

BULLETIN N° 107
ACADÉMIE EUROPÉENNE INTERDISCIPLINAIRE
DES SCIENCES



Séance du Mardi 12 septembre 2006
Conférence de notre Collègue Saadi LAHLOU :
"Représentations, cognition, technique.
Le double mécanisme de l'évolution culturelle"

Prochaine séance : le Mardi 10 octobre 2006

Assemblée générale de l'AEIS
Conférence du Pr. Bernard WALLISER:
"Cognition et émergence"

ACADEMIE EUROPEENNE INTERDISCIPLINAIRE DES SCIENCES
FONDATION DE LA MAISON DES SCIENCES DE L'HOMME

PRESIDENT : Michel GONDRAN
SECRETAIRE GENERAL : Irène HERPE-LITWIN
SECRETAIRE GENERAL ADJOINT : Noëlle CAGNARD
TRESORIER GENERAL : Bruno BLONDEL
CONSEILERS SCIENTIFIQUES :
SCIENCES DE LA MATIERE : Pr. Gilles COHEN-TANNOUDJI.
SCIENCES DE LA VIE ET BIOTECHNOLOGIES : Pr. François BEGON
PRESIDENT DE LA SECTION DE NICE : Doyen René DARS
PRESIDENT DE LA SECTION DE NANCY : Pierre NABET

PRESIDENT FONDATEUR
DOCTEUR Lucien LEVY (†).
PRESIDENT D'HONNEUR
 Gilbert BELAUBRE
SECRETAIRE GENERAL D'HONNEUR
 Pr. P. LIACOPOULOS

Jun 2006

N°107

TABLE DES MATIERES

- P. 3 Compte-rendu de la séance du 12 septembre 2006 avec la conférence de notre nouveau collègue, Saadi LAHLOU «*Représentations, cognition, technique. Le double mécanisme de l'évolution culturelle* »
 P. 5 Compte-rendu de la section Nice Côte d'Azur du 20 juillet 2006
 P. 8 Documents

Prochaine séance : Mardi 10 octobre 2006
 MSH, salle 215 à 18h
Assemblée générale de l'AEIS
Conférence du Pr. Bernard WALLISER :
«Cognition et émergence''

ACADEMIE EUROPEENNE INTERDISCIPLINAIRE DES SCIENCES
Fondation de la Maison des Sciences de l'Homme, Paris.

Séance du
Mardi 12 septembre 2006

Fondation de la Maison des Sciences de l'Homme, salle 215, à 18 h.

La séance est ouverte à 18 h. 00 sous la Présidence de Michel GONDRAN et en présence de nos collègues, Bruno BLONDEL, Alain CARDON, Françoise DUTHEIL, Jean Pierre FRANCOISE, Manuel GALAN, Irène HERPE-LITWIN, Jacques LEVY, Pierre MARCHAIS,

Etaient excusés : Gilbert BELAUBRE, Gilles COHEN-TANNOUDJI, Marie Louise LABAT, Gérard LEVY, Pierre SIMON

I) Informations générales

Notre Président nous fait un bref compte-rendu de la cérémonie de remise de médaille d'officier de la légion d'honneur au Pr. Benoît MANDELBROT le 11 septembre 2006 à l'Ecole Polytechnique à Palaiseau par le Sénateur Pierre LAFFITTE, camarade de promotion de ce dernier. Cette cérémonie, marquée par les interventions de MM. Les Professeurs Jean-Pierre KAHANE, membre de l'Académie des Sciences (mathématiques), Nassim TALEB (Economie) et Bernard SAPOVAL (Physique) , et Benoît MANDELBROT a accueilli une très vaste assistance. Elle a également permis de présenter le livre consacré aux Actes du congrès de novembre 2004, « Fractales en progrès » , édité par les éditions de l'AEIS sous le titre « *Irruption des géométries fractales dans les sciences* » .

Par ailleurs, il nous informe d'une possibilité de visite du site du CERN à Genève fin octobre 2006 ou en janvier 2007 (à confirmer).

Le mardi 10 octobre à 18heures, salle 215 de la Maison des Sciences de l'Homme , 54 boulevard Raspail 75006 PARIS , se tiendra **l'Assemblée générale annuelle de l'Académie**.

II) Conférence de notre collègue Saadi LAHLOU : «Représentations,cognition, technique. Le double mécanisme de l'évolution culturelle »¹

Avant d'aborder le thème proprement dit de sa conférence, notre nouveau collègue nous donne un bref tracé de son parcours très interdisciplinaire de statisticien économiste ayant évolué vers la biologie et les sciences sociales, et travaillant actuellement chez EDF pour le service de Recherche et Développement. Son cheminement l'a conduit à s'interroger sur les mécanismes d'évolution des représentations dans une population humaine.

Première assumption : il n'existe pas d'individu isolé de son environnement. Il faut donc raisonner sur des populations humaines dans un contexte ici et maintenant. Il y aurait donc au niveau des représentations quasiment l'émergence d'une espèce vivante supérieure, d'un superorganisme résultant d'une symbiose entre les aspects biologiques, symboliques et artefacts matériels .

Ce superorganisme est en évolution permanente à l'image des mutations qui surviennent dans les espèces vivantes . Il est également soumis à des lois de sélection: certaines représentations sont sélectionnées, d'autres sont abandonnées. Notamment il existe un lien fonctionnel entre les objets physiques et leurs représentations aboutissant à des lois d'évolution technique des objets représentés. Il existe des interactions rétroactives entre représentations sociales et production des objets matériels.

Un nouveau concept d'importance est issu : celui d' « affordance » qui caractérise ce qui est fourni par l'environnement.

Pour finir, notre conférencier nous donne une application à la gestion de l'environnement dans la gestion des tâches dans un grand groupe industriel.

Quelques questions surgissent :

- Cette modélisation ne présente-t-elle pas un caractère téléonomique ?
- Ne présente-t-elle pas des points communs avec la systémique ?
- Différence entre objets et concepts ?

Après ces discussions, la séance est levée à 20 heures.

Bien amicalement à vous,

Irène HERPE-LITWIN

¹ Pour plus de détails nous vous invitons à relire les articles du bulletin n°106 relatifs à ce domaine.

Compte -Rendu de la Section Nice-Côte d'Azur

Le savoir est le seul bien qui s'accroisse à
le partager. Comprendre est bien sans
limite qui apporte une joie parfaite.
Baruch SPINOZA (1632-1677)

Compte-rendu de la séance du 20 juillet 2006 (96^{ème} séance)

Présents :

Sonia Chakhoff, Pierre Couillet, Patrice Crossa-Raynaud, Guy Darcourt, René Dars,
Jean-Pierre Delmont, Jean-Paul Goux, Yves Ignazi, Gérard Iooss, Maurice Papo.

Excusés :

Jean Aubouin, René Blanchet, Emile Girard, Jacques Lebraty, Jean-François Mattéi,
Jacques Wolgensinger.

1- Approbation du compte-rendu de la 95^{ème} séance.

Le compte-rendu est approuvé à l'unanimité des présents.

2- Le mois écoulé.

a- Les champs magnétiques.

Une discussion s'instaure au sujet d'un article vantant l'utilisation des champs magnétiques pour soigner les dépressions nerveuses.

Guy Darcourt précise qu'il s'agit encore d'expérimentations et que les résultats sont discutables.

Maurice Papo élargit le débat en rappelant qu'il existe, dans le monde, des quantités de laboratoires qui font de la recherche sur l'influence du magnétisme sur la santé humaine, que ce soit les lignes à haute tension, les antennes de téléphone, les portables, etc. Aux Etats-Unis, pendant un certain temps, se sont vendus des petits appareils destinés à être promenés devant les fours à micro-onde pour vérifier qu'il n'y avait pas de fuite nocive !

Il manque donc à la Science en général une coordination interdisciplinaire pour donner un avis sur ces champs magnétiques, qu'ils soient faibles ou forts.

Or, les recherches actuelles sont faites dans des buts limités précis par des spécialistes. Par exemple, les uns cherchent à prouver que les téléphones portables sont ou ne sont pas dangereux, les autres voudraient démontrer qu'ils peuvent être utiles en psychiatrie. La fameuse « machine »

de Priore, financée un temps par l'Etat, était censée guérir le cancer, sans parler de la « mémoire de l'eau » de M. Benveniste.

Il y a peu, rappelle Delmont, la progression de la Science se faisait uniquement avec des publications sérieuses à comité de lecture. Mais malheureusement, on a ainsi publié ces derniers temps des articles qui se sont révélés être des faux. Les derniers en date étant ceux d'un chercheur coréen sur le clonage. Par contre, on a longtemps refusé les articles sur l'apoptose !

Actuellement, on est inondé par des publications sans contrôle paraissant sur Internet, dans des délais très courts, alors qu'il faut six mois pour une publication dans une revue sérieuse.

Nous rejoignons ainsi le sujet de notre colloque sur « les peurs » car si l'on retire précipitamment un médicament dont certains effets se révèlent dangereux (Vioxx), ou un insecticide qui tue les abeilles (Régent), la presse s'en empare et le public n'a plus confiance dans ses savants. Rappelons ce que le Docteur Dor écrit dans notre colloque sur la faible fiabilité des publications sur les médicaments financées par les laboratoires.

b- Relations entre le Conservatoire National des Arts et Métiers (CNAM) et l'Université.
(Yves Ignazi)

Une réunion a été organisée avec la participation de Daniel Toulouse qui est responsable de la recherche au CNAM afin de définir comment on peut établir des partenariats, des actions en commun entre les deux organismes.

Notre Académie interdisciplinaire est bien dans l'air du temps ...

c) L'arc en ciel. (Pierre Coulet)

Avec l'aide du Conseil Général, il est prévu de créer un lieu d'attraction où, en plein jour, on puisse créer des arcs en ciel, par un jeu de miroirs réfléchissant le soleil vers un mur d'eau. Cela se produit tous les jours à la cascade du château ou à Gairaut lorsque l'orientation est bonne.

Cette attraction serait couplée à un petit musée consacré à la lumière et qui aurait bien sa place sur la Côte d'Azur.

d) L'Institut Robert Hooke.

❖ Il est prévu, au CUM, l'an prochain, six conférences sur « les origines de la vie » avec des personnalités de premier plan invitées.

L'Académie est associée à ces manifestations.

❖ Le Ministère de l'Education Nationale souhaite développer des liaisons entre l'enseignement supérieur universitaire et des « collèges ambition réussite ». Dans notre département, ce sont trois collèges des zones prioritaires, l'Ariane, les Moulins, Bon Voyage qui sont retenus. Ce projet intéresse beaucoup Pierre Coulet car il veut faire participer ses étudiants à cette action.

3- Les Amis de la Fondation Sophia Antipolis.

L'Association Sophia Antipolis (A.S.A.), qui est à l'origine, sous la présidence de M. Pierre Laffitte, de la première technopole de notre pays, se transforme, sur une idée de notre confrère Maurice Papo, membre du Conseil d'Administration de l'A.S.A., en une Association des Amis de la Fondation Sophia Antipolis. Cette nouvelle structure aura pour objet, entre autres, d'appuyer et de rendre plus visible l'action de la Fondation Sophia Antipolis, qui est une Fondation de recherche reconnue d'utilité publique. Il nous paraît important que nous aidions cette

initiative soit individuellement, soit en tant que personne morale, susceptible de conseiller scientifiquement l'Association des Amis de la Fondation Sophia Antipolis en tant que de besoin.

4- Projet de colloque sur la modélisation et la simulation.

Ce colloque, qui est d'une actualité évidente, sera organisé dans le cadre du laboratoire Dieudonné de la Faculté des Sciences par Pierre Couillet et Gérard Iooss en 2007.

Le programme retenu concerne 22 laboratoires de l'Université. Notons qu'il y a beaucoup de domaines de modélisation à côté des modélisations élémentaires. On peut ainsi modéliser la cicatrisation tissulaire avec des biologistes, le trafic routier, l'influx nerveux, le management, etc.

Pour Crossa-Raynaud, il serait nécessaire d'évoquer la modélisation des phénomènes naturels qui implique de très nombreux facteurs dont on connaît mal l'importance relative et surtout leurs influences réciproques. Les prévisions varient fortement d'un modèle à l'autre. Annoncer par exemple que la température moyenne sur la Terre augmentera d'un degré C° (réchauffement mesuré pour le dernier siècle) à six degrés en 2100 nous incite à penser que l'on n'en sait rien.

Prochaine réunion

le jeudi 21 septembre 2006 à 17 heures
au siège
Palais Marie-Christine
20 rue de France
06000 NICE

Entreprise et mondialisation

le mardi 26 septembre 2006 à 18 heures
au CUM
1^{ère} conférence du cycle

L'insertion du bassin méditerranéen dans la mondialisation

par Albert Marouani,
Président de l'Université de Nice-Sophia Antipolis

Présence de tous indispensable à ce cycle de conférences dont nous sommes les initiateurs.

Invitez vos amis

Documents

En vue du futur colloque sur l'Émergence nous vous proposons différents articles de Bernard WALLISER , Professeur d'Économie à l'École Nationale des Ponts et Chaussées et à l'EHESS , dont les travaux dans le domaine font autorité :

P . 9 : « *Cognition et émergence* » Un texte de Bernard WALLISER résumant ses propres thèses

P . 10 : « *Phénomènes émergents.* » par Bernard WALLISER

P. 24: « *Les justifications des notions d'équilibre de jeux* » par Bernard WALLISER

Cognition et émergence

(B. Walliser, ENPC, EHESS)
10 octobre 2006

Tout comme les sciences de la nature, la théorie des jeux et la théorie économique s'interrogent sur l'émergence de phénomènes collectifs à partir de comportements individuels. Les phénomènes émergents prennent la forme de distributions statistiques particulières des propriétés des agents, de réseaux de relations durables entre les agents ou encore d'entités originales comme les institutions. En sciences de la nature, deux distinctions classiques sont traditionnellement à l'œuvre quant aux phénomènes émergents: l'émergence ascendante (passage du niveau micro au niveau macro) ou bilatérale (avec retour du niveau macro vers le niveau micro), l'émergence synchronique (coexistence instantanée des niveaux micro et macro) ou diachronique (émergence du niveau macro à partir du niveau micro). En sciences sociales, le phénomène est encore compliqué par le fait que les agents peuvent se faire une représentation mentale du social, qui va agir sur leurs comportements.

Traditionnellement, deux approches sont utilisées en théorie des jeux pour justifier l'apparition d'un état d'équilibre entre les agents, un tel équilibre pouvant être interprété comme une structure émergente. L'« approche éductive » suppose que les agents ont une rationalité très forte qui s'exerce instantanément à partir d'hypothèses croisées (« je sais que tu sais que tu sais ... ») sur leurs connaissances ou leurs comportements respectifs. Ils peuvent se coordonner sur un état d'équilibre de façon consciente et voulue en anticipant leurs actions réciproques par un processus de simulation mentale. L'« approche évolutionniste » suppose que les agents ont une rationalité limitée, mais se rencontrent en temps réel et profitent de leur expérience passée, c'est-à-dire mettent en œuvre des règles variées d'apprentissage. Ils peuvent aboutir à un état d'équilibre de façon inconsciente et spontanée, celui-ci n'étant autre qu'un état asymptotique du processus d'apprentissage.

Ces deux visions sont extrêmes, l'une supposant des agents hyper-intelligents, l'autre des agents purement réactifs. Elles ont été appliquées à d'autres structures que les équilibres, par exemple les structures d'échanges marchands ou les institutions économiques (en particulier la monnaie). Des approches intermédiaires sont en construction, qui considèrent que les agents, en situation passive d'apprentissage, sont néanmoins capables de mettre en œuvre des capacités cognitives plus fortes (par exemple en matière de catégorisation de leur environnement).

LES PHENOMENES EMERGENTS

(Bernard WALLISER, ENPC, EHESS)

mars 2006²

Résumé : Les phénomènes émergents sont des phénomènes macroscopiques imprévisibles voire inexplicables, qui prennent la forme de régularités statistiques, de structures relationnelles ou même d'entités originales. L'émergence unilatérale naît de la seule influence des attributs et des relations des entités microscopiques alors que l'émergence bilatérale suppose une action en retour du phénomène macroscopique sur le niveau microscopique. L'émergence synchronique se produit lorsque les influences ascendante et descendante entre niveaux s'exercent simultanément alors que l'émergence diachronique allie une influence ascendante à court terme et descendante à long terme. Pour les systèmes sociaux, l'émergence spontanée est à l'œuvre lorsque les agents sont inconscients des phénomènes qu'ils contribuent à forger alors que l'émergence réflexive postule la prise en compte par les agents des phénomènes qu'ils créent.

A différents niveaux d'organisation du monde naturel et social, des entités microscopiques qui interagissent dans un certain contexte donnent naissance à des phénomènes macroscopiques saillants et robustes. Ces 'phénomènes émergents' prennent des formes diverses, allant de simples combinaisons des propriétés des entités microscopiques ou des relations entre ces entités à l'apparition de caractéristiques ou d'entités macroscopiques originales. Les phénomènes émergents les plus manifestes qui se sont exprimés au cours de l'évolution de l'univers physique et social sont indéniablement l'apparition de la vie, l'irruption de la pensée et la genèse d'institutions (langage, monnaie). Cependant, on peut observer une variété de phénomènes émergents d'allure plus modeste comme les turbulences dans les fluides, les amas cellulaires, les colonies d'insectes ou les prix des biens. On peut dès lors s'interroger sur le point de savoir si ces phénomènes relèvent ou non de modèles conceptuels semblables.

