



HAL
open science

Rationalité contextuelle et auto-organisation

Bernard Ancori, Patrick Cohendet, Pierre Garrouste, Jean-Alain Heraud,
Gilles Lambert, Patrick Llerena, Marc Willinger

► To cite this version:

Bernard Ancori, Patrick Cohendet, Pierre Garrouste, Jean-Alain Heraud, Gilles Lambert, et al.. Rationalité contextuelle et auto-organisation. [Rapport de recherche] Centre national de l'entrepreneuriat(CNE). 1990, 219 p., figures, graphiques. hal-02185157

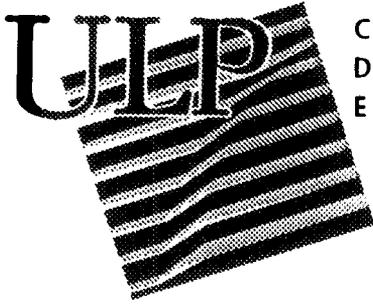
HAL Id: hal-02185157

<https://hal-lara.archives-ouvertes.fr/hal-02185157>

Submitted on 16 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CENTRE DE RECHERCHE
DES SCIENCES ECONOMIQUES
ET DE GESTION



Bureau
d'Economie
Théorique et
Appliquée



URA 1237

UNIVERSITE
LOUIS PASTEUR
STRASBOURG

BETA

38, boulevard d'Anvers
BP 1032/F
67070 Strasbourg Cedex
(France)

Tél. (33) 88 41 52 09

Fax (33) 88 61 37 66

Télex ULP 870 260 F

**Rationalité contextuelle et
auto-organisation**

Décembre 1990

RATIONALITE CONTEXTUELLE
ET
AUTO-ORGANISATION

Dirigé par

Bernard ANCORI

avec

Bernard Ancori, Parick Cohendet, Pierre Garrouste*, Jean-Alain Heraud, Gilles Lambert,
Patrick Llerena et Marc Willinger

S O M M A I R E

	Page
INTRODUCTION GENERALE	1
PARTIE I : VERS UNE RATIONALITE CONTEXTUELLE	
CHAPITRE 1 : L'HYPOTHESE D'INVARIANCE DES PREFERENCES: LES ENSEIGNEMENTS DE L'ECONOMIE EXPERIMENTALE (par Marc WILLINGER)	10
CHAPITRE 2 : MEMOIRE ET APPRENTISSAGE : DE LA NEUROBIOLOGIE A L'AUTO-ORGANISATION (par Bernard ANCORI)	35
CHAPITRE 3 : AUTO-ORGANISATION ET CREATION TECHNOLOGIQUE (par Pierre GARROUSTE et Jean-Alain HERAUD)	81
PARTIE II : AUTO-ORGANISATION, APPRENTISSAGE ET THEORIE DE LA FIRME	
CHAPITRE 4 : APPRENTISSAGE AUTO-ORGANISATIONNEL ET PROBLEMATIQUE DU CHOIX INDIVIDUEL (par Bernard ANCORI)	111
CHAPITRE 5 : NATURE DE L'INFORMATION, COUTS DE TRANSACTION ET ORGANISATION DE L'ENTREPRISE DE L'ENTREPRISE (par Patrick COHENDET et Patrick LLERENA)	144
CHAPITRE 6 APPRENTISSAGE INFORMATIONNEL ET RATIONALITE ECONOMIQUE DES CRITERES D'INVESTISSEMENT (par Gilles LAMBERT)	178
CONCLUSION GENERALE	207
TABLE DES MATIERES	214

INTRODUCTION GENERALE

La théorie économique contemporaine voit coexister, souvent en les opposant, plusieurs formes de rationalité: la "rationalité substantielle" de l'agent néo-classique, cet automate parfaitement individualiste que n'effraie aucun problème de calcul et qui, parfois omniscient et parfois non, est toujours au fait de ses propres préférences ordonnées de manière immuable; celle des acteurs décrits par la théorie des jeux, que certains traits rapprochent de la précédente (jeux non coopératifs) alors que d'autres l'en éloignent (jeux non coopératifs et répétés, débouchant sur une coopération à l'issue d'un processus d'apprentissage); celle connue sous le nom de "rationalité limitée" depuis les travaux de H. Simon [1976, 1978], qui oppose aux capacités de calcul infinies de l'agent optimisateur le caractère restreint de celles d'un décideur se contentant plus modestement, et sur la base d'informations toujours partielles, d'atteindre son objectif de manière jugée par lui "satisfaisante"; celle enfin que J. March [1978] qualifie de "contextuelle", et qui met l'accent sur le fait que le comportement de choix participe toujours d'un système complexe d'interrelations entre les différents processus cognitifs et centres d'intérêt de chaque agent, l'ensemble des agents et les structures sociales considérées.

Chacun de ces concepts présente exactement le même degré de pertinence que le cadre analytique dans lequel il vient fonctionner, eu égard à la nature du problème posé. Veut-on, par exemple, mettre en évidence les différences d'optimalités existant entre les équilibres respectifs de diverses formes de marché? C'est dans le monde de la rationalité substantielle qu'il faut résoudre ce problème, puisque seules les formes de marché sont ici à prendre en compte et qu'il est donc inutile de faire intervenir d'autres différences, telles celles qui porteraient sur les capacités de calcul respectives des micro-sujets présents sur un type donné de marchés. Par contre, pareilles différences de capacités devraient impérativement être explicitées dans d'autres classes de problèmes, tels ceux qui traiteraient de différences de pouvoirs de marchés précisément mises en correspondance avec les capacités de calcul respectives des unités économiques concernées. Bref, il serait totalement irrationnel de préférer tel type de rationalité à tel autre indépendamment de la nature du problème posé. Il n'existe pas plus de concept de rationalité absolument supérieur à tous les autres qu'il n'existe de rationalité en soi, de sorte que le choix entre les différents concepts doit être une affaire de pertinence opératoire, et non l'expression d'un inébranlable parti pris: la solution rationnelle du choix entre les différents concepts de rationalité relève de l'opportunité instrumentale, non d'une métaphysique.

Ainsi, dès lors que nous analysons certains aspects de la décision économique en situation d'incertitude et d'irréversibilité, nous devons rejeter le concept de rationalité substantielle. En effet, au contraire du risque, l'incertitude vraie (non probabilisable) échappe par définition à toute rationalité substantielle, et, comme l'a montré O. Favereau [1989], l'irréversibilité, qui n'existe pas dans une telle conception de la rationalité, peut néanmoins y être réintroduite par le biais du concept de valeur d'option: certes, cette dernière devient ici l'expression monétaire de l'effet-irréversibilité, mais le paradoxe est qu'il faille confronter rationalités procédurale et substantielle pour parvenir à un tel résultat!

C'est dire qu'au contraire de la précédente, la "rationalité limitée" à la Simon ne doit jamais être perdue de vue par une analyse de la décision irréversible en incertain, centrée sur l'étude de phénomènes d'apprentissage qui se trouvent également au coeur de ce type de rationalité. Tout en gardant pareille exigence présente à l'esprit, c'est cependant à un dépassement du modèle simonien que le présent ouvrage convie le lecteur: au-delà de la reconnaissance des limites de capacité de calcul des agents, et de la prise en compte d'un processus de décision dont les effets d'apprentissage sont souvent plus importants pour l'agent que le résultat de la décision lui-même, c'est un concept de *rationalité contextuelle* qui est ici exploré, et pareille exploration implique de donner un traitement théorique explicite à l'impact de l'environnement de l'agent sur les choix effectués par lui, aussi bien que, en retour, à l'impact des choix de cet agent sur son environnement.

A chaque fois que nous choisissons, c'est *hic et nunc* que nous le faisons. La banalité de ce constat est telle que ses conséquences n'ont jamais été réellement prises au sérieux. Ainsi la théorie traditionnelle juge-t-elle inutile de prendre explicitement en compte l'influence du contexte du choix sur l'expression de celui-ci. Sauf à affirmer, par exemple, que cette influence est incorporée d'emblée dans la description des préférences de l'agent et que l'identité analytique de ce dernier devient, par suite, variable selon le contexte, une telle attitude revient à considérer qu'une telle influence est parfaitement nulle, soit parce que le contexte n'en exercerait jamais aucune sur l'expression du choix individuel, soit parce qu'il exercerait toujours la même et que celle-ci pourrait dès lors être laissée dans l'implicite de la théorie.

Sous des formes diverses, c'est toujours à l'un ou l'autre de ces trois types de présupposés que l'on peut associer l'hypothèse d'invariance du système de préférences de l'agent économique - une hypothèse absolument fondamentale pour la théorie traditionnelle du choix. Ne nous attardons pas aux deux derniers, pour dire plutôt quelques mots du premier, beaucoup plus sophistiqué qu'un individualisme si outré qu'il en devient naïf et qu'un fixisme contextuel que l'observation la plus immédiate suffit à démentir. Ainsi ce que nous pouvons appeler le "programme intégrationniste" considère-t-il que toute déviation constatée par rapport aux comportements prédits par la théorie résulte d'une insuffisante description aux agents des conséquences de leurs choix: ces agents étant supposés dotés d'une rationalité préexistante à l'expression de leurs choix, et uniquement préoccupés des conséquences ultimes de ceux-ci, il suffirait de mieux leur détailler ces dernières pour voir toute anomalie disparaître. Ces dernières subsistent-elles néanmoins? Ce sont alors les capacités cognitives des agents eux-mêmes qu'il faut faire figurer au nombre des contraintes inscrites dans le programme d'optimisation décrivant leur comportement, et l'optimum est défini comme une solution qui minimise l'"effort mental" de l'agent individuel. De même, les difficultés liées à l'acquisition et au traitement de l'information seraient assimilables à des "coûts de transaction", et les décisions effectives des agents interprétables comme autant de "choix ajustés" compte tenu de ces coûts de transaction. Et voilà pourquoi votre fille est muette!

Pareille démarche suggère une rationalité substantielle rigoureusement infalsifiable au sens de K. R. Popper [1959]: à condition d'en respécifier au besoin les termes avec autant de précision que nécessaire, elle serait capable d'expliquer n'importe quelle anomalie venue démentir une description auparavant donnée avec un degré de précision tout juste inférieur. Outre cet argument épistémologique qui montre les efforts déployés pour assurer un monopole au monde de la rationalité substantielle aboutir à le faire voisiner avec celui de l'astrologie, deux autres raisons nous incitent à rejeter la voie "intégrationniste": celle-ci demande d'exposer les conséquences du choix avec un degré de détail à la limite infini et le plus souvent inobservable en pratique, ce qui élimine quasiment toute possibilité d'application; de plus, elle est fondée sur l'idée selon laquelle la rationalité préexisterait au choix. Or les travaux de H. Simon et d'autres ont montré que les agents ne naissent pas rationnels, mais apprennent à le devenir, notamment à travers les conséquences inattendues des choix qu'ils effectuent quotidiennement.

Au contraire, l'hypothèse de rationalité contextuelle que nous voulons défendre implique que les systèmes de préférences des agents sont liés aux contextes de leurs expressions selon une dépendance conçue elle-même comme le résultat d'un processus d'apprentissage, la rationalité qui en découle n'étant qu'un point de convergence de ce processus. La pertinence d'une telle hypothèse émerge alors de la variabilité dont font simultanément preuve le contexte de l'agent individuel et l'expression de ses choix: le lieu privilégié de son exploration consiste en l'espace théorique offert à l'évolution d'un organisme doté d'une rationalité limitée au sein d'un monde incertain et dont la trajectoire individuelle apparaît scandée par un apprentissage non programmé, effectué lors d'une succession ininterrompue d'interactions entre cet organisme et son environnement pertinent, et ayant pour effet d'alimenter l'incertitude propre à d'autres organismes individuels partageant le même monde.

La rationalité contextuelle ainsi entendue sera analysée en deux étapes dans cet ouvrage. Dans une première partie (**Vers une rationalité contextuelle**), nous visons à asseoir très précisément la pertinence théorique d'un tel concept, et dans une seconde partie (**Auto-organisation, apprentissage et théorie de la firme**), à faire fonctionner ce concept ainsi légitimé au sein d'espaces théoriques définis selon un degré décroissant d'abstraction.

Il importe en effet de montrer tout d'abord que le concept de rationalité contextuelle possède un contenu positif à la fois réellement observable et offert à une amorce de théorisation. Dans ce but, nous avons divisé notre première partie en trois chapitres. Dans le premier, **M. Willinger met à l'épreuve l'hypothèse d'invariance des préférences à travers les enseignements de l'économie expérimentale**: il s'agit ici de recenser les principaux biais de comportements mis en évidence par les psychologues, ainsi que d'analyser les implications de ces biais pour la théorie économique.

Inversion des préférences, révélation de préférences auto-contradictoires lors de présentations différentes des mêmes options, biais liés au traitement de l'information par l'agent victime d'illusions ou d'erreurs de perception quant à la nature réelle des options présentées: tels sont les trois principaux types de biais de comportements exhibés par les agents individuels testés, et ils sont systématiques. Sauf à remettre en cause la représentativité des expériences, ils apportent ainsi un démenti flagrant à l'hypothèse d'invariance des systèmes de préférences des agents, et encouragent l'expression d'alternatives à l'hypothèse d'optimisation. Plutôt que d'optimiser au sens traditionnel du terme, les agents utiliseraient-ils des règles empiriques simples leur permettant d'aboutir rapidement à une décision "satisfaisante" dans un univers cognitif dominé par une rationalité limitée à la Simon? S'éloignant plus radicalement encore du paradigme de l'optimisation, verraient-ils leurs préférences individuelles dépendre très étroitement du contexte de présentation et d'expression des choix? Quelles que soient les réponses données à de telles questions, les résultats expérimentaux ici exposés et discutés invitent l'économiste à entrer dans la "boîte noire" du décideur, afin de ne plus laisser au seul psychologue le soin de définir une "technologie de la décision" cohérente avec l'observation.

Notre deuxième chapitre, **Mémoire et apprentissage: de la neurobiologie à l'auto-organisation**, constitue précisément une tentative d'entrée dans la "boîte noire" mentionnée à l'instant, puisque **B. Ancori** y rapproche les conceptions de la mémoire et de l'apprentissage propres à I. Rosenfield et à H. Atlan, dont les traits partagés se révèlent aptes à expliquer rationnellement l'ensemble des biais précédents. Sur des bases neurobiologiques, ces deux conceptions s'entendent en effet pour rejeter toute notion de mémoire humaine fonctionnant sur un mode invariable comme une sorte de réservoir dans lequel la performance de remémoration viendrait mécaniquement puiser des souvenirs fixes afin de les comparer au perçu actuel selon une recherche de correspondance point par point. A rebours de cette lecture de notre mémoire comme une structure statique, de pareilles vues la donnent pour fondamentalement inventive, au sens où nous ne nous souvenons qu'en percevant dans le contexte actuel d'approximatives coïncidences avec un perçu ancien ayant fait l'objet d'une catégorisation. Dépendant du contexte présent et variant avec lui, ce fonctionnement de la mémoire est nécessairement dynamique: d'être associée au sein d'un geste unique avec la perception, la performance de remémoration est ici saisie comme un véritable processus d'apprentissage, projetant sans cesse sur le contexte présent des catégories anciennes que cette projection elle-même vient mobiliser, comme elle modifie le mode même de catégorisation opéré par le cerveau. Non programmé, issu d'interactions entre le système observé et son environnement, producteur de catégories toujours plus différenciées, ce type d'apprentissage ressortit à une logique connue pour être celle de l'auto-organisation. Les données expérimentales évoquées au chapitre précédent cessent alors de représenter des anomalies (elles ne sont telles qu'aux yeux des partisans de l'invariance du système de préférences individuelles) pour apparaître parfaitement normales dans le cadre de la rationalité contextuelle ainsi esquissée.

Pour avoir dès lors commencé de révéler sa fécondité potentielle, le concept d'auto-organisation ne se réduit pas pour autant à la seule interprétation qu'en a donné H. Atlan. Le troisième et dernier chapitre de notre première partie, **Auto-organisation et création technologique**, établit un point critique des différentes approches existant actuellement en cette matière. Après avoir recensé les différents concepts que l'on peut aujourd'hui réunir sous l'appellation générique "auto-organisation", P. Garrouste et J.-A. Héraud évaluent les portées respectives de leurs applications à l'étude des changements technologiques. Utilisés explicitement ou non dans l'analyse de la dynamique d'une industrie, du processus de compétition technologique, du système autonome d'une industrie, de celui de la technologie ou de l'évolution globale de la société, les concepts d'auto-organisation introduits par H. Atlan ou par I. Prigogine et I. Stengers, ou encore ceux d'autopoïèse et d'autonomie proposés par H. Maturana ou par F. Varela, n'ont ni la même signification, ni la même portée théorique. De leur réunion sous une même rubrique émergent néanmoins deux idées liées: d'abord, celle qui affirme la nécessité de ne pas se satisfaire d'un processus qui verrait l'évolution d'un système auto-organisateur se traduire par l'épuisement progressif de celui-ci sous les chocs répétés provenant d'un environnement lui-même inerte, mais demande au contraire de penser l'auto-organisation comme co-évolution d'un système et de son environnement selon des trajectoires sans limites assignées a priori; ensuite, celle qui, dans le prolongement de l'idée précédente qu'elle vient préciser, insiste sur la double dimension de tout phénomène d'auto-organisation, celui-ci se traduisant par l'auto-information du système considéré sous l'impact de son environnement mais aussi par l'auto-organisation de ce dernier auquel le système imprime sa marque propre.

Au terme de cette première partie, la pertinence du concept de rationalité contextuelle se voit ainsi démontrée: non seulement le type de problèmes qu'elle permet d'envisager n'a rien de périphérique eu égard aux comportements observés des agents, mais encore tous les éléments permettant de traiter ces problèmes sur le plan théorique sont-ils d'ores et déjà présents dans notre épistémè - fût-ce de manière relativement dispersée, d'être situés en des régions différentes de celle-ci. Constatée au niveau expérimental par les psychologues, l'absence d'invariance du système des préférences individuelles trouve un début d'explication théorique sur la base conjointement fournie par la neurobiologie et la théorie des systèmes auto-organisateurs. Il reste à l'économiste à faire son miel de l'ensemble de ces éléments, offerts comme autant de matériaux disponibles pour un renouvellement de la théorie économique de l'apprentissage et de l'évolution.

Tel est l'objet de la seconde partie de cet ouvrage: la pertinence du concept de rationalité contextuelle étant assurée par la conjonction d'un indéniable contenu positif et d'une direction théorique fermement tracée, il convient de faire fonctionner ce concept en l'associant plus étroitement à la double dimension de l'apprentissage auto-organisationnel.

C'est ainsi qu'un quatrième chapitre, **Apprentissage auto-organisationnel et problématique du choix individuel**, propose une formalisation à la fois très simple et très abstraite de l'ensemble des phénomènes précédemment mis en évidence: co-évolution du système et de son environnement, auto-information du premier par le second et auto-organisation du second par le premier, signification et création de l'information, etc., ce sont toutes ces dimensions de l'apprentissage et de l'évolution économiques qui sont ici analysées par le modèle de **B. Ancori**. Utilisant à un niveau élémentaire le langage de la théorie des graphes et des réseaux, ce modèle est celui d'un système complexe composé de ces sous-systèmes cognitifs que figurent des agents économiques pleinement saisis dans leurs activités de sujets connaissants, ainsi que des interactions potentielles et effectivement établies par ces mêmes agents à l'occasion de leurs communications. Lors de chacune de ces dernières, chaque sous-système constitue l'environnement (le contexte) de l'autre, et c'est précisément la communication alors établie qui est le moteur de leur co-évolution comme de celle du système considéré globalement. C'est parce que les agents accordent tous des significations différentes à une information néanmoins communicable (lors d'une communication, ce qui est reçu n'est pas strictement identique à ce qui est émis) que leurs interactions, loin de se réduire à une simple transmission d'information qui laisserait inchangée la mémoire globale du système, se traduit, au niveau de ce dernier comme à celui de ses composants, par une véritable création d'information. Ce modèle débouche sur une esquisse de problématique du choix individuel, étant donnée la forme imprimée à pareille évolution - un choix dont les termes respectifs se rapprochent des deux lectures possibles du concept de valeur d'option, puisqu'il s'établit entre une stratégie passive de rétention d'information, analogue à l'attitude d'un décideur reportant l'option irréversible dans l'attente d'informations supplémentaires, et une stratégie active d'émission d'information, proche de l'attitude d'un décideur optant pour la décision irréversible en pariant sur les gains qu'il en attend en termes d'apprentissage.

Un cinquième chapitre, **Nature de l'information, Coûts de transaction et Organisation de l'entreprise**, descend d'un cran dans l'abstraction pour analyser les implications théoriques de l'évolution récente des systèmes de production dans les économies industrielles modernes: l'environnement (le contexte) étant identifié à la réunion des conditions de la demande, de la concurrence et de la technologie, **P. Cohendet** et **P. Llerena** nous invitent à pénétrer plus profondément dans la "boîte noire" de la firme - ici le système observé de manière privilégiée. Contraint de repenser sa stratégie par le passage d'un régime d'information stable à un régime d'information perturbée en provenance de son environnement, ce système voit se poser de manière inédite le double problème de son adaptation optimale à l'évolution de ce dernier, et de l'évaluation de la contribution de chacun de ses sous-systèmes à sa performance globale. Chaque modèle de production ne supprimant pas les précédents mais venant s'y rajouter, le modèle de la standardisation et le monde du certain prévalant jusqu'au début des années soixante, comme celui de la variété et le monde du divers alors venu le compléter, s'accompagnent aujourd'hui du modèle de variété et du monde de l'incertain.

Alors que les deux précédents s'inscrivaient globalement dans un régime d'information stable, le troisième ouvre l'ère d'un régime d'information perturbée. Outre la réorganisation de la firme autour d'une stratégie de flexibilité dynamique remplaçant désormais ceux de productivité et de différenciation retardée, l'intégration conçue comme un principe d'allocation est à présent d'ordre informationnel et non plus spatiale ou physique: réduire la complexité engendrée par une meilleure flexibilité dynamique synonyme de multiplication des interactions internes (des "transaction") conduit à minimiser les coûts de transaction entre sous-systèmes tout en maximisant les effets d'apprentissage à l'intérieur de ceux-ci. Mais l'un des enseignements essentiels de ce chapitre est de montrer qu'à l'inverse d'une indépendance acceptable entre organisation de la firme et évolution de l'environnement en régime d'information stable, donc d'une séparation nette entre économie des relations externes de la firme et gestion de son organisation interne, l'avènement d'un régime d'information perturbée s'accompagne de l'apparition d'un certain flou dans la frontière du système constituée par la firme, donc d'un relatif effacement de la distinction entre économie et gestion.

Il était donc intéressant de clore cet ouvrage par un sixième et dernier chapitre, **Apprentissage informationnel et rationalité économique des critères d'investissement**, donnant le point de vue du gestionnaire sur les outils théoriques disponibles ou devant être forgés, dans la perspective d'une firme conçue comme un système auto-organisateur. C'est ainsi que G. Lambert met successivement l'accent sur les deux dimensions que nous avons reconnues être celles de l'auto-organisation, en se situant dans le cadre d'une rationalité procédurale à la Simon - et non d'une rationalité substantielle. Le système ici considéré est l'investisseur potentiel confronté à l'incertitude d'un environnement défini sur le même mode que dans le chapitre précédent, est l'opportunité de l'investissement est considérée successivement du point de vue de la capacité d'adaptation de la firme et de celui de sa dimension organisationnelle. Le premier point de vue, qui se situe dans l'optique globale de l'impact de l'environnement sur le système, débouche sur une conception stratégique de l'information, dont la seule perspective d'acquisition peut suffire à justifier le report d'un projet d'investissement. L'analyse bayésienne permet en effet de formaliser un critère de décision axé sur la préoccupation de l'agent de maintenir intact son espace de choix possibles futurs, rompant ainsi avec l'analyse financière classique en montrant qu'il est parfois opportun de rejeter un projet d'investissement malgré une valeur actualisée nette strictement positive. Après avoir été établi dans un cadre *a posteriori* de décisions d'investissement avec irréversibilité et croissance de l'information, le critère est complété par la définition d'un prix-limite consenti par le décideur en échange du service informationnel que lui rendrait un expert, chargé par lui de déterminer, avec un degré donné de fiabilité, l'état qui se réalisera effectivement parmi les états du monde possibles. Dans ce cadre *pre posteriori*, l'investissement informationnel précède donc l'investissement proprement dit. Plutôt que d'être associée au paramètre d'incertitude, comme c'est le cas lorsque l'information est révélée par la nature, la valeur de l'information révélée par l'expert lui est conférée par son caractère dissuasif vis-à-vis des différentes options.

Le second point de vue examine la dimension organisationnelle de l'investissement à travers les effets d'apprentissage qu'il permet éventuellement d'obtenir, non pas à l'intérieur de la firme selon une direction de recherches traditionnelle en sociologie des organisations, mais plutôt en étant créateur d'options technologiques nouvelles au niveau de la structure productive globale de l'économie. Cette perspective est bien celle qui correspond à l'étude de l'impact du système sur son environnement, et il est montré qu'il est possible, sous certaines conditions, d'incorporer de tels changements de structure globale dans le calcul de rentabilité effectué *ex ante* par l'investisseur individuel. Enfin, le chapitre s'achève avec l'illustration de cette dimension organisationnelle de l'investissement par quelques exemples d'investissements en rationalité procédurale.

