1986 ; Orléan 1998a-c).

En résumé, dans un jeu de coordination, l'approche standard conduit à une multiplicité d'équilibres de Nash (de deux à une infinité, selon le jeu considéré). Du point de vue *comportemental*, les agents parviennent cependant assez bien à se coordonner, et les expérimentalistes font apparaître des « *saillances à la Schelling* » (ou « point focaux ») dans cet ensemble de solutions, ce qui facilite la sélection d'un équilibre dans des jeux itérés. Le problème du *point focal* est alors de mobiliser des *connaissances extérieures* au cadre « standard » du jeu afin de servir de référence dans la sélection des équilibres (des « stéréotypes »). La « culture » et/ou « l'histoire » des joueurs peuvent ainsi jouer un rôle. Mais à y regarder de plus près, ce sont les « croyances sociales » auto-référentielles qui en sont les supports cognitifs. Ce sont elles qui jouent un rôle décisif. En particulier, en l'absence d'un stéréotype « culturel » pré-existant, la forme auto-référentielle (4) des croyances sociales est certainement la plus appropriée pour faciliter l'émergence par un processus « historique » de tels « points focaux ». Nous allons maintenant essayer de proposer une explication cognitive à cette focalisation sur un objet collectif : « les croyances du groupe », plutôt que la construction d'une chaîne récurrente de références croisées sur l'ensemble des agents, selon l'approche de la « croyance commune ».

### 3. « concours de beauté », hiérarchie cognitive et profondeur stratégique.

Les premiers exemples habituellement cités de la formalisation d'un jeu non coopératif reprenant le raisonnement central du « concours de beauté » de Keynes remontent aux années quatre-vingt (Moulin, 1986, Simonsen, 1988). La première expérimentation publiée basée sur ce jeu est due à Nagel (1995). L'auteur cite ses sources d'inspiration (en particulier Roger Gesnerie) en introduction de l'article, qui se situe clairement dans une perspective *post anticipations rationnelles* (cf. Orléan 2002, sur le lien avec les anticipations rationnelles). Ce jeu se présente comme une « stylisation » du thème évoqué par Keynes. Il évacue en particulier la dimension « culturelle » dont on a eu l'occasion de souligner l'importance au sujet de la coordination sur des « point focaux », ce qui nous permet de nous concentrer sur les mécanismes d'auto-référence et de « *hiérarchie cognitive* » (Camerer 2003, pour une définition dans le champ des jeux expérimentaux, Phan, 2004a, pour une application aux SMA).

#### 3.1. Le cadre du jeu et les résultats de l'approche (standard) par la dominance itérée

Dans le « concours de beauté » de Rosemarie Nagel, un groupe de joueurs, qui ne peuvent pas communiquer entre eux doit choisir un nombre (réel) entre 0 et 100. Le vainqueur du jeu sera celui qui choisira la valeur la plus proche de la moyenne des choix, multipliée par un nombre  $p$  tel que  $p < 0 < 1$  (pour  $p = 1$ , on retrouve le problème de coordination précédent, où les joueurs doivent se coordonner sur un point focal). Si plusieurs joueurs ont choisi cette valeur, ils se partagent également le gain. Prenons  $p = 1/2$  et gain = 900€ pour se fixer les idées.

Suivant Camerer (2003) on peut classer les joueurs dans une *hiérarchie cognitive* selon la « profondeur stratégique » de leur raisonnement. Au « niveau zéro », le joueur joue n'importe quoi sans réfléchir, au « niveau un » le joueur est tenté de jouer 50, qui est la

moitié, mais s'il s'aperçoit que les autres font comme lui, il devrait jouer 50/2 et l'on se trouve alors au « niveau deux ». Si l'on considère maintenant un automate qui aurait été programmé pour appliquer itérativement le principe de la sélection des stratégies dominées, ce dernier va appliquer ce raisonnement de manière récurrente jusqu'à ce que la variation de la solution soit plus petite qu'un seuil fixé à priori (pour éviter une chute sans fin dans la récurrence infinie). Cette série converge vers l'unique équilibre de Nash, égal à zéro. Notons que zéro est aussi la seule solution de l'équation de récurrence linéaire d'ordre un :  $X(t) = 0.5.X.(t-1)$ , qui nous sert à vérifier qu'aucun joueur n'a intérêt à dévier unilatéralement. Supposons que dix joueurs dotés d'une grande profondeur stratégique et appliquant mécaniquement la dominance itérés aient tous trouvé le bon résultat. Ils se partagent alors les 900€, soit 90 € chacun. Que se passe-t-il si un joueur « dévie unilatéralement » et joue par exemple 1 ? La moyenne pondérée sera maintenant de 0,05, qui est plus proche de zéro que 1 (le lecteur vérifiera que ce raisonnement est vrai pour toute valeur positive, même inférieure à un). Les neuf joueurs restant se partageront donc les 900€, ce qui fait 100€ chacun, et le déviant touche zéro !