Bien que repérables et descriptibles par l'observateur, les phénomènes émergents restent difficilement explicables et prévisibles au regard des propriétés et des relations des entités sous-jacentes. Dans une perspective holiste, les phénomènes émergents sont des régularités macroscopiques dont l'observateur ne peut rendre compte à partir des seules caractéristiques du niveau microscopique. Dans une perspective individualiste, ces mêmes phénomènes sont considérés comme réductibles, en droit sinon en fait, à travers une méta-relation liant le niveau microscopique et le niveau macroscopique. Cependant, selon que cette méta-relation entre les niveaux est purement ascendante ou aussi descendante, deux formes d'émergence peuvent être distinguées. L'«émergence unilatérale» suppose que le niveau microscopique engendre à lui seul les phénomènes macroscopiques alors que l'«émergence bilatérale» naît d'une boucle entre les niveaux microscopique et macroscopique.

² Je remercie M. Kistler et R. Topol pour leurs nombreuses observations qui ont permis d'améliorer ce texte. Ce dernier a également bénéficié des remarques des participants au colloque 'prospective' du CNRS (2003), du séminaire de philosophie économique de l'Université d'Aix-Marseille (2004), du séminaire 'complexité' de l'IHPST (2005), ainsi que des travaux de l'atelier 'structures émergentes' réuni dans le cadre du GDR 'économie cognitive'

Les phénomènes émergents possèdent naturellement une dimension temporelle dans la mesure où les niveaux d'organisation sont progressivement apparus au cours de la genèse de l'univers. Dans une perspective de court terme, les phénomènes émergents apparaissent comme simultanés à leur soubassement microscopique et peuvent dès lors être représentés comme le résultat d'une forme d'équilibre. Dans une perspective de long terme, ces mêmes phénomènes sont le résultat du fonctionnement d'un niveau microscopique préexistant et peuvent être conceptualisés par un processus d'évolution. Là encore, selon le temps est pris en compte de façon explicite ou non, deux types d'émergence peuvent être distingués. L'«émergence synchronique» suppose que les niveaux microscopique et macroscopique sont simultanément présents alors que l'«émergence diachronique» montre comment le niveau macroscopique est progressivement issu du substrat microscopique.

Les phénomènes émergents acquièrent une dimension nouvelle au niveau des acteurs humains dans la mesure où ces derniers peuvent prendre conscience des régularités qu'ils contribuent à créer. Dans une perspective spontanée, les acteurs agissent de façon inconsciente par rapport au processus d'émergence, mais peuvent néanmoins reconnaître le résultat collectif émergent de leurs actions et y réagir en le renforçant ou en l'inhibant. Dans une perspective constructive, les acteurs reconnaissent la réalité du processus qui conduit à un phénomène émergent, et peuvent ainsi participer à sa régulation voire à sa construction volontaire. Finalement, selon le degré de réflexivité qui est attribué aux acteurs, deux modalités d'émergence sont distinguées. L'«émergence spontanée» suppose que les acteurs microscopiques interagissent de façon purement causale alors que l'«émergence réflexive» suppose que les entités microscopiques prennent conscience de certains aspects du phénomène émergent.

1. Ontologie des phénomènes émergents

1.1. *Niveaux d'emboîtement*

L'ontologie classique de l'univers naturel et social articule classiquement des niveaux d'organisation emboîtés, chaque niveau étant caractérisé par des entités modulaires, définies par des propriétés originales. Chaque entité d'un niveau, si elle est suffisamment autonome et stable, constitue une totalité qui regroupe comme 'briques élémentaires' des entités du niveau inférieur. Se succèdent ainsi des atomes munis de propriétés physiques, des molécules munies de propriétés chimiques, des cellules munies de propriétés biologiques, des organes/tissus munis de propriétés physiologiques, des organismes/individus munis de propriétés psychologiques et des collectivités munies de propriétés sociales. La succession des niveaux est universellement acceptée, les 'transitions' entre niveaux procédant (alternativement) par juxtaposition d'entités homogènes ou par intégration d'entités hétérogènes

Cette hiérarchie soulève de délicats problèmes de frontières entre niveaux et manifeste l'existence de bifurcations en hiérarchies parallèles. D'abord, des courts-circuits peuvent apparaître entre niveaux non consécutifs, une entité de niveau n faisant appel à titre principal ou secondaire à des entités de niveau $n-2$. Ainsi, si un organisme biologique peut être directement formé de cellules indifférenciées, un groupe humain peut incorporer des objets matériels. Ensuite, des structures intermédiaires peuvent apparaître entre niveaux consécutifs, une entité de niveau n dépendant du niveau $n-1$ à travers un pseudo-niveau. Ainsi, si les molécules se regroupent en cristaux ou en colloïdes, les individus se regroupent en syndicats ou en villes. Enfin, des hiérarchies distinctes peuvent débiter et se poursuivre à partir des mêmes entités de base. Ainsi, des molécules sont regroupées en cellules, mais aussi en étoiles (puis en groupements d'étoiles) ou en pièces (puis en machines).

Chaque niveau d'emboîtement possède son propre langage de description qui caractérise, à l'aide de concepts originaux, les propriétés des entités pertinentes. Si les 'propriétés intrinsèques'

permettent de les définir et de les différencier, les ‘propriétés extrinsèques’ permettent de déterminer leur comportement ultérieur dans leur milieu. Ainsi, définis par leur masse volumique ou leur dureté, les métaux se manifestent par leur coefficient de dilatation ou leur valence; définis par leurs caractères morphologiques ou leurs modes de locomotion, les animaux s’expriment par leur régime alimentaire ou leur mode de reproduction. Ces grandeurs sont d’autant plus faciles à expliciter que les entités sont denses, au sens où leurs liaisons internes sont plus importantes que leurs liaisons externes (‘quasi-décomposabilité’ au sens de Simon). C’est ainsi que les particules, les atomes, les molécules sont liées par des forces d’interaction de plus en plus faibles qui permettent de mieux ‘séparer’ les niveaux.

Chaque niveau d’emboîtement possède ses propres techniques d’investigation pour apprécier les propriétés des entités concernées. Si les ‘grandeurs observables’ sont directement évaluées sur des échelles appropriées à l’aide d’instruments de mesure, les ‘grandeurs cachées’ sont révélées qu’à partir des précédentes en s’appuyant sur des définitions théoriques. Ainsi, les atomes électrisés sont détectés par un ampèremètre, les molécules sont soumises à des analyses chimiques, les cellules sont examinées au microscope, le cerveau est exploré par l’imagerie cérébrale, l’individu est soumis à des questionnaires et la société est mise en statistiques. Ces propriétés sont d’autant plus faciles à mesurer que les entités concernées sont plus isolables de leur contexte. Ainsi, contrairement aux molécules, à certaines cellules (organismes unicellulaires) ou aux individus (ermites) qui existent de façon autonome, les atomes sont rarement dans un état isolé et les organes renvoient toujours à un organisme.

1.2. Niveau microscopique

A un certain niveau (niveau microscopique), dans un certain contexte, des entités élémentaires entrent en interaction avec d’autres entités de même niveau, pour former un système de niveau supérieur (niveau macroscopique). Les entités de base, leurs relations mutuelles et le contexte global sont les trois composantes qui sont censées épuiser la description d’un système et exprimer la transition entre deux niveaux. Ces trois composantes ont une importance relative qui peut varier d’un système à l’autre, les relations pouvant l’emporter sur les propriétés individuelles ou inversement. Ces trois composantes ne sont pas indépendantes, ainsi lorsque les relations sont engendrées par des propriétés individuelles ou lorsque les propriétés individuelles sont influencées par le contexte global. Enfin, ces trois composantes ne sont pas intrinsèques au niveau d’organisation considéré, car elles peuvent fort bien être déjà influencées par l’appartenance du système à une entité de niveau supérieur.

Les entités de base, homogènes ou hétérogènes quant à leurs propriétés essentielles, sont caractérisées, au sein d’une population, par des ‘lois de répartition’ des entités selon ces propriétés. Les propriétés sont décomposables en ‘propriétés structurelles’, liées au comportement propre des entités, et en ‘propriétés relationnelles’, s’exprimant dans l’interaction des entités avec d’autres entités. Ainsi, les génomes ont une organisation propre, mais possèdent une valeur adaptative en relation avec d’autres; les consommateurs sont dotés de préférences ou d’anticipations propres, mais ont des réputations dans leur rapport aux autres. De plus, les propriétés des entités de base peuvent être indépendantes au départ ou manifester une corrélation préalable du fait de l’insertion des entités dans un environnement plus vaste. Ainsi, si les atomes peuvent voir leur spin couplé par une magnétisation préalable, les individus peuvent être pré-coordonnés par le moule scolaire.

Les relations mobilisées, de diverses natures et plus ou moins symétriques, sont définies à partir des propriétés des entités de base et de réseaux permanents par des ‘lois d’interaction’. Les relations sont décomposables en ‘relations sub-induites’ dérivant des propriétés relationnelles des entités, en ‘relations autonomes’ définies de façon exogène et en ‘relations super-induites’, résultant de contraintes globales. Ainsi, si les liaisons électriques entre atomes résultent de leurs charges électriques ou les relations sexuelles entre animaux de leurs caractéristiques hormonales, les cellules sont ‘préprogrammées’ pour certaines associations ou les individus sont conditionnés

par leur milieu social. De plus, les entités sont ancrées sur un ‘substrat’ préalable, qu’elles soient distribuées aléatoirement ou disposées sur un réseau de structure variable (ligne, cercle, treillis, sphère, tore, ...). Ainsi, les molécules d’un corps peuvent être piégées dans des structures réticulaires qui les accueillent et les acteurs individuels peuvent être disposés dans une ville à laquelle ils sont a priori attachés.

Le contexte global, qui admet des composantes diversifiées et agit de façon plus ou moins isotrope, peut être défini en fonction de grandeurs non forcément reliées aux entités de base par des ‘lois de contexte’. Ce contexte peut être décomposé en un ‘contexte englobant’ qui correspond aux entités de niveau supérieur dont dépend le système et un ‘contexte externe’ qui traduit des influences en provenance d’entités extérieures au système. Ainsi, si une cellule de la peau est influencée par l’organisme auquel elle appartient, elle subit aussi des agressions extérieures; si un individu est conditionné par son environnement social, il subit aussi des influences de son environnement écologique. De plus, le contexte peut agir de façon épisodique ou constituer un milieu encadrant permanent. Ainsi, si les atomes peuvent être soumis à un champ extérieur provisoire qui polarise leurs actions, les animaux peuvent être immergés dans un milieu naturel fixe qui oriente leurs comportements.

1.3. Niveau macroscopique

Au niveau supérieur (niveau macroscopique), le système formé d’entités microscopiques en interaction donne naissance à des propriétés plus ou moins originales. Parmi les propriétés macroscopiques, les ‘phénomènes émergents’ sont dotés de caractéristiques particulières du point de vue de l’observateur. Une première caractéristique concerne la «saillance» du phénomène, qui se manifeste comme un ‘pattern’ contrasté et original, directement extrait de l’appréhension du milieu sous-jacent. Une seconde caractéristique concerne la «robustesse» du phénomène, qui se présente comme un ‘invariant’ durable et systématique, malgré les variations des entités microscopiques et du contexte global. Si les attributs d’un phénomène émergent restent très subjectifs, ces phénomènes peuvent néanmoins être classés en trois types selon leur degré de détachement du substrat microscopique.

Tout d’abord, un phénomène émergent est ‘statistique’ s’il traduit une distribution particulière des propriétés des entités microscopiques, qu’il s’agisse de propriétés spatiales (localisation) ou qualitatives (couleur). Ainsi, des points lumineux font apparaître des couronnes d’interférence; des particules magnétisées font apparaître des lignes de champ; des animaux se regroupent en bancs de poisson ou en vols d’hirondelles; des individus engendrent des distributions spécifiques de revenu (distribution de Pareto) ou d’habitat (zones urbaines). En particulier, les ‘lois d’échelle’ relient, sur un ensemble d’entités microscopiques, une grandeur (quantitative) quelconque à sa taille à travers une fonction puissance. Ainsi, la fréquence des mots dans le langage peut être reliée à la longueur des mots ou la performance des villes dans un pays peut être reliée à la population de ces villes.

Ensuite, un phénomène émergent est ‘structurel’ s’il se présente sous la forme d’un réseau de relations nouvelles qui s’établissent entre les entités microscopiques, qu’il s’agisse de relations spatiales ou qualitatives. Ainsi, des molécules se combinent en polymères parfaitement structurés; des neurones s’associent pour former des boucles de mémorisation; des animaux sociaux se regroupent en ruches ou en fourmilières; des acheteurs et des vendeurs d’un bien entretiennent des relations de fidélité ou des foules se déchaînent sur certains espaces. En particulier, des ‘structures hiérarchiques’ peuvent apparaître entre les entités de base, dans lesquelles certaines entités privilégiées sont dotées d’un pouvoir de contrôle sur d’autres. Ainsi, le système nerveux est un système hiérarchique arborescent dominé par le cerveau qui commande aux mouvements du corps alors que la structure décisionnelle d’une entreprise est pilotée par des dirigeants qui contrôlent les agents.

Enfin, un phénomène émergent est ‘qualitatif’ s’il se manifeste sous la forme de propriétés nouvelles directement définies au niveau macroscopique, qu’il s’agisse d’attributs macroscopiques originaux ou d’entités originales. Ainsi, des molécules d’eau peuvent créer des tourbillons de Bénard dans un liquide confiné; le cerveau permet de construire une vision des formes ou des couleurs; les individus engendrent des normes sociales comme les normes de politesse ou des conventions de coordination comme la priorité à droite. En particulier, des ‘holons’ apparaissent comme des entités autonomes, qui ont des frontières nettes avec leur milieu, qui s’auto-entretiennent en puisant sur leur milieu et qui même sont capables de s’auto-reproduire. Ainsi, la cellule est une entité qui a une identité propre et est à même de se dupliquer alors que l’organisme végétal ou animal a une autonomie propre qui lui permet de vivre en autarcie relative.

2. Explication des phénomènes émergents

2.1. *Modes d’explication*

Du point de vue du modélisateur et non plus du seul observateur, le problème posé consiste à rendre compte d’un phénomène macroscopique à partir des entités microscopiques, de leurs relations et du contexte global. De ce point de vue, les phénomènes émergents sont certes des phénomènes macroscopiques, mais ils font l’objet d’un traitement particulier du fait de la présence de caractéristiques spécifiques. Une première caractéristique est l’« inexplicabilité » du phénomène, au sens où il ne peut être justifié a posteriori par un schéma explicatif satisfaisant. Une seconde caractéristique est l’« imprévisibilité » du phénomène, au sens où il n’a pu être prédit a priori par une voie reconnue. Cependant, ces traits communs traduisent un état de l’art à une certaine époque et expriment des impossibilités de fait plus que de droit, si bien qu’elles peuvent toujours être considérées comme transitoires.

Toute explication d’un phénomène émergent fait appel à un principe de « réduction », qui affirme précisément que l’on peut rendre compte d’un phénomène d’un niveau donné par les seules propriétés des entités du niveau inférieur et de leurs relations. Cette réduction suppose d’abord que les grandeurs macroscopiques peuvent être exprimées à partir des attributs des entités de base et des relations entre ces entités. Elle suppose ensuite que les lois ou les phénomènes macroscopiques peuvent eux-mêmes s’exprimer à partir des lois de répartition, des lois d’interaction et des lois de contexte. La réduction est ‘naturaliste’ si le niveau humain est ramené à des propriétés des systèmes naturels ou ‘psychologique’ si le niveau social est rabattu sur des propriétés individuelles. C’est ainsi qu’on s’efforce de réduire la valence chimique à un niveau atomique, le potentiel d’un gène à un niveau moléculaire, la douleur à un niveau neurophysiologique ou la ségrégation ethnique à un niveau individuel.

Toute explication d’un phénomène émergent peut être soumise plus avant à un principe de «survenance» («*supervenience*») (Davidson, 19), qui affirme que les propriétés microscopiques sont suffisantes, mais non forcément nécessaires pour rendre compte d’un phénomène macroscopique. Plus précisément, alors qu’une combinaison de propriétés microscopiques donne toujours naissance au même phénomène macroscopique, ce dernier peut résulter de différentes combinaisons de propriétés microscopiques (multi-réalisabilité du phénomène émergent). Par contraposition, deux phénomènes macroscopiques différents ne peuvent être obtenus que par des propriétés microscopiques différentes alors que différentes propriétés microscopiques peuvent conduire au même phénomène macroscopique. Ainsi, une même cellule peut résulter de molécules différentes ou un même organisme peut résulter de cellules différentes (qui d’ailleurs se renouvellent sans cesse).