D'un système de communication sociale formalisé en termes de la théorie des graphes à l'ingénierie simultanée comme exemple de la dimension organisationnelle de l'investissement, en passant par une redéfinition théorique de la stratégie optimale de la firme confrontée à un régime d'information perturbée: c'est à des niveaux d'abstraction très différents, et dans des domaines d'investigation très divers, que cette seconde partie fait fonctionner notre concept de rationalité contextuelle et la conception globale de l'apprentissage comme phénomène d'auto-organisation. Loin d'exprimer un impressionnisme non désiré, pareilles hétérogénéités et ruptures de tons témoignent de la multidimensionnalité des niveaux de discours et des domaines d'application potentiels d'un concept que nous invitons maintenant le lecteur à découvrir plus en détail.

Bibliographie

FAVEREAU O. [1989], "Valeur d'option et flexibilité: de la rationalité substantielle à la rationalité procédurale", in Cohendet P. et Llerena P. éd., "*Flexibilité, Information et Décision*", Economica, pp. 121-182

MARCH J. [1978], "Bounded Rationality, Ambiguity, and the Engineering of Choice", *Bell Journal of Economics*, 9, pp. 587-608

POPPER K. R. [1959], "*The Logic of Scientific Discovery*", Harper Torchbooks

SIMON H. [1976], "From Substantive to Procedural Rationality", in Latsis S. ed., "*Method and Appraisal in Economics*", Cambridge University Press

SIMON H. [1978], "Rationality as process and as product of thought", *American Economic Review*, Vol 68

Introduction

La théorie économique est en grande partie fondée sur l'analyse des comportements. De ce fait, elle est parfois considérée comme une science du comportement, au même titre que la psychologie. Les deux disciplines ont toutefois des projets fondamentalement différents. La première cherche à déterminer *comment* les agents devraient se comporter, alors que la seconde s'intéresse au *pourquoi* des comportements observables. Du point de vue de leurs objectifs, elles apparaissent donc comme largement complémentaires. Par le passé, ces complémentarités ont rarement donné lieu à des échanges. Depuis quelques années cependant, l'état d'ignorance mutuelle a fait la place à des controverses, parfois très vives. C'est le cas pour l'un des thèmes cher aux économistes, celui de la rationalité des choix. En dépit des désaccords sur le sujet, des échanges entre les deux disciplines ont eu lieu. Ainsi, certains psychologues ont adopté la théorie de la maximisation de l'utilité comme hypothèse de travail (H. Rachlin & alii, [1981]), alors que certains économistes ont eu recours aux méthodes expérimentales, utilisées par les psychologues, pour tester certaines hypothèses théoriques (cf. C. Plott, [1982]).

L'introduction de méthodes expérimentales (1) dans la panoplie des outils à la disposition de l'économiste, a un intérêt évident : celui de pouvoir simuler les comportements réels. Grâce à cet outil, il est possible d'identifier les variables explicatives de ces comportements dans un environnement contrôlé. On peut donc isoler, comme dans un laboratoire, les variables-clés et tester l'action de différents paramètres sur le comportement, grâce à une reproductibilité presque parfaite de cet environnement contrôlé. Bien sûr, les expériences possèdent des imperfections, et pour l'instant, leur champ d'application est très limité en économie. Elles peuvent cependant aider à mieux comprendre la portée de certaines hypothèses de comportement, et parfois à formuler des hypothèses alternatives plus satisfaisantes.

C'est dans cette perspective que ce texte aborde la question des préférences. L'analyse économique repose sur l'hypothèse qu'il existe un système de préférences "invariant". Sur la base de ces préférences les agents évaluent les options de choix et prennent des décisions. Les décisions sont donc supposées cohérentes entre elles et en accord avec les préférences. Dans le cas de la théorie du consommateur par exemple, la propriété d'invariance des préférences correspond au fait que les préférences ne dépendent pas du revenu. Si un consommateur préfère A à B étant donné une certaine contrainte de budget, il continuera à préférer A à B si le budget change. Cette modification de l'environnement n'a aucune incidence sur son ordre de préférence.

Plus généralement, nous admettrons dans ce texte qu'un système de préférence est invariant lorsqu'il satisfait aux trois propriétés ci-dessous, où a_i désigne une option de choix pouvant représenter un vecteur de biens, une distribution de probabilités sur un ensemble de conséquences,

une séquence de choix consécutifs, etc..., et \geq^* est une représentation des préférences sur un ensemble d'options, A.

- (P₁) Supposons $a_1, a_2 \in A$ et $a_1, a_2 \in A'$. Si $a_1 \geq^* a_2$ pour A, alors $a_1 \geq^* a_2$ pour A'.
- (P₂) Supposons que \geq^* admet une représentation $v(\cdot)$, telle que $a_1 \geq^* a_2 \Leftrightarrow v(a_1) \geq v(a_2)$.
Si $f(\cdot)$ est une transformation admissible de la représentation $v(\cdot)$, telle que $f(v(\cdot)) = w(\cdot)$, alors on doit avoir $w(a_1) \geq w(a_2)$.
- (P₃) Si $a_i \in A$ et $a_i' \in A'$ sont deux descriptions possibles pour les options, alors :
 $a_1 \geq^* a_2 \Rightarrow a_1' \geq^* a_2'$.

P₁ implique que les préférences sont indépendantes de l'ensemble des options; P₂ implique, en particulier, que lorsqu'on peut choisir librement l'unité de mesure pour les préférences, l'ordre est indépendant de l'unité retenue. Enfin, d'après P₃, les préférences dépendent seulement des conséquences terminales des options, et sont indépendantes des autres caractéristiques telles que le nombre d'étapes dans le processus de décision, la présentation sous forme "réduite" ou sous forme "extensive", etc... correspond au principe d'"extensionnalité" de K. Arrow [1982] et au "conséquentialisme" de P. Hammond [1988].

L'hypothèse d'un système de préférences invariant a été rarement mise en cause par les économistes. Depuis quelques années cependant, les doutes se sont accumulés (2) quant à l'invariance des préférences, et quant à l'existence même d'un système de préférences. Un nombre croissant de travaux expérimentaux montre en effet la nature fondamentalement "versatile" des préférences. Ces travaux montrent notamment une forte dépendance par rapport à la manière dont les options de choix sont présentées aux sujets et par rapport au mode de réponse retenu pour l'expression de ces préférences.

Ces résultats, en apparence peu surprenants pour un non-économiste, sont en profond désaccord avec toutes les théories normatives de la décision qui supposent l'existence d'un système de préférence invariant. En effet, ces théories admettent généralement les trois propositions suivantes : 1) La description des options n'influence pas les choix. Par exemple, lorsqu'un agent doit choisir entre deux risques, il importe peu que ces risques soient décrits par des fonctions de répartition cumulatives ("probabilité de réaliser un gain au plus égal à") ou des fonctions de répartition décumulatives ("probabilité de réaliser un gain au moins égal à"). Le choix doit être le même dans les deux cas. 2) L'évaluation coïncide avec le choix. Demander à un agent de choisir entre deux options, ou lui demander d'indiquer la valeur qu'il attribue à chacune de ces options, devrait conduire à sélectionner la même option. Pour l'économiste, la réponse à ces deux questions est

nécessairement la même dès lors que les agents sont supposés être rationnels. 3) Les préférences sont indépendantes de la méthode utilisée pour les exprimer. Par exemple, lorsqu'on demande à l'agent d'exprimer ses préférences en attribuant une note de 1 à 10, on devrait obtenir le même classement des options que dans l'hypothèse où on lui demande de définir un prix de réservation pour chacune de ces mêmes options. Ces trois prédictions simples, nécessaires pour la cohérence des modèles de décision, sont toutes infirmées par les expériences.

Ces contradictions ont des conséquences très importantes. Premièrement, elles démontrent de façon radicale l'absence de vertu prédictive de nos modèles de comportement. Deuxièmement, elles conduisent à s'interroger sur la validité normative de ces mêmes modèles et sur les recommandations pratiques qui en découlent. Enfin, troisièmement, elles mettent en évidence le jeu de variables non identifiées ou non contrôlées dans les processus d'évaluation et de prise de décision. De nombreuses critiques ont été adressées à ces résultats, en particulier leur manque de réalisme ou de représentativité des situations réelles, l'absence d'apprentissage et le manque de motivations ou d'incitations pour les agents. Les tests ont été réitérés pour tenir compte de ces critiques et pour mieux simuler la réalité économique. La plupart des effets persistent, malgré les nombreux stratagèmes imaginés en vue de les éliminer. Il s'agit donc de phénomènes robustes, reproductibles, et suffisamment massifs pour qu'ils interpellent aussi bien le théoricien que le décideur.

Ce texte se propose de discuter les résultats expérimentaux qui selon nous mettent en cause la propriété d'invariance des préférences, en mettant en évidence leurs implications théoriques et pratiques, en particulier en ce qui concerne l'évaluation et la prise de décision. Dans les trois premiers paragraphes nous montrerons la portée théorique et pratique des principaux résultats expérimentaux qui mettent en cause l'hypothèse d'invariance des préférences. Le premier des trois résultats discutés ci-dessous, le phénomène d'inversion des préférences, se rapporte à la manière dont les individus expriment leurs préférences. Le second, appelé "framing effect", est relatif à la façon dont les agents perçoivent leurs options de choix. Enfin, le troisième effet, s'applique à la manière dont les agents traitent l'information. Après cette présentation nous discuterons les arguments qui sont généralement invoqués à l'encontre de l'approche expérimentale en économie. Le dernier paragraphe abordera la question de la modélisation des préférences en tenant compte des résultats nouveaux issus de l'approche expérimentale.

1. L'inversion des préférences (Preference reversal)

Nous commençons par présenter le résultat correspondant à ce type d'expériences, puis nous en donnerons diverses interprétations avant d'en montrer les implications pour la théorie économique.

1.1 Le résultat expérimental

Ce phénomène a été tout d'abord mis en évidence par S. Lichtenstein & P. Slovic [1971] et par H. Lindman [1971]. Il s'agissait de montrer que le comportement face au risque, obéit généralement à d'autres règles que celle que préconisent les économistes. Plus exactement, deux règles de choix, considérées comme identiques en termes de préférences, la maximisation de l'utilité espérée et la "volonté de payer" (offre d'achat), peuvent aboutir selon ces auteurs à des ordres de préférences distincts contrairement à ce que prédit la théorie économique. Le test est le suivant : deux types d'options, ayant la même espérance de gain, sont présentées aux sujets de l'expérience. La première (appelée P-bet) a une probabilité élevée (proche de l'unité), de faire gagner une certaine somme. La seconde option (appelée \$-bet) a une probabilité plus faible de gagner, mais offre un gain beaucoup plus élevé en cas de succès. Ces deux types d'options sont illustrées sur la figure 1.

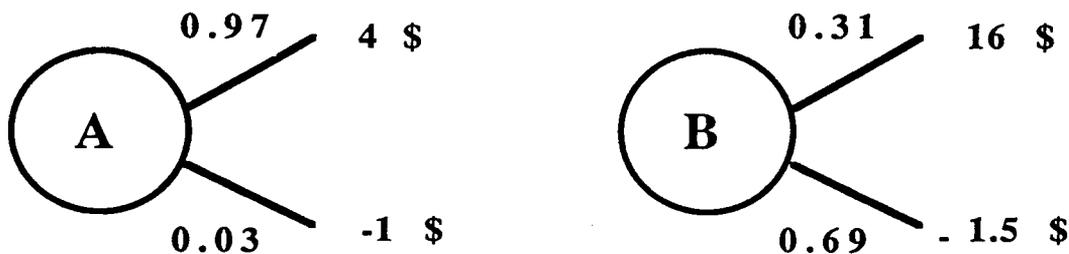


figure 1 : Effet d'inversion des préférences

Dans une première phase de l'expérience, on demande aux candidats de choisir l'une de ces options, et dans une seconde phase on leur demande de fixer le prix minimum auquel ils sont disposés à céder chacune de ces options (ou alternativement, le prix maximum qu'ils sont disposés à payer pour leur acquisition). La plupart des gens choisissent l'option A de préférence à B, mais

placent une valeur plus grande sur B. En d'autres termes on observe une inversion des préférences entre le choix et le jugement de valeur.

Ce résultat troublant a suscité bien des scepticismes et des critiques, de sorte que l'expérience a été réitérée un grand nombre de fois en modifiant les paramètres (notamment les enjeux) et le dispositif expérimental. A la suite de S. Lichtenstein & P. Slovic [1971] et de H. Lindman [1971], l'expérience a été menée par D. Grether & C. Plott [1979], W. Pommerehne, F. Schneider & P. Zweifel [1982], R. Reilly [1982], P. Slovic & S. Lichtenstein [1983], J. Berg, J. Dickhaut & J. O'Brien [1985] et par J. Cox & S. Epstein [1989]. L'étude de S. Lichtenstein & P. Slovic [1973] est particulièrement intéressante car elle montre l'existence de l'effet même en-dehors du contexte expérimental, c'est-à-dire dans une situation mettant en jeu de vrais joueurs dans un Casino de Las Vegas.

1.2 Des interprétations divergentes

Selon D. Grether & C. Plott [1979], cet effet révèle une incohérence fondamentale du comportement humain, qui ne semble obéir à aucune règle d'optimisation. D'un point de vue économique, la valeur la plus grande doit être attachée à l'option préférée. Il importe peu que l'ordre de préférence soit révélé par le choix ou par le jugement de valeur. Il faut que prix et préférences coïncident. Pour D. Grether & C. Plott le phénomène d'inversion des préférences contredit l'axiome de transitivité des préférences. De ce fait, les principes de rationalité du comportement des agents, qui servent habituellement à expliquer les "lois du marché", ne s'appliquent pas. Des principes différents sont à la base des régularités que l'on observe dans le fonctionnement des marchés.

La conclusion de D. Grether et C. Plott n'a pas fait l'unanimité comme on pouvait s'y attendre. D'une part, les économistes proposent un diagnostic très différent de l'effet d'inversion, et de l'autre, les psychologues font appel à d'autres variables de comportement pour expliquer ce phénomène. Pour les économistes, étant donné l'importance de l'approche normative du comportement, l'axiome de transitivité reste un des postulats-clés pour définir la rationalité. De ce fait, ils ont tenté de démontrer que les résultats observés mettaient en cause d'autres axiomes que la transitivité. Pour les psychologues au contraire, davantage préoccupés par l'aspect positif que normatif de ces expériences, il s'agissait avant tout d'identifier les variables susceptibles d'être à l'origine des comportements observés.

Du côté des économistes, C. Holt [1986] et E. Karni & Z. Safra [1987] ont montré que l'expérience menée par D. Grether et C. Plott ne permettait pas de conclure quant au postulat de rationalité réellement incriminé dans la théorie de l'utilité espérée. Au mieux, selon eux, on peut en conclure que c'est l'hypothèse de la maximisation de l'utilité espérée qui est en cause. Ils montrent

alors que l'effet d'inversion n'est pas en contradiction avec l'axiome de transitivité des choix, qui est à la base de la plupart des théories de la décision, mais qu'il infirme l'axiome d'indépendance, le postulat central de la théorie de l'utilité espérée. Par la suite U. Segal [1988] a conforté cette position en montrant que le phénomène d'inversion pouvait très bien s'analyser comme une réfutation de l'axiome de combinaison des loteries (compound lottery axiom), un axiome qui est lui aussi spécifique à l'hypothèse de la maximisation de l'utilité espérée. Ces conclusions, tout en offrant un mobile supplémentaire de rejet de la théorie de l'utilité espérée, permettent simultanément de valider les nouvelles constructions théoriques qui continuent à s'appuyer sur un axiome de transitivité (cf. M. Willinger [1990], pour un survey). L'objectif était de montrer que les procédures expérimentales ne permettent pas de discriminer entre les causes du phénomène d'inversion, mais autorisent le rejet de l'hypothèse de Bernoulli. Toutefois, l'étude expérimentale menée par J. Cox & S. Epstein [1989] semble confirmer les conclusions pessimistes de D. Grether & C. Plott. Les deux auteurs montrent en effet, que le phénomène d'inversion contredit l'axiome d'antisymétrie, un axiome qui n'est pas spécifique à l'hypothèse de l'utilité espérée, et qui sous-tend un grand nombre de théories de la rationalité des choix.

Certains économistes ont adopté une attitude plus positive. Ainsi, pour tenter d'expliquer le phénomène d'inversion et l'existence de non-transitivités, G. Loomes & R. Sugden [1982, 1983] se fondent sur la "théorie du regret" élaborée par D. Bell [1982]. L'idée est que les agents évaluent leurs options non pas isolément, en référence à un système de préférence invariant et préexistant, mais relativement aux autres options présentes au moment du choix. Lorsqu'un agent compare deux options, il tient compte pour chaque éventualité, du manque à gagner d'une option par rapport à l'autre. En d'autres termes il anticipe l'"intensité de son regret" d'avoir choisi une option plutôt qu'une autre. L'intérêt de cette théorie est qu'elle possède précisément la propriété d'intransitivité qui, selon D. Grether & C. Plott, est la cause du phénomène.

Les psychologues ont une attitude très différente de celle des économistes quant à l'interprétation du phénomène. Leurs explications se fondent généralement sur la différence de mode de réponse entre jugement et choix. La différence de mode de réponse fait appel à des processus cognitifs différents qui conduisent à des décisions différentes (S. Lichtenstein & P. Slovic, [1971], [1973]). La question "Que préférez-vous ?" appelle une réponse qualitativement différente que la question "Quelle est la valeur de chacune des options ?". Il n'y donc aucune raison a priori de retrouver le même ordre à partir de processus différents. W. Goldstein & H. Einhorn [1987] ont montré qu'il y avait non seulement des différences de mode de réponse (jugement ou choix), mais également des différences dans les modes d'expression (prix, ordre, désirabilité, etc...). Les expériences habituellement menées confondent mode de réponse et mode d'expression, puisque les jugements sont exprimés sur une échelle de valeur en termes de prix de réservation, et les choix sur une échelle de valeur en termes de désirabilité. Il n'est donc pas possible de conclure quant à

l'origine de l'effet (mode de réponse ou mode d'expression). En partant de ce constat, W. Goldstein & H. Einhorn [1987] montrent qu'il existe plusieurs types de phénomènes d'inversion, certains s'expliquant par des différences de modes de réponse et d'autres par des différences de modes d'expression.

1.3 Les implications du phénomène d'inversion

Que peut-on conclure de ces expériences et de la justification de leurs résultats? En premier lieu, il faut admettre que ces résultats démontrent l'existence de variables intervenant dans le processus de prise de décision, et habituellement ignorées de l'économiste. Cet oubli porte-t-il à conséquence ? On serait tenté de répondre que non. Le rôle de l'économiste n'est pas de prédire le comportement d'individus isolés. Ceci relève de la psychologie. Son but est de décrire et de prédire le comportement de groupes à partir d'une théorie élémentaire raisonnable du comportement humain. Si l'on ne peut nier l'existence de biais et d'incohérences de comportement au plan individuel, il est fort possible que ces erreurs se compensent à un niveau agrégé, et qu'en moyenne le comportement des agents économiques soit cohérent. Après tout, on constate bien des régularités de comportement à l'échelle des marchés, et on les explique assez bien à partir de l'hypothèse que les agents maximisent leur satisfaction. Mais cela ne suffit pas pour démontrer que l'hypothèse est juste.

L'argument de la compensation des erreurs au niveau agrégé a d'ailleurs été mis en cause également. Plusieurs expériences (J. Hershey, H. Kunreuther & P. Schoemaker [1982], J. Hershey & P. Schoemaker [1985], P. Delquié & R. de Neufville [1988]), ont montré que beaucoup d'"erreurs" de choix étaient systématiques. Par conséquent il ne s'agit plus d'erreurs, mais de biais de comportement. Dans ces expériences les sujets avaient toujours la même "tâche" (task) à effectuer (ajuster une variable pour rendre deux options équivalentes), afin de maintenir invariant le processus cognitif. La structure générale de ce type d'expériences est la suivante : dans une première étape on demande aux sujets de déterminer la valeur Y_2^* qui les laisse indifférents entre les options (X_1, Y_1) et $(X_2, ?)$; dans une étape ultérieure on leur demande quel est la valeur X_1^* qui les laisse indifférents entre les options $(?, Y_1)$ et (X_2, Y_2^*) . La différence entre X_1 et X_1^* est en général très significative, et de même signe pour tous les sujets. Ce biais semble être dû au changement de point de référence (Y_1 dans l'étape n° 1, et X_2 dans l'étape n° 2). Ces expériences tendent ainsi à démontrer l'inexistence d'"un système de valeur stable, préexistant dans la tête des agents", qu'il serait possible de dévoiler à partir d'un processus de révélation.

2. Le "Framing"

Contrairement aux effets engendrés par le mode de réponse, les "framing effects" sont liés à la manière dont les agents perçoivent les options. D'après la théorie économique standard, les préférences ne doivent pas être influencées par la manière dont les options de choix sont décrites. Par exemple, lorsqu'on doit choisir entre deux installations de sécurité, il importe peu que leurs caractéristiques soient décrites en termes de probabilité de panne ou en termes de fiabilité (probabilité que l'installation ne soit pas en panne). Cependant, plusieurs expériences ont montré que le "framing" des options exerçait une influence significative sur les choix. Ces expériences ont été largement discutées par A. Tversky & D. Kahneman [1986]. L'une des plus intéressantes d'entre elles a été réalisée par B. McNeil & al. [1982].

2.1 L'expérience de B. McNeil & al. [1982]

Dans cette expérience, les sujets disposent d'une information statistique sur les thérapies du cancer du poumon. Les mêmes statistiques furent présentées à deux reprises, la première fois en termes de survie, et la deuxième fois en termes de décès. Dans chacun des cas on demandait aux personnes interrogées d'exprimer leurs préférences entre les thérapies.

Problème n°1 (survie)

Chirurgie : Sur 100 personnes opérées 90 survivent à l'opération, 68 sont en vie à la fin de la première année et 34 sont en vie après 5 ans.

Radiothérapie : Sur 100 personnes ayant suivi le traitement, toutes sont en vie à la fin du traitement, 77 sont en vie à la fin de la première année et 22 survivent après 5 ans.

Problème n°2 (décès)

Chirurgie : Sur 100 personnes opérées 10 meurent au cours de l'opération ou durant la phase postopératoire, 32 meurent au bout de la première année et 66 meurent dans les 5 ans.

Radiothérapie : Sur 100 personnes ayant suivi le traitement, aucune n'est décédée en cours de traitement, 23 meurent au bout de la première année et 78 meurent dans les 5 ans.

Les résultats obtenus auprès de différents groupes de personnes, se répartissent comme suit (N indique le nombre des réponses)

Problème n°1 : (N = 247) 18% sont favorables à la radiothérapie

Problème n°2 : (N = 336) 44% sont favorables à la radiothérapie

En d'autres termes, l'avantage de la radiothérapie par rapport à la chirurgie, devient plus important lorsqu'il est formulé en termes de "réduction de décès" immédiat de 10% à 0% plutôt qu'en termes d'"accroissement de la survie" de 90% à 100%. Il faut bien noter que dans les deux cas les probabilités sont les mêmes (ce sont les probabilités complémentaires), et que seule la manière de les présenter change. Ce résultat était vérifié aussi bien pour des médecins confirmés que pour les étudiants et pour les patients.

Avant de tirer les conclusions de cette expérience, examinons les résultats de A. Tversky & D. Kahneman [1986] qui généralisent ce type d'effet.

2.2 Le "framing" des conséquences

A. Tversky & D. Kahneman ont réalisé plusieurs expériences destinées à mettre en évidence l'effet de "framing". L'une d'entre elles est la suivante. Dans chacun des problèmes suivants, les agents doivent effectuer un choix.

Problème n°1 (N = 126) Supposez que vous êtes de 300 \$ plus riche qu'aujourd'hui. Que choisissez-vous ?

- un gain certain de 100\$ (72%)
- 50% de chances de gagner 200\$ et 50% de chances de ne rien gagner (28%)

Problème n°2 (N = 128) Supposez que vous êtes de 500 \$ plus riche qu'aujourd'hui. Que choisissez-vous ?

- une perte certaine de 100\$ (36%)
- 50% de chances de ne rien perdre et 50% de chances de perdre 200\$ (64%)

Les résultats de ce test révèlent plusieurs choses. Tout d'abord un résultat classique : la majorité des agents manifeste une aversion à prendre des risques en présence de gains (problème 1) et au contraire une propension à prendre des risques en présence de pertes (problèmes 2). Mais le résultat le plus intéressant est que, dans les deux problèmes, les enjeux sont exactement les mêmes : 400\$ avec certitude contre un jeu "50-50" d'avoir 500\$ ou 300\$. Il s'agit par conséquent d'un "framing effect", puisqu'une proportion importante de sujets inverse ses choix en passant du problème n° 1 au problème n° 2. De plus, contrairement à l'exemple des traitements du cancer, le cadre est ici relativement "neutre" en termes de charge émotionnelle puisqu'il s'agit de gains et de pertes monétaires.