Il n'y a donc, selon l'approche standard, qu'un seul équilibre de Nash égal à zéro. Les résultats expérimentaux montrent que cet équilibre n'est jamais atteint dans le cadre d'un jeu unique, mais que les joueurs peuvent s'en rapprocher dans une certaine mesure si le jeu est itéré (apprentissage). L'explication la plus simple est que les joueurs n'appliquent pas la méthode de la dominance itérée, ou ne « descendent » pas très profond. Une explication de second niveau suggère que des joueurs peuvent anticiper que d'autres joueurs ne descendront pas très loin. Au troisième niveau, le problème se ramène maintenant à former des croyances sur la profondeur stratégique moyenne des autres joueurs. Mais on pourrait alors dire à un quatrième niveau qu'il suffit de tout le monde croit que les autres croient que le niveau de profondeur stratégique ne soit pas très élevé pour que cette prévision se réalise (c'est le phénomène de prophétie auto-réalisatrice connu chez les sociologues au moins depuis Merton (1948), mais développé largement en économie dans le cadre de la théorie des anticipations rationnelles (cf. par exemple : Azariadis, 1981 ; Azariadis, Guesnerie 1982 cf. aussi pour la psychologie sociale: Snyder, 1984). Qu'en est-il réellement ?

### **3.2. Résultats expérimentaux, hiérarchie cognitive et « croyances sociales »**

Les résultats des (économistes) expérimentalistes montrent qu'en fait, dans un premier temps, un peu plus de 50% des sujets ne descendent pas au dessous du niveau 1 de la hiérarchie cognitive définie précédemment. Les choix de la grande majorité de la population laisse apparaître une profondeur cognitive apparente de 1 ou 2. Si l'on retient notre exemple précédent de  $p = 0,5$ , la distribution des résultats obtenus par Nagel et Camerer est reproduite sur le tableau 3. Dans les expérimentations de Camerer, la valeur modale s'établit à 25. Chez Nagel, le mode correspond au choix de ceux qui se situent au niveau deux, soit la classe [8-10].

Tableau 3 - Hiérarchie cognitive et profondeur stratégique dans le concours de beauté

	Niveau 0	Niveau 1	Niveau 2	Niveau 3
Nagel (1995)	24%	30%	41%	6%
Camerer (2003)	16%	38%	47%	0%

Source : Camerer (2003) p.211 – La somme des deux lignes fait 101% à cause des arrondis utilisés par Camerer

Dans le problème de pure coordination, les joueurs doivent se coordonner entre eux dans des rencontres bilatérales. En l'absence de croyance sociale dominante de type (3) en un stéréotype reconnu, la *subsumption* d'un ensemble de croyances individuelles par une « croyance » de type (4) attribuée à une entité d'emblée « collective » - le groupe - peut améliorer la coordination en facilitant l'émergence d'un point focal. Ici la situation, bien que très proche (un paramètre  $p < 1$  au lieu de  $p = 1$ ) présente immédiatement une dimension collective dans la mesure où la coordination ne doit pas se faire avec un autre choix individuel, mais avec *un objet collectif inclus dans les règles du jeu*, sous multiple de la moyenne du choix de l'ensemble des joueurs du groupe. Par contre, même si une part significative des joueurs ont une hiérarchie cognitive limitée, les joueurs plus « raffinés » qui vont faire des raisonnements de niveau 3 et former des croyances sur la profondeur stratégique moyenne des autres joueurs, ne peuvent faire l'impasse sur la formation d'une croyance sur la part de joueurs dans la population qui descendra également jusqu'à ce niveau de profondeur stratégique, ce qui introduit de la réflexivité dans le processus de formation des croyances. Il faut donc être « un niveau au-dessous » du niveau moyen, mais se garder d'aller en dessous. Les résultats mentionnés par Camerer (2003) à propos des élèves de l'Université de Caltech et d'autres groupes particuliers confirment ce dernier point : la moyenne des résultats obtenus par ces groupes est « un demi niveau » au dessous de celui obtenu en général (comme dans le tableau 3, où la totalité de la population se trouve à un niveau hiérarchique inférieur à trois), ce qui montre qu'une part significative de ces groupes est certainement descendu au « niveau 3 ».