Toute explication d’un phénomène émergent renvoie enfin à un principe d’ «individualisme méthodologique», qui affirme encore que l’on peut rendre compte d’un phénomène d’un niveau donné par les seules caractéristiques du niveau inférieur. En fait, des versions de plus en plus

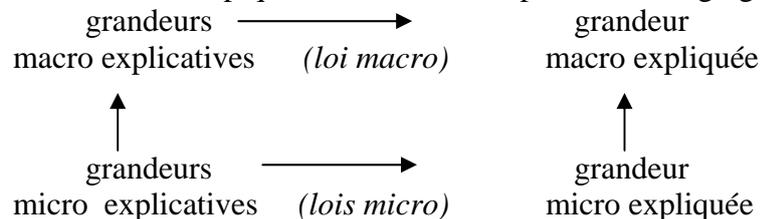
faibles du principe d'individualisme méthodologique ont été proposées. D'abord, le phénomène macroscopique peut être expliqué par les seuls attributs et relations des entités microscopiques ou nécessiter la prise en compte du contexte. De façon plus fine, l'explication du phénomène peut ne prendre en compte que le contexte externe ou nécessiter aussi la prise en compte du contexte englobant. Ainsi, au niveau socio-économique, si les phénomènes collectifs sont expliqués à partir des comportements individuels, du contexte matériel ainsi que du contexte institutionnel, les institutions peuvent elles-mêmes être expliquées par les actions individuelles pour se ramener à un individualisme plus strict.

2.2. Emergence unilatérale

L'émergence unilatérale suppose que tout phénomène macroscopique peut être expliqué exclusivement à partir des propriétés des entités, des relations entre ces entités et du contexte externe. La méta-relation qui lie les deux niveaux d'organisation successifs considérés (et non plus les entités de base) est alors une relation à sens unique. Il s'agit de déduire de lois générales (lois de répartition, d'interaction, de contexte) et de conditions particulières (état initial du système, paramètres des lois) des phénomènes macroscopiques spécifiques. Cela suppose en particulier que les relations entre entités individuelles sont sub-induites ou autonomes, mais non super-induites. Ainsi, les particules ont des charges électriques qui ne dépendent pas du champ électrique lui-même et les individus ont des états mentaux (en particulier des préférences) qui ne sont pas influencés par le milieu social.

Le passage d'un niveau à l'autre passe par la considération de 'principes de pontage' (*'bridge principles'*) qui lient une grandeur macroscopique à des grandeurs microscopiques. Même s'ils peuvent traduire des principes théoriques ou des corrélations empiriques, ces principes doivent être considérés comme de pures définitions, qui ne sont valables que dans des contextes précis. Cependant, certaines grandeurs sont directement définies au niveau macroscopique et considérées comme non réductibles aux propriétés microscopiques. Dans la théorie cinétique des gaz, la température du gaz peut être définie comme l'énergie cinétique moyenne des molécules constituantes; en revanche, le volume du gaz est directement défini par l'enceinte qui contient le gaz. Dans la théorie macroéconomique, la consommation globale d'un bien est simplement définie par la somme des consommations individuelles; en revanche, le prix du bien est défini d'emblée au niveau global.

En s'appuyant sur les principes de pontage, il est possible, dans certains cas, de dériver une loi macroscopique des lois microscopiques en vertu d'une 'procédure d'agrégation' :



En théorie cinétique des gaz, la loi de Boyle-Mariotte (qui relie la température, la pression et le volume du gaz) résulte des propriétés cinétiques des molécules de gaz. En théorie macroéconomique, la fonction de consommation globale (qui relie la consommation totale du bien aux prix et au revenu total) résulte des comportements individuels de consommation.

La focalisation de l'explication sur deux niveaux successifs repose sur l'hypothèse implicite de «séparabilité» entre niveaux, à savoir qu'il n'existe pas d'influence directe entre des niveaux non successifs. Plus précisément, si deux niveaux sont séparés par un niveau intermédiaire, les entités du niveau inférieur influencent les entités du niveau supérieur exclusivement par le biais des entités intermédiaires. Ainsi, les perceptions du cerveau sont expliquées à partir des propriétés des neurones constituants, elles-mêmes expliquées à partir des propriétés électriques et chimiques

des molécules qui les composent, sans avoir à faire intervenir de relation directe entre les molécules et les perceptions. De même, les comportements économiques des agents sont expliqués par leurs croyances et leurs préférences, ces états mentaux étant eux-mêmes expliqués par les propriétés neuro-physiologiques du cerveau, sans avoir à faire intervenir de relation directe entre le cerveau et les comportements.

2.3. *Emergence bilatérale*

L'émergence bilatérale considère que tout phénomène macroscopique doit être expliqué à partir des propriétés des entités, des relations entre ces entités, du contexte externe ainsi que du contexte englobant. La méta-relation qui lie les deux niveaux d'organisation successifs est alors une relation à double sens formant une boucle. Le sens 'ascendant' (*'bottom up'*) décrit classiquement comment les propriétés des entités microscopiques se combinent pour donner naissance à des propriétés macroscopiques. Le sens 'descendant' (*'top down'*) montre, de façon moins classique, comment les propriétés macroscopiques conditionnent en retour les propriétés des entités microscopiques. En particulier, les relations entre entités individuelles peuvent cette fois être super-induites. Ainsi, un corps chimique est transformé par les produits même de la transformation à laquelle il participe (autocatalyse), un individu évolue du fait même du milieu social qu'il a contribué à modifier.

Le point fondamental concerne les canaux par lesquels le niveau supérieur peut agir sur le niveau inférieur. En effet, seules les relations entre entités microscopiques ont été considérées jusque là comme de véritables relations actives. De fait, si ces relations sont sub-induites, la relation du niveau macroscopique sur le niveau microscopique est juste une agrégation de ces relations qui peut faire intervenir une grandeur agrégée. Si ces relations sont super-induites, on suppose qu'il peut exister une influence directe du niveau macroscopique sur le niveau microscopique, qui est irréductible à des relations bilatérales entre entités microscopiques. Ainsi, des particules électrisées vont définir conjointement un champ électrique et chaque particule va réagir à ce champ comme elle réagirait en moyenne à l'ensemble de ses congénères. De même, des individus vont définir conjointement des normes sociales et chacun va réagir à cette norme comme il réagirait en moyenne à ses condisciples.

Le principe d'«auto-organisation» évoque souvent cette possibilité de relation bouclée entre deux niveaux d'organisation successifs. Les entités microscopiques entretiennent des relations bilatérales tant matérielles qu'informatives incluant des composantes aléatoires. Leurs interactions font apparaître un phénomène émergent qui exerce une contrainte en retour sur les entités considérées. Les entités microscopiques sont à l'origine de caractéristiques émergentes qui rétroagissent instantanément sur ces mêmes entités. Ainsi, lors du développement de l'embryon, les cellules se démultiplient progressivement à partir d'une cellule initiale et constituent un groupement qui exerce des contraintes spatiales et physiques sur ses membres. De même, les travailleurs s'agglomèrent peu à peu à partir d'un germe pour former un syndicat qui exerce des contraintes en retour sur ses adhérents.

Là encore, on suppose une séparabilité entre les niveaux d'organisation, au sens où il n'existe pas d'influence directe entre des niveaux non successifs, dans un sens ou un autre. Si deux niveaux sont séparés par un niveau intermédiaire, les entités du niveau inférieur influencent les entités du niveau supérieur exclusivement par le biais des entités intermédiaires et inversement. En particulier, il est supposé qu'il n'existe pas d'action en retour du phénomène macroscopique sur des entités situées à un niveau infra-microscopique. Ainsi, le fonctionnement du cerveau n'exerce pas d'influence sur les molécules constituantes du cerveau sans transiter par les neurones qui tissent la matière cervicale. De même, les phénomènes sociaux n'exercent pas d'influence sur les caractéristiques physiologiques des agents concernés, sans transiter par les états mentaux de ces mêmes agents.

3. Temporalité des phénomènes émergents

3.1. Echelles temporelles

Pour tenir compte des transformations subies par tout système, des ‘échelles temporelles’ d’occurrence des phénomènes sont introduites. Si à chaque échelle sont associées des grandeurs spécifiques qui possèdent des vitesses d’évolution semblables, les échelles sont imbriquées, au sens où les grandeurs sont classées des plus rapides aux plus lentes. De plus, les grandeurs de chaque échelle influencent les grandeurs des échelles voisines de façon asymétrique selon des temporalités elles-mêmes emboîtées. La situation la plus simple comprend trois échelles temporelles, respectivement liées à des grandeurs rapides, lentes et fixes, et définissant deux régimes. Si seules les grandeurs rapides se modifient à ‘court terme’ (sous l’influence des grandeurs rapides passées et des grandeurs lentes), les grandeurs lentes peuvent elles-mêmes s’ajuster à ‘long terme’ (sous la contrainte des grandeurs fixes).

Les échelles temporelles des systèmes sont en correspondance plus ou moins biunivoque avec les niveaux d’emboîtement de ces mêmes systèmes. Les grandeurs rapides sont généralement associées aux propriétés et aux relations des entités microscopiques, qui se trouvent engagées dans des processus parallèles d’apprentissage. Les grandeurs lentes sont plutôt relatives au phénomène macroscopique qui subit une adaptation progressive aux modifications des entités sous-jacentes. Ainsi, en mécanique, des billes peuvent se déplacer à court terme à l’intérieur d’un bol tandis qu’à long terme, la forme du bol peut elle-même varier sous l’influence des billes ou d’autres facteurs. En économie, les quantités de biens offertes et demandées s’adaptent aux prix fixés à court terme tandis que les prix évoluent eux-mêmes à long terme sous l’influence des offres et demandes ou de facteurs extérieurs.

Le principe de réduction s’avère d’autant plus aisé à satisfaire que le phénomène macroscopique résulte d’une dynamique plus affirmée. En effet, il est dès lors possible de suivre la manière dont le phénomène prend naissance sans avoir à supposer qu’il est déjà là de tout temps. Ainsi, la réduction naturaliste peut être soutenue par une compréhension fine de la manière dont la conscience est apparue et s’est développée chez les primates supérieurs. En revanche, la survenance peut poser problème dès lors que des entités microscopiques en interrelation donnent naissance à des ‘solutions multiples’ au niveau macroscopique. Il s’agit d’états d’équilibre multiples dans un cas statique ou d’attracteurs suivant des bifurcations dans un cas dynamique. Ainsi, des prix de marché distincts peuvent être obtenus instantanément à partir des mêmes offres et demandes ou résulter dynamiquement d’un même processus d’apprentissage (phénomène d’hystérésis ou de ‘dépendance au chemin’).

Pour expliquer des phénomènes émergents apparus à différents niveaux d’organisation de l’univers, il est possible de faire appel à des classes analogiques de modèles s’appliquant à des champs disciplinaires différents. Les ‘analogies structurales’ se contentent de mettre en relation les concepts de base et les lois relationnelles concernant les deux champs, tout en modifiant leur interprétation. Les ‘analogies substantielles’ vont jusqu’à postuler que des mécanismes explicatifs homologues s’avèrent pertinents dans les deux champs, avec des interprétations semblables. Tout naturellement, la démarche analogique conduit à utiliser les mêmes outils formels dans les deux champs. Ces outils introduisent pour la plupart des aléas sur diverses grandeurs ou relations, aléas dont l’interprétation peut ou non changer en passant d’un champ à un autre. Ils doivent répondre à des difficultés communes, en particulier de faire apparaître des phénomènes nouveaux sans supposer qu’ils sont potentiellement déjà là.

3. 2. *Emergence synchronique*

L'«*émergence synchronique*», la plus anciennement étudiée, postule qu'un phénomène émergent est 'structurellement' explicable par une boucle reliant instantanément les niveaux macroscopique et microscopique. Comme on s'est placé dans le cas le plus général de l'*émergence bilatérale*, les échelles temporelles sont semblables dans le sens ascendant et dans le sens descendant de la méta-relation:



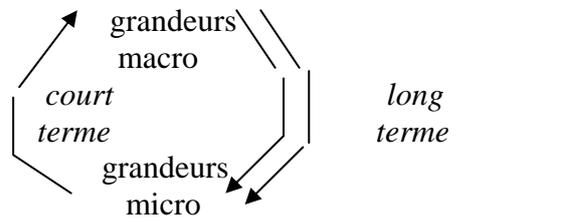
Le 'bouclage' entre les niveaux microscopique et macroscopique ne peut se stabiliser qu'en un point fixe de la boucle, à savoir un état d'«*équilibre*». En un tel état, le système n'est pas enclin à dévier en l'absence de perturbations extérieures, ce qui rend le phénomène émergent auto-renforçant. Cette émergence n'en soulève pas moins un double problème: le 'problème d'implémentation' apparaît du fait de la non explicitation du processus par lequel s'effectue le bouclage, le 'problème de sélection' apparaît dès lors que plusieurs états d'équilibre sont possibles. En physique statistique, les spins des particules (grandeurs micro) interagissent et engendrent un potentiel; les atomes réagissent alors globalement à ce potentiel (grandeur macro) plutôt que localement à chacun de leurs voisins. En microéconomie, les producteurs et les consommateurs expriment des offres et des demandes de biens (grandeurs micro) en fonction du système de prix (grandeur macro); ces prix sont fixés par une entité fictive, le 'commissaire-priseur walrasien', de façon à ajuster les offres et les demandes.

Dans ce processus, des propriétés symétriques au niveau des entités microscopiques peuvent conduire à un phénomène asymétrique au niveau macroscopique. Ainsi, des propriétés symétriques des atomes induisent des 'brisures de symétrie' au niveau du corps chimique alors que des choix mimétiques réalisés par des firmes entre des technologies qu'elles jugent équivalentes peuvent conduire à l'adoption finale par toutes d'une même technologie. Ensuite, des relations structurellement discrètes au niveau des entités microscopiques peuvent conduire à un phénomène qualitativement continu au niveau macroscopique. Ainsi, des transmissions électriques discontinues entre neurones donnent naissance à des perceptions parfaitement continues alors que des achats et ventes discrètes au niveau des agents élémentaires peuvent provoquer des changements continus des prix (mais aussi des ruptures brutales).

Les modèles statiques d'*émergence* d'un phénomène sont essentiellement de deux types. Les 'modèles d'optimisation' décrivent le phénomène comme résultant de la maximisation d'une grandeur globale unique définie à partir des caractéristiques des entités microscopiques. Ainsi, en physique, une configuration spécifique des spins des particules d'un corps résulte de la minimisation d'un 'hamiltonien' (qui traduit l'énergie globale du système); en économie, la configuration d'un réseau social peut résulter de la maximisation d'une 'fonction d'utilité collective' (qui accepte d'agréger les utilités individuelles). Les 'modèles d'équilibration' rendent compte du phénomène par l'équilibre résultant de la boucle d'interaction qui le relie aux entités microscopiques. Ainsi, en physique, l'ordre émergent d'un système de petites cubes aimantés résulte de l'équilibre entre les forces qui les relient; en économie, la distribution des prix pratiqués par divers producteurs résulte d'un équilibre de concurrence imparfaite entre ces producteurs et ces consommateurs.

3.3. *Emergence diachronique*

L'«*émergence diachronique*», plus récemment étudiée, suppose qu'un phénomène émergent est explicable par un processus séquentiel d'adaptation entre les niveaux macroscopique et microscopique. Cependant, toujours dans une perspective d'émergence bilatérale, ces influences obéissent à des temporalités différentes, à court terme dans le sens ascendant, à long terme dans le sens descendant :



Un tel processus fait apparaître des états transitoires (assimilés à des équilibres à court terme) et d'éventuels états asymptotiques (assimilés à des équilibres à long terme). Les deux problèmes soulevés en statique sont résolus en dynamique : le problème d'implémentation par l'explicitation du processus d'équilibration, le problème de sélection par la convergence du processus, s'il converge, vers un attracteur (éventuellement en probabilité). En biologie évolutionnaire, les animaux se reproduisent à court terme en fonction de leur capacité adaptative (*'fitness'*); cette capacité adaptative se modifie à long terme en fonction de l'évolution du milieu naturel et de la démographie animale. En microéconomie du travail, chaque employé prospecte à court terme sur le marché du travail et accepte un nouveau poste si le salaire offert est supérieur à son salaire de réservation; il adapte, à long terme, son salaire de réservation vers le haut ou le bas en fonction du salaire auquel il aurait pu accéder.

Dans ce processus, des comportements parfaitement réversibles au niveau microscopique peuvent engendrer peu à peu un phénomène irréversible au niveau macroscopique. Ainsi, des molécules ayant des lois de mouvement réversibles donnent naissance à des gaz évoluant dans le sens de l'entropie croissante ou des échanges de titres parfaitement réversibles sur un marché financier donnent naissance à des crashes boursiers sans possibilité de retour. De même, des évolutions saltistes au niveau microscopique peuvent conduire à un phénomène temporellement continu au niveau macroscopique. Ainsi, des impulsions électriques discontinues au niveau des neurones donnent naissance à des images mentales continues ou des déplacements individuels discrets au niveau des individus conduisent à des mouvements de foule continus.