2.3 Les implications du "framing"

Le "framing" pose un problème majeur à l'analyse économique, dans la mesure où il révèle l'incohérence des choix lorsqu'on modifie la présentation des options. Dans ces conditions, il semble difficile de faire des prédictions à partir de choix connus, sur les choix que feront les agents dans d'autres contextes. Si la théorie économique perd ainsi de son pouvoir prédictif, cela a de graves conséquences quant aux applications du calcul économique, en particulier pour l'analyse bénéfice-coût et l'analyse bénéfice-risque (K. Arrow, [1983]). En effet, les méthodes d'évaluation économiques se basent sur l'observation des choix réalisés par les agents économiques, pour déduire ce que seraient leurs choix dans d'autres domaines ou circonstances, pour lesquels on ne dispose généralement d'aucune observation. L'hypothèse-clé à la base de ces extrapolations est celle de rationalité des agents, c'est-à-dire de la transitivité de leurs choix et de l'invariance de leurs préférences. Une technique, couramment utilisée en pratique, consiste à estimer les coûts ou les bénéfices d'un projet, à partir de comportements observés dans d'autres circonstances. Par exemple, dans le cas du risque de destruction de la couche d'ozone stratosphérique, la valeur de la préservation peut s'exprimer en termes de gains en espérance de vie. Il existe en effet une relation directe entre espérance de vie et exposition au rayonnement ultra-violet. Connaissant le montant qu'un agent est disposé à payer en moyenne pour un accroissement de son espérance de vie (en calculant par exemple la différence de salaire entre les professions à "haut risque" et les professions à "faible risque", toutes choses égales par ailleurs), on peut en déduire la valeur de la protection de l'ozone stratosphérique, ou du moins une estimation du consentement à payer par la collectivité. Outre les approximations pratiques que l'on est obligé de faire pour obtenir ce genre d'estimations (cf. N. Ashford & R. Stone, [1988]), l'enseignement tiré des expériences menées dans le domaine de la prise de décision, remet complètement en question la validité même de ce type de méthodes.

En effet, ce genre de raisonnement perd de son sens en présence de "framing". J. Knetsch & J. Sinden [1984] par exemple, ont montré que certaines méthodes, considérées comme équivalentes du point de vue de l'analyse bénéfice-coût, pouvaient conduire à des estimations très différentes de la volonté de payer des agents. Sur le plan expérimental il existe un biais systématique entre la volonté de payer pour le maintien d'un droit et la compensation monétaire exigée pour la perte de ce droit. On sait qu'en théorie les deux évaluations ne sont pas équivalentes à cause des effets-revenus. La volonté de payer est en effet contrainte par le revenu disponible. En général, la compensation demandée est alors supérieure à la volonté de payer. Cependant, dans le cas de l'expérience de J. Knetsch & J. Sinden, la propriété précédente ne résulte pas d'un effet-revenu, car ce type d'effet a été éliminé dans la conception du protocole expérimental. S'agit-il alors d'un comportement irrationnel de la part des sujets interrogés ? Pas nécessairement. Une perte a davantage d'impact psychologique qu'un gain de même montant (D. Kahneman & A. Tversky [1979]). Les gens sont moins enclins à dépenser des revenus existants que des opportunités de revenus, c'est-à-dire des

revenus dont ils ne disposent pas, mais qu'ils ont la possibilité d'obtenir (3).

Les résultats de J. Knetsch & J. Sinden présentent également l'intérêt de donner un éclairage nouveau au paradoxe de T. Scitovsky [1941]. D'après ce paradoxe classique de l'économie publique, le passage d'une allocation A vers une allocation B et un mouvement en sens inverse, peuvent être simultanément optimaux au sens de Pareto. Dans les deux cas en effet, les bénéfices des gagnants permettent de compenser les pertes des perdants, comme le préconise le critère de Hicks-Kaldor (4). D'après J. Knetsch & J. Sinden ce type de paradoxe se traduit par l'absence de réversibilité des mouvements de long des courbes d'indifférence. Ceci est illustré par la figure 2 : échanger X contre Y n'est pas équivalent à échanger Y contre X. Il existerait en réalité deux cartes d'indifférence, l'une décrivant les substitutions de Y vers X, et l'autre les substitutions de X vers Y.

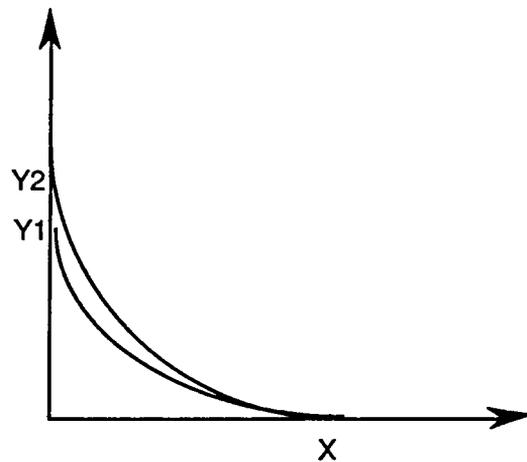


Figure 2 : Non-réversibilité des courbes d'indifférence.

Cette hypothèse semble confirmée par l'expérience réalisée par J. Knetsch [1989]. L'expérience était destinée à confirmer les résultats obtenus antérieurement par J. Knetsch & J. Sinden [1984], et à vérifier l'hypothèse de non-réversibilité des courbes d'indifférence, sur la base d'échanges de troc. Cette expérience faisait intervenir trois groupes distincts. Dans le groupe 1 les participants recevaient un bol qu'ils avaient la possibilité d'échanger immédiatement et sans efforts contre une tablette de chocolat. Dans le groupe 2, les individus étaient placés exactement dans la situation inverse, et dans le groupe 3, on demandait aux participants de choisir entre la tablette de

chocolat ou le bol. Le tableau ci-dessous résume les choix effectués dans les différents groupes.

GROUPE	Bol préféré à chocolat	Chocolat préféré à bol	
1. Dotation en bol	89 %	11%	76
2. Dotation en Chocolat	10 %	90 %	87
3. Pas de dotations	56 %	44 %	55

Ces résultats montrent clairement que le sens de l'échange n'est pas indifférent (5). Il existe une asymétrie dans les choix qui résulte ici de l'aversion à se séparer de ses dotations initiales. C'est pourquoi on a appelé ce résultat "effet de dotation" ("endowment effect"). Ce résultat contredit évidemment la proposition théorique selon laquelle en un point d'une courbe d'indifférence, le taux marginal de substitution est le même dans l'une ou l'autre direction.

Les résultats précédents ont également le mérite de faire clairement ressortir que le framing n'est pas spécifique aux options comportant un risque, de type "loteries", mais persiste en situation de certitude. L'explication de ce type de phénomène n'est donc pas liée à la "capacité" de traitement des sujets de dimensions abstraites, tels que la probabilité, mais plutôt à la "manière" dont ils traitent l'information, comme nous allons le voir dans le paragraphe suivant.

3. Biais liés au traitement de l'information

Un certain nombre d'expériences ont en effet montré que les agents exploitent de manière très insatisfaisante l'information dont ils disposent. Loin d'utiliser l'information rationnellement, les agents semblent être souvent victimes d'illusions et d'erreurs de perception.

3.1 L'expérience de J. Ronen [1971]

Dans cette expérience, on propose aux participants plusieurs options ayant toutes le même résultat (en général un gain monétaire), mais caractérisées chacune par une probabilité de succès distincte. La probabilité de succès est en réalité une probabilité jointe de succès, c'est-à-dire qu'elle dépend du succès à chacune des étapes d'une séquence. Par exemple le succès d'une innovation technologique peut être considéré ex ante comme le cumul de plusieurs succès, celui de la recherche, celui du développement et celui de la commercialisation.

Lorsqu'on présente les options en termes de probabilité jointe de succès, les sujets choisissent à l'unanimité celle qui possède la probabilité la plus élevée (ils maximisent les chances de succès). Au contraire, lorsqu'on leur présente une information "plus fine", ils ont tendance à choisir les options dont la probabilité de succès est la plus forte à la première étape, même si la probabilité jointe de succès, est plus faible. On constate par conséquent que dans ce cas, les agents violent même l'axiome de dominance d'après lequel ils devraient toujours choisir l'option ayant la plus grande chance de succès. Une forte probabilité de succès au premier tour n'est pas une condition suffisante pour gagner la partie. Le fait que les sujets surestiment celle-ci serait une indication d'une certaine "myopie" dans les problèmes de choix dynamiques.

Selon J. Ronen ce comportement s'explique par la présence d'un effet de séquence. Les agents "actualisent" les sensations futures attendues. Ils préfèrent par conséquent une séquence temporelle du type - "plaisir-déplaisir" (ou "confort-inconfort moral", "facilité-difficulté") - à la séquence "déplaisir-plaisir" (ou "inconfort-confort moral", "difficulté-facilité"). Toutefois, cette explication bien que séduisante n'en demeure pas moins insuffisante. A. Tversky & D. Kahneman [1986] obtiennent également des violations de dominance en faisant simplement varier l'information disponible, sans que la variable temps n'intervienne dans l'expérience.

3.2 L'expérience de A. Tversky & D. Kahneman [1986]

Dans ce problème, la question était formulée de la façon suivante :

Supposez que vous soyez confrontés simultanément au couple de décisions ci-dessous. Examinez tout d'abord chacune des décisions, puis choisissez l'option préférée (N = 150).

Décision 1 : choisissez entre

A. un gain certain de 240 \$ (84%)

B. 25% de chances de gagner 1000\$ et 75% de chances de ne rien gagner (16%)

Décision 2 : choisissez entre

C. une perte certaine de 750 \$ (13%)

D. 75% de chances de perdre 1000\$ et 25% de chances de ne rien perdre (87%)

A priori ces résultats n'ont rien d'étonnant. Les agents ont une aversion au risque pour la première décision qui implique des gains, et une propension à prendre des risques pour la décision 2 qui entraîne des pertes. Ce résultat est classique. Ce qui l'est moins, c'est la violation de la dominance dans ce test. En effet, il était bien précisé aux participants que les deux décisions étaient simultanées. Par conséquent il faudrait, en toute logique, tenir compte de la combinaison des gains et de leurs probabilités pour les deux décisions. Si on réalise cette combinaison pour les choix les plus fréquemment observés, on obtient les résultats suivants :

A+D : 25% de chances de gagner 240\$ et 75% de chances de perdre 760\$

B+C : 25% de chances de gagner 250\$ et 75% de chances de perdre 750\$

A l'évidence, B+C domine A+D. Pour A. Tversky & D. Kahneman [1986], les résultats de cette expérience apportent une preuve supplémentaire à l'hypothèse de l'absence d'invariance des préférences. En effet, lorsque les agents sont confrontés aux options A+D et B+C, ils choisissent sans hésiter B+C. Au contraire, lorsqu'on leur fournit plus d'information sur la structure de ces options il choisissent le couple A&D. Comme dans le cas du "framing", en faisant simplement varier la quantité d'information relativement aux options, les préférences s'inversent. Ainsi, contrairement à la perception visuelle, ou à la compréhension du langage, en matière de risques, l'homme ne traduit pas spontanément les options de choix en termes canoniques, afin de les rendre aisément comparables.

D'autres sources de biais se rapportant à la manière dont les agents traitent l'information existent (cf. R. Hogarth, [1980]), en particulier le "biais de confirmation" (cf. P. Wason, [1966]), selon lequel la plupart des gens cherchent à confirmer leurs opinions et non à les infirmer, comme le voudrait la rigueur scientifique (critère d'infirmité de Popper). Toutefois, les comportements en cause dans ces expériences dépassent largement le cadre de l'hypothèse d'invariance auquel nous nous sommes volontairement limités. Nous nous contenterons donc d'évaluer la portée des résultats évoqués dans les développements précédents.

4. Le problème de la représentativité de l'expérience

Les expériences que nous avons discutées ont fait l'objet de nombreuses critiques. Ces critiques peuvent se résumer à trois points. 1) les contradictions observées ne s'appliquent qu'à des choix insignifiants. Il s'agit de "curiosités de laboratoire" ne possédant pas d'équivalent dans le monde réel de la prise de décision; 2) Les expériences ne permettent aucun apprentissage, alors qu'en réalité les agents améliorent progressivement leurs décisions, car il ont la possibilité d'apprendre par la pratique et de corriger leurs biais; 3) les mécanismes du marché ont des vertus correctrices à travers les "sanctions" pour les mauvaises décisions et les "récompenses" pour les bonnes décisions. En d'autres termes, les mécanismes d'incitation et l'expérience améliorent la qualité quant aux comportements réels.

Il est toujours possible de rétorquer à ces arguments, qu'il existe, malgré tout, des limites aux capacités cognitives des agents. Mais une telle proposition ne remet pas en question l'hypothèse que le comportement d'optimisation est une bonne approximation des comportements économiques réels. Il existe néanmoins des raisons de douter des trois arguments précédents.

Prétendre qu'il s'agit de tests de laboratoire, sans valeur descriptive, pose de façon générale le problème de la représentativité du comportement expérimental, et peut-être de l'expérimentation scientifique toute entière. On ne peut donc considérer qu'il s'agit d'une critique spécifique à l'expérimentation en sciences sociales. Admettons cependant que les expériences ne sont pas représentatives. Il faut dès lors accepter le principe qu'aucun test n'est représentatif, pas même ceux qui confirment la théorie. Or, comme l'a fait très justement remarquer R. Thaler [1986], les économistes ont souvent cité des études expérimentales qui tendaient à confirmer leurs théories. C'est le cas par exemple de l'étude de J. Kagel et alii [1975], qui a montré, à partir d'expériences sur des rats et des pigeons, que les courbes de demande étaient décroissantes et que les courbes d'offre de travail étaient croissantes ! C'est également le cas pour la théorie des jeux appliquée à l'économie industrielle (C. Plott, [1982]) ou à la théorie de la négociation et des enchères [A. Roth, 1988]. Il faut donc, ou bien accepter l'expérimentation comme méthode d'investigation à part entière, ou bien la rejeter.

L'argument de l'apprentissage restreint a priori la portée des résultats expérimentaux aux situations non-répétitives impliquant des décisions non routinières. Toutefois, là encore on peut faire plusieurs observations. Premièrement, les situations répétitives supposent des environnements décisionnels stables. La question de l'intérêt théorique et expérimental de cette classe de décisions revient donc à se poser la question de l'intérêt d'étudier les environnements stables. Dans de tels environnements peut-on encore raisonner en termes de décision, pour un acte qui au terme d'un processus d'"essais-erreurs" est une simple répétition automatique ? Deuxièmement, à l'exception

du cas idéalisé d'un environnement immuable, la temporalité de l'apprentissage est toujours limitée par celle de l'environnement décisionnel. Les expériences que nous avons mentionnées appartiennent également à cette catégorie, avec la particularité d'avoir une durée très limitée. Troisièmement, même lorsque les temps d'apprentissage sont explicitement pris en considération, les biais ne disparaissent pas pour autant. Le problème est lié à plusieurs facteurs : les agents ont tendance à rechercher l'information qui confirme leurs opinions (biais de confirmation); le feedback qu'ils reçoivent en réponse à leurs décisions, n'est pas nécessairement représentatif de l'environnement : "le serveur pense que les jeunes gens laissent peu de pourboire, en conséquence il leur consacrerait moins d'attention qu'aux autres consommateurs; donc il recevra moins de pourboire de la part des jeunes; ceci peut même modifier la structure par âge de l'établissement dans lequel il travaille". Cet exemple de "prophétie autoréalisatrice", emprunté à R. Hogarth [1980], montre que l'apprentissage est loin d'être optimal, même s'il peut paraître efficace en apparence (le serveur maximise finalement son profit parce que l'environnement a changé). Mais la répétition des succès, du fait de l'apprentissage, peut conduire à un "excès de confiance" (overconfidence) qui fausse l'évaluation et le jugement , et par suite aboutir à des choix "sous-optimaux" (H. Einhorn et R. Hogarth, [1978]).

L'argument de l'insignifiance (voire l'inexistence) des stimuli a fait l'objet d'une analyse dans certaines expériences. Par exemple, dans l'expérience du phénomène d'inversion, les enjeux ont été augmentés de manière substantielle afin de motiver davantage les participants, c'est-à-dire de stimuler leur attention et leur réflexion. L'idée sous-jacente est qu'il existe une relation "stimulus-réponse" : plus les enjeux sont élevés et "mieux" les agents vont optimiser. Certes, en augmentant le montant des enjeux, la fréquence de certains biais tend à se réduire, mais statistiquement les biais de comportement observés restent significatifs. Rendre l'expérience plus attractive ne suffit pas à les éliminer. Cette conclusion est bien illustrée par les controverses entre économistes et psychologues sur le phénomène de "inversion des préférences". Beaucoup d'efforts ont été consacrés à imaginer des mécanismes d'incitation, sous forme de paiements (W. Pommerehne, F. Schneider & P. Zweifel [1982], R. Reilly [1982]) et/ou de pénalités (J. Berg, J. Dickhaut & J. O'Brien, [1985]), avec l'espoir d'éliminer significativement, voire complètement, le phénomène. Le succès de ces tentatives a été très limité, de sorte que l'effet d'inversion est aujourd'hui de plus en plus considéré comme une "illusion cognitive" (A. Roth, [1988]).

Reste enfin la question de la pertinence de ces biais au niveau macroscopique. Ces "irrationalités" peuvent-elles engendrer des effets significatifs à l'échelle du système économique, ou sont-elles négligeables à ce niveau d'agrégation ? Nous avons déjà abordé ce point à propos de la compensation au niveau agrégé des erreurs individuelles. Puisqu'il s'agit de biais et non pas d'erreurs, la compensation a peu de chances de se produire ! Mais existe-t-il une influence sensible sur le système économique ? A priori cette question est difficile à trancher d'un point de vue

empirique. Il existe cependant des réponses sur le plan théorique. G. Akerlof et J. Yellen [1985] ont montré que dans le cas où une fraction de la population "maximise à peu près", même lorsque les conséquences de ce comportement sont négligeables au niveau individuel car insignifiantes (6), elles peuvent engendrer des conséquences de premier ordre au niveau agrégé, c'est-à-dire une sous-optimalité au sens de Pareto du fait de pertes de bien-être ou de "pertes sèches".

5. Vers une alternative à l'hypothèse d'optimisation ?

Les sections précédentes ont montré que, lorsqu'on désire évaluer de façon empirique ou expérimentale la valeur que les agents attribuent à différentes situations ou objets, on se heurte à de sérieuses difficultés. Il existe en effet plusieurs sources de biais dans les jugements, résultant de l'absence d'invariance des préférences. Ces biais, lorsqu'ils se cumulent, peuvent fausser complètement l'évaluation des options. Prenons le cas d'un individu auquel on demande d'évaluer un risque technologique, par exemple le risque d'accident nucléaire. Le processus d'évaluation subjectif de ce risque est le suivant. Dans un premier stade (stade de l'acquisition d'information), l'individu interrogé fait d'abord appel à sa mémoire ou aux informations qui lui sont présentées (par exemple des statistiques sur les accidents, des données sur la sécurité des installations, des avis d'experts, etc...). Dans un second stade (stade du traitement), il traite ces données en les comparant à des risques qui lui sont familiers (risque d'accident de voiture, risque d'incendie dans son voisinage, risque de décès par cancer, etc...). Enfin, dans un troisième temps (stade de l'expression), il exprime son jugement. A chacune de ces étapes, des éléments peuvent biaiser le jugement qui sera porté. Au premier stade, l'individu sera très influencé par la manière dont les informations lui sont présentées (framing); de même, il aura généralement tendance à attacher plus d'importance aux informations qui confirment son opinion initiale (biais de confirmation). Au stade du traitement de ces informations, le biais peut provenir par exemple d'une attitude conservatrice par rapport aux données nouvelles qui sont présentées, ou encore du type d'heuristique employée afin de réduire l'effort mental (par exemple, se fonder sur l'estimation d'un risque connu et ajuster cette estimation à la situation proposée). Enfin, au stade de l'expression du jugement, le biais peut provenir du mode de réponse (jugement, désirabilité, ...) ou du mode d'expression (en termes de coûts, de probabilité, etc...).

Le schéma ci-dessous, proposé par R. Hogarth [1980], est un schéma de référence plausible du processus de jugement qui permet d'identifier les différentes sources possibles de biais dans le comportement.

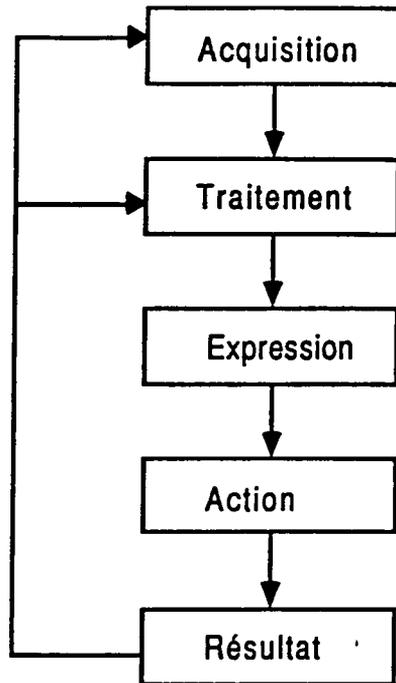


Figure 3 : Modèle conceptuel du processus de jugement

En plus des trois stades que nous avons mentionnés ci-dessus, R. Hogarth intègre dans son schéma la possibilité d'effets de rétroaction (feed-back) des jugements sur la perception ou sur le traitement, engendrés par les résultats de l'action adoptée à la suite d'un jugement. Ce "feed-back" est à la base des phénomènes d'apprentissage. Un exemple de biais fréquent relatif au feed-back, est la confusion entre "cause" et "aléa". Les individus attribuent les résultats qu'ils observent, non pas à des phénomènes du hasard, mais à leurs actes. Ils peuvent ainsi être victime d'une "illusion de contrôle" sur leur environnement.

Ce schéma descriptif du comportement est malheureusement trop général pour servir de fondement à une théorie économique du comportement. Il a le mérite d'attirer l'attention sur un certain nombre de problèmes qui peuvent se poser dans l'application du calcul économique. Ainsi, lorsqu'on cherche à mesurer la valeur d'un bien pour une population à partir d'un questionnaire, il est nécessaire d'identifier toutes les sources potentielles de biais, et d'analyser le sens dans lequel ces biais peuvent jouer, afin de pouvoir les corriger le cas échéant. Par exemple, appréhender l'attitude face à un risque d'irréversibilité en termes de volonté de payer (valeur d'option) peut entraîner des conclusions contradictoires avec les choix effectifs des agents. Les agents peuvent être disposés à payer une forte somme pour préserver une espèce, et faire néanmoins des choix qui favorisent la disparition de cette même espèce. Par exemple, condamner le massacre des éléphants, mais acheter des bijoux en ivoire, ou être opposé à la chasse aux félins, mais porter des vêtements de fourrure.

Reste que le principal problème posé par les résultats expérimentaux et le modèle de R. Hogarth, est celui de l'invariance des préférences. Dans la mesure où les réponses varient selon le mode d'expression et le type de "framing", on est en droit de se poser la question de l'existence d'un système de préférences invariant chez les agents économiques. Sans aller jusqu'à nier l'existence des préférences, il faut reconnaître que les résultats expérimentaux discutés ici peuvent donner lieu à au moins deux interprétations.

La première serait que les agents n'ont pas un comportement d'optimisation au sens habituel du terme, mais appliquent des règles empiriques simples qui leurs permettent d'aboutir rapidement et/ou facilement à une réponse ou à une décision. Ils utilisent alors des heuristiques du type "anchoring-and-adjustment" (cf. par exemple R. Dawes [1988]), ou des "raisonnements approximatifs". Cette interprétation est en accord avec l'idée de "rationalité limitée" proposée par Simon [1956], et de ce fait elle ne contredit pas a priori l'hypothèse d'un comportement d'optimisation prenant en compte des contraintes supplémentaires (capacités cognitives limitées, coûts de l'effort mental, coûts relatifs à l'exploration du champs des possibles, etc...).

La seconde interprétation consiste à dire que les préférences dépendent très fortement de l'environnement de la décision ou du *contexte* de l'évaluation : les jugements et les choix sont influencés par des facteurs externes non pertinents à la prise de décision. Nous sommes en présence dans ce cas d'un phénomène comparable aux effets externes. De même que les effets externes modifient les contraintes de la décision (ensembles de budget, ensembles de consommation), et par conséquent le niveau de profit ou d'utilité, les effets discutés ici exercent une influence sur l'ordre de préférence des agents et de ce fait sur l'utilité. A la différence des effets externes toutefois, la dépendance entre le choix et l'environnement de ce choix n'est pas imposée de l'extérieur (comme dans le cas "pollueur-pollué" par exemple), mais est "créée" par l'agent lui-même. Il s'agit ici plutôt

d'une "externalité subjective", personnelle à l'agent, qui résulterait de l'interaction entre l'évaluation et le choix. Cette idée est conforme à la théorie de Rosenfield, selon laquelle la mémoire n'est pas localisée et invariante, mais est "créée" ou "inventée" par l'agent. La théorie de Rosenfield s'oppose à la théorie de la localisation fonctionnelle du cerveau selon laquelle le cerveau humain serait "une mosaïque de zones (de centres) où seraient imprimées et déposées en permanence des images et symboles formant ... la mémoire" (cf. le texte de B. Ancori). Pour Rosenfield la remémoration dépend très largement du contexte, ce qui est en contradiction avec l'hypothèse d'invariance.