## Conclusion

L'introduction de *croyances* attribuées aux agents en situation d'interactions stratégiques peut conduire à des raisonnements récurifs très éloignés des pratiques de coordination effectives, comme le montre la confrontation des résultats de la théorie des jeux standard avec ceux de l'économie expérimentale. Nous avons montré que l'utilisation de « *raccourcis cognitifs* » tels que l'usage dans le raisonnement d'*entités sociales* pouvait aider à simplifier certains problèmes de coordination. Dans les cas sous revue, c'est précisément parce que la théorie des jeux standard n'utilise pas ce type d'entités que l'on peut aboutir à des résultats en apparence surprenants d'un point de vue pragmatique, comme la multiplicité d'équilibres, ou des raisonnements récurifs infinis. Un usage de ces *raccourcis cognitifs* consiste en effet à subsumer la représentation d'un ensemble d'entités par une « *entité sociale* » unique qui permet alors de limiter la profondeur stratégique des raisonnements, la complexité du problème et/ou la quantité d'information à traiter par les agents. Cette subsumption pourrait également donner l'illusion de *favoriser la compréhension* de l'agent en remplaçant d'un ensemble collectif au fonctionnement complexe par une entité unique dotée d'attributs

anthropomorphiques porteurs de sens. En conséquence, d'une part, la subsumption favorise la coordination par une *simplification des processus sélectifs*, mais elle n'est pas sans conséquences, d'autre part, sur l'issue de ces processus. En effet, les dynamiques auto-référentielles associées aux *entités sociales* ainsi introduites dans le processus jugement peuvent conduire à des dérives collectives ou encore stabiliser le système dans des situations sous-optimales, en particulier lorsque les dynamiques collectives produisent des effets complexes qui rendent la fiction commode de l'entité représentative inopérante ou même fallacieuse.

Dans les deux classes jeux considérés, la connaissance qu'ont les joueurs de « l'arrière plan culturel » des autres joueurs peut être considérée comme un réducteur d'incertitude, qui aide (le plus souvent) à la coordination, à travers le support de plusieurs niveaux de « croyances collectives ». Ces dernières facilitent l'élimination des solutions qui apparaissent alors comme peu (ou pas) pertinentes et la convergence éventuelle vers un point focal. La question de l'*émergence* de telles croyances reste cependant posée.

L'apparition d'un ensemble de comportements similaires au sein d'une population d'individus disposant d'un indicateur leur permettant de s'identifier mutuellement, peut être à la source de la formation de « croyances collectives » de ce type. L'étude de la nature de l'émergence d'un *comportement partagé* chez des ensembles d'agents dotés des signes distinctifs extérieurs et les conditions de formation de croyance associées à ce phénomène émergent sont traités dans des travaux complémentaires qui portent sur les cadres cognitifs de l'observation et de la détection des phénomènes émergents (Phan 2004b, Dessalles, Phan, 2005). Plus spécifiquement, il s'agit d'identifier les conditions selon lesquelles une approche en terme de « *croyances individuelles partagées* » pourrait conduire par « *émergence* » (et *subsumption*) à une approche en terme de « *croyances sociales* ». On montre alors que l'émergence de « croyances sociales » au sens défini précédemment peut être modélisée grâce à la conceptualisation / formalisation d'une autre forme de « hiérarchie cognitive », conçue pour formaliser à la fois des phénomènes « d'*émergence forte* » (au sens de Muller, 2002) et des processus d'observation et de *conceptualisation* de ces phénomènes par « *réduction de la complexité* » (Dessalles, 1992, Dessalles, Bonabeau, 1997).