Les modèles dynamiques d'émergence sont encore de deux types. Les 'modèles de diffusion' supposent qu'un certain médium (objet, trait personnel) peut être transmis d'une entité à une autre, chaque entité ayant une certaine probabilité de l'adopter ('degré d'acceptation') en fonction de ses caractéristiques propres, ce qui conduit à l'émergence d'une distribution limite du médium. Ainsi, si un microbe peut se transmettre à un membre d'une population selon sa réceptivité, une opinion (information, innovation) peut se diffuser aux membres d'un groupe selon sa pertinence pour eux. Les 'modèles d'adaptation' supposent que les entités se transforment individuellement face à leur environnement en fonction de leurs résultats passés ('index de performance'), ce qui conduit à l'émergence de règles de comportements adaptées. Ainsi, des animaux chasseurs peuvent s'adapter à leur milieu de façon à converger vers une norme de partage d'une proie; des propriétaires et des métayers peuvent co-évoluer de façon à dégager une norme de partage des récoltes.

4. Réflexivité dans les phénomènes émergents

4.1. Réflexivité des acteurs

Lorsque l'on progresse dans les niveaux d'organisation, les propriétés des systèmes évoluent vers une plus grande complexité et subtilité. Les entités de base deviennent plus adaptatives et

conscientes. Ainsi, si les atomes ont des comportements complètement rigides, les organes sont susceptibles de s'adapter à leur milieu, les animaux acquièrent une certaine finalité et les humains sont capables de raisonnements stratégiques. Les relations deviennent plus informationnelles et symboliques. Ainsi, si les relations entre atomes sont essentiellement matérielles, les relations entre cellules ont un contenu plus informatif et les relations entre humains passent par un langage qui soulève des problèmes d'interprétation. Les contraintes globales deviennent enfin plus abstraites et incitatives. Ainsi, on passe de contraintes physiques incontournables à des contraintes sociales molles qui peuvent être contournées moyennant un certain coût.

Tout acteur humain est doté d'intentionnalité au sens où il est capable de représenter dans son esprit un objet qui lui est extérieur. Il se forge des croyances primaires sur son environnement matériel, mais aussi sur les autres acteurs et sur lui-même. Tout acteur humain est de plus doté de réflexivité au sens où il peut analyser sa propre intentionnalité comme celle d'autrui. Il adopte des croyances secondaires sur ses propres croyances primaires (croyances auto-hiérarchiques) comme sur celles de l'autre (croyances hétéro-hiérarchiques). Les relations entre les acteurs sont symboliques au sens où ces acteurs s'expriment à partir d'un cadre conceptuel qu'ils partagent. Toute communication véhicule un contenu direct sur leur environnement commun, mais possède aussi une signification indirecte concernant leur rapport. Ces relations sont aussi stratégiques au sens où les acteurs agissent en anticipant le comportement des autres. Toute communication vise à faire passer de l'émetteur au récepteur un message qui crée une situation favorable au premier.

Le passage du niveau individuel au niveau social se décompose alors en deux transitions complémentaires. La 'transition comportementale' fait passer des états mentaux d'un acteur à ses actions, la 'transition systémique' fait passer des actions des acteurs aux phénomènes sociaux. Ainsi, la méta-relation ascendante comme la méta-relation descendante peuvent être de nature matérielle ou informative. L'intentionnalité d'un acteur peut lui faire prendre conscience du phénomène émergent qu'il crée et la réflexivité d'un acteur peut lui faire prendre conscience de la conscience que prennent les autres de ce phénomène. Cette prise de conscience est instantanée si elle concerne le processus d'émergence lui-même ou différée si elle ne concerne que son résultat. Le principal phénomène social émergent concerne les institutions, en particulier les 'croyances collectives' (croyances partagées par tous) et les 'normes sociales' (règles de comportement suivies par tous).

Le transfert d'un modèle d'émergence d'un système simple à un système complexe de niveau supérieur n'est valide que sous des conditions restrictives. D'abord, les propriétés complexes peuvent être négligeables au regard des propriétés déjà présentes dans les systèmes simples. Ainsi, la chute d'un chat ou d'un homme des étages d'un immeuble est indépendante de leurs propriétés biologiques ou psychiques. Ensuite, les propriétés complexes peuvent s'avérer formellement réductibles à des propriétés de systèmes simples, en dépit d'interprétations différentes. Ainsi, la trajectoire de la lumière, l'orientation d'une plante et le comportement d'un joueur de billard obéissent tous à un même modèle d'optimisation, moyennant des mécanismes sous-jacents différents. Enfin, des propriétés complexes peuvent être réductibles à des propriétés simples, dans certains contextes et avec certains aménagements. Ainsi, une population animale et une société humaine évoluent selon des principes de transmission, de mutation et de sélection, bien que les mutations soient plus dirigées dans le second cas.

4.2. Emergence spontanée

L'émergence spontanée suppose que les acteurs ont un comportement essentiellement réactif, entretiennent des relations principalement matérielles, et surtout ne prennent pas conscience du phénomène macroscopique qu'ils contribuent à engendrer. Ces agents participent à un processus généralement dynamique, qui est qualifié d'«évolutionniste» dans la mesure où il correspond à une évolution conjointe du comportement des acteurs. Il peut converger vers un 'équilibre d'actions', équilibre qui est auto-renforçant car il est dans l'intérêt des acteurs de le maintenir. Cet

équilibre est interprété comme un phénomène émergent non intentionnel, à savoir un phénomène non voulu, non anticipé et même non reconnu. Il en est ainsi de l'apparition progressive de phénomènes socio-institutionnels comme une langue naturelle (chinois), un marché particulier (marché médiéval du blé) ou une monnaie spécifique (coquillage).

Se contenter d'une analyse en termes d'émergence spontanée repose sur les trois types de justifications énoncés. D'abord, les capacités intentionnelles et réflexives des agents peuvent être considérées comme négligeables pour le phénomène considéré. Ainsi, dans un contexte complexe où leur rationalité est limitée, les acteurs mettent en œuvre des règles de prévision et de décision très simples. Ensuite, les acteurs peuvent obéir à des schémas de comportement simples, moyennant une interprétation sophistiquée. Ainsi, un acteur peut être amené à en imiter d'autres parce qu'il pense qu'ils en savent plus que lui. Enfin, les capacités supérieures des agents sont supposées réductibles à des capacités plus modestes s'ils disposent d'un environnement favorable. Ainsi, les acteurs ont peu de difficultés d'interprétation de leur environnement s'ils ont une catégorisation et une vision communes du système social.

Sans prendre conscience du déroulement même du processus d'émergence, les acteurs peuvent néanmoins prendre conscience de son résultat. A travers un 'processus de naturalisation', les acteurs vont reconnaître le phénomène émergent, puis conditionner leur comportement sur lui plutôt que sur les autres acteurs. Le phénomène émergent est isolé de son substrat et s'avère robuste dans la mesure où il résulte d'un équilibre auto-renforçant. Par un 'processus de normalisation', les acteurs vont légitimer le phénomène émergent, voire lui donner une reconnaissance légale. Le phénomène émergent finit par imposer sa pertinence aux agents et se trouve stabilisé par des contraintes sociales. Ainsi, la norme sociale 'conduire à droite' découle d'un processus d'interaction dynamique entre conducteurs, est reconnue comme une entité distincte et vient leur servir de référence, puis se trouve légitimée par eux et fait même l'objet d'une loi.

Les modèles d'émergence utilisés en sciences sociales sont souvent analogues aux modèles dynamiques proposés dans les sciences physiques ou biologiques. Les modèles de contagion se décomposent en 'modèles de diffusion' où une caractéristique extérieure aux acteurs se propage et en 'modèles d'agglomération' où les acteurs eux-mêmes se déplacent ou se regroupent. Ainsi, la formation d'une ville résulte de la pure attraction des individus vers certains sites du fait de rendements croissants ou de leur propension à se rapprocher en vertu d'externalités positives. Les modèles d'adaptation se décomposent en 'modèles d'apprentissage' où des caractéristiques des acteurs s'adaptent au contexte et en 'modèles d'évolution' où les acteurs eux-mêmes subissent un processus de sélection. Ainsi, si un crash boursier peut advenir d'un comportement purement réactif des agents à la séquence passée des prix des actifs, il résulte plus sûrement d'un comportement mimétique des agents qui alignent leurs comportements les uns sur les autres.

4.3. Emergence réfléchie

L'émergence réfléchie suppose que les acteurs ont un comportement finalisé, entretiennent des relations surtout symboliques fondées sur des croyances croisées, et prennent conscience du processus bouclé qui conduit au phénomène macroscopique. Ces agents participent à un processus aussi bien statique que dynamique, qualifié d'«éductif» dans la mesure où il correspond à un raisonnement sophistiqué des acteurs. Il peut converger vers un 'équilibre de croyances', qui s'avère 'auto-réalisateur' dans la mesure où sa conception par les acteurs provoque ipso facto sa réalisation et son maintien. Cet équilibre est encore interprété comme un phénomène émergent, cette fois intentionnel, au sens où il est sinon voulu, du moins anticipé et reconnu par les acteurs. Il en est ainsi de la détermination consciente de phénomènes socio-institutionnels comme une langue artificielle (esperanto), un marché nouveau (marché financier d'options) ou une monnaie originale (euro).

Entreprendre une analyse en termes d'émergence réfléchie pose la question du statut des croyances simultanées des acteurs. Les croyances sont 'collectives' quand tous les acteurs partagent les mêmes croyances (directes et croisées) sur un environnement extérieur. De ce point de vue, une croyance commune (je sais que tu sais que je sais ...) est collective au sens elle se compose d'un empilement de croyances croisées qui est le même pour tous. Les croyances sont 'sociales' quand elles engagent les relations mêmes entre les acteurs et peuvent être attribuées à une entité sociale qui les regroupe. Elles forment alors des institutions auxquelles tous les acteurs adhèrent par différents mécanismes. Cependant, on considère parfois que les croyances collectives n'existent pas pour le modélisateur, mais n'expliquent que dans la tête des acteurs. Ainsi, il se peut qu'aucun acteur ne croit en la rentabilité d'un actif financier, bien que chacun croit qu'il s'agit d'une croyance sociale.

Cas limite d'émergence réfléchie, l'«émergence volontaire» suppose que le phénomène émergent résulte d'un processus de choix explicite entre plusieurs acteurs. Elle s'applique déjà à une montre, un corps chimique ou un organisme génétiquement modifié. Mais elle concerne surtout les institutions qui sont décidées par un groupe restreint avant d'être appliquées à un groupe plus large. Les opportunités sont traduites par un ensemble d'institutions possibles limité par diverses contraintes (faisabilité, acceptabilité sociale, coût de mise en place). Les croyances portent sur l'impact que vont avoir les institutions sur les actions des agents et leurs conséquences conjointes. Les préférences sont définies sur les effets attendus des institutions et sont agrégées en cas de décision collective. C'est ainsi qu'en économie, la littérature sur la 'conception des mécanismes' (*'mechanism design'*) postule l'existence d'un planificateur omniscient dont le rôle est de mettre en place des mécanismes institutionnels susceptibles d'inciter les agents à réaliser un optimum social.

Les modèles d'émergence utilisés en sciences sociales sont des modèles originaux faisant intervenir explicitement les croyances et les raisonnements des acteurs. Les croyances considérées peuvent se limiter à des anticipations sur des faits que les acteurs contribuent à provoquer. C'est ainsi que la formation d'une ville peut résulter de comportements opportunistes des agents qui anticipent les revenus qu'ils peuvent obtenir isolément en s'y installant voire les effets à long terme des externalités positives qu'ils engendrent. Mais, ces croyances sont le plus souvent des croyances croisées sur leurs caractéristiques et leurs comportements, qui culminent dans une croyance commune. C'est ainsi qu'un crash boursier peut résulter de l'hypothèse faite par chaque acteur qu'il est plus intelligent que les autres dans ses anticipations de prix, c'est-à-dire qu'il raisonne à un niveau supérieur d'anticipations croisées.

Conclusion

Le thème des phénomènes émergents apparaît comme un thème analytiquement peu exploré et donc objet de controverses. Les philosophes insistent sur l'irréductibilité des phénomènes émergents, à savoir qu'en plus d'être non prévisibles a priori, ils sont même impossibles à expliquer a posteriori. Les scientifiques, tout en reconnaissant que les phénomènes émergents sont actuellement non prévisibles, mettent en avant que leur explication est un défi non perdu d'avance. De fait, comme la seconde attitude est plus exigeante que la première, elle s'avère aussi plus constructive. Pour l'instant, elle a surtout été entreprise pour rendre compte de la genèse des phénomènes émergents, en particulier des institutions sociales. Elle doit être poursuivie pour rendre compte, selon des mécanismes éventuellement différents, de leur maintien, de leur transformation ou de leur disparition.

Le thème des phénomènes émergents a l'avantage de soulever un grand nombre de problèmes ontologiques. Il mobilise, de façon transdisciplinaire, les notions de niveaux d'emboîtement, d'échelles temporelles ou de boucles d'interaction. En revanche, le concept de

‘fonction’ comme caractérisant la contribution de la partie au tout, n’est mis en œuvre qu’à certains niveaux. De même, la dimension ‘historique’ des phénomènes, mêlant contingence, nécessité et volonté, n’est introduite que dans les disciplines les plus sociales. Par ailleurs, des concepts originaux sont requis pour rendre compte de chaque niveau organisationnel, concepts qui ne sont pas pertinents avant que les phénomènes émergents ne soient apparus. De même, des mécanismes variés sont introduits pour rendre compte du phénomène émergent, mécanismes qui mettent en œuvre diverses versions du principe d’individualisme méthodologique.

Le thème des phénomènes émergents soulève également un grand nombre de problèmes épistémologiques. Il s’exprime dans nombre de modèles formels qui sont chargés d’expliquer localement l’origine de tels phénomènes. Cependant, les données d’observation issues du terrain ou de l’expérimentation sont très variables d’un niveau à l’autre. De même, les principes théoriques se modifient selon les disciplines, en particulier lorsqu’on en vient à introduire la réflexivité humaine. D’ores et déjà, des modèles exploratoires très généraux sont proposés pour rendre compte de façon générale de l’émergence, fondés par exemple sur l’idée d’une baisse brutale de ‘complexité systémique’ en changeant de niveau. Enfin, le modèle que construit le modélisateur du phénomène émergent peut différer profondément du modèle attribué à l’acteur, bien que tous deux soient contraints par des considérations de ‘complexité algorithmique’.

Bibliographie

- Binmore, K. (1987-88) : Modeling rational players, *Economics and Philosophy*, 3 : 179-214, 4 : 9-59.
- Davidson, D. ()
- Dessalles, J.-L.- Phan, D. (2005) : Emergence in multi-agent systems, in P. Mathieu ed : *Artificial Economics*, Springer.
- Holland, J. (1998) : *Emergence, from chaos to order*, Addison Wesley.
- Humphreys, P. (1997) : How properties emerge, *Philosophy of Science*, 64, 1-20.
- Kim, J. (1999) : Making sense of emergence, *Philosophical studies*, 95, 3-36.
- Lane, D.- Maxfield, R. (2005) : Ontological uncertainty and innovation, *Journal of Evolutionary Economics*, 15, 3-50.
- Lewis, D. (1969) : *Convention, a philosophical study*, Harvard University Press.
- Moulines, U. (2000) : Ontologie, réduction et unité des sciences, *Philosophie*, 60, 3-15.
- Muller, J.P. (2002) : Des systèmes autonomes aux systèmes multi-agents : interactions, émergence et systèmes complexes, HDR.
- Rueger, A. (2000) : Robust supervenience and emergence, *Philosophy of science*, 67, 466-489.
- Schelling, T.S. (1978) : Models of segregation, *American Economic Review*, 59(2).
- Schelling, T.S. (1978) : *Micromotives and macrobehavior*, W.W. Norton
- Searle, J.R. (1995) : *The construction of social reality*, Free Press.
- Sethi, R. (1999) : Evolutionary stability and media of exchange, *Journal of Economic Behavior and Organization*, 40(3), 233-54.
- Stephan, A. (2002) : Emergence, in L. Nadel ed. , *Encyclopedia of cognitive science*, 1, 1108-1115.
- Walliser, B. (2005) : Game theory and emergence of institutions, mimeo.
- White, L., Wilson, A., Wilson, D. eds (1969) : *Hierarchical structures*, Elsevier.
- Young, P.H. (1996) : The economics of conventions, *Journal of Economic Perspectives*, 10(2), 105-22.
- Intellectica 25 : émergence et explication

LES JUSTIFICATIONS DES NOTIONS D'ÉQUILIBRE DE JEUX

Bernard WALLISER

Ecole Nationale des Ponts et Chaussées
Ecole des Hautes Etudes en Sciences Sociales

Version finale 13.08.02

***Résumé :** Les notions d'équilibre statiques ou dynamiques proposées en théorie des jeux peuvent être justifiées selon deux perspectives. Du point de vue éductif, les équilibres résultent du seul raisonnement de joueurs hyper-intelligents qui ont une connaissance commune de la structure du jeu et de leurs rationalités respectives. Si l'équilibre rationalisable ou l'équilibre corrélé sont aisément justifiés, l'équilibre de Nash n'est obtenu que sous des conditions très drastiques; quant à l'équilibre parfait, sa justification est très sensible aux hypothèses faites. Du point de vue évolutionniste, les équilibres résultent de la convergence d'un processus d'apprentissage ou d'évolution de joueurs en rationalité limitée, mais observant le déroulement passé du jeu. L'équilibre de Nash, du moins en stratégies pures, est souvent obtenu comme état asymptotique et certains de ses raffinements peuvent même être sélectionnés; l'équilibre parfait est aussi justifié sous des conditions très étendues.*

A l'image de l'équilibre walrassien de la théorie économique, la théorie des jeux est fondée sur diverses notions d'équilibre, qui traduisent chacune la manière dont des acteurs rationnels coordonnent leur actions sur un état doté d'une certaine stabilité. Cependant, toute notion d'équilibre est postulée en surplomb du jeu par le modélisateur, un état d'équilibre étant soumis à la condition nécessaire que, si les acteurs s'y trouvent, ils ne perçoivent aucun intérêt à en dévier unilatéralement. En revanche, aucun processus concret d'atteinte d'un état d'équilibre, qui serait fondé sur les seules délibérations et actions des acteurs sans intervention extérieure, n'est décrit par le modélisateur. Par analogie avec le commissaire-priseur walrassien qui fournit les prix d'équilibre aux agents économiques, on peut certes introduire une entité fictive, le 'régulateur nashien', qui calcule un état d'équilibre de jeu et suggère aux acteurs de l'adopter. Encore faut-il que les acteurs l'adoptent effectivement, ce qui n'est le cas que s'ils ont de bonnes raisons de penser que leurs adversaires l'adopteront aussi, la condition de stabilité postulée à l'équilibre n'étant pas forcément suffisante.