Quelle que soit l'interprétation retenue, il apparaît que les questions relatives à la formation des préférences et aux processus de prise de décision, mériteraient une attention plus grande. Un parallèle avec la théorie de la firme est tentant. Pendant longtemps la théorie économique a considéré le processus de production comme une "boîte noire", l'étude du processus lui-même étant réservée à l'ingénieur, alors que l'économiste devait se borner à l'étude des flux d'inputs et d'outputs. Depuis quelques années cette conception du rôle de l'économiste a profondément évolué, et il est aujourd'hui admis que la compréhension des phénomènes de progrès technique et d'innovation passe impérativement par l'étude des transformations des processus de production. Il est dès lors indispensable de dépasser le niveau de détail de la microéconomie traditionnelle afin de décrire ce qui se passe à l'intérieur de la boîte noire. Pour des raisons analogues, il nous semble nécessaire de faire un pas dans la même direction pour la théorie du consommateur. Les préférences ne sont pas données mais sont modelées par différents paramètres. Elles ne préexistent pas nécessairement à l'évaluation mais peuvent se former conjointement. Décider n'est pas synonyme de préférer. Pour comprendre ces propositions et en tenir compte, il est indispensable que l'économiste étudie ce qui se passe dans la "boîte noire" du décideur, afin de comprendre la "technologie de la décision", un domaine traditionnellement réservé aux psychologues.

NOTES

(1) Selon A. Roth [1988], la première expérience a été menée par Bernoulli en 1738 à propos du célèbre jeu de Saint-Petersbourg. Si Bernoulli manifestait ainsi une préoccupation expérimentale indéniable, c'est en fait à E. Chamberlin [1948] que revient le mérite d'avoir réalisé la première expérience véritable, à propos de l'équilibre concurrentiel.

(2) Le "paradoxe d'Allais" apparaît comme une contradiction fondatrice de ce point de vue. En effet, il a joué un rôle important pour la multiplication des recherches expérimentales.

(3) Il est intéressant de remarquer qu'une interprétation similaire peut être faite pour l'un des paradoxes d'Allais, l'effet de "conséquence commune" (cf. M. Machina, [1987]).

(4) Cette compensation peut être réalisée en prélevant par exemple une taxe sur les bénéficiaires du changement. Le critère de Hicks-Kaldor est généralement invoqué dans les projets publics dont l'évaluation est faite selon la méthode coûts-avantages.

(5) L'expérience de J. Knetsch a été conçue de façon à éviter l'existence d'effets revenu, d'incertitude et de coûts de transactions, de sorte qu'il n'y a pas d'ambiguïtés quant à l'interprétation des résultats.

(6) L'argumentation de G. Akerlof et de J. Yellen est fondée sur le théorème d'enveloppe. Soit $f(x,a)$ la fonction objectif à maximiser, où x est la variable de contrôle et a un vecteur de paramètres. Soit $x(a)$ la solution (supposée unique) du programme : $\max f(x, a)$ slc $x \in X$. Soit $M(a) = f(x(a), a)$. D'après le théorème d'enveloppe on a :

$$dM(a)/da = \partial f(x(a), a)/\partial a .$$

"A la marge" la variation de la fonction objectif engendrée par une variation des paramètres, est la même que les agents ajustent ou non la variable de contrôle (ce résultat reste valable dans le cas de l'optimisation sous contraintes). La différentielle totale de $M(a)$ s'écrit :

$$dM(a)/da = [\partial f(x, a)/\partial a].dx(a)/da + \partial f(x(a), a)/\partial a.$$

Le premier membre à droite de l'égalité est nul car $\partial f(x, a)/\partial a = 0$, en vertu de la condition de premier ordre de l'optimum. De cette propriété résulte la proposition d'Akerlof et de Yellen, à savoir que les "erreurs" de premier ordre commises par un agent n'ont généralement que des effets de deuxième ordre en termes de variation d'utilité ou de profit.

BIBLIOGRAPHIE

- AKERLOF G., YELLEN J. [1985], "Can small deviations from rationality make significant differences to economic equilibria?", *American Economic Review*, vol. 75, 1985, pp. 708-720.
- ALLAIS M. [1953], "Le comportement de l'homme rationnel devant le risque : critique des postulats de l'école américaine", *econometrica*, vol.21, n°4, 1953, pp. 503-546.
- ASHFORD N. A., STONE R. F. [1988], "*Cost-benefit analysis in environmental decision-making*", MIT, Cambridge, 1988.
- ARROW K.J. [1982], "Risk perception in psychology and economics", *Economic Inquiry*, 20, 1982, pp. 1-9.
- ARROW K.J. [1983], "Behavior under uncertainty and its implications for policy", in "*Foundations of Utility and Risk Theory with Applications*", B. STIGUM & F. WENSTOP (eds), D. Reidel (Publ.), 1983, pp. 19-32.
- BELL D. [1982], "Regret in decision-making under uncertainty", *Operations Research*, 30, 1982, pp. 961-981
- BERG J. E., DICKHAUT J.W., O'BRIEN J. R. [1985], "Preference reversals and arbitrage", in V.L. Smith (ed), *Research in Experimental Economics*, vol. 3, Greenwich : JAI Press, 1985.
- COX J. C., EPSTEIN S. [1989], "Preference reversals without the independence axiom", *american economic review*, 79, 1989, pp. 408-426.
- DAWES R. [1988], "*Rational choice in an uncertain world*", Harcourt Brace Jovanovich Publ., 1988.
- DELQUIE P., DE NEUFVILLE R. [1988], "*Response-modes and inconsistencies in preference assessments*", TPP, MIT, Cambridge, september 1988.
- EINHORN H., HOGARTH R. [1978], "Confidence in judgement : persistence in the illusion of validity", *Psychological review*, 85, 1978, pp. 395-516.
- GOLDSTEIN W., EINHORN H. [1987], "Expression theory and the preference reversal phenomena", *psychological review*, 94, 1987, pp. 236-254.
- GRETHER D., PLOTT C. [1979], "Economic theory of choice and the preference reversal phenomenon", *american economic review*, 69, 1979, pp. 623-638.
- HAMMOND P. [1988], "Consequentialism and the independence axiom", in "*Risk, decision, and rationality*", B. MUNIER (ed), D. Reidel (Publ.), 1988, pp. 503-516.
- HERSHEY J., KUNREUTHER H., P. SCHOEMAKER P. [1982], "Sources of bias in assessment procedures for utility functions", *Management science*, 28, 1982, pp. 936-954.
- HERSHEY J., SCHOEMAKER P. [1985], "Probability versus certainty equivalence methods in utility measurement : are they equivalent ?", *Management Science*, 31, 1985, pp. 1213-1231.

HOLT C. A. [1986], "Preference reversals and the independence axiom", *American Economic Review*, 76, 1986, pp. 508-514.

HOGARTH R. [1980], "*Judgement and choice*", second edition, Wiley.

KAGEL J., RACHLIN H., GREEN L., BATTALIO R., BASMANN R., KLEMM W. [1975], "Experimental studies of consumer demand using laboratory animals", *Economic Inquiry*, 13, 1975, pp. 22-38.

KAHNEMAN D., TVERSKY A. [1979], "Prospect theory : an analysis of decision under risk", *Econometrica*, 47, 1979, pp. 263-291.

E. KARNI E., SAFRA Z. [1987], "Preference reversals and the observability of preferences by experimental methods", *Econometrica*, 55, 1987, pp. 675-685.

KNETSCH J. L. , SINDEN J. A. [1984], "Willingness to pay and compensation demanded : experimental evidence of an unexpected disparity in measures of value", *Quarterly Journal of Economics*, 99, 1984, pp. 507-21.

KNETSCH J. L. [1989], "The endowment effect and evidence of nonreversible indifference curves", *American Economic Review*, 79, 5, 1989, pp. 1277-1284.

LICHTENSTEIN S., SLOVIC P. [1971], "Reversals of preferences between bids and choices in gambling decisions", *Journal of Experimental Psychology*, 89, 1971, pp. 46-55.

LICHTENSTEIN S., SLOVIC P. [1973], "Response-induced reversals of preferences in gambling : an extended replication in Las Vegas", *Journal of Experimental Psychology*, 101, 1973, pp. 16-20.

LINDMAN H. R. [1971], "Inconsistent preferences among gambles", *Journal of Experimental Psychology*, 89, 1971, pp. 390-397.

LOOMES G., SUGDEN R. [1982], "Regret theory : an alternative of rationale choice under uncertainty", *Economic Journal*, 1982, pp. 805-824.

LOOMES G., SUGDEN R. [1983], "A rationale for preference reversal", *American Economic Review*, 1983, pp. 428-432.

MACHINA M. [1987], "Choice under uncertainty : problems solved and unsolved", *Economic Perspectives*, 1, 1, 1987, pp. 575-594.

McNEIL B., PAUKER S., SOX H., TVERSKY A. [1982], "On elicitation of preferences for alternative therapies", *New England Journal of Medicine*, 306, 1982, pp. 1259-1262.

POMMEREHNE W.W., SCHNEIDER F. & ZWEIFEL P. [1982], "Economic theory of choice and the preference reversal phenomenon : a reexamination", *American Economic Review*, 72, 1982, pp. 569-74.

PLOTT C. [1982], "Industrial organization theory and experimental economics", *Journal of Economic Literature*, 20, 1988, pp. 1485-1527.

RACHLIN H., BATTALIO R., KAGEL J., GREEN L., [1981], "Maximization theory in behavioral psychology", *The Behavioral and Brain Sciences*, 4, 1981, pp. 371-417.

REILLY R.J. [1982], "Preference reversal : further evidence and some suggested modifications in experimental design", *american economic review*, 72, 1982, pp. 576-84.

RONEN J. [1971], "Some effects of sequential aggregation in accounting on decision-making", *Journal of Accounting research*, 9, 1971, pp. 307-332.

ROSENFELD I. [1989], "*L'invention de la mémoire. Le cerveau, nouvelle donne*", Eshel.

ROTH A. [1988], "Laboratory experiments in economics : a methodological overview", *Economic Journal*, 98, 1988, pp. 974-1031.

SCITOVSKY T. [1941], "A note on welfare propositions in economics", *Review of economic studies*, 9, 1941, pp. 77-88.

SEGAL U. [1988], "Does the preference reversal phenomenon necessarily contradict the independence axiom ?", *american economic review*, 78, 1988, pp. 233-36.

SLOVIC P. & LICHTENSTEIN S. [1983], "Preference reversals : a broader perspective", *american economic review*, 73, 1983, pp. 596-605.

THALER R. [1986], "The psychology and economics conference handbook : comments on Simon, on Einhorn and Hogarth, and on Tversky and Kahneman", in *Rational choice : the contrast between economics and psychology*; R. HOGARTH & M. REDER eds., University of Chicago Press, 1986, pp. 95-100.

TVERSKY A., KAHNEMAN D. [1986], "Rational choice and the framing of decisions", in *Rational choice : the contrast between economics and psychology*; R. HOGARTH & M. REDER eds., University of Chicago Press, 1986, pp. 67-94.

VON NEUMANN J., MORGENSTERN O. [1947], "*Theory of games and economic behavior*", Princeton university press, 1947.

WASON P.C. [1960], "On the failure to eliminate hypotheses in a conceptual task", *Quarterly Journal of Experimental Psychology*, 12, pp. 129-140.

WILLINGER M. [1990], "La rénovation des fondements de l'utilité et du risque", *Revue Economique*, vol. 41, n° 1, pp. 5-47.

CHAPITRE 2

MEMOIRE ET APPRENTISSAGE

DE LA NEUROBIOLOGIE A L'AUTO-ORGANISATION

par

Bernard ANCORI

Contrairement au postulat qui sous-tend l'ensemble de l'analyse théorique du choix au niveau microéconomique, le texte de M. Willinger vient de suggérer que le système de préférences individuelles pourrait ne pas être invariant. L'expérience montre les agents réagir de manières différentes aux alternatives proposées, selon le mode de présentation de celles-ci par l'expérimentateur ou celui de l'expression même des préférences des agents. Face à une telle situation, le théoricien n'a que deux possibilités: retoucher partiellement un cadre analytique dont l'essentiel lui paraît demeurer valide en fournissant une explication *ad hoc* à chaque comportement venu infirmer localement une théorie à laquelle il continue d'adhérer globalement, ou tenter de produire un nouveau paradigme en reformulant entièrement les bases mêmes de la théorie - deux attitudes mutuellement exclusives qui sont celles que T.S. Kuhn [1962] nous a appris jadis à considérer respectivement comme celles de la science ordinaire et de la science extraordinaire.

Naturellement, le degré de rationalité propre à chacune de ces attitudes dépend étroitement du matériel épistémologique disponible pour une reconstruction plus ou moins radicale de la théorie. Tant que ce matériel demeure relativement pauvre, la première attitude est la plus rationnelle, puisqu'elle refuse légitimement de rejeter un acquis qu'aucun apport nouveau n'est assez consistant pour remplacer. Par contre, dès lors qu'existent des éléments suffisamment solides pour offrir les matériaux d'une construction alternative, intégrant les exceptions ou anomalies inexplicables par le paradigme précédent, il devient davantage rationnel d'adopter la seconde. Bien que la première attitude continue d'informer majoritairement la littérature, il semble que la théorie du choix individuel se trouve aujourd'hui dans la seconde situation: une série de contributions récentes, initialement ancrées dans le domaine de la biologie mais de portée bien plus générale, convergent pour offrir à l'économiste les ingrédients d'une nouvelle approche de l'*homo oeconomicus*. Le présent texte a pour ambition de présenter deux de ces contributions, qui semblent propres à étayer un modèle de comportement de l'homme confronté au problème du choix dans un contexte bien précis, donc à apporter de premiers fondements à la notion de *rationalité contextuelle*.

Dans le sillage de I. Rosenfield [1989, 1990], la première fournit les bases biologiques d'une conception relativement inédite du fonctionnement de la mémoire humaine. Selon cette conception, qui rejoint en cela certains acquis fondamentaux de la théorie et de la pratique psychanalytiques issues des travaux de S. Freud, toute mise en acte de la mémoire est invention liée au contexte présent autant que remémoration du passé. C'est dire que son premier intérêt pour l'économiste est d'apporter une explication d'ensemble aux comportements apparemment incohérents des agents, recensés dans le texte précédent. Comme l'indiquait la conclusion de M. Willinger, l'influence de facteurs externes non pertinents (eu égard à une situation de choix idéalement exempte de tout contexte particulier) nous met en présence d'un phénomène comparable à ce que nous avons pour habitude de désigner par l'expression d'"effets externes". Ainsi que nous le verrons, le moindre mérite de la théorie rosenfieldienne de la mémoire n'est

pas de nous permettre d'avancer dans la voie de l'endogénéisation de tels effets dans la théorie économique du choix.

Mais le modèle de I. Rosenfield présente un autre intérêt pour notre discipline, en s'inscrivant dans un cadre épistémologique que nous connaissons depuis une bonne décennie pour être celui de l'auto-organisation. Plus précisément, bien que le mot lui-même ne figure pas une seule fois dans les ouvrages de I. Rosenfield, la manière dont celui-ci conçoit le fonctionnement de la mémoire, en particulier de la mémoire humaine, évoque irrésistiblement le concept d'auto-organisation introduit naguère par H. Atlan [1972, 1979]. Or il est désormais patent que ce dernier concept revêt une importance absolument cruciale pour toute théorisation de la dimension évolutionniste de l'économie (cf. B. Ancori et al., [1987]). Et une telle dimension n'est-elle pas d'emblée présente dans toute situation réelle de choix, toujours marquée par l'incertitude dans laquelle se prennent les décisions et par l'irréversibilité plus que fréquente des conséquences de celles-ci? Il apparaît donc logique d'examiner - et c'est la seconde contribution proposée ici - le concept d'auto-organisation proposé par H. Atlan, ainsi que la problématique qui sous-tend ce concept.

La relative ancienneté de la première expression de celui-ci (les années 1971-1972) ne doit pas nous dissuader d'en donner une présentation quelque peu détaillée: même s'il peut apparaître aujourd'hui bien connu, le concept d'auto-organisation n'a probablement pas encore révélé toute sa fécondité, et pour ce qui nous concerne directement, c'est précisément dans le détail de son exposition que nous trouverons d'intéressantes correspondances avec le détail de la conception rosenfieldienne de la mémoire. Du reste, ce n'est pas à la seule fin de mettre en évidence de telles correspondances que nous nous livrons ici à cette confrontation de la pensée de I. Rosenfield avec celle de H. Atlan: nous attendons de cette rencontre provoquée qu'elle produise un certain nombre de propositions minimales, qui sont autant de réquisits auxquels devrait satisfaire tout modèle théorique dont l'ambition serait d'analyser de manière réaliste le choix effectué par l'*homo computans* - ce produit obtenu par le croisement, si l'on peut dire, de l'*homo cogitans* avec l'*homo oeconomicus*. Après avoir présenté la théorie rosenfieldienne de la mémoire (1), nous serons ainsi amené à rappeler le "noyau dur" de la conception atlantienne de l'auto-organisation (2), pour conclure en montrant les convergences de ces deux approches (3).

1. L'invention de la mémoire

C'est sous ce titre à double sens - mémoire inventive autant qu'inventée - que le mathématicien, médecin et philosophe I. Rosenfield [1989] a proposé de nouvelles vues sur le fonctionnement du cerveau humain, et du substrat neurophysiologique de ce fonctionnement., avant d'étendre la perspective ainsi ouverte à une théorie de la conscience et du langage [1990]. Ces vues sont nouvelles en ceci qu'elles contredisent la conception traditionnelle de la mémoire,

CHAPITRE 2

MEMOIRE ET APPRENTISSAGE

DE LA NEUROBIOLOGIE A L'AUTO-ORGANISATION

par

Bernard ANCORI

associée à la théorie de la localisation fonctionnelle du cerveau. Après avoir exposé les grandes lignes de cette dernière théorie (1.1), nous mentionnerons certaines données expérimentales susceptibles de lui apporter un démenti (1.2), puis nous attacherons à réinterpréter ces données dans la perspective du "darwinisme neuronal" de G.M. Edelman, dans laquelle s'inscrit explicitement la contribution propre de I. Rosenfield (1.3).

1.1 Mémoire et théorie de la localisation fonctionnelle

De manière plus ou moins explicite, le cerveau humain est généralement conçu comme une mosaïque de zones (de centres) où seraient imprimées et déposées en permanence des images et symboles formant ce qu'il est convenu d'appeler la mémoire. Ainsi, chacun de nous disposerait constamment d'un stock de souvenirs consistant en traces identifiables, classées et enregistrées, de perceptions passées représentant elles-mêmes autant d'images exactes des événements survenus dans son environnement. C'est à partir de ce stock de souvenirs que s'effectuerait cette fonction cérébrale distincte de la perception - la reconnaissance: à tout instant, notre cerveau découperait le monde en données sensibles qu'il comparerait, pour les reconnaître, aux images acquises précédemment et enregistrées. Bien qu'elle ne réponde pas à la question de savoir comment sont gravées les *premières* images auxquelles comparer les nouvelles données sensibles - alors même qu'elle implique qu'il n'y ait pas de perception sans apprentissage préalable - cette théorie bénéficie de deux types de soutiens, l'un relativement ancien et l'autre très moderne.

Le soutien ancien est celui de P. Broca, qui montra en 1861 que la perte de certains usages de la parole, rendue manifeste par des troubles d'élocution, correspondait à une petite lésion d'une certaine zone du cerveau gauche - appelée depuis l'aire de Broca. Plus précisément, à la suite de ces travaux la perte de l'image auditive des mots apparut tout naturellement due à l'altération d'un centre où serait localisée la mémoire des mouvements nécessaires au langage articulé. Par extension,

"on découvrit par la suite d'autres centres du langage, possédant une action spécifique sur la parole, et des aires cérébrales contrôlant la motricité de certaines parties du corps: les mains, les doigts, la langue, etc. (...) Au terme du XIXème siècle, de nombreux neurologues en avaient déduit que le cerveau était constitué d'un ensemble de régions hautement spécialisées exerçant un contrôle sur la parole, l'activité motrice ou la vision. A la localisation et à la spécialisation fonctionnelle vint s'ajouter un découpage de la mémoire en de nombreuses sous-unités spécialisées. Il y eut des centres de la mémoire se rapportant aux 'images visuelles des mots', aux 'images auditives des mots'; la perte de la mémoire peut alors s'expliquer par la disparition d'une image (ou d'un centre) mnésique spécifique, ou bien par l'incapacité du cerveau à 'consulter ses fichiers', due à une rupture des voies de conduction nerveuses, (...) Sur le plan théorique, la plus importante conséquence de ces recherches fut la doctrine de la localisation fonctionnelle," [I. Rosenfield, 1989, p.20].

La théorie de la localisation fonctionnelle considère donc comme établie l'existence de toute une série de relations entre certains lieux précis du cerveau et certaines grandes fonctions remplies par l'organisme. Une telle conception fut d'emblée à l'origine de brillantes études cliniques prédisant la localisation exacte de lésions cérébrales d'après les symptômes de dysfonctionnement présentés par les patients - et voyant leurs prédictions confirmées à l'autopsie.

Ainsi, s'il est vrai que

" sans la croyance en des souvenirs constamment présents dans la mémoire, la théorie de la localisation fonctionnelle n'aurait pas été concevable" [I. Rosenfield, op. cit., p.21],

il est tout aussi vrai que l'effet de ces études cliniques fut de renforcer et d'étendre une telle croyance. Remarquons immédiatement que ce concept de "mémoire" - sous-tendant une théorie de la localisation fonctionnelle qui vient à son tour le renforcer - connaît une extension très large: ce concept recouvre aussi bien le type de mémoire nécessaire pour (re)produire une suite de sons (articulés en une chaîne parlée dotée de signification) à partir de l'image auditive des mots, que la mémoire dont chacun a besoin pour retrouver son chemin dans une ville déjà visitée par lui. Bien que tel puisse être le cas, il ne s'agit donc pas forcément ici d'une mémoire mise en acte de façon délibérée et consciente. Il reste que, tous types confondus, cette mémoire est idéalement considérée par la théorie de la localisation fonctionnelle comme étant parfaitement fidèle - une fidélité *point par point*, pourrait-on dire: une telle mémoire garde en stock l'image *exacte* d'un stimulus-événement perçu, et conserve indéfiniment cette image sans altération aucune. Soigneusement classée et enregistrée, celle-ci est donc prête à tout moment à être confrontée - par une fonction de reconnaissance présumée distincte de la fonction de perception - à l'image d'un stimulus-événement éventuellement candidat à être reconnu comme identique au précédent. Un stock de souvenirs spécifiques, présents en permanence dans notre cerveau de manière indépendante du contexte actuel de celui-ci, et immédiatement mobilisables au service de la fonction de reconnaissance: telle est la conception de la mémoire, inhérente à la théorie de la localisation fonctionnelle.

C'est manifestement une conception de ce genre qui fonde implicitement le postulat d'existence d'un système invariant de préférences chez l'agent dont le comportement est analysé par la théorie microéconomique: l'ordre dans lequel sont classées les préférences d'un tel agent est supposé parfaitement indépendant du contexte dans lequel ce dernier se voit soumettre les termes du choix, puis exprime ses préférences. En effet, dès lors qu'il est mis en demeure de choisir entre deux paniers de biens A et B et qu'il révèle sa préférence pour le panier A, cet agent est réputé se tenir à un tel choix dans l'avenir - en préférant toujours ce panier au panier B quand les deux lui sont simultanément présentés. L'adverbe "toujours" signifie en particulier ici "quel que soit le contexte dans lequel le choix entre A et B est soumis à l'agent", de sorte que tout se passe comme si celui-ci gardait en permanence en mémoire un ordre stable de préférences, dont ses choix successifs viendraient révéler (et inmanquablement confirmer) la

nature. Demeurant inchangé à travers les divers contextes présidant à l'expression du choix comme à la formulation de ses termes, ce système invariant de préférences répond étroitement à la mémoire parfaitement fidèle de la théorie de la localisation fonctionnelle: il apporte toujours la même réponse à une question donnée, indépendamment du contexte précis dans lequel cette question est posée.

Pour être ainsi prêtées à la mémoire de l'*homo cogitans* qu'abrite l'*homo oeconomicus*, de telles propriétés évoquent fortement celles que nous tenons habituellement pour caractéristiques de la mémoire des ordinateurs. C'est la raison pour laquelle la théorie de la localisation fonctionnelle trouve un soutien très moderne dans le développement des activités liées à l'informatique - plus particulièrement, note I. Rosenfield [1989, p.22], dans l'extension récente de la simulation par ordinateur. Un tel soutien apparaît si naturel qu'il demande à peine à être explicité. Avec P. Lévy [1987, p.90], il suffit en effet d'étendre le contenu du concept de *calcul* au delà de son sens mathématique restreint d'ensembles d'opérations arithmétiques, en rassemblant sous ce terme toute opération de tri, de classement, de permutation, de combinaison, de comparaison, de substitution ou de transcodification, pour qu'il devienne évident que l'*homo computans* de la théorie économique représente la figure emblématique de ce que cet auteur appelle le *paradigme du calcul* [op.cit., pp.88-118]. A l'instar de l'ordinateur, notre *homo computans* serait donc un automate, c'est-à-dire un dispositif transformant des messages d'entrée en messages de sortie suivant la structure logique préalablement imprimée en lui par un algorithme de tri, de classement, etc. Dans la mesure où l'observation montre des agents économiques réels se livrer très concrètement à de telles transformations, l'on peut légitimement se demander si la description théorique du comportement de l'*homo computans* ne traduit pas de manière entièrement fidèle l'ensemble des processus se déroulant dans les méandres du cerveau humain:

"En 1943, Mac Culloch et Pitts mettent en évidence l'isomorphie entre la dynamique des états d'un réseau de neurones idéalisés, ou formels, et le calcul des propositions dans le système de Russell. Ils démontrent en outre qu'un réseau de neurones formels (c'est-à-dire un modèle simplifié du cerveau) possède les mêmes capacités de calcul qu'une machine universelle de Turing pourvu qu'on lui adjoigne un ruban de longueur indéfinie et une tête de lecture-écriture. Cette condition semble raisonnable, puisque la plupart des hommes savent tracer des signes mnémotechniques sur différents supports et les relire. En bref, des automates physiques peuvent réaliser toutes les opérations intellectuelles descriptibles de façon finie et non ambiguë, et le cerveau pourrait bien être un de ces automates. Cela ouvre à la fois le programme d'une partie de la neurobiologie contemporaine et celui de l'intelligence artificielle". [P. Lévy, op. cit., p.95].