## Bibliographie

- Anderson P.W., Arrow K.J., Pines D. eds. (1988) *The economy as an evolving complex system*, Santa Fe Institute, Studies on the Sciences of Complexity, Addison-Wesley Pub.Co, Reading Ma.
- Amy B., Bonabeau E. Dessalles J.-L. Ducret J.-J. Giacometti A., Grumbach A. (1992) *Emergence dans les modèles de cognition*, séminaire ENST 15-16 Avril S003, ENST Paris.
- Arthur W.B., Durlauf S.N., Lane D.A. eds. (1997) *The Economy as an Evolving Complex System II*, Santa Fe Institute, Studies on the Sciences of Complexity, Addison-Wesley Pub.Co, Reading Ma .
- Azariadis, C. (1981), "Self-Fulfilling Prophecies", *Journal of Economic Theory*, 25, p. 380-396
- Azariadis C., Guesnerie R. (1982) « Prophéties créatrices et Persistance des Théories », *Revue Economique*, 33, 5, p. 787-806
- Blume L.E. (1997) "Population Games", in : Arthur, Durlauf, Lane eds., op cit. , p. 425-460.
- Bonabeau E., Dessalles J.-L. (1997) "Detection and Emergence", *Intellectica*, 25 p.89-94.

- Boudon R. (1995) *Le juste et le vrai, Etudes sur l'objectivité des valeurs et de la connaissance*, Fayard, Paris.
- Boudon R. (1997) « L'explication cognitiviste des croyances collectives », in Boudon, Bouvier, Chazel dir. *Cognition et Sciences Sociales*, Presses Universitaires de France, Paris, p. 19-54.
- Boudon R. (2003) *Raison, bonnes raisons*, Col. « Philosophe en sciences sociales », Presse Universitaires de France, Paris.
- Bourgine P., Nadal J.P. eds. (2004) *Cognitive Economics*, Springer Verlag
- Camerer C. (2003) *Behavioural Game Theory: Experiments on Strategic Interactions*, Princeton University Press.
- Descombes V. (1996) *Les institutions du sens*, Les éditions de minuit, Paris
- Dessalles J.L. (1992) « Aspects cognitifs de l'émergence » in Amy *et al.* op cit., p.47-59
- Dessalles J.L., Phan D. (2005) "Emergence in multi-agent systems: cognitive hierarchy, detection, and complexity reduction" (*in preparation*)
- Douglas M. (1986) *How Institutions Think*, Syracuse University Press. trad. Fr. *Comment pensent les institutions*, La Découverte, Paris, 1999
- Durkheim E. (1895) *Les règles de la méthode sociologique*, rééd. Presses Universitaires de France, Paris, 1973
- Ferber J., Phan D. (2005) "Cognitive Hierarchy in Multi-Agent Games: formal, conceptual and artefactual aspects" (*in preparation*)
- Galam S. (1986) "Majority rule, hierarchical structure and democratic totalitarianism", *Journal of Mathematical Psychology*, 30, p.426.
- Gilbert N. (1995) "Emergence in social simulation", in Gilbert, Conte eds. *Artificial Societies, The computer simulation of social life*, UCL Press, London, p.144-189.
- Gilbert M. (1990) "Rationality, coordination and conventions", *Synthese* 84, p.1-21.
- Gilbert M. (2000) « A propos de la socialité : le sujet pluriel comme paradigme » in Livet, Ogien, op cit. p.107-126.
- Gilbert M. (2003) *Marcher ensemble Essais sur les fondements des phénomènes collectifs*, Col. « Philosophe en sciences sociales », Presse Universitaires de France, Paris.
- Kaufmann L. Clement F. (2003) « Les formes élémentaires de la vie sociale » séminaire « *Formes symboliques* » : [http://formes-symboliques.org/article.php3?id\\_article=49](http://formes-symboliques.org/article.php3?id_article=49)
- Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*, London: Macmillan, trad. française : *Théorie générale de l'emploi, de l'intérêt et de la monnaie*, Payot, Paris, 1939.
- Lesourne J., Orléan A., Walliser B. dir. (2002) *Leçons de microéconomie évolutionniste*, Odile Jacob, Paris.
- Lewis D.K. (1969) ; *Convention, a philosophical Study*, Harvard University Press Cambridge Mass
- Livet P., Ogien R. eds. (2000) *L'enquête ontologique, du mode d'existence des objets sociaux*, Editions de l'EHESS, Paris.
- Moulin H. (1986) *Game theory for social science*, New York press, NY.
- Müller J.P. (2002) « Des systèmes autonomes aux systèmes multi-agents : interactions, émergence et systèmes complexes » *Rapport présenté pour l'obtention de l'HDR*, Université de Montpellier, novembre, 58 p.
- Orléan A. (1998a) "Informational Influences and the Ambivalence of Imitation", in Lesourne, Orléan (eds.), *Advances in Self-Organization and Evolutionary Economics*, p. 39-56.
- Orléan A., (1998b) "The Ambivalent Role of Imitation in Decentralised Collective Learning", in Lazaric, Lorenz (eds.), *Trust and Economic Learning*, Elgar Publishers, p. 124-140.
- Orléan A. (1998c) "The Evolution of Imitation", in Cohendet, Llerena, Stahn, Umbhauer (eds.) *The Economics of Networks. Interaction and Behaviours*, Springer-Verlag, p. 325-339.
- Orléan A. (1999) *Le pouvoir de la finance*, Odile Jacob, Paris
- Orléan A. (2002) « Le tournant cognitif en économie », *Revue d'Economie Politique* N°5 p. 717-738.
- Orléan A. (2004) "What is a Collective Belief" in Bourgine, Nadal, op.cit., p.199-212.