L'« économie cognitive » a pour objet d'étudier les croyances et les raisonnements que les acteurs économiques mobilisent pour s'adapter à des situations dynamiques d'interaction mutuelle (Walliser, 2000). Un de ses thèmes majeurs est de rendre compte de processus concrets par lesquels les acteurs, dotés d'une rationalité tant instrumentale que cognitive, sont susceptibles de se coordonner par leurs propres forces sur un état d'équilibre. Les processus exhibés doivent permettre, dans un même mouvement, de justifier telle ou telle notion d'équilibre posée a priori, et de sélectionner tel ou tel état d'équilibre associé en cas de multiplicité des états. Une première approche vise à donner des « justifications épistémiques » aux équilibres, c'est-à-dire à fonder les équilibres sur les seuls raisonnements d'acteurs autonomes dotés d'une rationalité extrêmement forte. Une seconde approche vise à donner des « justifications évolutionnistes » aux équilibres, c'est-à-dire à faire apparaître les équilibres comme des états asymptotiques de processus dynamiques entre acteurs dotés d'une rationalité très limitée.

La première section traite des justifications épistémiques des seules notions usuelles d'équilibre statique (équilibre de Nash, équilibre rationalisable, équilibre corrélé). La deuxième section traite des justifications épistémiques de la notion dominante d'équilibre dynamique (équilibre parfait en sous-jeux) en revenant sur le « paradoxe de la rétroduction » (*backward induction paradox*). La troisième section traite des justifications évolutionnistes des notions d'équilibre, aussi bien statiques que dynamiques, pour divers processus d'apprentissage et d'évolution. Chaque section est subdivisée en trois parties : la première est consacrée aux outils analytiques nécessaires, la deuxième à leur application aux jeux considérés et la troisième à l'énoncé des principaux résultats.

1. Justifications épistémiques des équilibres statiques

1.1. Principes logiques

La formalisation de la structure des croyances d'un acteur s'effectue dans le cadre de la logique épistémique (une variété de logique modale), sous les deux formes syntaxique et sémantique dont on peut montrer l'équivalence. En syntaxe, l'univers physique de référence de l'acteur est décrit par des « propositions » et le savoir de l'acteur à son sujet est traduit par un « opérateur de croyance » qui indique s'il sait ou non telle ou telle proposition. En sémantique, les états physico-psychiques de l'univers (associant propriétés matérielles et croyances des acteurs) sont décrits par des « mondes possibles » et le savoir de l'acteur est décrit par un « domaine d'accessibilité », indiquant en tout monde (l'un étant singularisé comme étant le monde réel) quels sont les mondes entre lesquels il ne peut discriminer. Le passage de la syntaxe à la sémantique se fait simplement en associant à chaque proposition un événement, sous-ensemble de mondes où elle est vraie. De plus, un acteur sait une proposition en un certain monde si l'événement associé est vrai dans tous les mondes accessibles à partir de ce monde.

La représentation syntaxique permet de définir un ensemble exigeant d'axiomes auxquels sont soumises les croyances de l'acteur sur son environnement et sur lui-même (croyances auto-hiérarchiques). Il s'agit des axiomes d'omniscience logique (l'acteur sait toutes les conséquences de ce qu'il sait), de véridicité (ce que l'acteur sait est vrai), d'introspection positive (l'acteur sait ce qu'il sait) et d'introspection négative (l'acteur sait ce qu'il ne sait pas). Ces propriétés ont des contreparties en sémantique sous forme de propriétés auxquelles sont soumis les domaines d'accessibilité ou encore la relation d'accessibilité associée (un monde est relié à un autre s'il est dans son domaine d'accessibilité). De fait, l'omniscience logique est automatiquement satisfaite dans la sémantique retenue alors que les trois autres propriétés renvoient respectivement à la réflexivité, à la transitivité et à l'euclidianité de la relation d'accessibilité. En ce qui concerne la véridicité, propriété la plus controversée, le savoir est qualifié de « connaissance » lorsqu'il s'avère vrai (au sens du modélisateur) et de « croyance » lorsqu'il est susceptible d'être faux.

Les croyances précédentes ont été définies dans un cadre propositionnel (ou ensembliste), au sein duquel le savoir s'exprime en tout ou rien, l'acteur sachant ou ne sachant pas une certaine proposition. En sémantique, lorsque toutes les propriétés sont simultanément satisfaites, l'acteur dispose d'une « structure d'information » partitionnelle sur les mondes, au sens où les domaines d'accessibilité forment une partition sur l'ensemble des mondes possibles. Alternativement, les croyances peuvent être définies dans un cadre probabiliste, au sein duquel le savoir est plus nuancé, l'acteur sachant cette fois une proposition avec une certaine probabilité. En sémantique, compte tenu de propriétés généralisées du savoir, l'acteur dispose d'une distribution de probabilités en chaque monde sur l'ensemble de tous les mondes. En pratique, un cadre sémantique mixte tend à s'imposer, défini d'une part par une distribution de probabilités initiale (*prior*) sur les mondes, commune à tous les acteurs et traduisant une information publique objective, et d'autre part par une partition d'information, propre à chacun des acteurs et traduisant son information privée.

Par ailleurs, même dans un contexte de quasi certitude (Monderer-Samet, 1989 ; Börgers, 1994 ; Brandenburger, 1992), le passage du cadre ensembliste au cadre probabiliste peut être réalisé en considérant toute une gamme de croyances (Stalnaker, 1996). A l'un des pôles, l'acteur dispose d'une connaissance ensembliste que rien ne saurait remettre en cause (aucun message contradictoire n'est susceptible d'advenir). A l'autre pôle, l'acteur est doté d'une 1-croyance (croyance avec probabilité 1) qui peut néanmoins s'avérer fausse (un message peut intervenir comme une surprise). Dans une première situation intermédiaire, l'acteur a une 1-croyance, dans le monde réel, que ce qu'il croit est vrai (aucune surprise n'est possible dans le monde réel). Dans la seconde situation intermédiaire, l'acteur a une 1-croyance 'robuste par rapport à la vérité', au sens où sa croyance est confirmée s'il reçoit un message vrai dans le monde considéré (une surprise devient alors possible, même dans le monde réel). Dans les cas probabilistes, lorsqu'une révision des croyances s'avère nécessaire au regard d'un message non contradictoire, la règle de révision supposée utilisée n'est autre que la règle de Bayes.

Dans un contexte syntaxique cette fois multi-acteurs, ceux-ci vont adopter des croyances croisées (croyances hétéro-hiérarchiques) du type 'je sais que tu sais que je sais ...'. La distribution inter-individuelle de ces croyances fait passer, par force croissante, de la croyance partagée ('chacun sait X') à la croyance commune ('chacun sait X, sait que l'autre sait X et ainsi de suite jusqu'à l'infini'), cette dernière introduite par Lewis (1969). Chacune est symbolisée par un opérateur de croyances autonome, muni de propriétés remarquables déduites des axiomes des opérateurs individuels. En sémantique, les croyances distribuées s'expriment par des relations d'accessibilité obtenues simplement à partir des relations d'accessibilité individuelles. En particulier, si les croyances individuelles sont partitionnelles, la croyance commune est également traduite par une partition d'information, à savoir la partition la plus fine parmi les partitions plus grossières que celles des acteurs. En passant d'un cadre ensembliste à un cadre probabiliste, la croyance commune s'avère là encore plus ou moins exigeante, depuis la connaissance commune jusqu'à la 1-croyance commune avec des positions intermédiaires (Stalnaker, 1996).

1.2. Hypothèses

Dans un contexte de jeu (statique ou dynamique), les croyances de chaque joueur sont de différents types et sont affectées d'une incertitude plus ou moins forte, exprimée sous une forme ensembliste ou probabiliste. Les « croyances structurelles » portent sur la structure du jeu, c'est-à-dire sur les « caractéristiques » décisionnelles des autres joueurs (opportunités, croyances, préférences) et sur la rationalité qui les animent; mais le joueur est supposé connaître ses propres caractéristiques. Les « croyances factuelles » portent sur le déroulement passé du jeu, c'est-à-dire sur les actions passées des autres joueurs; le joueur connaît là encore ses propres actions passées. Les « croyances stratégiques » portent sur le déroulement futur du jeu, c'est-à-dire sur les actions futures anticipées des autres joueurs, et sont appelées « conjectures »; le joueur connaît à nouveau ses propres intentions d'action. Les joueurs étant dans une situation d'interaction stratégique (les effets de l'action d'un joueur dépendent de celles des autres), les joueurs sont naturellement engagés dans un système de croyances croisées sur leurs actions futures respectives, elles-mêmes ancrées dans un système de croyances croisées sur leurs caractéristiques respectives.

Pour un jeu statique, on définit les opportunités d'un joueur par un ensemble d'actions (ou de stratégies pures) qu'il peut mobiliser et ses préférences par une fonction d'utilité qui dépend à la fois de sa propre action et de celle des autres. Quant aux croyances d'un joueur, elles portent tant sur les éléments précédents d'autrui que sur les croyances d'autrui; elles sont en fait résumées par le « type » d'autrui, choisi dans un ensemble de types possibles. L'acteur adopte de plus une 'rationalité bayésienne', à savoir qu'il choisit la règle de décision qui maximise son espérance d'utilité, compte tenu de ses croyances sur le type des autres. Il effectue des choix individuels face aux diverses sources d'incertitude sans recourir à une médiation extérieure, toutes les influences extérieures étant intégrées de façon plus ou moins fidèle dans ses croyances. L'état défini par les actions conjointes des joueurs (ou plutôt par leurs intentions d'action conjointes) apparaît dès lors

comme un ‘équilibre de croyances’, au sens où aucun joueur n’est incité à modifier ses croyances en un tel état.

Dans un cadre sémantique, on considère un espace universel d’incertitude où les mondes possibles incorporent les combinaisons de types des joueurs comme leurs intentions d’action; un joueur est encore supposé connaître son propre type. Chaque joueur étant doté d’une structure d’information mixte sur les mondes (la probabilité a priori leur étant commune ou non), toutes les incertitudes sont ainsi supposées probabilisables (et même partiellement partagées) en dernière instance. De plus, chaque joueur est amené à choisir une stratégie (ou règle de décision), chaque règle définissant l’action qu’il met en œuvre en chacun des mondes possibles; le joueur est censé connaître sa propre action retenue, si bien qu’elle doit être identique en deux mondes entre lesquels il ne peut discriminer. La rationalité (ex post) du joueur est définie par le choix de la stratégie maximisant l’espérance d’utilité du joueur face aux stratégies envisagées d’autrui. Chaque joueur définit enfin une conjecture, à savoir une distribution de probabilités sur les actions d’autrui; toute stratégie mixte d’un joueur est alors interprétable comme une croyance probabiliste des autres sur ses propres actions.

1.3. Résultats

Le premier résultat -naturellement attendu- affirme que, sous l’hypothèse de connaissance commune de la structure du jeu et de la rationalité (bayésienne) des joueurs, les stratégies itérativement fortement dominées sont éliminées (Tan-Werlang, 1988). Une stratégie d’un acteur est dite (fortement) dominée lorsqu’il existe une autre stratégie de cet acteur qui lui donne une utilité (strictement) meilleure pour toutes les stratégies de ses adversaires. Une stratégie est dite itérativement dominée si elle est éliminée dans le processus suivant appliqué au jeu de départ: on élimine d’abord les stratégies dominées pour chaque joueur, puis dans le jeu résiduel, on élimine à nouveau les stratégies devenues dominées et ainsi de suite. L’élimination séquentielle des stratégies dominées dans le processus épistémique s’explique par le fait que chaque joueur va éliminer ses stratégies dominées, sait que l’autre va en faire autant et ainsi de suite. Bien entendu, l’élimination séquentielle des stratégies dominées laisse généralement subsister un grand nombre d’issues, entre lesquelles une sélection ne peut s’effectuer qu’à l’aide de conditions supplémentaires.

Un deuxième résultat affirme que si, en sus de la connaissance commune de la structure du jeu et de la rationalité des joueurs, on postule la connaissance commune de l’indépendance des joueurs, on obtient un «équilibre rationalisable» (Bernheim, 1984, 1986; Pearce, 1984). L’équilibre rationalisable est défini en considérant que chaque joueur définit sa meilleure réponse à la stratégie anticipée des autres, elle-même anticipée comme une meilleure réponse aux stratégies adverses, et ainsi de suite jusqu’à boucler. Il est construit par élimination itérée des stratégies inférieures des joueurs, une stratégie inférieure étant une stratégie qui n’est jamais une meilleure réponse. L’obtention de l’équilibre s’explique par le fait que la connaissance de l’indépendance des joueurs (leurs intentions d’action ne sont pas corrélées) permet de décomposer les conjectures sur l’ensemble des stratégies en conjectures relatives à chacune des stratégies. Toute stratégie rationalisable est itérativement non dominée, mais non l’inverse; il n’y a identité que dans le cas de deux joueurs car la condition d’indépendance entre joueurs est alors automatiquement satisfaite.

Un troisième résultat affirme que si, en plus de la connaissance commune de la structure du jeu et de la rationalité des joueurs, on postule que les croyances des joueurs sur leurs stratégies respectives résultent d’une distribution a priori commune, on obtient un «équilibre corrélé» (Aumann, 1987). Un équilibre corrélé est caractérisé par une certaine distribution de probabilités sur l’ensemble des issues (pures) du jeu; une entité externe, le corrélateur, tire au sort une issue du jeu, en conformité avec cette distribution de probabilités, et indique à chaque joueur la stratégie correspondante; chaque joueur a alors intérêt à suivre la recommandation du corrélateur si les autres

le font. Du point de vue épistémique, les joueurs sont cette fois précoordonnés par leurs croyances a priori communes sur les mondes, ces derniers reflétant des états de la nature exogènes conditionnant les issues possibles. Là encore, une stratégie d'un équilibre corrélé est itérativement non dominée, mais non l'inverse; en revanche, elle n'est pas directement comparable à une stratégie rationalisable.

Une quatrième vague de résultats porte sur l'équilibre de Nash, équilibre usuel défini par le fait que chaque stratégie d'un joueur est la meilleure réponse aux stratégies d'équilibre des autres. Pour les jeux réduits à deux joueurs, un premier résultat (Tan-Werlang, 1988 ; Brandenburger-Dekel, 1989), bientôt affaibli (Aumann-Brandenburger, 1995), affirme que, sous l'hypothèse de connaissance partagée de la structure du jeu, de la rationalité des joueurs et des conjectures des joueurs, les conjectures constituent un équilibre de Nash (en stratégies mixtes). Pour les jeux à nombre quelconque de joueurs, un résultat plus exigeant (Aumann-Brandenburger, 1995) affirme que, sous l'hypothèse de connaissance partagée de la structure du jeu et de la rationalité des joueurs, sous l'hypothèse que les croyances des joueurs résultent d'une distribution a priori commune et que leurs conjectures soient de connaissance commune, les conjectures de tous les joueurs sur un même joueur s'accordent et définissent encore un équilibre de Nash.

L'équilibre de Nash apparaît à la fois comme un affaiblissement de l'équilibre rationalisable (les croyances sur les actions d'autrui bouclent dès le deuxième niveau) et de l'équilibre corrélé (les probabilités affectées aux issues se décomposent en probabilités sur les stratégies de chaque joueur). Cependant, les résultats obtenus sont très restrictifs puisqu'ils supposent que les joueurs peuvent connaître les conjectures d'autrui, qui ne sont pas des caractéristiques structurelles immuables de ces joueurs, mais des croyances conjoncturelles dont l'origine n'est à nouveau pas décrite. Dans le cas de plus de deux joueurs, des conditions supplémentaires plus drastiques encore sont nécessaires pour s'assurer que les conjonctures de deux joueurs sur un même troisième sont identiques. La difficulté de coordination épistémique entre les joueurs provient de ce que l'équilibre de Nash repose fondamentalement sur une boucle d'interaction entre ces joueurs. Il apparaît en fait comme un équilibre auto-réalisateur, à savoir que les anticipations des joueurs sur leurs stratégies (de Nash) provoquent leurs réalisations.