Dans la suite de son article, la perspective de P. Lévy est clairement critique vis-à-vis de certains prolongements d'un tel néo-mécanisme, qui conçoit l'univers physique comme étant exclusivement composé d'automates se livrant tout uniment à des opérations de traitement de données sur le mode impliqué par le paradigme du calcul, et qui donne par là-même l'univers entier pour omnicalculant. Cependant, malgré sa visible réticence à l'égard de cette métaphysique

computationnelle, cet auteur n'en souligne pas moins que

"Force est de constater que le débat scientifique et philosophique tend de plus en plus à se polariser autour des problèmes et dans le langage que le néomécanisme a réussi à imposer." [Ibid., p.99].

Or ce type de problèmes, comme le langage utilisé par le néomécanisme, renvoient directement à la conception de la mémoire qui est associée à la théorie de localisation fonctionnelle. Désormais soutenue par l'évidence des succès opératoires des outils construits sur son modèle comme par certaines interprétations cliniques de la pathologie de la mémoire, une telle conception n'en prête pas moins le flanc à la critique. A vrai dire, il est possible de réinterpréter de manière totalement différente les observations cliniques sur lesquelles la théorie de la localisation fonctionnelle fut initialement fondée, voire d'expliquer sur ces bases nouvelles bon nombre d'observations négligées par cette dernière.

1.2 Contre la localisation fonctionnelle

Plus de soixante pages: un bon tiers de I. Rosenfield [1989] tient en un premier chapitre [pp.27-93], consacré à une réinterprétation critique des données cliniques venant apparemment appuyer, depuis plus d'un siècle, la conception de la mémoire liée à la théorie de la localisation fonctionnelle. Il n'est certes pas question d'énumérer ici exhaustivement les éléments du long débat opposant les partisans de cette dernière, ou les tenants de la modularité fonctionnelle (1), à d'autres auteurs ayant défendu en ordre plus dispersé, qui une théorie holistique du fonctionnement du cerveau, qui une approche mettant l'accent sur l'importance du contexte et de la signification pour l'activité de remémoration. En fait, de la lecture de ce chapitre ressortent essentiellement deux propositions liées: 1) la mémoire activée par le cerveau humain ne peut pas consister en traces mnésiques spécifiques, organisées en une collection d'images fixes que la fonction de reconnaissance permettrait de comparer à celle des objets nouvellement perçus; 2) la remémoration n'est pas indépendante du contexte dans lequel le cerveau est invité à (re)produire un souvenir: au contraire, elle dépend étroitement des aspects affectifs entourant aussi bien la "mise en mémoire" que "l'extraction du souvenir".

Pour se convaincre de la pertinence de ces deux propositions, il suffit d'envisager un corpus de données rassemblées autour de trois points. Le premier examine brièvement le cas de "l'homme qui ne savait pas lire ce qu'il avait écrit", rapporté en 1892 par J. Dejerine (cf. I. Rosenfield, [1989, pp.42-66]). Le deuxième évoque la polémique ayant opposé le neurologue anglais J. Hughlings-Jackson à S. Freud [ibid., pp.69-80]; les vues de ce dernier furent indirectement confirmées par les "flash-back" mnésiques mis en évidence par W. Penfield dès 1933 et dont l'exploitation analytique ne fut complétée qu'en 1982 par P. Gloor [ibid., pp.152-156]. Quant au troisième point, il rappelle les résultats expérimentaux obtenus par A.A. Low et A.R. Luria [ibid., pp.83-87].

1.2.1 l'homme qui ne savait pas lire ce qu'il avait écrit

Bien qu'il considère la théorie de la localisation fonctionnelle comme une affaire entendue, J. Dejerine s'oppose à la version radicale de cette doctrine, qui donne chaque faculté de l'homme et de l'animal pour déterminée et contrôlée au niveau cérébral par une aire anatomiquement distincte. Plus précisément, il conteste la réalité du prétendu centre responsable de l'écriture mis en avant par S. Exner, en faisant remarquer que l'agraphie n'est jamais un symptôme isolé, mais se trouve généralement associée à une alexie: selon lui, écrire et lire sont deux activités étroitement liées, car placées sous la dépendance d'un seul centre mnésique contenant les images visuelles des mots. A l'appui de son affirmation, J. Dejerine cite le cas d'un malade devenu incapable de lire et d'écrire (l'autopsie révéla ensuite qu'une lésion unique était à l'origine de cette double incapacité). Ce malade continuait cependant de pouvoir lire certaines graphies: son nom, par exemple, ou encore des nombres à un ou deux chiffres qu'il pouvait également tracer sous la dictée - ce que Dejerine explique par la ressemblance de ces graphies avec des dessins.

Un an plus tard, J. Dejerine présente le tableau clinique d'un homme qui écrit et parle correctement, mais se révèle incapable de lire. Comme le malade précédent, Oscar C. semble reconnaître à leurs formes certaines graphies: il identifie les lettres isolées en s'aidant du geste qui dessine leurs contours, dit du nombre 112 qui lui est présenté que c'est un 1, un 1 et un 2, et ne parvient à dire cent douze qu'en écrivant. Après quatre ans de cécité verbale pure, l'incapacité de lire manifestée par Oscar C. se complique d'agraphie et de paraphasie qui aboutissent finalement, durant les dix derniers jours de sa vie, à une perte totale de l'expression, parlée et écrite. A l'autopsie, Dejerine découvre deux lésions, dont l'une d'apparition récente: selon lui, la première n'avait pas atteint le "centre visuel des mots", de sorte que le patient pouvait encore écrire et évoquer les images des lettres en les traçant mécaniquement avec ses doigts; par contre, la seconde lésion avait touché ce centre - d'où l'agraphie et la paraphasie observées (2). Car, ajoute Dejerine, certaines formes de notation équivalent à des dessins, et demeurent reconnaissables quand les images mnésiques relatives au langage sont détruites ou inaccessibles à la suite d'une lésion cérébrale: avec les nombres et les signatures, c'était de tels dessins que percevaient ses patients, et les lettres qu'ils reproduisaient n'étaient qu'autant de tracés graphiques. I. Rosenfield propose à cet égard une tout autre explication:

"En fait, ces patients pouvaient lire des nombres, à l'exclusion des lettres et des mots, parce que les nombres d'un chiffre, comme une signature, sont toujours lus de la même manière, quel que soit le contexte. Un 3 est un trois dans les expressions 'trois pommes' ou 'trois pour cent de rabais'. Mais les chiffres trois du nombre 333333 sont lus chacun différemment: trois cent mille, trente-trois mille, trois cents, etc. La signification d'un chiffre dans le cas d'un nombre qui en comporte plusieurs dépend de son emplacement, c'est-à-dire de son contexte. Il ne faut pas oublier qu'Oscar pouvait lire le nombre 112 en épelant 'un 1, un 1 et un 2'". [1989, pp. 55-56].

Il en va de même pour les mots, bien que s'ajoute ici une complication supplémentaire par rapport aux nombres comportant plusieurs chiffres: alors que dans un tel nombre la signification d'un 3 n'est conditionnée que par son emplacement, et non par le chiffre 3 ou le chiffre 4 qui le précèdent ou le suivent éventuellement, le changement d'une seule lettre dans un mot peut modifier prononciation et signification de celui-ci, qui elle-même dépend de ce qui précède et de ce qui suit. En somme,

"Si nous n'étions pas en mesure d'apercevoir de quelle manière les lettres d'un mot (ou, plus simplement, les chiffres d'un nombre) sont reliées entre elles, elles ne voudraient plus rien dire, puisqu'en elles-mêmes elles ne véhiculent aucun sens. C'est leur impuissance à saisir cette organisation globale - dans laquelle des stimuli identiques, des lettres, changent constamment de signification - qui est caractéristique des patients souffrant de cécité verbale. Ceux-ci ne peuvent pas combiner les stimuli de manière à donner un sens aux symboles (...) La faculté de combiner diversement les stimuli conditionne la reconnaissance et, jusqu'à un certain point, doit être au principe de la mémoire. Un visage, nous le reconnaissons en organisant les stimuli selon des modalités comparables à celles dont nous usions dans le passé (non nécessairement identiques, puisque la personne a sans doute vieilli). Cette analogie d'organisation fait le lien entre le passé et le présent. Naturellement, certaines images peuvent paraître 'immuables' - tels les nombres d'un chiffre -, mais leur caractère immuable n'est qu'apparent, puisqu'elles changent de signification dans des nombres de plusieurs chiffres, tout comme les lettres changent de sens d'un mot à l'autre" [I. Rosenfield, 1989, p.56].

Là où, voyant Oscar C. reproduire les lettres une à une en les référant à divers objets matériels (le "A" à un chevalet, le "Z" à un serpent, le "P" à une boucle, etc.), Dejerine pensait qu'il les "dessinait", I. Rosenfield ajoute une précision qui modifie entièrement l'interprétation que l'on peut donner d'un tel geste: plutôt que de "dessiner" les lettres, Oscar les reproduisait de manière particulière:

"On ne reproduit pas un dessin ligne après ligne (à moins de ne pas en avoir une vue d'ensemble), mais on se fait une idée approximative du tout qu'il constitue ou des parties que l'on veut reproduire, sans s'embarasser des détails exacts de l'original. Quand nous essayons de reproduire une image bidimensionnelle, nous cherchons d'instinct à en retrouver l'organisation. C'est seulement si nous ne parvenons pas à en extraire un sens que nous agissons comme Oscar, nous efforçant de reproduire servilement la longueur et l'emplacement de chaque ligne de l'original." [Ibid., p.58].

De même pour l'inaptitude désormais totale d'Oscar (auparavant aussi mélomane que sa femme était musicienne) à lire des notes de musique, ou pour son incapacité à percevoir les couleurs dans la partie droite de son champ visuel: ce type d'opérations demande de percevoir des relations complexes entre éléments dont la signification change selon la relation perçue qui, elle-même, change de signification suivant le contexte, musical ou chromatique, dans lequel elle se trouve. Selon I. Rosenfield, ces symptômes traduisent un déficit portant sur l'organisation des éléments (et non une perte mnésique spécifique et localisée), carence également responsable de l'inaptitude à lire des lettres et des nombres de plusieurs chiffres. L'on voit donc ici se profiler une conception de la mémoire très différente de celle associée à la localisation

fonctionnelle: capacité à reproduire une organisation plutôt qu'à convoquer des éléments spécifiques, la notion de mémoire y apparaît inséparable de celle de contexte. Pour l'instant, celui-ci n'est intervenu que dans la définition du contenu remémoré, mais nous allons voir qu'il est également opérant dans la situation de remémoration elle-même.

1.2.2 Hughlings-Jackson, Freud, Penfield et Gloor: l'importance du contexte pour la remémoration

Avec la description du cas de surdit  verbale de Marie Bouquet par C. Giraudeau vers la fin du si cle dernier, et son interpr tation par F. Bateman huit ans plus tard, il aurait d   tre clair - mais les neurologues de l' poque ne le virent pas - que le contexte dans lequel avait (ou n'avait pas) lieu la rem moration  tait tout   fait crucial pour la capacit  de produire cette derni re: le protocole des questions pos es   cette patiente montre que ses difficult s   comprendre la parole s'estompent d s lors que cette parole et ses questions s'expriment dans un cadre qui lui est familier - ou qui lui est devenu tel au cours de l'entretien lui-m me:

"Des questions ambigu s ne sont compr hensibles que dans le contexte des questions ant rieures. Pour Marie Bouquet, la difficult  semble se situer   ce niveau. Elle a beaucoup de peine    tablir un contexte lui permettant de comprendre ce qu'on lui demande. La r it ration permet de restaurer un contexte et de contraindre le patient   se concentrer sur la question pos e,   l'exclusion de celles qui l'ont pr c d e. La plus grande promptitude que manifeste Marie Bouquet dans certains cas souligne  galement   quel point le contexte conditionne la compr hension. Plac es en t te de l'interrogatoire, certaines questions, qui portent sur son nom ou sur la dur e de sa maladie, d pendent moins de ce qui a pu pr c der (...) C'est pourquoi il semble peu probable que les troubles de la compr hension pr sent s par Marie Bouquet soient dus   une disparition d'images mn siques verbales; cela provient plut t d'une incapacit    s'appuyer sur un contexte au sein duquel les mots et les phrases qu'elle entend retrouveraient un sens." [I. Rosenfield, op. cit., p.41].

Le nom, mais aussi les jurons et certains mots tels que "oui" ou "non": ce sont l  autant d'exemples de signifiants relativement ind pendants de leur contexte d' mission, donc susceptibles d' tre rem mor s dans n'importe quel contexte. Quel que soit ce dernier, mon nom reste le mien, un juron garde son sens ind pendamment du contexte *pr cis* dans lequel il est prof r , et le "oui" comme le "non" expriment le consentement et le refus en n'importe quelle occasion. D'apr s I. Rosenfield, c'est la raison pour laquelle tous les cas cliniques rapport s jusqu'  pr sent montrent des patients, pourtant atteints de troubles graves de la m moire, demeur s capables de reconna tre et d' mettre de tels signifiants:

"Des patients porteurs de l sions c r brales sont souvent r duits    mettre quelques mots susceptibles d' tre employ s dans n'importe quelle circonstance, parce qu'ils sont incapables d'organiser une r ponse verbale plus pr cise." [1989, p.71].

Dès 1879, J. Hughlings-Jackson attirait l'attention sur l'importance du contexte pour l'activité cérébrale: selon lui, les stimuli ne sont pas imprimés dans le cerveau, mais résultent d'une activité cérébrale se rapportant à l'objet extérieur. Contrairement au dogme central de la théorie de la localisation fonctionnelle, il n'existe donc pas d'état de conscience spécifiquement lié aux objets:

"Un état interne n'a de sens qu'en relation avec un certain contexte; des activités cérébrales analogues peuvent avoir des significations très différentes. Pouvoir forger des significations originales ou nouvelles, en réorganisant des stimuli dans de nouveaux contextes, est le propre du langage en particulier, et de la fonction cérébrale en général. Et c'est bien cette capacité, et son importance clinique, que Hughlings-Jackson a mise en évidence dans sa description de ce qu'il appela, chez un patient aphasique, 'les degrés supérieurs de l'articulation.' " [I. Rosenfield, *ibid.*, p.70].

Selon Hughlings-Jackson en effet, il existerait des niveaux plus ou moins profonds d'activité cérébrale: s'il remarque bien, chez les patients porteurs de lésions cérébrales, la persistance d'une capacité à proférer des jurons, ou d'exprimer certains énoncés appropriés dans un contexte émotionnel particulier (3), c'est, selon lui, parce qu'il s'agit là de réponses primitives, persistant à un niveau plus profond chez ces patients. En somme, les réponses "spontanées" (liées aux niveaux les plus "profonds" de l'activité cérébrale) subsisteraient malgré les lésions, alors que les réponses "volontaires" (liées à des niveaux plus superficiels d'activité cérébrale) seraient détruites par ces mêmes lésions. I. Rosenfield s'inscrit en faux contre une telle conception:

"Cette distinction néglige l'indépendance contextuelle relative des jurons, qui s'opposent à la plupart des expressions, dont le sens diffère selon le contexte (...) Des personnes porteuses de lésions cérébrales sont dans l'impossibilité de concevoir l'emploi de la plupart des mots, parce qu'il leur est difficile de relier les mots entre eux et de saisir les variations de sens qui surviennent nécessairement à chaque nouvelle phrase. Il n'y a pas eu perte de souvenirs, mais déficit portant sur la faculté de former des corrélations. Ne subsistent que les expressions ayant le moins possible recours à cette aptitude, c'est pourquoi on s'imagine que certains souvenirs ont été conservés, alors que manquent tous les autres mots. L'expression verbale est censée s'adapter au moment présent; et les souvenirs et les rappels sont probablement aussi l'expression adaptée de ce qu'on voit, entend et ressent à tout moment. Bien sûr, le déroulement des événements détermine en premier lieu ce qui est ou n'est pas approprié; cette capacité d'appréciation faisait défaut à Marie Bouquet, tout comme à Oscar, qui était incapable de relier les lettres entre elles. Les travaux de Hughlings-Jackson montrent la place que tient la capacité d'organiser dans la production des souvenirs, mais aussi celle qu'occupe l'affectivité, bien qu'il ait supposé à tort que les 'souvenirs émotionnels' étaient un élément primitif de la personnalité. L'acte de jurer n'est pas primitif, mais simplement affranchi du contexte jusqu'à un certain point." [1989, pp.72-73].

Il revient à Freud d'avoir montré toute l'importance du rôle joué par les émotions dans l'organisation de la mémoire. Chacun sait que le sens d'une expression donnée peut se différencier de ceux de la même expression employée dans des énoncés antérieurs, uniquement du fait d'une intonation à présent différente dans la voix de l'énonciateur. De manière très

générale, les affects ou émotions associées à un souvenir influent considérablement sur l'importance et la signification de celui-ci pour le sujet considéré:

"Si l'affectivité est absente, l'interlocuteur ne relève ni ne reconnaît le simple énoncé de l'épisode vécu. Il n'y a pas de mémoire sans affect. Les émotions sont indispensables à la constitution du souvenir, parce qu'elles le rattachent à un ensemble organisé et qu'elles le situent dans un séquence d'événements, de la même manière que les notions de temps et d'ordre, qui conditionnent la reconnaissance du souvenir en tant que tel, et non comme la pensée ou la vision d'un seul instant sans lien avec le passé." [I. Rosenfield, 1989, pp.73-74].

Une preuve empirique de l'importance cruciale du contexte émotionnel pour la remémoration nous est fournie par un exemple bien connu: celui des témoins oculaires d'un acte criminel, incapables d'identifier après coup l'auteur de l'acte dès lors qu'ils ne l'ont pas "situé" comme étant un criminel au moment exact du déroulement de cet acte. Si la personne en cause leur est initialement apparue hors de tout contexte criminel, ces témoins ne peuvent se souvenir d'elle ensuite comme étant inscrite dans un tel contexte. Plus généralement:

"Soulignant l'importance de cette question, Sigmund Freud observa que les individus ne se souviennent pas d'expériences passées qui ont été séparées de leur contexte émotionnel initial. Pour lui, l'expérience persiste, isolée pour ainsi dire, et n'est plus susceptible de reproduction dans le cours de l'activité intellectuelle. L'effet de cette isolation est identique à celui du refoulement avec amnésie". [I. Rosenfield, op. cit., p.74].

Il est parfaitement inutile de s'attarder ici à commenter un principe qui se trouve au fondement même de l'efficacité opératoire de la pratique psychanalytique, dont le succès paraît reposer sur la réalité et la qualité de l'investissement affectif lié à ce qu'il est convenu d'appeler le "transfert". Nous ne discuterons pas davantage de l'affirmation de I. Rosenfield selon laquelle Freud aurait été contraint d'expliquer l'aspect fragmentaire de la mémoire des rêves à l'aide des notions de condensation et de déplacement uniquement parce qu'il était convaincu de l'existence de souvenirs fixes, alors qu'en réalité, à condition de rejeter une telle conception de la mémoire, il serait possible, et davantage pertinent, d'expliquer l'aspect fragmentaire de cette dernière, dû à l'absence relative de contexte durant le sommeil [op. cit., pp.75 sq.]. Plus immédiatement pertinent pour notre propos est le fait que les vues de S. Freud sur le rôle de l'affectivité dans le processus de remémoration furent indirectement confirmées en 1933 par les expériences du neurochirurgien W. Penfield, complétées un demi-siècle plus tard par celles de P. Gloor et son équipe.

Au cours d'une opération sur un patient entièrement conscient, W. Penfield stimula électriquement la surface du cerveau de celui-ci. A la surprise générale - celle de Penfield comme celle du patient -, le sujet perçut alors une sorte de "flash-back" mnésique. L'expérience s'avéra réitérable (même fréquemment) quand l'on replaçait l'électrode peu après l'avoir ôtée, et le phénomène s'interrompait dès lors que l'électrode était retirée. D'après la description donnée par Penfield, cette évocation de souvenir, déclenchée électriquement, était tout à fait aléatoire, et, dans la plupart des cas, l'événement remémoré n'était guère important: simplement, le sujet "se

souvenait", sur un mode onirique, de scènes relativement quelconques qu'il se trouvait simultanément en mesure de commenter. W. Penfield pensait avoir découvert ainsi la preuve irréfutable de l'existence, dans la mémoire, de souvenirs permanents.

"En fait, ces souvenirs n'étaient souvent que l'observation d'autres personnes en train d'agir, et il était peu probable que le patient ait réellement pu être témoin de ces actes (...) En réalité, sur les cinq cent vingt patients qui reçurent une stimulation électrique au niveau des lobes temporaux, la zone du cerveau qui, selon Penfield, emmagasinait ces souvenirs, seulement quarante ont présenté ce qu'il a appelé une 'réponse vécue'. Des études ultérieures ont montré que ces réponses ne survenaient que lorsque les structures limbiques (généralement considérées comme indispensables aux expériences émotionnelles) étaient activées." [I. Rosenfield, 1989, p. 153].

En 1982 en effet, P. Gloor répéta l'expérience de Penfield sur trente-cinq épileptiques, et conclut sans ambiguïté qu'à moins d'activer les structures limbiques de ces patients, le phénomène de "réponse vécue" ne survenait pas. Tout semble donc se passer comme si la stimulation électrique des centres limbiques permettait d'amener à un niveau conscient des percepts élaborés dans le néocortex. En somme, même après élaboration dans le néocortex temporal, ce qui est éprouvé par les sens paraît devoir être transmis aux structures limbiques pour que l'expérience ait un caractère immédiat. Tel que le rapporte I. Rosenfield [ibid., p. 154], le passage suivant est souligné dans le texte de Gloor:

"La contribution limbique propre à ce processus pourrait être d'attribuer un rôle d'affect ou de motivation à un percept. Cette continuité limbique est probablement la condition préalable permettant à un percept d'être vécu ou évoqué consciemment; et cela pourrait impliquer que tout événement consciemment perçu possède une part de dimension affective, aussi petite soit-elle."

Pareille observation rejoint l'affirmation de Freud, selon laquelle il est impossible de reconnaître des souvenirs dépourvus d'affect. Naturellement, la seule activation limbique est insuffisante à procurer une sensation pleine et entière de mémoire: nous savons que les évocations de souvenirs restent fragmentaires tant qu'elles ne sont pas délibérément organisées au sein d'un contexte. Les observations effectuées dans les années trente par A. A. Low, comme celles rapportées trente années plus tard par A. R. Luria, sont tout à fait significatives à cet égard.

1.2.3 Contexte et signification

Il ne s'agit plus ici de mettre en évidence l'importance des aspects affectifs du souvenir dans le processus de remémoration, mais de montrer sous deux angles complémentaires l'importance du contexte (en général) pour la signification: d'abord en évoquant, avec A. A. Low, les modifications de compréhension et d'usage du langage entraînées, chez une personne porteuse d'une lésion cérébrale, par des changements de contexte; ensuite, en rapportant le cas, décrit par A. A. Luria, d'un patient, doué d'une mémoire remarquable, qui créait des contextes artificiels afin de se souvenir de listes de mots aléatoires [cf. I. Rosenfield, 1989, pp. 83-87].