- Mehta J., Starmer C., Sugden R. (1994a) "Focal Point in Pure Coordination Games: An Experimental Investigation", *Theory and Decision*, 36 p.163-185.
- Mehta J., Starmer C., Sugden R. (1994b) "The nature of Salience: An Experimental investigation of Pure Coordination Games" *American Economic Review*, 84, p. 658-673.
- Merton R. K. (1948) "The self-fulfilling prophecy", *Antioch Review*, 8, p. 193-210..
- Nagel R. (1995) "Unravelling in Guessing Games: An experimental Study", *American Economic Review*, 85, p. 1313-1326.
- Pettit P. (1993) *The Common Mind, An essay on Psychology, Society, and Politics* Oxford University Press
- Pettit P. (2000) « Défense et définition du holisme social » in Livet, Ogien, op.cit p.65-85.
- Pettit P. (2004) *Penser en société, Essais de métaphysique sociale et de méthodologie*, Col. « penser en sciences sociales », Presse Universitaires de France, Paris.
- Phan D. (2004) "From Agent-Based Computational Economics towards Cognitive Economics" in Bourguine, Nadal op.cit., p. 371-398.
- Phan D. (2004a) "Hierarchy of cognitive interactive agents and statistical mechanics : how Object Oriented Programming highlights the connection", in Coelho, Espinasse. Seidel eds., *5th Workshop on Agent Based Simulation*, Lisbon, Portugal, May 2004 SCS Pub. House, Erlangen, San Diego, p. 69-76.
- Phan D. (2004b) "L'émergence dans les Systèmes Multi-Agents (SMA) : Modéliser et interpréter l'émergence avec des SMA, Support de discussion pour le séminaire « émergence » du GDR « économie cognitive », CREA, Paris, 4 juin et pour le séminaire « Complexité » de l' I.H.P.S.T, Paris, 2 décembre. <http://perso.univ-rennes1.fr/denis.phan/slides/emergenceSMA.pdf>
- Phan D., Gordon M.B, Nadal J.P. (2004) « Social Interactions in Economic Theory: an Insight from Statistical Mechanics » in Bourguine, Nadal op cit. p. 225-358.
- Schelling T.S. (1960) *The strategy of Conflict*, Harvard University Press, Cambridge, Ma.. trad. française : *La stratégie du conflit*, PUF 1986.
- Schuster H.G. (2001) *Complex Adaptive Systems, An Introduction*, Scator Verlag, Saarbrücken.
- Searle J.R. (1995) *The Construction of Social Reality*, Free Press ; trad. française : *La construction de la réalité sociale*, NRF-essais, Gallimard, Paris.
- Simonsen M.H. (1988) "Rational Expectations, Game Theory and Inflationary Inertia" in Anderson, Arrow , Pines eds. *The economy as an evolving complex system*, Santa Fe series Vol5 Addison-Wesley CA., p. 205-241
- Snyder, M. (1984) "When belief becomes reality", in Berkowitz ed., *Advances in experimental social psychology*, 18, Orlando, FL : Academic Press, p. 247-305..
- Weisbuch G. (1989) *Dynamique des Systèmes Complexes : une introduction aux réseaux d'automates*, InterEditions / CNRS