Si le problème de la définition d'une notion d'équilibre est traité au travers des résultats précédents, il demeure le problème de la sélection d'un état d'équilibre dans le cas (fréquent) de multiplicité des états associés. Pour l'équilibre rationalisable, comme chaque joueur sélectionne indépendamment l'une de ses stratégies rationalisables, aucune coordination sur tel ou tel état n'est envisageable. Pour l'équilibre corrélé, c'est le corrélateur fictif (ou en son absence, la distribution de probabilités a priori commune aux deux joueurs) qui sélectionne un état particulier. Pour l'équilibre de Nash, le fait de connaître (sans préciser comment) les conjectures de l'autre définit là encore un état parmi d'autres. Toutefois, dans les jeux de coordination (pour lesquels toutes les issues d'équilibre sont utilité-équivalentes pour les joueurs), la sélection peut s'effectuer grâce à des « conventions » qui sont de connaissance commune entre les joueurs³. Ces conventions portent directement sur des 'états focaux' du jeu (Schelling, 1960), dont la saillance reflète encore des phénomènes culturels hors modèles. Elles concernent plus ambitieusement des critères de choix entre les issues, comme le critère de symétrie ou le critère de Pareto-optimalité. Enfin, si le jeu est répété, c'est son histoire même qui est susceptible de faire apparaître des caractères saillants (Crawford-Haller, 1990).

³ Voir l'article de A. Orléan dans ce même numéro

2. Fondements épistémiques des équilibres dynamiques

2.1. Principes logiques

Lorsqu'on se place dans un cadre dynamique, il se pose un problème supplémentaire, celui de la genèse et de l'évolution des croyances de l'acteur. En fait, la logique épistémique n'aborde pas de front le problème de la formation des croyances, mais se cantonne au problème de la révision d'une croyance a priori (dont l'origine n'est pas précisée). Traditionnellement, on distingue deux contextes de révision, caractérisés par le type de message qui fait passer d'une croyance initiale à une croyance finale à propos d'un univers de référence. Dans le contexte de « rectification » (*revising*), le message vient préciser voire infirmer la croyance initiale concernant cet univers considéré comme fixe. Dans le contexte de « actualisation » (*updating*), le message indique dans quelle direction se modifie l'univers considéré cette fois comme évolutif. Un troisième contexte, celui de « focalisation » (*focusing*), concerne un message relatif à un objet tiré au sort au sein d'un univers constitué d'une population d'objets, mais il est réductible à un principe de rectification associé à un principe de projection.

En syntaxe, des axiomatiques relatives à chacun des deux principaux contextes ont été développées, d'abord pour des croyances propositionnelles (Alchourron-Gärdenfors-Makinson, 1985; Katsuno-Mendelzon, 1992). Une famille d'axiomes est commune aux deux contextes. Ainsi, l'axiome de succès affirme que la croyance finale doit valider le message, ce qui attribue une priorité au message (supposé vrai) sur la croyance initiale (éventuellement fausse). De même, l'axiome d'inclusion affirme que la part de la croyance initiale compatible avec le message est conservée dans la croyance finale, ce qui traduit un 'principe de conservation' faible (on conserve ce qui peut l'être). D'autres axiomes sont spécifiques de chaque contexte. Pour la rectification, l'axiome de préservation affirme que, si le message est compatible avec la croyance initiale, la croyance finale est limitée à leur partie commune, ce qui traduit un 'principe de conservation' fort (on change le moins possible ce qui doit être changé). Pour l'actualisation, l'axiome de monotonie exprime que, pour un message donné, si la croyance initiale est affaiblie, la croyance finale l'est aussi, ce qui traduit encore un principe de changement minimal.

En sémantique, des règles de révision sont inférées des axiomatiques par des théorèmes de représentation. Pour la rectification, tout se passe comme s'il existe un ensemble de couronnes concentriques autour de la croyance initiale, reflétant des mondes de plus en plus éloignés de cette croyance initiale. La croyance finale s'obtient alors simplement comme intersection entre le message et la première couronne qui intersecte le message. Réinterprétée en syntaxe, cette règle de révision revient à doter chaque proposition d'un « degré d'enracinement épistémique », à rajouter le message à la croyance initiale, et à retirer du système ainsi constitué les propositions les moins enracinées jusqu'à restaurer sa cohérence logique. Pour l'actualisation, tout se passe comme s'il existe un ensemble de couronnes concentriques autour de chaque monde, reflétant des mondes de plus en plus éloignés de ce monde. La croyance finale s'obtient ici comme l'union, pour chacun des mondes de la croyance initiale, des intersections du message avec la première couronne qui intersecte le message.

Cette démarche peut être étendue d'un cadre ensembliste à un cadre probabiliste, un message (toujours ensembliste) faisant passer d'une distribution de probabilités a priori à une distribution a posteriori (Walliser-Zwirn, 2002). En syntaxe, les axiomes précédemment définis peuvent recevoir une transcription faible (en termes de support des distributions de probabilités) ou forte (en termes de valeurs numériques attribuées aux probabilités). En sémantique, des règles de révision des distributions de probabilités sont alors dérivées des axiomes par des théorèmes de représentation. Pour la rectification, avec une transcription faible des axiomes, la règle adéquate est la règle de *conditioning* généralisée. La règle la plus usuelle pour les théoriciens des jeux, la règle de conditionnement de Bayes, n'est singularisée que pour une transcription très forte des axiomes; elle ne reçoit de justification épistémique (par opposition à sa justification décisionnelle au travers

de paris effectués par l'acteur) que dans un contexte très restrictif. Pour l'actualisation, avec une transcription faible des axiomes, la règle adéquate est la règle d'*imaging* généralisée proposée par Lewis.

La révision des croyances peut être considérée comme la forme fondamentale du raisonnement dans la mesure où diverses autres formes de raisonnement peuvent s'y ramener (Walliser-Zwirn-Zwirn, 2002). Il en est ainsi du raisonnement non monotone du type 'des faits A, on infère normalement les faits B', qui affaiblit la déduction classique en envisageant des exceptions et qui peut être réinterprété comme une révision dans un contexte de rectification. Il en est de même du raisonnement abductif du type 'des faits A, on abduit l'hypothèse B' qui est une forme d'explication inversée et qui peut aussi être réinterprété de deux façons dans un contexte de rectification. Il en est surtout ainsi du raisonnement conditionnel, de type 'si l'antécédent A était le cas, alors le conséquent B serait vrai', qui est un renforcement de l'implication matérielle et qui peut être réinterprété cette fois dans un contexte d'actualisation. Il repose en effet sur une transformation physique de l'univers pour rendre l'antécédent vrai, même si cette transformation reste virtuelle.

En syntaxe, le raisonnement conditionnel distingue une proposition profactuelle dont l'antécédent est vrai d'une proposition contrefactuelle dont l'antécédent est faux. Une proposition conditionnelle $A \triangleright B$ se distingue de l'implication matérielle $A \rightarrow B$ dans la mesure où cette dernière est (abusivement) vraie dès que son antécédent est faux. Le raisonnement conditionnel est soumis à une série d'axiomes dont on peut donner quelques exemples. L'axiome de réflexivité affirme qu'un conditionnel dont le conséquent est identique à l'antécédent est toujours valide. L'axiome d'infra-classicalité affirme qu'un conditionnel dont l'antécédent est vrai coïncide avec l'intersection de l'antécédent et du conséquent. L'axiome de monotonie prudente affirme que si deux conditionnels de même antécédent sont valides, on peut construire un conditionnel valide en prenant comme antécédent la conjonction de l'antécédent commun et du conséquent de l'un et comme conséquent le conséquent de l'autre. Des axiomes supplémentaires ont un rôle plus 'topologique' (*et* à droite, *ou* à gauche).

En sémantique, le raisonnement conditionnel est traduit par une fonction de sélection qui, en chaque monde, associe à chaque événement hypothétique H un événement retenu K, éventuellement réduit à un monde unique (Stalnaker, 1968). Le pont entre syntaxe et sémantique est alors réalisé par la condition suivante : un conditionnel est valide en un monde si, pour ce monde, tous les mondes retenus pour son antécédent valident son conséquent. Moyennant les seuls axiomes topologiques, la fonction de sélection peut être dérivée d'un ensemble de couronnes concentriques regroupant des mondes de plus en plus éloignés du monde considéré (Lewis, 1973). Un conditionnel est alors valide si, dans les mondes les plus proches du monde considéré où l'antécédent est vrai, le conséquent est également vrai. Les axiomes plus substantiels correspondent dès lors à des propriétés assignées à la fonction de sélection (ou aux couronnes de mondes). La propriété de «satisfaction» impose à l'événement retenu à partir d'un monde et d'un événement hypothétique d'être contenu dans ce dernier. La propriété de «maintien» impose, pour un monde situé dans l'événement hypothétique, d'être situé dans l'événement retenu pour ce monde et cet événement hypothétique.

Une variante (Samet, 1996) crée un lien plus direct entre le raisonnement classique et le raisonnement conditionnel. En syntaxe, on adjoint à l'opérateur de croyance usuel un opérateur de croyance hypothétique (paramétré par une proposition hypothétique); le second se réduit au premier lorsque la proposition hypothétique est la tautologie. Les deux opérateurs sont soumis à une liste d'axiomes qui à la fois généralisent les propriétés du premier (omniscience logique, introspection positive et négative) et les combinent. En sémantique, on définit à nouveau, en se calant sur la partition associée à la croyance usuelle, une fonction de transformation d'hypothèse qui associe à tout monde et à tout événement hypothétique un événement retenu. La fonction de transformation d'hypothèse satisfait alors aux deux propriétés classiques de satisfaction et de

maintien. On peut montrer (Halpern, 1999) que l'on peut se ramener au raisonnement conditionnel classique en interprétant l'opérateur de croyance hypothétique comme 'si l'antécédent est accepté, le conséquent est connu' (une proposition est dite acceptée s'il n'est pas connu qu'elle est fausse).

2.2. Hypothèses

Les conditionnels ont d'abord été mobilisés en théorie de la décision dans une configuration où les actions sont susceptibles d'influencer les états de la nature. Les deux critères de décision proposés ont en commun de conserver la règle de maximisation de l'espérance d'utilité, mais différent quant à la probabilité à affecter aux états (Gibbard-Harper, 1978). La «théorie de la décision évidentielle» retient la probabilité de l'état conditionnelle à l'action, traduisant une simple corrélation probabiliste sans dépendance causale entre action et état (due par exemple à un facteur commun qui influence les deux entités). La «théorie de la décision causale» retient la probabilité du conditionnel 'si telle action, alors tel état', exprimant cette fois une causalité directe entre action et état. Ces deux théories ont été mobilisées dans l'analyse du 'problème de Newcomb' pour montrer que la première conduit l'acteur à prendre une boîte (action maximisant l'espérance d'utilité) alors que la seconde le conduit à prendre deux boîtes (action dominante).

Les conditionnels ont été tôt évoqués, sinon vraiment utilisés, pour traiter des jeux sous forme extensive, exprimés sous la forme d'un arbre du jeu (décrivant les coups successifs possibles et les utilités résultant de toute trajectoire). Selten-Leopold (1982) font en effet remarquer que c'est ce qui se passerait hors de la trajectoire d'équilibre du jeu qui justifie le choix de la trajectoire elle-même, et discutent en conséquence de la pertinence de diverses théories des conditionnels. Harsanyi-Selten (1988) observent qu'une stratégie doit être exprimée par un conditionnel plutôt que par une implication matérielle. Ces réflexions se sont poursuivies plus récemment en s'efforçant de décrire par des conditionnels divers aspects du jeu. Les travaux se restreignent à la classe des jeux génériques, à savoir des jeux sans égalités entre les issues pour un même joueur (ou tels que, lorsqu'il y a égalité pour un joueur, il y a aussi égalité pour les autres) et à information parfaite (chaque joueur sait à tout moment où il se trouve dans l'arbre du jeu). De plus, la structure du jeu est considérée comme de connaissance commune.

Tout d'abord, le raisonnement conditionnel peut être appliqué à la structure du jeu, c'est-à-dire à l'arbre du jeu. Il indique qu'en tout nœud intermédiaire, le joueur doit retenir une action et une seule, et qu'en tout nœud terminal, le joueur reçoit une certaine utilité; il indique à rebours que tout nœud ne peut être atteint que si un coup spécifique a été joué précédemment. En fait, le raisonnement conditionnel ne diffère pas ici de l'implication matérielle et ne fait que traduire des contraintes physiques et des règles utilitaristes relatives au jeu. Par ailleurs, le raisonnement conditionnel trouve à s'appliquer à la définition d'une stratégie d'un joueur, à savoir l'action que jouerait ce joueur en un quelconque nœud intermédiaire de l'arbre du jeu. Il peut s'agir d'un conditionnel profactuel si le nœud se trouve effectivement sur la trajectoire suivie, ou d'un contrefactuel si le nœud se trouve en dehors de cette trajectoire. Se limiter à une implication matérielle est ici insuffisant car cela autoriserait n'importe quelle action en un nœud hors équilibre.

En combinant les deux usages précédents, le raisonnement conditionnel sert enfin à traduire de façon non ambiguë la rationalité du joueur, toujours analysée d'un strict point de vue individuel. Cette rationalité est appréhendée à la fois en chaque nœud intermédiaire de l'arbre du jeu et en chacun des mondes possibles. Comme les mondes possibles traduisent ici encore les types des joueurs (essentiellement leurs croyances), les stratégies retenues par les joueurs sont naturellement conditionnelles à ces mondes. La rationalité d'un joueur est plus précisément définie en considérant les stratégies des autres joueurs comme fixées par ses croyances et en examinant s'il a intérêt à dévier de sa propre stratégie, en fonction de ce qui se passe en aval du nœud considéré (principe conséquentialiste). Mais elle reçoit néanmoins de nombreuses définitions alternatives, plus ou moins fortes, en fonction du critère individuel de choix retenu, du type de nœud où elle est effectivement évaluée et des croyances attribuées au joueur en ce nœud.

Tout d'abord, exprimée en termes ensemblistes, la rationalité peut se contenter de dire que le joueur n'a pas connaissance qu'il peut faire mieux avec une autre action ou vouloir affirmer que le joueur a pleine connaissance qu'il ne peut pas faire mieux. En termes probabilistes, elle peut s'exprimer par le fait que le joueur sait presque sûrement qu'il ne peut améliorer son espérance d'utilité. Cette dernière hypothèse est intermédiaire entre les deux précédentes. Ensuite, la rationalité est dite 'substantielle' (Aumann, 1995) si elle est définie en tout nœud, conditionnellement à l'atteinte de ce nœud, c'est-à-dire au travers d'un conditionnel. Elle est dite 'matérielle' si elle n'est définie qu'aux nœuds atteints par la trajectoire d'équilibre, c'est-à-dire par une implication matérielle. La rationalité matérielle est naturellement plus faible que la rationalité substantive. Enfin, la rationalité est définie *ex ante* avec les croyances disponibles au début du jeu ou *ex post* avec les croyances disponibles au moment de jouer. La seconde est d'évidence plus forte que la première.

En sémantique, le raisonnement conditionnel va être traduit par une fonction de sélection définie en tout monde, avec pour événement hypothétique l'atteinte d'un certain nœud. Les conditions générales imposées à cette fonction reçoivent dès lors une interprétation plus concrète, en relation avec les stratégies suivies par les joueurs en ce monde (Halpern, 2001). La propriété de satisfaction indique que les stratégies adoptées dans les mondes les plus proches du monde considéré conduisent à une trajectoire qui passe par le nœud considéré. La propriété de maintien indique que si le nœud considéré est atteint par les stratégies relatives au monde considéré, le monde le plus proche se réduit à ce monde. En outre, une propriété supplémentaire d'«unicité» impose que, sur le sous-arbre commençant au nœud considéré, les stratégies relatives au monde considéré et au monde le plus proche coïncident; elle assure que les stratégies dictent précisément ce que les joueurs vont faire si le nœud est effectivement atteint.

2.3. Résultats

Pour un jeu sous forme extensive et à information parfaite, la notion fondamentale d'équilibre est celle d'équilibre parfait (de sous-jeux). Cette notion est plus forte que celle d'équilibre de Nash, qui reste cependant applicable si le jeu sous forme extensive est ramené sous forme normale grâce au concept de stratégie. L'équilibre parfait a l'avantage d'exister et d'être unique pour tout jeu générique car il est obtenu par une procédure constructive, la procédure de rétroduction (*backward induction*). En un nœud terminal, le joueur qui a le trait choisit sa meilleure action; en un nœud précédent, le joueur qui a le trait choisit sa meilleure action, compte tenu de celle du suivant; la procédure se poursuit de nœud en nœud jusqu'au nœud initial. Si cette notion d'équilibre a d'abord paru quelque peu paradoxale, le paradoxe s'est progressivement dissout lorsque l'équilibre parfait a été justifié ou infirmé par une variété de résultats, qui sont fondés sur des hypothèses épistémiques qui ne diffèrent que de façon très subtile .