Le patient présenté par A. A. Low avait été victime d'un accident vasculaire cérébral entraînant une paralysie partielle et la perte de la parole. Quelques mois après cet accident, il était à nouveau capable de parler, tout en manifestant certaines difficultés à trouver les mots, ainsi qu'un handicap de lecture bien particulier: il sautait de nombreux mots et combinaisons de mots dans une phrase ou un paragraphe qu'on lui donnait à lire. C'est afin de déterminer la nature et la cause de ce déficit précis que A. A. Low le soumit à toute une batterie de tests, dont nous ne mentionnerons ici que les trois les plus immédiatement pertinents pour notre propos. En premier lieu, sans pouvoir reconnaître vraiment ce qui était effectivement écrit dans une liste de mots renvoyant à des référents concrets, le patient parvenait à rendre la signification de ces mots (par exemple, pour *dad* il "lisait" *father*), ou "traduisait" la signification d'un mot écrit par une autre, voisine de la première (par exemple, il "lisait" *girl* pour *child*). Deuxièmement, le patient manifestait une incapacité quasi-totale à lire de courts mots-outils: *at*, *to*, *as*, *in*, etc., alors que d'autres, plus longs, étaient lus par lui avec moins de difficulté: *beyond*, *above*, etc. Au vu de ces résultats, Low conclut que les capacités de son client étaient limitées parce que sa lecture était entièrement dépendante de la signification: c'est bien parce que les significations de *dad* et *father* sont voisines, de même que - dans une mesure moindre - celles de *girl* et *child*, que ce patient pouvait "lire" l'un de ces mots pour un autre. Par ailleurs, il est clair que la longueur des mots-outils n'était nullement en cause dans sa plus ou moins grande capacité à les lire: selon Low, il lisait plus difficilement *at* que *beyond* parce que le premier possède un large éventail de significations: *at home*, *at ease*, *at your service*, etc., et n'acquiert un sens spécifique que dans un contexte donné, alors que le second a une signification définie, indépendante du contexte. Une telle interprétation apparut encore plus pertinente quand Low soumit son patient à un troisième type de tests: il lui donna à lire des syllabes dépourvues de significations: *sto*, *fal*, *tla*, etc., que le patient "lut" en les complétant ou les modifiant de manière que le résultat obtenu participât d'un réel contexte sémantique. C'est ainsi qu'il lut *story* à la place de *sto*, *far* à la place de *fal*, *serve* à la place de *ser*, *atlas* à la place de *tla*, etc. De même pour des phrases entières, dont il faisait une lecture fautive en essayant de donner un sens à des propositions imprimées qui n'en avaient aucun: *serve booksellers books* était lu *waiters sell books*, et *make, stop, listen* était lu *stop, look and listen*.

Il apparaît donc que ce qu'avait perdu ce patient du fait de la lésion cérébrale dont il était atteint, loin d'être des traces mnésiques spécifiques de mots qu'il aurait eu à confronter aux percepts nouveaux représentés par les signes écrits qui lui étaient donnés à lire, était une capacité (voire une appétence) à reconnaître des mots abstraits isolés dont la signification est étroitement indépendante de leur contexte: il "reconnaissait" bien, en un sens, certains mots concrets qu'il pouvait situer dans un champ sémantique familier (*father* pour *dad*), "lisait" plus facilement les mots-outils de signification indépendante du contexte (*beyond*) que ceux qui ne prennent sens que dans un contexte précis (*at*), et surtout s'efforçait spontanément d'inscrire les mots et phrases dépourvus de sens, qui lui étaient donnés à lire, dans le contexte sémantique lui

paraissant en l'occurrence le plus adéquat (4).

Les enseignements de la description du patient de A.R. Luria vont tout à fait dans le même sens: ce patient construisait délibérément un contexte artificiel pour se souvenir de listes de mots aléatoires. Plus précisément, il utilisait le support du plan fictif d'une petite ville pour ce faire, situant chaque mot à un emplacement différent de cette ville, et percevant alors la liste comme étant une partie du plan. Dans ces conditions, oublier un mot n'est pas un déficit de la mémoire, mais un déficit de la perception. Ainsi, le patient oublie le mot *crayon* parce qu'il l'a placé près de la barrière qui se trouve au bas de la rue: l'image du crayon s'est confondue avec celle de la barrière, si bien qu'il est passé devant sans la remarquer. De même pour le mot *oeuf*, qu'il a commis l'erreur de placer sur un mur blanc, et qui s'est fondu dans l'arrière plan: comment aurait-il pu repérer un oeuf blanc sur un mur blanc? (5).

L'évocation d'un souvenir apparaît ici comme un type de perception, de sorte que chaque contexte perçu modifie la nature de ce qui est évoqué. De manière plus générale, le réexamen des données cliniques effectué par I. Rosenfield débouche sur une conception de la mémoire très différente de celle associée à la théorie de la localisation fonctionnelle, à laquelle elle s'oppose fondamentalement sur quatre points liés. Premièrement, la perception et la reconnaissance ne sont pas conçues ici comme relevant de deux fonctions cérébrales indépendantes - ainsi que c'était le cas avec la théorie de la localisation fonctionnelle - mais comme deux aspects différents d'une seule et même fonction: c'est par un geste unique que le sujet perçoit et reconnaît. Deuxièmement, ce geste met en oeuvre une capacité d'organiser les stimuli perçus en éléments d'information cohérents, et non une aptitude à confronter chacun de ces stimuli à une trace mnésique spécifique, enregistrée et stockée en permanence dans notre cerveau. Troisièmement, la performance du sujet en situation de remémoration n'est jamais indépendante du contexte précis dans lequel elle est produite - contrairement à la mémoire telle qu'elle est conçue depuis P. Broca. Il est clair que ces trois premiers points sont étroitement liés entre eux: c'est bien parce que des souvenirs ne peuvent à la fois être fixes et varier en fonction du contexte que I. Rosenfield est conduit à abandonner la notion de mémoire permanente, et c'est un tel abandon qui l'incite à attribuer reconnaissance et perception à une fonction cérébrale unique. De la combinaison de ces trois points émerge alors une quatrième différence, opposant à l'aspect rigide de la mémoire absolument exacte associée à la théorie de localisation fonctionnelle - une mémoire qui ne nous permettrait guère de survivre dans un monde en perpétuel changement, fait remarquer I. Rosenfield [1989, p.79] - le caractère essentiellement dynamique de la mémoire telle que la conçoit ce dernier. Selon lui en effet, loin d'être immuables, les souvenirs doivent être considérés comme des reconstitutions opérées sur le passé et en perpétuel remaniement, qui nous donnent un sentiment de continuité, la sensation d'exister dans le passé, le présent et le futur: plutôt que des unités discrètes se perpétuant à travers le temps, les souvenirs constituent un système dynamique - au sens le plus fort de ce terme, puisque ce système ne se perpétue qu'en évoluant. La mémoire ainsi conçue procède par généralisations, en mobilisant des

catégories, elles-mêmes directement à l'origine du sentiment de continuité précité - que des souvenirs spécifiques seraient impuissants à suggérer. Ayant toujours lieu dans le contexte de l'instant même, *hic et nunc*, la remémoration consiste en la mise en oeuvre de processus opératoires visant à intégrer continuellement de nouveaux stimuli à des catégories préexistantes qui s'en trouvent elles-mêmes plus ou moins profondément modifiées et remaniées: produite par l'imagination, une telle reconstitution est une vision du passé adaptée au moment présent. Telle est la conception qui se dégage globalement des éléments *négatifs* de preuve issus du réexamen par I. Rosenfield d'expériences ou de données cliniques jusque-là destinées à conforter la conception associée à la théorie de la localisation fonctionnelle. Après cette réfutation d'une mémoire pour ainsi dire *inventée*, il reste à apporter les preuves *positives* militant en faveur de la mémoire *inventive* dont cet auteur veut promouvoir la notion.

1.3 Darwinisme neuronal et mémoire inventive

Ce n'est certes pas un hasard si c'est *aussi* en raison de l'inadéquation de la notion de mémoire exacte et fixe à un monde en perpétuel changement que I. Rosenfield rejette une telle notion: ce rejet traduit l'influence de la problématique darwinienne de l'évolution sur sa conception de la mémoire humaine. En fait, la problématique évolutionniste se trouve au fondement même de cette conception, et ceci de deux manières complémentaires. En premier lieu, c'est en contrastant l'aptitude manifeste de notre cerveau à créer continuellement de nouveaux processus opératoires avec l'incapacité avérée des automates à nous imiter sur ce point, que I. Rosenfield souligne les limites des tentatives les plus récemment effectuées en ce domaine par l'informatique. En second lieu, cet auteur met une telle aptitude en relation très étroite avec la nécessité, continuellement éprouvée par notre organisme, de s'adapter à un environnement changeant: selon I. Rosenfield, reconnaître en percevant, c'est percevoir ce qui peut être utile à notre survie dans l'environnement (le contexte) actuel, en organisant ce perçu de manière cohérente avec nos expériences passées. Examinons brièvement le premier de ces deux aspects complémentaires, pour développer davantage le second - qui débouche directement sur la théorie rosenfieldienne de la mémoire.

1.3.1 Limites de la reconnaissance automatique

Après avoir montré dans un très court deuxième chapitre [1989, pp.95-108], que la signification va toujours du général au particulier lors d'un processus de perception et d'apprentissage (notamment celui de la parole par l'enfant), et que la perception des stimuli dépend de la catégorie que nous leur attribuons et de leur organisation en fonction des autres stimuli- et non de la structure propre à chaque stimulus (catégorisation *versus* localisation) -, I. Rosenfield consacre entièrement son chapitre trois [1989, pp.109-145] à la reconnaissance automatique des formes et à ses limites. Selon lui, tant en ce qui concerne les "machines de vision" construites par D. Marr que ce qu'il appelle un PDP ("parallel distributed processor"), la

limite principale de l'approche informatique en ce domaine consiste à faire travailler les processus opératoires accomplissant une tâche donnée sur un mode en lui-même invariable: s'il résoud certaines difficultés créées par la notion de mémoire fixe et prend en compte l'importance du contexte, le modèle de D. Marr continue de traiter perception et reconnaissance comme des processus dissociés se déroulant toujours de la même manière [op. cit., pp.110-133]; quant au PDP, fonctionnant sur un modèle de mémoire associative suscitant aujourd'hui beaucoup d'espoirs chez les promoteurs des "machines neuronales" (cf. L. Personnaz, G. Dreyfus et I. Guyon [1988]), ils sont capables de généraliser (de classer un élément inconnu dans une catégorie connue), mais ceci uniquement à l'intérieur du cadre prédéfini par le programmeur:

"Si, par exemple, on entre le code d'une fleur qu'il n'a jamais *appris*, il associera, malgré tout, une couleur à cette fleur. Mais cela n'a rien de vraiment surprenant, puisque le programmeur codera la nouvelle fleur de la même manière que les autres, dont les codes auront déjà été appris. Il s'agit d'une généralisation préfabriquée. Les schémas d'activité, qui se chevauchent et qui permettent à la machine d'associer une couleur à une fleur inconnue, dépendent des codes du programmeur et ne sont pas dans la nature des fleurs ou des couleurs." (I. Rosenfield, op. cit., p.136).

Le type de mémoire associative utilisé par le PDP ou la machine neuronale ne fonctionne donc pas dans les conditions qui sont celles des processus de perception-reconnaissance formant ce qu'il est convenu d'appeler la mémoire de l'organisme humain. Ici,

"la plupart des stimuli ne sont pas organisés en groupes ordonnés d'information, tels que les mots ou le langage en général. Le cerveau doit d'abord organiser les stimuli en ensembles significatifs. Trouver des sous-traités communs entre eux et observer leurs relations contextuelles ne suffit pas à produire de nouvelles catégories, qui dépendent de différentes combinaisons de stimuli ne présentant par elles-mêmes aucun sens pour l'organisme, mais qui l'aideront à affronter l'environnement. En fait, il est impossible de prévoir à quoi serviront les stimuli ou leurs combinaisons - cela représente une des restrictions de toute machine programmée. Les réseaux du PDP généralisent à partir d'échantillons d'information déjà existants. Mais, pour un organisme, il est tout aussi impossible de connaître à l'avance la composition de l'information. Le cerveau doit dans la mesure du possible essayer des combinaisons de stimuli afférents, puis sélectionner celles qui aideront l'organisme à se rattacher à son environnement. D'où l'importance fondamentale du contexte présent pour déterminer la signification d'un stimulus ou d'un groupe de stimuli. En fait, les stimulations issues du PDP sont réellement fondées sur les principes de la localisation et sur la confrontation entre stimuli et mémoire fixe. Les généralisations réalisées par ces dispositifs ne sont rien de plus qu'un chevauchement de schémas en codes prédéterminés. Une réelle généralisation crée de nouvelles catégories d'informations; et, du point de vue d'un organisme, cette élaboration est la conséquence d'éléments imprévus issus de l'environnement." [I. Rosenfield, 1989, pp.136-137].

En somme, alors que la mémoire de l'ordinateur traditionnel permet de reconnaître, en confrontant un stimulus perçu à un stimulus préalablement enregistré par lui, mais non de généraliser (de "reconnaître" et de classer un stimulus qu'il n'a jamais "vu"), la mémoire associative du PDP ou de la machine neuronale permet de généraliser, mais ceci toujours de la même manière, pré-établie. Au sens des catégories de l'apprentissage définies par G. Bateson

[1977, pp.253-282], le premier est capable d'apprendre, et le second d'apprendre à apprendre, mais aucune machine de ce type ne manifeste *cette aptitude apparemment propre au cerveau humain: apprendre à apprendre à apprendre*. A suivre I. Rosenfield, c'est pourtant sur ce troisième niveau hiérarchique d'apprentissage et de fonctionnement de la mémoire humaine qu'il faut faire porter notre effort de compréhension de celle-ci. Pour notre survie, il est nécessaire que notre mémoire nous permette de décider, selon le contexte, que tel cercle représente le soleil, une roue, un plateau de table, ou autre chose encore que nous ne tardons pas à désigner d'un nouveau nom, *absent jusque-là de tout lexique* - ce dont aucun ordinateur n'est capable. C'est de cette capacité qu'il nous faut rendre compte dans une théorie pertinente de la mémoire humaine contribuant à fonder

"notre faculté d'adaptation à des environnements aussi différents que celui du troglodyte il y a trente mille ans et le monde de l'informatique." [op. cit., p.119].

Nous débouchons ainsi directement sur la théorie du "darwinisme neuronal" de G.M. Edelman, dont I. Rosenfield s'inspire très largement (et le nom même de cette théorie, allié aux considérations précédentes, indique assez les raisons d'une telle démarche) pour construire sa propre théorie de la mémoire.

1.3.2 Darwinisme neuronal et mémoire inventive

C'est en combinant la théorie darwinienne de l'évolution avec ses propres travaux sur le système immunitaire humain - travaux pour lesquels il reçut le prix Nobel en 1972 - que G.M. Edelman fut conduit à formuler une explication inédite du fonctionnement de notre cerveau (6). Cette explication est centrée sur l'idée selon laquelle le cerveau fonctionnerait comme un système sélectif - ce que nous appelons apprentissage serait une forme de sélection -, et elle repose sur trois postulats: 1) au cours du développement cérébral de l'embryon, se formerait entre les neurones un schéma de connexions extrêmement variable et individualisé; 2) après la naissance, certaines combinaisons de connexions de ce schéma, plutôt que d'autres, seraient sélectionnées en réponse aux stimulations sensorielles reçues par le cerveau; 3) cette sélection surviendrait plus spécifiquement au niveau de groupes de neurones reliés en couches - au niveau de "cartes" dialoguant entre elles afin de constituer des catégories de choses et d'événements. (Cf. I. Rosenfield, [1989, p.159]).

Loin de tout déterminisme génétique attribuant comportement et souplesse d'adaptation à un apprentissage produisant des schémas d'organisation cérébraux en réponse aux stimuli externes à la manière de programmes fournissant des instructions à des ordinateurs, il s'agit d'expliquer la pensée et l'action en tant que réponses efficaces et infiniment variées aux conditions inattendues présentées par le milieu: le fond du problème consiste en l'ampleur de la variabilité de la structure du cerveau, suffisante, malgré une série limitée de gènes, pour permettre à l'organisme de s'adapter à un environnement imprévisible.

Selon G.M. Edelman, ceci n'est possible que parce qu'au cours du développement embryonnaire la différenciation cellulaire n'est pas uniquement déterminée par les mécanismes génétiques, mais dépend également de l'endroit où chaque cellule se trouve au début du processus de différenciation, ainsi que des positions qu'elle occupait auparavant. L'inévitable variation de ses formes et mouvements selon les individus rend la position d'une cellule donnée à un moment donné impossible à prévoir exactement: une différenciation complètement déterminée par les gènes impliquerait des conséquences désastreuses pour l'organisme dès lors qu'une seule cellule serait mal située.

A rebours de la rigidité architecturale associée à la seule détermination génétique, G. Edelman envisage donc un certain degré de flexibilité en ce qui concerne les schémas d'organisation cellulaire. Ces derniers résultent de processus façonnant les groupes cellulaires en structures aptes à des fonctions plus complexes, et dont la forme est déterminée par les signaux en provenance d'autres groupes. Des limites se créent, au fur et à mesure que les tissus se développent dans l'embryon, et déterminent les différentes zones fonctionnelles de l'organisme grâce à des colles intercellulaires appelées "molécules d'adhérence cellulaire" ou CAM. Il existe différents types de CAM, dont la structure moléculaire est déterminée par des gènes précis, mais dont la quantité exacte et le degré d'adhésivité (qui varie continuellement au cours du développement embryonnaire) dépendent des localisations actuelle et passée des cellules qui les portent. Les frontières se forment entre groupes de cellules agglutinées par différentes CAM, et ces cellules se différencient une fois ces limites constituées. Au fur et à mesure de cette différenciation, de nouvelles frontières apparaissent, puis de nouvelles modifications sont induites par les signaux émis par les groupes cellulaires en direction des gènes activant les CAM et de ceux responsables de la différenciation. Selon les étapes par lesquelles sont passés deux groupes cellulaires ainsi constitués, les signaux échangés entre eux déterminent la formation ultérieure de types cellulaires extrêmement variés de part et d'autre de la frontière. Il apparaît finalement que

"La fonction de cette limite située entre les groupes à CAM différents dépend donc du contexte: l'environnement des cellules et leur passé. De plus, les règles qui gouvernent la réponse des CAM sont généralement les mêmes pour les neurones du cerveau et pour d'autres structures corporelles. Parce que les frontières des groupes cellulaires dépendent de la dynamique des mouvements, il y aura des variations individuelles qui ne seront pas simplement déterminées par les gènes, et dont la diversité garantira une structure propre à chaque cerveau. Cependant, la structure d'ensemble qui en résulte et les conséquences similaires du développement embryonnaire pourraient expliquer qu'individuellement les cerveaux des membres d'une espèce se ressemblent." (I. Rosenfield, *ibid.*, p.163).

A l'infinie diversité potentielle du contexte répond donc l'infinie diversité potentielle des individus de l'espèce. C'est pourquoi il est impossible de prévoir dans le détail la morphologie définitive d'un animal à partir de la connaissance de la totalité de son patrimoine génétique, et c'est aussi ce qui permet de comprendre que deux cerveaux ne peuvent être absolument identiques: même avec un patrimoine génétique commun, comme c'est le cas des vrais jumeaux,

il faudrait pour cela que le contexte et le développement cellulaire de chacun déterminent des mécanismes de CAM strictement identiques, ce qui est pratiquement impossible.

Au contraire, ces mécanismes sont vecteurs de diversité parmi les fonctions cérébrales des différents individus, et ceci, ajoute Edelman, par le biais d'un processus de sélection des groupes de neurones. Cette sélection ne joue pas de la même manière avant et après la naissance. En effet, nous avons vu qu'au stade embryonnaire la sélection portait sur le schéma d'organisation en groupes neuronaux distincts: l'unité de sélection correspondant à un jeu de neurones interconnectés fonctionnant simultanément, les changements qui surviennent dans la dynamique des CAM déterminent la variété des groupes, dont chacun répondra différemment à un même stimulus capté par les récepteurs sensoriels de la lumière, du toucher et du son, situés au niveau de l'oeil, de la peau et de l'oreille. Après la naissance, la sélection intéresse le *degré* des connexions entre neurones, et non plus leur schéma d'organisation: un vaste répertoire de différents groupes neuronaux ayant déjà été créé, les modifications d'intensité des connexions neuronales déterminent à présent les circuits qu'emprunteront les signaux neuronaux. Tel groupe, répondant plus activement que tel autre à un stimulus provenant de l'environnement, verra les jonctions synaptiques entre ses éléments s'en trouver renforcées: tout se passe comme si le stimulus *sélectionnait* des variantes particulières au sein de la population de groupes neuronaux:

“en effet, la réponse d'un groupe est susceptible d'être amplifiée. Comme ses connexions sont renforcées, il lui est possible de modifier l'intensité de ses liaisons avec d'autres groupes et, en entrant en compétition avec eux, d'intégrer leurs neurones dans sa propre activité de réponse. Le renforcement des zones synaptiques crée ce qu'Edelman appelle un répertoire secondaire, constitué de groupes neuronaux qui répondent mieux à certains stimuli parce qu'ils ont été sélectionnés et leurs connexions renforcées.” [I. Rosenfield, 1989, p. 166].

Reste que les stimuli ne sont pas organisés d'emblée en éléments cohérents d'information: ceux qui intéressent plus précisément les fonctions cérébrales supérieures, telle la fonction nous permettant de ranger deux tulipes concrètes différentes dans la même classe abstraite “tulipe”, *doivent être organisées par l'organisme afin de présenter un sens pour lui*. Selon Edelman, c'est le rôle de *cartes cérébrales* constituées de groupes neuronaux que de réaliser une telle organisation:

“Une carte rassemble plusieurs de ces groupes disposés dans le cerveau, de manière à préserver le schéma des relations qui unissent soit une couche de récepteurs sensoriels (tels que ceux de la surface cutanée de la main) et une couche de tissu nerveux cérébral recevant les stimuli sensoriels, soit deux couches de tissu nerveux entre elles. Les groupes sont disposés en cartes ‘interactives’ pour créer des catégories d'objets ou d'événements. On trouve divers types de cartes dans différentes zones du cerveau, et l'étude de leurs interactions - appelées réintroductions - représente ce que la théorie d'Edelman a d'essentiel et de définitif.” [I. Rosenfield, 1989, p. 167].

Ces cartes trient les stimuli afférents en fonction de leurs similitudes et en combinant leurs propriétés: elles *organisent* les stimuli en schémas aidant l'organisme à faire face à son environnement. Selon Edelman, il doit exister une interaction permanente entre elles pour qu'il y

ait catégorisation: des expériences récentes montrent par exemple que la localisation des objets dans l'espace s'obtient par l'interaction d'un grand nombre de cartes sensorielles classant et combinant les stimuli sonores, visuels, etc, et débouchant sur un schéma d'activité qui déclenche une réponse motrice.

C'est ainsi que le cerveau *crée* ses catégories et ses généralisations, et remodèle celles-ci si nécessaire - l'organisation de cette cartographie n'étant pas immuable, mais susceptible de se réarranger en cas de besoin. En effet,

"Le cerveau possède de nombreux types de cartes, et il en constitue beaucoup d'autres à l'origine de multiples catégories destinées aux afférences. L'objectif des cartes est la création de catégories perceptives qui permettront à l'animal d'agir de manière appropriée. Les environnements auxquels il est confronté se modifient nécessairement, et cette modification affecte les catégories perceptives. Mais telle est précisément la fonction qu'opèrent ces multiples cartographies: interagir au moyen de cartes réorganisant constamment l'information. Et en rapportant les schémas opératoires (*mappings*) formés plus fréquemment aux cartes sensorielles primaires, qui entretiennent une relation continue avec les stimuli externes, le cerveau peut efficacement suivre les progrès de ses divers regroupements d'entrées sensorielles." (I. Rosenfield, 1989, p.170).

Ce darwinisme neuronal fournit des preuves *positives* en faveur de la théorie rosenfieldienne de la mémoire humaine. Selon I. Rosenfield en effet, l'explication de G.M. Edelman constitue le fondement biologique d'une nouvelle psychologie, de sorte qu'il est possible d'étendre aux fonctions cérébrales supérieures, gouvernant les représentations mentales et leur expression par le langage articulé, les résultats obtenus sur des formes moins sophistiquées de perception-reconnaissance et d'expression (7).

En définitive, ce que nous percevons ne résulte pas d'une analyse de données, mais de ce que nous avons perçu et vécu dans le passé ou à l'instant même: nos premiers contacts avec le monde extérieur nous incitent à mettre à l'épreuve différentes façons d'organiser nos sensations, et celles qui permettent un comportement rationnel ou utile sont renforcées par la suite. Expériences, sentiments et pensées diffèrent radicalement d'une personne à l'autre. Inexplicables en termes de processus aussi immuables que le traitement de données, ces variations relèvent de constructions de combinaisons toujours singulières organisant les divers stimuli en provenance d'un environnement en perpétuel changement: nous devons être capables de faire face à ce dernier, non au moyen d'images invariables mémorisées à l'avance, mais d'une manière qui prenne en compte le nouveau et l'inattendu aussi bien que le caractère individuel de nos expériences passées. Nous ne nous servons donc pas d'images enregistrées, mais de processus opératoires permettant de comprendre le monde et de le transformer. Mobilisée par notre confrontation à un environnement variable, la mémoire est davantage invention de catégories dans lesquelles ranger le perçu actuel que mise en correspondance d'images anciennes avec des stimuli d'emblée organisés.

Tel qu'il est décrit par I. Rosenfield, le mode de fonctionnement de la mémoire humaine se révèle donc propre à expliquer rationnellement certaines incohérences apparentes d'agents économiques concrets, soumis à des tests montrant que leurs systèmes de préférences n'ont rien d'invariant. En réalité, l'ensemble des comportements observés lors de ces tests (inversion des préférences, "framing effect" et biais liés au traitement de l'information) ne sont incohérents que dans le cadre dessiné par la théorie de la localisation fonctionnelle et l'assimilation correspondante de notre mémoire à un fichier informatique consulté selon des procédures prédéterminées et immuables. Au contraire, dès lors que le fonctionnement de la mémoire est identifié à une mise en acte de processus opératoires redéfinissant continuellement les catégories (et les combinaisons de catégories) dans lesquelles ranger les stimuli perçus, de tels comportements deviennent parfaitement cohérents: si la dimension d'invention dont est toujours empreinte la mémoire (qui reconnaît en percevant) dépend étroitement et simultanément du contexte actuel et de l'expérience passée, il n'est guère étonnant que les agents inversent leurs préférences entre choix et jugement de valeur, ni qu'ils lient leurs réponses à la manière dont ils perçoivent les options, ni enfin qu'ils utilisent l'information d'une manière biaisée - eu égard à la stricte rationalité économique.