Binmore (1987) considère la possibilité de justifier l'équilibre parfait par la croyance commune de la rationalité des joueurs, et met en évidence le 'paradoxe de la rétroduction' (*'backward induction paradox'*). Ce paradoxe s'appuie sur le fait que le raisonnement du joueur doit s'exercer dans les deux sens de la flèche du temps : dans le sens inverse pour définir les stratégies en fonction des croyances (rationalité instrumentale), dans le sens direct pour réviser les croyances en fonction des informations recueillies (rationalité cognitive). Il exprime que, puisque la rationalité de l'acteur est fondée sur son action en des nœuds hors équilibre, il faut envisager concrètement la possibilité d'une déviation hors de la trajectoire d'équilibre et examiner les croyances révisées de l'acteur aux nœuds déviants. Or, les croyances du joueur dans un nœud hors équilibre ne sont plus compatibles avec la connaissance commune de la rationalité qui a servi à définir l'équilibre, du moins si les autres hypothèses structurelles sont supposées conservées (Reny ,1992).

Un premier résultat (Aumann, 1995) montre cependant que, si les joueurs ont une connaissance commune de leurs rationalités respectives, les stratégies retenues déterminent un équilibre parfait. Il repose sur l'hypothèse que la rationalité du joueur est ensembliste faible, substantielle et *ex ante*; aucune révision des croyances et aucun conditionnel n'est considéré. Le résultat est a fortiori conservé avec une forme plus forte de rationalité, en particulier la rationalité bayésienne ou la rationalité *ex post*. Cependant, la rationalité est supposée de connaissance commune au sens fort (elle n'est jamais remise en cause en vertu du principe de véridicité) et cette connaissance ne saurait être affaiblie. Le résultat indique en fait que, sous l'hypothèse de connaissance commune de la rationalité en tout noeud, aucun joueur ne saurait dévier de sa trajectoire d'équilibre parfait sous peine de la violer. On est proche du théorème des travaux virtuels en mécanique, qui indique quelle trajectoire va suivre un système matériel dans l'ensemble de trajectoires virtuelles, mais ces dernières ne sont jamais réalisées.

Ce résultat a déclenché une polémique entre Aumann (1996) et Binmore (1996), le premier considérant que l'on peut maintenir l'hypothèse de connaissance commune de la rationalité contre vents et marées, le second affirmant qu'on ne peut s'affranchir de considérer et d'interpréter ce qui se passe hors équilibre. S'il se trouve en un noeud hors équilibre, un joueur doit en effet choisir, dans l'ensemble des hypothèses qui justifient l'équilibre, laquelle doit être remise en cause (Walliser, 1996). Cette hypothèse peut résulter du degré d'enracinement épistémique que le joueur attribue à chacune d'entre elles, degré qui dépend éventuellement du jeu considéré (dans le jeu d'échecs, même si les stratégies d'équilibre parfait ne sont pas connues, des déviations seront attribuées à des défauts de rationalité cognitive de l'adversaire plutôt qu'à une méconnaissance de la structure du jeu). Les degrés d'enracinement épistémiques et plus généralement les règles de révision des croyances font alors partie des caractéristiques des joueurs et peuvent elles-mêmes faire l'objet d'une connaissance commune. On peut montrer que, selon l'hypothèse remise en cause dans la révision, c'est l'équilibre parfait ou d'autres notions d'équilibre qui s'avèreront épistémiquement justifiées.

La première hypothèse que l'on peut considérer est celle de «non velléité» des joueurs, hypothèse qui affirme que les actions retenues par un joueur sont effectivement celles qu'il met en œuvre. Il est facile de voir que si l'on remet en cause cette hypothèse sous forme d'une 'main tremblante', à savoir que tout joueur est affecté de tremblements successifs aléatoires et indépendants lorsqu'il implémente ses intentions d'action (Selten, 1975), l'équilibre parfait va subsister. En effet, toute déviation d'autrui sera alors interprétée par un joueur comme purement accidentelle et ne révélant rien quant au comportement futur de ce joueur. Une deuxième hypothèse porte sur la connaissance commune de la structure du jeu. Kreps *et al.* (1982) ont montré (dans un cadre non formalisé en logique épistémique) que, si un joueur a une incertitude sur le type des autres joueurs, on peut obtenir d'autres solutions que celle d'équilibre parfait. Toute déviation d'autrui est alors interprétée par un joueur par le fait qu'il joue un jeu différent que le jeu réel ou encore qu'il croit qu'un tiers joue un jeu différent.

L'hypothèse la plus sensible est cependant celle de connaissance commune de la rationalité des joueurs car elle n'autorise aucune surprise. Si l'on n'envisage plus qu'une 1-croyance commune que ce que les agents croient est vrai, l'équilibre parfait subsiste car aucune surprise ne reste envisagée (Stalnaker, 1996). En revanche, si l'on affaiblit encore en une 1-croyance commune avec robustesse par rapport à la vérité, l'équilibre parfait n'est généralement plus garanti (Stalnaker, 1996). Enfin, s'il y a simplement 1-croyance commune, l'équilibre parfait n'est a fortiori pas garanti car une surprise est explicitement envisagée. Par exemple, Ben Porath (1992) montre que, s'il y a 1-croyance commune de la rationalité au début du jeu, l'issue réalisée est celle obtenue par une étape d'élimination des stratégies faiblement dominées, puis par élimination séquentielle des stratégies fortement dominées; Stalnaker (1998) obtient une notion semblable

d'équilibre lorsqu'il est de connaissance commune que les joueurs sont rationnels et appliquent un 'principe de rationalisation'. Par ailleurs, dans un jeu de longueur m , si l'équilibre parfait reste justifié avec une m -connaissance commune de la rationalité (Bicchieri, 1988), il n'est en revanche plus garanti par une ε -croyance commune.

On peut enfin s'intéresser à l'hypothèse de rationalité elle-même, malgré le fort degré d'enracinement à son égard. Le résultat de base justifiant l'équilibre parfait ne survit pas si la rationalité substantive est remplacée par la rationalité matérielle, sauf pour certaines classes particulières de jeux (Aumann, 1998). Plus précisément, si l'on évalue la rationalité non pas en chaque monde considéré, mais au monde le plus proche au sens d'une fonction de sélection, le résultat est invalidé (Stalnaker, 1998). La fonction de sélection, supposée identique pour tous les joueurs et satisfaisant aux trois conditions préalablement énoncées, sert en fait à réviser les croyances lorsqu'un nœud hors équilibre est atteint. Même si, dans le monde considéré et dans le monde le plus proche, les stratégies aval coïncident, les croyances vont alors différer (Halpern, 2001). Un résultat analogue (Samet, 1996) montre que, si l'on introduit un raisonnement conditionnel à l'aide d'un opérateur de croyances hypothétiques, l'équilibre parfait n'est plus garanti. Il n'est restauré que sous des conditions supplémentaires très fortes (croyances hypothétiques séquentielles).

Une hypothèse reliée à celle de rationalité des joueurs est celle d'indépendance entre les choix des joueurs, considérée par Stalnaker sous deux aspects. L'hypothèse d'indépendance causale, inhérente à l'idée de jeu non coopératif, considère simplement que les joueurs jouent de façon indépendante. L'hypothèse d'indépendance épistémique, plus exigeante, exprime que ce qu'un joueur apprend sur un autre (en particulier ce que cet autre pense d'un tiers) n'affecte pas ce que lui-même pense d'un tiers. L'indépendance causale n'entraîne pas l'indépendance épistémique, la première n'empêchant pas les croyances d'être éventuellement corrélées. Vu sous cet angle, Aumann (1995) justifie l'équilibre parfait par une hypothèse d'indépendance causale, associée à une hypothèse d'indépendance épistémique renforcée, à savoir l'indépendance entre coups successifs d'un même joueur (insensibles aux messages reçus). Stalnaker (1998) montre toutefois que, si les croyances des joueurs satisfont à une condition d'indépendance épistémique entre joueurs, mais différent à chaque nœud, le concept adéquat d'équilibre reste bien celui d'équilibre parfait.

La condition d'indépendance épistémique joue en fait le même rôle qu'une quatrième propriété (jointe à celles de satisfaction, de maintien et d'unicité) assignée à la fonction de sélection (Halpern, 2001). Celle-ci indique que, si un premier monde est accessible à partir d'un monde retenu par la fonction de sélection, il existe un second monde accessible à partir du monde considéré tel que les stratégies en ces deux mondes soient identiques sur le sous-arbre commençant au monde considéré. Cette propriété très exigeante s'interprète par le fait qu'en un nœud intermédiaire, le joueur conserve les mêmes croyances sur les autres dans le monde considéré et dans le monde le plus proche. Comme conséquence, les stratégies qu'un joueur envisage aux mondes retenus par la fonction de sélection sont un sous-ensemble de celles envisagées au monde considéré. Si, comme il a été affirmé, les trois premières propriétés de la fonction de sélection ne suffisent pas à garantir l'équilibre parfait, la quatrième permet à nouveau de l'assurer.

3. Justifications dynamiques des équilibres

3.1. Principes généraux

La théorie des jeux évolutionniste considère un jeu de base qui est joué séquentiellement sur un ensemble généralement infini de périodes. Elle est fondée sur cinq grands principes de modélisation. Seul le *principe de satisfaction* est commun avec la théorie des jeux classique et explicite les caractéristiques immuables des joueurs. Il détaille la structure du jeu de base, à savoir les actions possibles de chaque joueur et les utilités résultant de la combinaison de leurs actions. Selon que le jeu de base est lui-même statique ou dynamique, ces caractéristiques seront représentées sous forme de matrice du jeu ou d'arbre du jeu. Il exprime également la manière dont les joueurs agrègent les utilités obtenues lors d'occurrences successives du jeu (en général en actualisant les utilités successives). Les quatre autres principes concernent le processus dynamique, généralement aléatoire, selon lequel se répète le jeu de base, et décrivent aussi bien les interactions physiques entre les joueurs que les raisonnements psychiques qui les animent.

Le *principe de confrontation* précise la nature des interactions entre les joueurs, notamment leurs modes de rencontre. Le jeu de base peut être joué aussi bien par des joueurs singularisés que par des populations de joueurs. Lorsque le jeu est asymétrique, on introduit autant de populations que de joueurs du jeu de base (jeu multi-populations). Lorsque le jeu est symétrique, on peut en faire autant (jeu multi-populations) ou considérer que les joueurs ne forment qu'une seule et même population (jeu mono-population). Chaque joueur peut potentiellement rencontrer tout autre individu (interactions globales) ou seulement ceux situés dans son 'voisinage d'interaction' (interactions locales). De fait, les individus sont souvent situés sur un réseau (linéaire, planaire) sur lequel les voisinages d'interaction sont géométriquement définis. A chaque période, les interactions entre individus peuvent être systématiques ou seul un échantillon d'individus est actif, chaque individu rencontrant un autre échantillon d'individus de son voisinage.

Le *principe d'information* décrit l'information que recueille chaque joueur tant sur la structure du jeu que sur son déroulement passé. L'information structurelle est souvent très réduite, le joueur connaissant ses possibilités d'action, mais non forcément ses préférences et a fortiori les caractéristiques d'autrui. L'information factuelle est plus riche et se présente sous deux types. D'une part, il peut observer les actions passées jouées par les autres joueurs (les siennes lui étant connues). D'autre part, il peut observer les utilités qu'il obtient avec les actions qu'il a mises en œuvre (et plus exceptionnellement les utilités obtenues par autrui). Toutes ces informations sont collectées dans son 'voisinage d'information', généralement inclus dans son voisinage d'interaction. Là encore, il peut observer intégralement l'information ou ne recevoir les informations que d'un échantillon aléatoire de joueurs. Enfin, comme il sera vu plus loin, il obtient l'information soit comme sous-produit spontané du déroulement du jeu (expérimentation passive), soit de façon volontaire en déplaçant son action 'normale' (expérimentation active).

Le *principe d'évaluation* porte sur l'interprétation et le traitement de l'information afin d'obtenir une information condensée en vue de l'action. D'une part, le joueur peut concentrer son attention sur la distribution statistique des actions passées de ses adversaires pour en dégager certains invariants. Il va alors utiliser ces indicateurs rétrospectifs pour former des anticipations (généralement probabilistes) sur les actions futures d'autrui. D'autre part, le joueur peut calculer des index relatifs aux performances passées de ses propres stratégies (ou de l'ensemble des stratégies). Il va alors supposer que ces indicateurs restent valables dans le futur et les comparer éventuellement à des seuils d'aspiration normatifs, eux-mêmes évolutifs (croissants ou décroissants selon qu'ils ont été ou non atteints dans le passé). Plus rarement, il va tenter de révéler, à partir des utilités qu'il a lui-mêmes éprouvées, ses propres préférences et même, à partir

des actions observées d'autrui, certaines caractéristiques structurelles des autres, en particulier leurs préférences.

Le *principe de décision* explicite les règles de choix adoptées par chaque joueur, à partir des informations agrégées précédentes. On distingue deux types de comportement, qui prennent en compte la dimension stratégique de l'information qu'il est susceptible d'acquérir. Le comportement d'exploitation consiste à utiliser au mieux de ses intérêts l'information déjà existante. Le comportement d'exploration consiste à déplacer l'action précédente de façon plus ou moins aléatoire pour tester les adversaires et acquérir une information originale. Le joueur doit en fait réaliser un arbitrage entre 'exploration' et 'exploitation', la seconde se traduisant par une perte d'utilité à court terme au profit d'un gain d'utilité à long terme (investissement informationnel). Cet arbitrage (non optimal en général) est souvent directement inscrit dans la règle de choix, qui combine les deux comportements. En particulier, l'exploration se doit d'apparaître comme plus importante en début de processus et l'exploitation en fin de processus.

3.2. Typologie

Trois grandes familles de modèles évolutionnistes sont généralement considérées, qui correspondent à des illustrations relativement contrastées des principes précités (Walliser, 1998). Elles attribuent aux joueurs des informations de plus en plus réduites et des formes de rationalité cognitive (adéquation entre moyens disponibles et objectifs poursuivis) et instrumentale (adéquation entre informations disponibles et croyances adoptées) de plus en plus faibles (Walliser, 1989). L'*apprentissage épistémique* est fondé sur un processus de révision des croyances des acteurs sur les stratégies de leurs adversaires. L'*apprentissage comportemental* est fondé sur le renforcement par les acteurs de leurs propres stratégies les plus performantes. Le *processus évolutionnaire* est fondé sur un mécanisme lamarckien de sélection naturelle des individus formant des populations. Ces trois types de modèles peuvent être combinés dans des modèles hybrides associant divers principes. De plus, la deuxième et la troisième familles présentent un isomorphisme formel, qui permet de se ramener aux deux premières familles, qui sont les plus réalistes.

L'apprentissage épistémique suppose que les acteurs connaissent leur fonction d'utilité et sont à même d'observer leurs actions respectives. S'appuyant sur ces observations, ils révisent leurs croyances sur les stratégies futures d'autrui. S'appuyant sur leur fonction d'utilité, ils choisissent leur meilleure réponse à ces croyances, en vertu du comportement d'exploitation. Cette meilleure réponse peut toutefois être 'perturbée' par des aléas, en vertu du comportement d'exploration. Le modèle le plus simple obéissant à ce schéma est celui de '*fictitious play*'. Chaque joueur calcule d'abord la fréquence observée d'utilisation de chaque action d'autrui; il transforme ensuite cette fréquence passée en probabilité future d'utilisation d'une action; il choisit enfin sa meilleure réponse au sens de l'espérance d'utilité de chacune de ses actions. Le modèle de '*fictitious play stochastique*' en est une variante qui suppose que le joueur n'optimise plus, mais joue une action avec une probabilité croissante avec son espérance d'utilité future.

L'apprentissage comportemental suppose que les acteurs n'observent que les performances (en termes d'utilité) liées à leurs propres actions passées. Ils calculent un index agrégé de performance de chaque action, éventuellement comparé à un seuil d'aspiration. Ils adoptent un choix probabiliste, la probabilité de jouer une action étant croissante avec son index; ce choix présente ainsi simultanément une composante d'exploration (par le fait que toute action a toujours une probabilité non nulle d'être utilisée) et d'exploitation (par le fait que les actions les plus performantes sont jouées plus souvent). Ils peuvent également imiter le comportement des joueurs les plus performants s'ils observent leurs performances. Le modèle le plus simple obéissant à ce

schéma est le modèle CPR. Chaque joueur se contente d'abord d'observer l'utilité obtenue avec l'action qu'il a mise en œuvre ; il calcule ensuite l'utilité cumulée obtenue par cette action depuis le début du jeu ; il prend enfin une décision probabiliste en choisissant chaque action avec une probabilité proportionnelle à son index.