En effet, le résultat apparemment troublant selon lequel le même agent préfère une option A à une option B tout en valorisant davantage B que A en termes monétaires, loin de refléter l'incohérence dénoncée par Grether et Plott, trouve une double explication: celle développée par Loomes et Sudgen sur la base de la "théorie du regret" introduite par Bell, qui fait explicitement appel à l'importance du contexte pour le choix en avançant que les agents évaluent leurs options en référence aux autres options présentes au moment de choisir; celle proposée par les psychologues, tout aussi rationnelle que la précédente qu'elle vient compléter, selon laquelle les processus cognitifs et les modes d'expression mobilisés par une question posée en termes de jugement diffèrent de ceux qu'active un autre type de question, posée en termes de valeur - ordonner n'est pas mesurer, et la désirabilité ne s'exprime pas nécessairement par un prix de réservation. Il en va de même pour le "framing effect": bien que la probabilité de panne d'un équipement soit mathématiquement complémentaire (à l'unité) de celle que ce même équipement ne tombe pas en panne, il est clair que ces deux événements contraires n'évoquent nullement le même contexte émotionnel, ou affectif, chez l'agent. Il n'y a donc aucune raison que ce dernier classe les événements qui lui sont présentés dans un contexte "positif" de la même manière que lorsque les différentes options offertes s'inscrivent pour lui dans un contexte "négatif". Il serait facile de montrer que le même type de phénomènes est à l'oeuvre dans la troisième catégorie d'expériences mentionnées dans le texte de M. Willinger - notamment dans le biais de confirmation.

Pour l'économiste, le premier intérêt de la théorie rosenfieldienne de la mémoire est donc qu'elle offre un cadre rationnel d'explication à certains comportements apparemment incohérents des agents. Pour n'être pas mince, cet intérêt n'est pas le seul, car la contribution de I.

Rosenfield évoque irrésistiblement un autre type de modèle: celui de l'auto-organisation, proposé naguère par cet autre biologiste qu'est H. Atlan.

2. L'auto-organisation d'après H. Atlan

Convergence née de leur commune familiarité avec un même ensemble de disciplines? A l'instar de I. Rosenfield, H. Atlan fonde sa recherche sur l'insatisfaction qu'il éprouve face à l'utilisation de la notion de programme dans l'étude des êtres vivants, mais ne dénie pas pour autant toute pertinence aux simulations du fonctionnement de ces mêmes êtres par des automates.

Tous deux demandent néanmoins à de tels automates d'être capables de modifier par eux-mêmes (sans instructions préalablement fournies par un programmeur, c'est-à-dire sans programme explicite) leurs processus d'apprentissage, afin de pouvoir prétendre imiter le vivant - notamment ce type particulier de vivant que représente l'être humain. Rapportée à ce dernier, la nébuleuse de notions associées au concept d'information (mémoire, perception, décision, apprentissage, etc.) doit en effet rendre compte de l'aptitude, manifestée par notre cerveau, à accéder au troisième niveau hiérarchique défini jadis par G. Bateson dans sa typologie des catégories de l'apprentissage (cf. *supra*): c'est bien parce que Darwin II semble capable de telles performances que I. Rosenfield le distingue de tous les automates construits auparavant, et nous allons le voir, c'est à la réalisation de performances de cet ordre que H. Atlan appelait depuis longtemps (8).

D'après H. Atlan en effet, les organisations biologiques, étant dynamiques et non totalement maîtrisées par l'homme (parce que non construites par lui), constituent une sorte de compromis entre deux extrêmes:

"un ordre répétitif parfaitement symétrique dont les cristaux sont les modèles physiques les plus classiques, et une variété infiniment complexe et imprévisible dans ses détails, comme celle des formes évanescents de la fumée."
[1979, p.5],

et il s'agit de comprendre ce type d'organisations et, plus largement, les systèmes non construits (naturels ou sociaux), en évitant deux obstacles: d'abord celui du vitalisme, pétition de principe niant que le phénomène de la vie soit explicable en termes physico-chimiques et mécanistes - une attitude purement négative et stérilisante, selon H. Atlan; ensuite, celui que constitue la notion téléonomique de programme, introduite par J. Monod [1970] afin d'éviter le finalisme implicitement présent dans le discours des biologistes jusqu'au début des années soixante, mais qui ne fait que déplacer le problème, sans le résoudre [H. Atlan, 1979, pp. 14-18].

En réalité, les questions posées par la biologie moléculaire, comme les métaphores cybernétiques appliquées à la biologie, témoignent d'un langage nouveau et appellent des réponses impliquant

"non pas une réduction du vivant au physico-chimique, mais un élargissement de celui-ci à une biophysique des systèmes organisés, applicable à la fois à des machines artificielles et naturelles" [Ibid., pp.23-24].

C'est en ce sens que vont tous les travaux sur la logique de l'auto-organisation appliquée aux êtres vivants. Dans un tel cadre, le postulat fondamental sur lequel s'appuie cette logique est que les performances les plus extraordinaires des organismes vivants sont le résultat de principes cybernétiques particuliers qu'il s'agit de découvrir et de préciser:

"En tant que principes *particuliers*, ils doivent rendre compte du caractère propre aux organismes vivants que présentent ces performances. Mais en tant que principes *cybernétiques*, ils sont postulés en continuité avec les autres domaines de la cybernétique, les mieux connus, ceux qui s'appliquent aux automates artificiels. Les conséquences de ce postulat sont doubles: a) la spécificité des organismes vivants est rattachée à des principes d'organisation plutôt qu'à des propriétés vitales irréductibles; b) ces principes une fois découverts, rien ne devrait empêcher de les appliquer à des automates artificiels dont les performances deviendraient alors égales à celles des organismes vivants. C'est dans cette perspective que les recherches formelles sur la logique de systèmes auto-organiseurs, qui sont à la fois hypothétiques, en ce sens que personne n'en a jamais réalisé, et pourtant bien réels, en ce sens que la nature en fournit abondamment, peuvent présenter un intérêt." [Ibid., p.24].

La dernière phrase de cette citation de H. Atlan est rendue passablement obsolète par l'existence de Darwin II [cf. note (7)], dont la mémoire fonctionne sur un mode à ce point proche du nôtre qu'elle crée comme la mémoire humaine les catégories à l'intérieur desquelles elle range les objets offerts à sa perception - faisant ainsi de chaque souvenir une nouvelle création, souligne I. Rosenfield [1989, p.174]. Mais l'ensemble de cette même citation montre bien que c'est précisément ce type de performance qu'attendait H. Atlan de l'application à des automates artificiels des principes cybernétiques particuliers qu'il voulait découvrir. Par exemple, la notion de "répertoire secondaire" introduite par G.M. Edelman constitue bien l'un de ces principes, puisque le renforcement des zones synaptiques, par lequel un tel répertoire est créé, représente un cas évident de rétroaction positive.

Entre H. Atlan et I. Rosenfield [1989, 1990]] étendant la théorie de G.M. Edelman aux fonctions mentales supérieures, la jonction conceptuelle présente dès lors l'aspect quasi-inéluctable de la réunion d'apports complémentaires à des projets analogues issus de présupposés identiques: au sein de cette problématique commune, l'un (Rosenfield) élabore, à partir d'un substrat neurophysiologique, une conception très globale de la nature et du fonctionnement de la mémoire du vivant en tant que système capable de se réorganiser constamment sous la pression de stimuli contextualisés, tandis que l'autre (Atlan) fournit un formalisme rigoureux à la notion d'auto-organisation d'un système perturbé par un "bruit" aléatoire, en provenance de son environnement et déclenchant un "processus d'augmentation de complexité à la fois structurale et fonctionnelle résultant d'une succession de désorganisations rattrapées, suivies chaque fois d'un rétablissement à un niveau de variété plus grande et de redondance plus faible." [H. Atlan, 1979, p.49].

La parenté conceptuelle de cette définition d'un processus auto-organisateur avec la conception rosenfieldienne de la mémoire se laisse d'emblée entrevoir au plan global.

Néanmoins, c'est au niveau du *détail* de chacun de ces modèles que les correspondances existant entre eux nous paraissent les plus fructueuses. Il convient donc d'exposer à un niveau relativement fin l'articulation interne du modèle de H. Atlan (2.1), exposition qui nous permettra de préciser ensuite la signification de l'information dans un système hiérarchisé (2.2).

2.1 L'auto-organisation selon H. Atlan

Chaque terme de la définition générale de l'auto-organisation donnée à l'instant prend un sens particulier dans le texte de H. Atlan, qui emprunte les notions de complexité, structure, fonction, organisation, variété et redondance à la théorie des systèmes, à la cybernétique, à la théorie de l'information, pour les réinterpréter dans le cadre du paradigme auto-organisationnel qu'il veut promouvoir. Il importe donc de bien saisir leurs significations précises dans un tel cadre.

A cette fin, il nous semble indiqué de partir d'une expression exprimant formellement la propriété auto-organisatrice présentée par certains systèmes, puis d'interpréter une telle formule. Soient H , la quantité d'information contenue dans un message; R , le degré de redondance de ce message, et H_{\max} , la quantité maximale d'information contenue dans ce message, correspondant à $R = 0$. Par définition, $R = (H_{\max} - H)/H_{\max}$, d'où $H = H_{\max} (1-R)$. Par hypothèse, la quantité H varie dans le temps sous l'effet de facteurs aléatoires: $dH/dt = d(H_{\max} - R \cdot H_{\max})/dt$. Après une série de transformations élémentaires, il vient: $dH/dt = -H_{\max} \cdot dR/dt + (1-R) \cdot dH_{\max}/dt$, qui est la formule que nous avons à interpréter.

Selon H. Atlan, nous devons pour cela retourner à C. Shannon, et plus particulièrement au théorème établi par ce dernier, selon lequel

"la quantité d'information d'un message transmis dans une voie de communication perturbée par du bruit ne peut que décroître d'une quantité égale à l'ambiguïté introduite par ce bruit entre l'entrée et la sortie de la voie. Des codes correcteurs d'erreurs, introduisant une certaine redondance dans le message, peuvent diminuer cette ambiguïté de telle sorte qu'à la limite la quantité d'information transmise sera égale à la quantité émise; mais en aucun cas elle ne pourra être supérieure." [1979, pp.44-45].

Ainsi appliqué à un message composé d'une suite de symboles et transmis dans une voie de communication perturbée par du bruit, ce théorème (9) exclut toute possibilité d'un rôle organisationnel de ce dernier. Qu'en est-il si l'on étend le formalisme de Shannon à la théorie des systèmes organisés, en mesurant par la quantité d'information correspondante le degré d'organisation d'un tel système? A priori, une telle extension paraît d'autant plus légitime que le théorème de la voie avec bruit est d'application très générale, au double sens où: 1) il porte sur toute suite de symboles composant un message comme autant de parties de celui-ci; 2) sa validité ne dépend en aucune manière de la nature physique de la voie de transmission considérée. Bien qu'il ait notoirement été établi par Shannon afin de résoudre certains problèmes de télécommunications, ce théorème semble donc s'inscrire tout naturellement dans une théorie de

l'organisation des systèmes généraux, en s'appliquant au cas particulier dans lequel la voie de transmission considérée s'établit entre le système considéré et l'observateur de celui-ci. Et dans ce cadre, rien ne s'oppose à l'assimilation de toute suite d'éléments de ce système à un message émis par une source en direction d'un destinataire-observateur, ni à la caractérisation mathématique de cette suite à l'aide de la célèbre formule de Shannon.

Néanmoins, nous allons voir qu'au contraire de ce qui se passe au niveau d'une voie élémentaire de transmission (niveau considéré par Shannon dans son théorème) où, en l'absence de codes correcteurs d'erreurs, le bruit a pour effet inévitable de *réduire* la quantité d'information transmise, au niveau de la voie établie entre le système et l'observateur considérant l'ensemble de ce dernier, l'ambiguïté introduite par des facteurs de bruits peut être *génératrice* d'information. Pour le montrer, commençons par relier de manière plus précise la complexité (ou le degré d'organisation) d'un système à la quantité totale d'information contenue dans un message. Cette quantité mesure, sur un grand nombre de messages écrits dans une langue donnée avec un alphabet donné, la quantité d'information moyenne par symbole de cet alphabet multipliée par le nombre de symboles du message. Or, la quantité moyenne d'information par symbole est celle que mesure la formule de Shannon: $H(x) = - \sum p_i \cdot \log_2 p_i$, où i est l'indice des n différents symboles x_i utilisés dans le message x , et où p_i est la probabilité d'apparition du symbole x_i . Etendons ce formalisme à un système composé de parties. La quantité d'information d'un tel système est alors définie

"à partir des probabilités que l'on peut affecter à chaque sorte de ses composants, sur un ensemble de systèmes supposés statistiquement homogènes les uns aux autres; ou encore à partir de l'ensemble des combinaisons qu'il est possible de réaliser avec ses composants, qui constitue l'ensemble des états possibles du système. Dans tous les cas, la quantité d'information d'un système mesure le degré d'improbabilité que l'assemblage des différents composants soit le résultat du hasard. Plus un système est composé d'un grand nombre d'éléments différents, plus sa quantité d'information est grande car plus grand est l'improbabilité de le constituer *tel qu'il est* en assemblant au hasard ses constituants. C'est pourquoi cette grandeur a pu être proposée comme une mesure de la complexité d'un système, en ce qu'elle est une mesure du degré de variété des éléments qui le constituent." [H. Atlan, 1979, p.45].

Bien qu'une telle mesure soit insatisfaisante, en raison du caractère statique et uniquement structural de la complexité qu'elle envisage en se bornant à l'assemblage des éléments du système en cause alors qu'il conviendrait de prendre également en compte la complexité fonctionnelle et dynamique due aux interactions entre ces éléments, elle met clairement en évidence le lien entre quantité d'information contenue dans un message et dans un système organisé. Appelons S un tel système, et supposons qu'il comporte deux sous-systèmes A et B , reliés par une voie de communication transmettant un flux d'information de A vers B . Pour que S possède réellement le caractère d'ensemble *organisé* propre à un système, il faut que cette transmission d'information soit telle qu'elle ne se corresponde à aucun des cas suivants: 1) une absence totale de contrainte de A sur B , au sens où toute l'information émise par A serait perdue

lors de sa transmission vers B; 2) l'existence d'une contrainte totale de A sur B, au sens où cette transmission s'effectuerait sans aucune perte d'information. En effet, le premier cas serait celui d'une indépendance totale de la structure de B par rapport à celle de A: lors de la transmission de A vers B, l'ambiguïté serait ici égale à la quantité totale d'information de A, et la quantité d'information de l'ensemble A et B s'obtiendrait en additionnant celle de A à celle de B. Dans la mesure où l'existence même du système S dépend de l'établissement effectif d'une relation entre A et B par l'intermédiaire de la voie de communication les reliant, la situation décrite à l'instant impliquerait la destruction de S en tant que système. Dans le second cas, B ne serait qu'une simple copie de A, et la quantité d'information de l'ensemble A et B serait égale à celle de A (ou à celle de B): au lieu du système S, nous n'aurions alors qu'un sous-système répliqué.

En somme, la notion de système organisé n'a de sens que pour les situations intermédiaires entre les deux cas extrêmes que nous venons d'évoquer: celles où la structure de A exerce une réelle contrainte sur celle de B, sans que cette contrainte ne soit totale pour autant. En d'autres termes, la notion d'organisation du système S suppose qu'il y ait transmission d'une quantité d'information non nulle entre A et B (on n'est pas dans le premier cas), et que cette transmission s'effectue avec une ambiguïté non nulle (on n'est pas dans le second cas).

Cette double condition étant satisfaite, on peut montrer que l'ambiguïté joue des rôles diamétralement opposés selon que l'on se place au niveau de la voie élémentaire de transmission de l'information (entre A et B) ou à celui du système considéré dans son ensemble. Il est en effet évident que le théorème shannonien de la voie avec bruit s'applique bien *au niveau de la voie élémentaire* de transmission d'information considérée: l'ambiguïté mesure ici l'incertitude qui subsiste au sujet du sous-système B quand l'on connaît le sous-système A, de sorte qu'elle traduit le manque de contrainte exercé par A sur B - un manque qui serait maximal si A et B étaient parfaitement indépendants, et nul si B n'était qu'une copie strictement conforme de A. En somme, assimilée à l'exercice d'une contrainte d'un sous-système sur un autre par H. Atlan, la transmission d'information de A vers B renvoie au sens philosophique premier du verbe "informer" - du latin *informare* qui signifie "donner une forme, mettre en une forme". Par définition, plus l'incertitude subsistant sur B sachant A est grande (plus l'ambiguïté est élevée), moins est importante la quantité d'information transmise dans la voie: puisque l'ambiguïté joue ce rôle négatif sur la quantité d'information transmise au niveau de la voie élémentaire entre A et B, nous pouvons convenir de l'associer ici à un signe algébrique négatif. C'est exactement l'inverse qui prévaut *au niveau du système considéré dans son ensemble* - à celui de la voie de transmission d'information établie entre le système et l'observateur, et non plus *à l'intérieur même* du système. Car à ce niveau la contrainte exercée par A sur B joue un rôle opposé au précédent: si cette contrainte était totale (ambiguïté nulle), la connaissance de B sachant A n'apporterait aucune information supplémentaire à l'observateur, puisqu'aux yeux de celui-ci le second sous-système ne constituerait qu'une copie à l'identique du premier; et si cette contrainte était au contraire nulle (ambiguïté maximale), la connaissance de B sachant A apporterait à

l'observateur une quantité supplémentaire d'information strictement égale à la quantité d'information contenue dans B. Dans le cas d'un véritable système organisé, où nous imposons à la contrainte de n'être ni totale ni nulle, il subsiste ainsi toujours une certaine ambiguïté sur B quand l'observateur ne connaît que A. Mais cette incertitude restante, mesurée par la quantité d'information de B indépendamment de A, ne représente plus ici de l'information perdue: la quantité d'information transmise à l'observateur considérant successivement A et B est égale à la somme de la quantité d'information de A et de celle que lui apporte l'observation de B *bien qu'il connaisse déjà* A. Loin de traduire une *perte* d'information, la différence entre A et B est *productive* d'information pour l'observateur du système dont elle augmente la variété. Nous devons donc donner ici un signe algébrique positif à l'ambiguïté.

Ce rôle bivalent du bruit, H. Atlan le marque par les concepts d'ambiguïté destructrice (< 0) et d'ambiguïté-autonomie (> 0), et c'est en termes de ces concepts que nous devons interpréter l'expression formelle, donnée plus haut, de la propriété auto-organisatrice de certains systèmes, car chaque élément de la somme figurant le membre de droite de cette expression correspond à l'un des rôles opposés de l'ambiguïté, mis en évidence à l'instant. Rappelons cette expression formelle, dans laquelle l'organisation du système concerné est définie par un caractère structural, mesuré par la quantité H, et par un caractère fonctionnel, exprimé par dH/dt . Nous avons écrit: $dH/dt = - H_{\max} \cdot dR / dt + (1 - R) \cdot dH_{\max} / dt$. Considérons le membre de droite. Son premier terme, $- H_{\max} \cdot dR / dt$, mesure la variation de la redondance du système au cours du temps. Si cette redondance est initialement suffisamment élevée (10), elle diminuera sous l'effet de l'ambiguïté-autonomie: $dR / dt < 0$. Puisque H_{\max} est par définition toujours positif, le produit: $- H_{\max} \cdot dR / dt$, sera positif, ce qui se traduira au cours du temps par un accroissement de H, qui mesure la quantité d'information contenue dans le système. Ainsi, ce premier terme rend compte d'un processus de différenciation des sous-systèmes, ou d'augmentation de la complexité structurale et fonctionnelle du système, en mesurant l'augmentation de variété de ce dernier: agissant à l'intérieur du système, le "bruit" exerce donc un effet positif sur la voie établie entre le système et l'observateur. Quant au second terme du membre de droite dans l'expression ci-dessus, $(1 - R) \cdot dH_{\max} / dt$, il mesure la variation de H_{\max} au cours du temps, c'est-à-dire la variation instantanée par rapport à l'état de complexité maximale que peut atteindre le système sans tenir compte de son degré de redondance. Selon H. Atlan, le processus de désorganisation par rapport à un état donné (instantané) d'organisation, sous l'effet de l'ambiguïté destructrice, s'exprime par une fonction décroissante de H_{\max} de telle sorte que dH_{\max} / dt et le second terme tout entier sont négatifs, contribuant à une diminution de la quantité d'information H du système. En somme, le "bruit" agit ici directement sur la voie de transmission du système à l'observateur en réduisant la quantité d'information véhiculée par elle: l'effet de l'ambiguïté destructrice est donc bien négatif du point de vue de l'observateur.

En définitive, l'organisation fonctionnelle d'un système peut être décrite par le taux de variation dH/dt , obtenu en additionnant deux fonctions du temps, dont l'une exprime le taux de

décroissance de la redondance du système, et l'autre le taux de décroissance de la quantité d'information maximale de ce dernier. Sous certaines conditions de redondance structurale initiale et de fiabilité du système, celui-ci se révèle capable d'intégrer à sa propre organisation tout "bruit" provoqué en son sein par des facteurs aléatoires en provenance de son environnement. La diminution consécutive de la redondance du système, et l'augmentation corrélative de sa complexité, traduisent alors le caractère auto-organisateur de ce système. L'auto-organisation n'est donc pas conçue ici comme un état, mais comme un processus ininterrompu de désorganisation-réorganisation.

Deux paramètres importants interviennent dans les deux fonctions du temps dont résulte par sommation le taux dH / dt qui décrit l'organisation fonctionnelle du système: la redondance structurale initiale R_0 , et la fiabilité, caractère fonctionnel exprimant l'efficacité de l'organisation dans sa résistance aux changements aléatoires. Plus précisément, si la fiabilité du système dépend de la redondance de celui-ci, elle ne s'y réduit pas. Nous avons vu en effet que la redondance initiale d'un système doit avoir une valeur minimale pour que ce dernier présente des propriétés auto-organisatrices consistant en une augmentation de complexité par destruction de redondance. Dans ces conditions, absolument nécessaires pour que la courbe de variation $H(t)$ ait une partie initiale croissante,

"La fiabilité mesurera aussi la *durée* de cette phase ascendante, soit le temps t_m au bout duquel le maximum est atteint, d'autant plus long que la fiabilité est grande." [H. Atlan, 1979, p.52].

On obtient ainsi une courbe représentative de l'évolution de la quantité d'information du système au cours du temps. Jusqu'à une date t_m , l'ambiguïté-autonomie l'emporte sur l'ambiguïté destructrice, de sorte qu'au total $dH / dt > 0$, et c'est l'inverse qui se produit ensuite. Le profil d'une telle courbe est tout à fait typique des organismes vivants: croissance et maturation ($dH/dt > 0$ pour $t < t_m$) puis, après un maximum ($dH / dt = 0$ pour $t = t_m$), vieillissement et mort ($dH / dt < 0$ pour $t > t_m$). On remarquera que ce sont toujours des facteurs *aléatoires* en provenance de l'environnement du système qui nourrissent le développement de celui-ci et le mènent ensuite à la mort. Ces facteurs sont aléatoires car, au moment où ils provoquent au sein du système le "bruit" dont l'effet est successivement organisateur ($t < t_m$) et désorganisateur ($t > t_m$), ils ne correspondent à aucun programme préétabli contenu dans l'environnement et destiné à organiser ou à désorganiser le système. Nous avons affaire ici à un bruit inintentionnel (dont l'occurrence est a priori dépourvue de sens), mais dont l'effet consiste, dans la partie croissante de $H(t)$, à *créer de l'information* en augmentant le degré d'organisation du système par un accroissement du degré de différenciation des sous-systèmes composant ce dernier. Laissé en suspens par la théorie strictement shannonienne de l'information, le problème de la création de celle-ci est donc pensable en termes de l'élargissement de cette théorie que constitue le paradigme auto-organisationnel atlanien.