Le processus évolutionnaire suppose que les acteurs n'observent plus rien (sauf les actions d'autrui si leurs stratégies en dépendent) et n'ont plus de croyances. Ils ont un comportement fixe, mais forment des populations d'individus symbolisant un même joueur. Ils subissent un processus de sélection favorisant la reproduction de ceux qui obtiennent la meilleure utilité (assimilée au '*fitness*' des biologistes), ce qui traduit le comportement d'exploitation ; ils connaissent éventuellement un processus de mutation modifiant aléatoirement leur stratégie, ce qui traduit un comportement d'exploration. Le modèle le plus simple obéissant à ce schéma est le modèle du réplicateur. Dans un cadre multi-populations ou mono-population, les individus se rencontrent tous aléatoirement ; ils se reproduisent de façon proportionnelle à l'utilité qu'ils retirent de leurs interactions. Une variante, le réplicateur stochastique, considère en plus des individus mutants introduits de façon aléatoire dans la population à chaque période.

3.3. Résultats

On s'intéresse à présent aux états asymptotiques des processus précédents, sachant que ces processus peuvent aussi bien converger vers des états stationnaires (attracteurs ponctuels) que ne pas converger (attracteurs cycliques ou chaotiques). On peut remarquer que le problème de la sélection d'un équilibre ne se pose plus car la trajectoire du système se dirige toujours (du moins en probabilité) vers une issue déterminée (si elle converge). Les résultats obtenus ne sont valables que pour de classes particulières de processus dynamiques et des catégories spécifiques de jeux. Ils sont aussi relatifs au type de stabilité imposé aux trajectoires asymptotiques ainsi qu'à l'échelle de temps à laquelle on se situe (long terme en l'absence de perturbations stochastiques, très long terme avec perturbations). Ces résultats s'avèrent peu robustes à de petites modifications des spécifications des modèles et sont en particulier très sensibles aux éléments stochastiques introduits sur différents éléments des modèles (modes d'interaction, échantillonnage d'information, règles de décision).

Les résultats concernent tout d'abord les dynamiques s'appuyant sur des jeux de base statiques. En fait, le seul résultat vraiment général concerne l'élimination des stratégies fortement dominées tout comme des stratégies itérativement fortement dominées. Ce résultat a été démontré pour des processus évolutionnaires assez étendus incluant le réplicateur standard (Weibull, 1995) ainsi que pour divers processus d'apprentissage. Cependant, il n'est pas maintenu pour les stratégies (itérativement) faiblement dominées. Samuelson-Zhang (1992) ont en effet montré qu'en cas de dynamique évolutionnaire, avec la dynamique du réplicateur et même en présence de bruit, les stratégies faiblement dominées peuvent être conservées. Ce résultat s'applique à des jeux de base dynamiques car ceux-ci peuvent se ramener à des jeux statiques grâce à la notion de stratégie. Ils assurent que les équilibres parfaits, qui ne sont jamais fortement dominés, mais sont parfois faiblement dominés, ne sont pas systématiquement éliminés.

Si l'on s'intéresse aux équilibres de Nash de jeux statiques, ils apparaissent comme des points de repos de la plupart des processus dynamiques (si l'on s'y trouve initialement, on y reste). Plus ambitieusement, les équilibres de Nash en stratégies pures stricts peuvent être obtenus comme états limites dans certains processus car toute déviation conduit à une perte stricte d'utilité. Ainsi, avec la règle CPR (Laslier-Topol-Walliser, 2002), les trajectoires convergent avec une probabilité positive vers n'importe quel équilibre de Nash en stratégies pures strict, s'il en existe. De même, avec la dynamique du réplicateur standard dans un cadre multi-populations, un état est

asymptotiquement stable si et seulement si c'est un équilibre de Nash strict. En revanche, les équilibres de Nash en stratégies mixtes sont beaucoup plus difficiles à obtenir. Toujours avec le modèle CPR, la convergence des trajectoires avec probabilité positive vers de tels équilibres n'est réalisée que pour des classes particulières de jeux, par exemple ceux à équilibre unique. Pour le modèle de *fictitious play stochastique*, une telle convergence avec stabilité globale est cependant assurée (Hofbauer-Sandholm, 2001) pour diverses classes de jeux (à somme nulle, potentiels, supermodulaires).

Des raffinements de l'équilibre de Nash peuvent parfois être obtenus, toujours pour des jeux de base statiques. Ainsi, Young (1993a) considère un modèle de *fictitious play stochastique* où chaque joueur est doté d'une mémoire bornée, n'observe qu'un échantillon des actions d'autrui et choisit son action soit comme meilleure réponse à ces observations (avec une certaine probabilité) soit de façon aléatoire (avec la probabilité complémentaire). Il montre que, pour les jeux de coordination 2×2 , les équilibres stochastiquement stables correspondent à une sélection des équilibres de Nash en stratégies pures, à savoir les équilibres 'risque-dominants'. Cette sélection est uniquement due aux perturbations aléatoires qui entraînent la trajectoire du système dans les bassins d'attraction les plus étendus. Un résultat voisin est obtenu par Kandori-Mailath-Rob (1993), qui considèrent un processus limite de *fictitious play stochastique* (chaque joueur définit sa meilleure réponse à la distribution des actions des autres sur la période précédente, avec possibilité d'une déviation aléatoire de l'action).

Des résultats plus récents portent sur des jeux de base statiques dans lesquels les joueurs sont situés sur un réseau et n'ont que des interactions locales. Plus précisément, ils sont distribués sur un cercle, un plan quadrillé ou un tore quadrillé et n'interagissent ou ne recueillent d'information qu'avec leurs voisins immédiats. Ellison (1993) reprend le modèle de Kandori-Mailath-Rob (1993) sans perturbations et montre que l'on ne peut qu'aboutir à un cycle limite ou converger vers un équilibre de Nash. En ajoutant une procédure d'échantillonnage stochastique (spatiale) des informations à la Young, Durieu-Solal (2000) montrent que l'on peut éliminer les cycles-limites pour ne garder que les équilibres de Nash; on peut même obtenir les équilibres risque-dominants si les joueurs sont à même d'expérimenter. En tout état de cause, on peut obtenir asymptotiquement diverses structures spatiales, en particulier une segmentation du domaine faisant apparaître tel équilibre dans certaines zones et tel autre dans d'autres zones.

Des résultats ont enfin été obtenus pour des jeux de base dynamiques (génériques et à information parfaite), à savoir une convergence vers l'équilibre parfait (unique). En s'appuyant sur un processus d'apprentissage épistémique particulier autorisant des mutations aléatoires, Nöldeke-Samuelson (1993) ont montré que toute issue 'localement stable' est bien un équilibre parfait, bien que tout équilibre parfait pouvait se trouver dans une composante localement stable avec d'autres équilibres. S'agissant de l'apprentissage par renforcement, il s'exerce à présent sur chaque action relative à un arc de l'arbre du jeu et non plus sur la stratégie globale, l'utilité globale obtenue suite à une trajectoire suivie étant affectée à chacune de ses actions constituantes (Pak, 2001). Avec le modèle CPR, la convergence du processus vers l'équilibre parfait est garantie par le fait que chaque action est jouée une infinité de fois (Laslier-Walliser, 2002). Des résultats semblables ont été obtenus avec des processus évolutionnaires.

Conclusion

Les résultats obtenus dans une perspective éductive et dans une perspective évolutionniste présentent, au regard de la dispersion des contextes considérés, à la fois de fortes similarités et de fortes différences. Les stratégies (itérativement) fortement dominées sont unanimement éliminées. Dans un cadre évolutionniste, les équilibres de Nash purs et stricts sont souvent justifiés, et même

les équilibres mixtes pour l'apprentissage épistémique. En revanche, dans un cadre éductif, les équilibres de Nash ne peuvent être obtenus, qu'ils soient purs ou mixtes, que sous des conditions très drastiques. Des notions d'équilibre plus faibles sont toutefois justifiées dans un cadre éductif, comme l'équilibre rationalisable ou l'équilibre corrélé, qui n'ont pas de contrepartie claire (actuellement) dans un cadre évolutionniste. Enfin, l'équilibre parfait est obtenu dans les deux perspectives, éventuellement avec d'autres équilibres, de manière peu robuste dans un cadre éductif, de manière plus robuste dans un cadre évolutionniste.

Les résultats obtenus s'avèrent très sensibles aux détails de la modélisation et exigent une formalisation très fine pour préciser les conditions de validité de chaque notion d'équilibre. Ceux effectués dans une perspective éductive ont même mis en évidence l'existence d'hypothèses de rationalité implicites jouant un rôle fondamental dans les conclusions. Les travaux réalisés ont conduit à développer les outils analytiques traditionnels et à mobiliser des outils nouveaux, cette quête se poursuivant actuellement. La perspective éductive peut s'enrichir de nouveaux outils logiques comme les logiques temporelles alors que la perspective évolutionniste s'appuie sur de nouveaux théorèmes concernant les processus stochastiques appliqués à des réseaux. Les efforts entrepris permettent enfin d'éclairer certains problèmes plus concrets, s'ils peuvent être exprimés en termes d'équilibres de jeux. Il en est ainsi du problème de la genèse des institutions, que celle-ci résulte de la coordination consciente des acteurs ou d'un processus d'apprentissage partiellement inconscient.

Bibliographie

- Alchourron, C.E.- Gärdenfors, P.- Makinson, D. (1985) : On the logic of theory change : partial meet contraction and revision functions, *Journal of Symbolic Logic*, 50, 510-530.
- Aumann, R.J. (1987) : Correlated equilibrium as an expression of Bayesian rationality, *Econometrica*, 55, 1-18.
- Aumann, R.J. (1995) : Backwards induction and common knowledge of rationality, *Games and Economic Behavior*, 8, 6-19.
- Aumann, R.J. (1996) : Reply to Binmore, *Games and Economic Behavior*, 17, 138-146.
- Aumann, R.J. (1998) : A note on the centipede game, *Games and Economic Behavior*, 23, 97-105.
- Aumann, R.J.- Brandenburger, A. (1995) : Epistemic conditions for Nash equilibrium, *Econometrica*, 63, 1161-80.
- Basu, K. (1990) : On the non-existence of a rationality definition for extensive games, *International Journal of Game Theory*, 19, 33-44.
- Bernheim, D. (1984) : Rationalizable strategic behavior, *Econometrica*, 52, 1007-1028.
- Bicchieri, C. (1988) : Strategic behavior and counterfactuals, *Synthese*, 76, 135-169.
- Bicchieri, C. (1989) : Self-refuting theories of strategic interaction : a paradox of common knowledge, *Erkenntnis*, 30, 69-85.
- Binmore, K. (1987) : Modeling rational players , *Economics and Philosophy*, 3, 9-55 ; 4, 179-214.
- Binmore, K. (1996) : A note on backward induction, *Games and Economic Behavior*, 17, 135-137.
- Binmore, K. (1997) : Rationality and backward induction, *Journal of Economic Methodology*, 4, 23-41.
- Binmore, K. - Samuelson, L. (1996) : Rationalizing backward induction, in K.J.Arrow et al. : *The rational foundations of economic behaviour* , Mc Millan.
- Binmore, K. - Brandenburger, A. (1990) : Common knowledge and game theory, in *Essays in the Foundations of Game Theory* , Blackwell, 105-150.

- Bonnano, G. (1991) : The logic of rational play in games of perfect information, *Economics and Philosophy*, 7, 37-65.
- Brandenburger, A.- Dekel, E. (1987) : Rationalizability and correlated equilibria, *Econometrica*, 55, 1391-1402.
- Brandenburger, A.- Dekel, E. (1989) : The role of common knowledge assumptions in game theory, in F. Hahn ed. : *The economics of missing markets, information and games*, Oxford University Press.
- Crawford, J.- Haller, M. (1990) : Learning how to cooperate : optimal play in repeated coordination games, *Econometrica*, 58(3), 571-596.
- van Damme, E. (1987) : *Stability and Perfection of Nash Equilibria*, Springer Verlag.
- Dekel, E.- Gul, F. (1997) : Rationality and common knowledge in game theory, in D. Kreps- M. Wallis eds. *Advances in Economics and Econometrics*, vol 1, 87-172.
- Durieu, J. –Solal, P. (2000) : Adaptive play with spatial sampling, *mimeo*.
- Ellison, G. (1993) : Learning, local interaction and coordination, *Econometrica*, 61, 1047-1071.
- Fudenberg, D. - Levine, D. (1998) : *The theory of learning in games*, MIT Press.
- Gibbard, A. – Harper, W. (1978) : Counterfactuals and two kinds of expected utility, in C.A. Hooker, J. Leach, E.F. Mac Clennen eds : *Foundations and Applications of Decision Theory*, Reidel.
- Halpern, J.Y. (1999) : Hypothetical knowledge and counterfactual reasoning, *International Journal of Game Theory*, 28, 315-330.
- Halpern, J.Y. (2001) : Substantive rationality and backward induction, *Games and Economic Behavior*, 37, 425-435.
- Harsanyi, J.C. – Selten, R. (1988) : *A general theory of equilibrium selection in games*, MIT Press
- Hofbauer, J. - Sandholm, W.H. (2001) : Evolution and learning in games with randomly disturbed payoffs, *mimeo*.
- Kandori, M.- Mailath, G.- Rob, R. (1993) : Learning, mutation and long run equilibria in games, *Econometrica*, 61, 29-56.
- Katzuno, A.- Mendelzon, A. (1992) : Propositional knowledge base revision and nonmonotonicity, in P. Gärdenfors ed. : *Belief revision*, Cambridge University Press.
- Kreps, D. -Milgrom, P. -Roberts, J. -Wilson, R. (1982) : Rational cooperation in the finitely repeated prisoner's dilemma, *Journal of Economic Theory*, 27, 245-52.
- Laslier, J.-F.- Topol, R.- Walliser, B. (2001) : A behavioral learning process in games, *Games and Economic Behavior*, 37, 340-366.
- Lewis, D. (1969) : *Conventions : a philosophical study*, Harvard University Press.
- Lewis, D. (1973) : *Counterfactuals*, Harvard University Press.
- Nöldeke, G.- Samuelson, L. (1993) : An evolutionary analysis of backward and forward induction, *Games and Economic Behavior*, 5, 425-454.
- Pak, M. (2001) : Reinforcement learning in perfect information games, *mimeo*.
- Pearce, D. (1984) : Rationalizable strategic behavior and the problem of perfection, *Econometrica*, 52, 1008-1050.
- Pettit, P.- Sugden, R. (1989) : The backward induction paradox, *Journal of Philosophy*, 4, 1-14.
- Rabinowicz, W. (1998) : Grappling with the centipede : defence of backward induction for bi-terminating games, *Economics and Philosophy*, 14, 98-126.

- Reny, P. (1992) : Rationality in extensive form games, *Journal of Economic Perspectives*, 6, 103-118.
- Samet, D. (1996) : Hypothetical knowledge in games with perfect information, *Games and Economic Behavior*, 17, 230-251.
- Samuelson, L. : *Evolutionary games and equilibrium selection*, MIT Press.
- Samuelson, L.- Zhang, J. (1992) : Evolutionary stability in asymmetric games, *Journal of Economic Theory*, 4, 364-91.
- Schelling, T. (1960) : *The strategy of conflict*, Harvard UP
- Selten, R. (1975) : Reexamination of the perfectness concept for equilibrium points in extensive games, *International Journal of Game Theory*, 4, 25-55.
- Selten, R. - Leopold, U. (1982) : Subjunctive conditionals in decision theory and game theory, in W. Stegmüller, W. Balzer, W. Spohn, eds. *Studies in Economics, Philosophy of Economics*, Springer Verlag.
- Stalnaker, R.C. (1968) : A theory of conditionals, in N. Rescher ed. *Studies in Logical Theory*, Blackwell.
- Stalnaker, R.C. (1996) : Knowledge, belief and counterfactual reasoning in games, *Economics and Philosophy*, 12, 133-163.
- Stalnaker, R.C. (1998) : Belief revision in games : forward and backward induction, *Mathematical Social Sciences*, 36, 31-56.
- Tan, T.- Werlang, S. (1988) : The Bayesian foundations of solution concepts in games, *Journal of Economic Theory*, 45, 370-391.
- Walliser, B. (1989) : Instrumental rationality and cognitive rationality, *Theory and Decision*, 27, 7-36.
- Walliser, B. (1996) : Comment to Binmore- Samuelson, in K.J.Arrow et al. : *The rational foundations of economic behaviour* , Mc Millan.
- Walliser, B. (1998) : A spectrum of equilibration processes in game theory, *Journal of Evolutionary Economics*, 8, 67-87.
- Walliser, B. (2000) : *L'économie cognitive*, Odile Jacob
- Walliser, B.- Zwirn, D. (2002) : Can Bayes rule be justified by cognitive rationality principles ?, *Theory and Decision*, à paraître.
- Walliser, B.- Zwirn, D.- Zwirn, H. (2002) : Abductive logics in a belief revision framework, *mimeo*.
- Weibull, J. (1995) : *Evolutionary game theory*, MIT Press.
- Young, H.P. (1993) : The evolution of conventions, *Econometrica*, 61, 57-84.
- Young, H.P. (1998) : *Individual strategy and social structure, an evolutionary theory of institutions*, Princeton University Press.