En effet, la possibilité d'un accroissement de la quantité d'information transmise du système à l'observateur, sous l'effet d'un bruit créateur d'ambiguïté, s'inscrit dans le prolongement direct du formalisme shannonien: sans contredire le théorème de la voie avec bruit, ce prolongement voit le changement de signe de la fonction ambiguïté résulter d'un simple *changement de point de vue* sur le rôle joué par l'ambiguïté vis-à-vis du système. De sorte que ce changement de signe possède deux implications tout à fait fondamentales pour la théorie de l'information. La première tient à *la prise en compte explicite de l'observateur* dans la définition même des systèmes non consciemment et délibérément construits par l'homme en tant que sujet individuel - tels les systèmes sociaux, économiques, symboliques, etc.: c'est bien en considérant alternativement le système de l'intérieur (du point de vue de la transmission entre les sous-systèmes A et B) et de l'extérieur (du point de vue de la transmission entre le système S et l'observateur), que nous avons pu opposer une ambiguïté destructrice (premier point de vue) à une ambiguïté-autonomie (second point de vue). La seconde implication donne une interprétation de la première, propre à en étendre la portée: elle consiste à considérer ces points de vue différents comme des *niveaux d'organisation différents* à l'intérieur de tout système intégré en niveaux hiérarchisés:

"L'introduction de la position de l'observateur ne constitue pas *seulement* une étape logique du raisonnement: *cet observateur, extérieur au système, c'est en fait, dans un système hiérarchisé, le niveau d'organisation supérieur* (englobant) par rapport aux systèmes éléments qui le constituent; c'est l'organe par rapport à la cellule, l'organisme par rapport à l'organe, etc. C'est par rapport à lui que les effets du bruit sur une voie à l'intérieur du système peuvent, dans certaines conditions, être positifs. Autrement dit: pour la cellule qui regarde les voies de la communication qui la constituent, le bruit est négatif. Mais pour l'organe qui regarde la cellule, le bruit dans les voies à l'intérieur de la cellule est positif (tant qu'il ne tue pas la cellule) en ce qu'il augmente le degré de variété, donc les performances régulatrices de ses cellules." [H. Atlan, 1979, p. 70].

Plus généralement, il faut donc s'attendre que l'ambiguïté associée à un bruit provoqué par un facteur aléatoire, provenant de l'environnement d'un système auto-organisateur que ce facteur vient perturber, exerce des effets diamétralement opposés sur un niveau hiérarchique donné d'un tel système, et sur le niveau immédiatement englobant comme sur le niveau englobé de ce niveau hiérarchique donné. Nous débouchons ainsi sur la question cruciale de la *signification* de l'information dans un système hiérarchisé, signification dont les développements précédents suggèrent qu'elle change suivant le niveau hiérarchique considéré à l'intérieur du système.

2.2 Signification de l'information dans un système hiérarchisé

Afin de mettre progressivement en place les éléments permettant de comprendre une telle signification, nous procéderons ici en trois temps: la différence entre le principe d'ordre par le bruit de Von Foerster et celui du bruit organisationnel de H. Atlan (2.2.1) nous permettra d'établir une distinction claire entre le concept de complication et celui de complexité (2.2.2), qui débouchera ensuite sur la notion de signification de l'information dans un système

hiérarchisé (2.2.3).

2.2.1 Ordre par le bruit *versus* bruit organisationnel

H. Von Foerster [1960] propose le modèle qualitatif d'un principe d'ordre par le bruit, dans lequel l'ordre d'un système est considéré comme une répétition (ou une redondance) qui augmente suite à la perturbation de ce système par un "bruit". Ce modèle est celui de cubes aimantés dont la disposition initiale au sein d'un récipient est celle d'un amas informe, et qui sont agités de manière aléatoire en même temps que le récipient les contenant. Au fur et à mesure d'une telle agitation, en laquelle consiste le bruit évoqué plus haut, ces cubes se disposent de manière de plus en plus "ordonnée" du fait de l'aimantation de leurs faces qui s'attirent. On remarque immédiatement que *ce processus suppose connu d'emblée le mécanisme de construction des formes*: l'observateur sait dès le départ de l'expérience que les cubes vont s'accoupler selon les forces d'attraction dues à l'aimantation. En ce sens, le bruit a pour fonction et pour effet de permettre aux contraintes potentielles, constituées par ces forces d'attraction, de se réaliser: le système finalement construit correspond exactement à la connaissance que l'on avait *a priori* des mécanismes de construction. Celle-ci s'effectue de telle sorte que c'est bien la redondance du système qui augmente du fait du bruit: par rapport à l'amas informe de départ, au sein duquel la position d'un cube donné ne peut, par définition, être déduite logiquement de celles que l'on connaît éventuellement s'agissant d'autres cubes, l'état du système évolue de manière de plus en plus prévisible pour l'observateur. Du point de vue de ce dernier, il est en effet de plus en plus facile de prévoir la position d'un cube donné à partir de la connaissance qu'il a de celles d'autres cubes.

Le principe du bruit organisationnel introduit par H. Atlan est inverse : l'un de ses intérêts essentiels réside précisément dans le fait qu'ici le bruit, loin d'augmenter la redondance, *diminue* celle-ci par augmentation de complexité. *Si l'on ne connaît pas le mécanisme de construction des formes*, comme c'est le cas dans les systèmes non construits, où l'apprentissage se fait sans programme préalable, la forme qui apparaît progressivement est plus complexe que l'amas de départ: dans celui-ci, les "morceaux" étaient parfaitement interchangeables sans modification de la forme globale, de sorte que le nombre d'éléments différents qu'il était nécessaire de spécifier pour reconstruire un amas statistiquement identique était relativement réduit par rapport au cas d'une forme géométrique déterminée, où chaque "morceau" occupe une place non quelconque - où les portions d'espace ne sont pas interchangeables quant à leurs probabilités respectives d'être vides ou occupées par tel "morceau". La redondance du système a diminué, et, corrélativement, sa complexité a augmenté, sous l'effet du bruit organisationnel.

Entre le principe d'ordre par le bruit de H. Von Foerster, et celui du bruit organisationnel de H. Atlan, la différence est donc patente: le premier s'applique à des systèmes évoluant de

manière de plus en plus prévisible aux yeux de l'observateur, dont la connaissance a priori des règles globales d'évolution de ces systèmes lui permet de prévoir de mieux en mieux l'évolution des sous-systèmes correspondants; le second envisage au contraire des systèmes dont la loi globale d'évolution n'est pas connue de l'observateur, de sorte qu'il est de plus en plus difficile pour lui de prévoir l'évolution des sous-systèmes correspondants.

2.2.2 Complexité et complication

Cette différence entre ordre par le bruit et bruit organisationnel débouche immédiatement sur une autre différence, fondamentale pour notre propos: celle qui distingue la complication de la complexité. A la limite, la complication d'un système n'exprime que le nombre d'étapes nécessaires pour décrire entièrement celui-ci à partir de ses constituants. C'est donc un attribut de systèmes construits, ou constructibles, par l'homme qui en connaît et comprend totalement la structure et le fonctionnement. En ce sens, s'il contient un nombre important de cubes, le système illustrant le principe d'ordre par le bruit de H. Von Foerster peut être dit compliqué. Par contre, selon H. Atlan ce même système ne peut pas être considéré comme étant complexe, car *"la complexité est reconnue comme une notion négative: elle exprime qu'on ne connaît pas, ou qu'on ne comprend pas un système, malgré un fond de connaissance global qui nous fait reconnaître et nommer ce système. Un système qu'on peut spécifier explicitement, dont on connaît la structure détaillée, n'est pas vraiment complexe. Disons qu'il peut être plus ou moins compliqué. La complexité implique qu'on en ait une perception globale, avec en même temps la perception qu'on ne la maîtrise pas dans ses détails. C'est pourquoi on la mesure par l'information qu'on ne possède pas et dont on aurait besoin pour spécifier le système en ses détails."* [1979, p. 76].

Avec H. Atlan, nous retrouvons ici la conception de l'information de L. Brillouin [1959] interprétant la quantité d'information définie par la formule de Shannon comme une mesure de l'information *qui manque* à l'observateur pour spécifier explicitement un système. A la limite, aussi compliqué soit-il, un système parfaitement connu peut être réduit à un seul élément: son programme de construction.

Dans ce cas, pour H. Atlan comme pour L. Brillouin, $H = 0$ (et $R = 1$): une complexité ordonnée, sous forme d'un programme de construction supposé connu, n'est plus une complexité mais seulement une complication. Réciproquement, un désordre n'est pas forcément synonyme de complexité, ou n'est complexe que par rapport à un ordre supposé exister et que l'on cherche. Bref, selon H. Atlan

"la complexité est un désordre apparent où l'on a des raisons de supposer un ordre caché; ou encore, la complexité est un ordre dont on ne connaît pas le code." [1979, p. 78].

Illustrons cette notion en donnant rapidement trois mesures de complexité d'un système - mesures qui diffèrent selon le degré de connaissance que l'on a de ce dernier. La première est celle de la complexité d'un système composé d'un grand nombre d'éléments et dont on ne connaît pas le mécanisme de construction. C'est en somme la situation canonique de complexité -

celle où cette dernière est la plus élevée. Dans cette situation, la complexité est mesurée par la variété du système, au sens d' Ashby : c'est un cas particulier de la fonction H de Shannon, celui où il existe n éléments différents, dont on ne prend pas réellement en compte les probabilités respectives car on ne les connaît pas. Plus précisément, l'on exprime une telle ignorance en considérant ces éléments comme équiprobables: $\forall i, \forall j, p_i = p_j = 1/n$. En partant de la formule de Shannon: $H = - \sum p_i \log_2 p_i$, il vient que $H = \log_2 n = H_1$. Imaginons à présent que l'on connaisse la distribution de probabilités des éléments du système, avec des p_i quelconques. On retrouve alors tout simplement la formule de Shannon: $H = - \sum p_i \log_2 p_i = H_2$, et il est clair que $H_2 < H_1$. En effet, l'information supplémentaire consistant en la connaissance de la distribution de probabilités des p_i , dont l'on ne disposait pas dans le cas précédent, correspond à notre meilleure connaissance du code du système envisagé. Elle implique donc une réduction de la complexité de ce dernier. En fait, nous sommes ici en face d'une redondance nulle (les n éléments du système sont parfaitement indépendants), c'est-à-dire que nous retrouvons ce que nous avons appelé H_{\max} plus haut : $H_2 = H_{\max}$. Enfin, il se peut que nous ayons connaissance de contraintes internes au système, établies entre les éléments de ce dernier. Par rapport au cas précédent, la complexité est encore réduite, du fait de la redondance qui mesure ces contraintes. Cette complexité est ici mesurée par $H_3 = H_{\max} (1 - R) = H_2 (1 - R)$. Finalement $H_3 < H_2 < H_1$, où H_1 correspond au maximum de H_2 (du fait de l'équiprobabilité de la distribution), alors que H_2 correspond au maximum de H_3 (du fait de l'indépendance des éléments), et que H_1 exprime un maximum maximorum d'ignorance ou de complexité.

Nous pouvons ainsi concevoir une première approche de la signification de l'information. De manière générale, nous venons de voir que l'information était mesurée par la formule de Shannon, au sens de quantité d'information qui manquait à l'observateur du système pour pouvoir reconstruire à l'identique celui-ci, dont le code (le principe de construction) était imparfaitement connu de cet observateur. Une telle mesure dépend donc de manière critique du choix de ce qui est considéré comme étant les éléments constitutifs du système. Par exemple, s'agissant d'un système biologique, cette quantité variera selon que l'on situe l'observation au niveau des particules élémentaires, des atomes, des molécules, des macromolécules, des organelles, des cellules, des organes, ou de l'organisme tout entier, considéré comme système englobant les sous-systèmes précédents. Selon H. Atlan, c'est précisément en ces termes qu'il nous faut poser le problème de la signification de l'information. En effet :

“c'est justement là que se loge la signification (pour le système) de l'information (pour l'observateur, c'est-à-dire l'information qu'il n'a pas), alors même que celle-ci est mesurée de façon probabiliste, qui ignore la signification: cela vient de ce qu'il s'agit, encore, d'une information qu'on n'a pas sur le système. Le choix, au départ, est laissé à la décision de l'observateur.” (1979, p. 75).

H. Atlan donne un exemple très simple, qui permet de mieux comprendre ce point: celui de l'observateur d'un système biologique, système que cet observateur choisirait de décrire, soit à partir de ses atomes constitutifs, soit à partir de ses molécules. Dans le premier cas,

“H mesurera l'information supplémentaire nécessaire quand on ne connaît que le type d'atomes rencontrés dans un

ensemble statistiquement homogène de systèmes identiques et leur fréquence dans cet ensemble. Cette information nécessaire est évidemment très grande par rapport à celle dont on aurait besoin si on décrivait le système à partir de ses molécules. C'est que dans ce cas on utiliserait une information supplémentaire qu'on possède déjà sur la façon dont les atomes sont associés en molécules. C'est autant de complexité qui disparaît par rapport au cas précédent." [Ibid.].

Il apparaît clairement ici que la signification de l'information est indissociable d'une conception d'un système hiérarchisé en niveaux d'organisation dont chacun est défini comme une totalité: par opposition à un agrégat, construit à partir d'éléments dont les significations propres précèdent ainsi celle de l'ensemble qu'ils constituent, dans une totalité la signification de chaque élément est strictement déterminée par celle de l'ensemble qui leur est logiquement antécédente. Bien que le terme de totalité n'apparaisse pas dans le texte de H. Atlan, probablement parce que l'on reconnaît généralement cette notion comme faisant partie intégrante de celle de système, c'est bien une notion de totalité qui transparaît dans la citation ci-dessus: se situer au niveau de la molécule pour décrire le système, c'est *ipso facto* considérer connu l'arrangement des atomes constituant celle-ci. Plus généralement, comme l'ont montré les trois mesures de complexité données précédemment:

"Plus on connaît la façon dont les éléments sont assemblés pour construire le système - plus on possède de véritable information, avec sa signification, sur le système - plus la fonction H diminue." (Ibid.).

Il est à présent facile de comprendre les liaisons conceptuelles existant entre les notions de redondance, complexité et codage dans le cadre de systèmes non construits par l'homme. Parfaitement duale de la notion de complexité, celle de redondance traduit en termes d'information l'effet sur l'observateur de l'existence de contraintes internes à un tel système: du fait de ces contraintes, établies entre ces sous-systèmes que figurent les éléments du système considéré, la connaissance qu'a l'observateur de l'un de ces éléments influe sur celle qu'il peut avoir des autres. Il s'ensuit que

"la redondance est une mesure de la simplicité et de l'"ordre". Ainsi l'ordre serait plutôt répétitif ou redondant. Il n'est pas nécessaire qu'il soit physiquement répétitif, comme dans un cristal, au sens d'un élément ou motif unique répété un grand nombre de fois. Il suffit qu'il soit redondant, c'est-à-dire déductivement répétitif: la connaissance d'un élément nous apporte une certaine information sur les autres (en diminuant l'incertitude à leur sujet) et c'est cela qui nous fait percevoir un ordre." (Ibid., p. 79).

A priori, la relation conceptuelle entre codage et complexité peut paraître plus déroutante que la précédente: il semble en effet paradoxal que la fonction H exprime la quantité d'information transmise dans une voie établie entre le système et l'observateur, alors même qu'elle mesure la quantité d'information que cet observateur ne possède pas - puisque c'est précisément celle dont il aurait besoin pour spécifier le système! En fait, ce paradoxe n'est qu'apparent : pour le dissoudre, il suffit de souligner à nouveau que ce que transmet le système à l'observateur est toujours un *déficit* d'information. Ajoutons que la quantification de ce qui est

ainsi transmis ne prend pas directement en compte le codage (à l'entrée de la voie) ni le décodage (à la sortie) du sens des messages. L'information mesurée est celle qui est transmise du système à l'observateur dans une optique strictement shannonienne, c'est-à-dire à l'intérieur d'une voie : dans le calcul lui-même de la quantité H, les bornes de la voie de transmission sont pour ainsi dire ignorées - de manière formellement analogue à un intervalle *ouvert* en mathématiques. C'est une information *sans code* qui est transmise à l'occasion de l'observation du système, ce qui n'implique évidemment pas que la notion de signification de l'information soit absente du paradigme de l'auto-organisation à la Atlan - bien au contraire, comme le montre la pleine explicitation de cette notion dans le cadre d'un système hiérarchisé.

2.2.3 Signification de l'information dans un système hiérarchisé

De manière tout à fait générale, la signification d'une information donnée peut être définie "comme l'effet de la réception de cette information par son destinataire. Cet effet peut apparaître soit sous la forme d'un changement d'état, soit sous celle d'un 'output' de ce destinataire lui-même envisagé comme un sous-système." [H. Atlan, 1979, p. 86].

Il est donc parfaitement évident qu'un événement donné, véhiculant une certaine information pour qui perçoit cet événement et fait ainsi figure de destinataire de l'information associée, revêtira des significations différentes pour des destinataires différents. En particulier, l'événement consistant en la survenue d'un "bruit" venant perturber un système auto-organisateur présentera des significations différentes selon les différents niveaux hiérarchiques auxquels l'on se placera lors de l'observation de ce système: un bruit organisationnel donné aura une signification différente selon que l'on considère les voies élémentaires constituant le système lui-même, ou le niveau englobant auquel ce système est considéré comme une totalité par l'observateur. Au niveau des voies élémentaires constituant le système lui-même, la connotation de ce bruit organisationnel est clairement positive, puisque le système utilise ce bruit pour se reconstruire de façon nouvelle - en augmentant sa complexité par diminution de redondance: bien que le bruit soit *a priori* aléatoire, au sens où sa survenue en provenance de l'environnement ne correspond ni à un programme préétabli dans celui-ci et destiné à organiser ou désorganiser le système, ni n'est prédéterminée par l'état actuel du système, *a posteriori* la notion d'aléa perd de sa pertinence en ce qui concerne ce dernier. En fait, le système est précisément auto-organisateur parce qu'il est capable d'intégrer ce bruit à son organisation propre, et dès lors qu'il est ainsi intégré, ce bruit n'a plus rien d'aléatoire: sa survenue signifie au contraire que le système possède désormais une meilleure information sur lui-même, ce qui lui permet éventuellement de mieux fonctionner au sein de son environnement:

"l'information qu'un système aurait sur lui-même, celle dont nous avons vu qu'elle peut augmenter sous l'effet de ce qui nous apparaît comme du bruit (et qu'on mesure alors par une information qui nous manque), c'est bien ce qui permet au système de fonctionner, et même d'exister en tant que système. Il s'agit donc de l'ensemble des *effets* structuraux et fonctionnels, de la réception de l'information transmise dans le système, sur les différents sous-systèmes et différents niveaux d'organisation du système. Il s'agit bien de la *signification* de cette

information pour le système.” [H. Atlan, 1979, p. 87].

En somme, bien que le système soit dépourvu de tout plan préalable, tout se passe comme s'il faisait son miel d'un bruit *a priori* aléatoire qui lui permet de faire l'apprentissage de lui-même en voyant se différencier davantage les sous-systèmes le composant. C'est pourquoi la signification de ce bruit doit être considérée comme étant positive pour le système lui-même. Il en va différemment au niveau englobant de l'observateur. Pour ce dernier, le bruit organisationnel est le résultat d'un hasard - au sens de A. Cournot, c'est-à-dire qu'il résulte de la rencontre de deux séries causales indépendantes *a priori*. La survenue de ce bruit ayant produit ses effets sur le système, ceux-ci sont tels que la connaissance qu'a l'observateur au sujet de ce système ne s'en trouve nullement accrue. Au contraire, cette connaissance a diminué: l'augmentation de complexité consécutive à la survenue du bruit est synonyme d'accroissement du déficit d'information de l'observateur concernant le système, car pour pouvoir spécifier ce dernier, il lui manque désormais davantage d'information qu'il ne lui en manquait précédemment. La connotation du bruit est considérée par lui comme étant clairement négative. La signification de l'information pour l'observateur est donc à l'opposé de celle que cette information revêt pour le système lui-même.

Plus généralement, une variation donnée de l'expression quantitative de l'information par la fonction H de Shannon peut donc présenter des sens radicalement opposés selon le niveau auquel on situe l'observation d'un système hiérarchisé - ici, en deux niveaux seulement: cette quantité étant toujours saisie comme un déficit, ce qui est destruction d'information à un niveau élémentaire (ou englobé) peut parfaitement apparaître comme une création d'information à un niveau englobant. Le principe de complexité par le bruit, qui se trouve au coeur du concept d'auto-organisation proposé par H. Atlan, représente donc une façon détournée d'introduire des effets de sens dans une théorie quantitative de l'organisation:

“Bien sûr, le sens n'est là que de façon négative, comme son ombre, puisqu'il n'y est théorisé qu'à travers les effets du bruit, c'est-à-dire d'une négation de l'information. Mais il est quand même là parce qu'il s'agit de la *négation d'une négation*, puisque tout se passe comme si l'information shannonienne niait le sens, autre façon de dire qu'elle mesure ce que nous ne comprenons pas du système.” [Ibid., p. 88].

Appliqué au cas le plus simple où n'existent que deux niveaux à l'intérieur d'un système hiérarchisé - celui des voies élémentaires entre les n éléments du système et celui de l'observateur -, le modèle de l'auto-organisation proposé par H. Atlan peut en définitive être résumé de la manière suivante. A la date initiale t_0 , l'état du système est caractérisé par une redondance R_0 , et par la complexité correspondante H_0 mesurée par la formule de Shannon, diversement interprétée suivant que l'observateur connaisse, ou non, la distribution de probabilités des éléments du système et/ou l'existence de contraintes initiales exercées par certains éléments sur d'autres. A cette même date, des facteurs aléatoires en provenance de l'environnement du système provoquent un bruit à l'intérieur de celui-ci. Ce bruit a pour effet de modifier le

système, de telle sorte qu'à la date suivante, $t_1 = t_0 + dt$, corresponde un nouvel état de celui-ci. Si le système est auto-organisateur et fiable au sens de Atlan (si R_0 est suffisant et si $t < t_m$), ce nouvel état sera tel que $R_1 < R_0$ et, corrélativement, $H_1 > H_0$. Mesurant l'excès net de la valeur de l'ambiguïté-autonomie sur la valeur absolue de l'ambiguïté destructrice, cet accroissement de complexité présente une signification positive pour le système, qui possède désormais une meilleure connaissance de lui-même, et une signification négative pour l'observateur de ce système, dont le déficit d'information relativement à ce dernier a augmenté.

L'idée générale de ce modèle est donc que, sous certaines conditions, un facteur *qualitatif*, tel le bruit organisationnel, induit une modification à la fois qualitative et quantitative dans la représentation d'un système non construit par l'observateur : cette modification est qualitative, puisqu'elle se traduit par les effets de sens rappelés à l'instant, mais elle est également quantitative, puisque ces effets de sens s'appuient sur une mesure exprimée dans le formalisme shannonien. Tout en conservant un tel formalisme, avec les avantages liés à l'expression quantitative d'une théorie quelle qu'elle soit, ce modèle ouvre à un traitement analytique de la *création* et de la *signification* de l'information au sein d'un système *hiérarchisé*. En permettant d'envisager ces trois notions solidaires que C. Shannon laissait en dehors de sa problématique, le modèle d'auto-organisation de H. Atlan constitue bien un dépassement de la théorie shannonienne de l'information. Mais ce modèle présente en outre un avantage non négligeable par rapport à d'autres modèles de systèmes dynamiques, tel celui du principe d'ordre par fluctuation de I. Prigogine: contrairement à ce dernier, ce n'est pas un modèle déterministe, mais un principe de représentation probabiliste de systèmes mal connus. En ce sens, il paraît bien adapté à la représentation des systèmes sociaux, économiques et symboliques qui, tous, sont historiques, et dont la structure et le fonctionnement nous restent inéluctablement imparfaitement connus. Pour mieux apprécier la fécondité de la conception atlanienne de l'auto-organisation, il convient à présent de la confronter à la théorie rosenfieldienne de la mémoire.

3. Vers un modèle d'apprentissage non dirigé

Le recensement des implications communes à la conception rosenfieldienne de la mémoire et au paradigme auto-organisationnel de H. Atlan permet d'esquisser une problématique de l'apprentissage prenant explicitement en compte la double dimension de l'incertitude et de l'irréversibilité - double dimension dont est inéluctablement empreint l'écoulement du temps historique. Ces implications communes sont au nombre de trois.

Le premier point d'accord entre les conceptions de I. Rosenfield et de H. Atlan consiste en leur commun rejet de toute notion d'apprentissage faisant appel à l'existence d'un programme préétabli au sein des entités constituant leurs objets d'étude respectifs. En se proposant d'analyser certains fonctionnements de systèmes non construits (naturels ou sociaux) à l'aide de son modèle d'auto-organisation, H. Atlan produit un modèle d'apprentissage non dirigé, au sens

le plus fort de cette expression: non seulement il écarte d'emblée tout finalisme, mais il construit explicitement son modèle *contre* le concept de téléonomie introduit par J. Monod afin de contrer un tel finalisme - vaine tentative, selon H. Atlan. Etrangère, plus encore qu'opposée, à toute notion de finalité, de plan ou de réalisation d'un programme préalable en ce qui concerne le fonctionnement des systèmes non construits, la contribution de H. Atlan se présente donc, fût-ce implicitement, comme étant parfaitement incompatible avec toute conception qui assimilerait le fonctionnement de notre cerveau - ce cas particulier de système non construit - à celui des ordinateurs traditionnels.

C'est évidemment sur ce point précis que H. Atlan rejoint - ou plutôt, annonce - la conception rosenfieldienne de la mémoire. Car nous avons vu l'ensemble de I. Rosenfield [1989] dresser un vibrant réquisitoire à l'encontre de la théorie de la mémoire associée à celle de la localisation fonctionnelle: loin de fonctionner sur le mode du traitement de données effectué par un ordinateur classique (PDP compris), qui recherche une information précédemment enregistrée, classée et stockée en permanence en un endroit bien précis de sa mémoire, pour la comparer à une information actuellement offerte à sa perception selon une liste d'instructions préalablement fournies par un programmeur, notre mémoire est inventive. Sans avoir reçu d'instructions préalables pour cela, elle combine les traits structuraux des stimuli qu'elle reconnaît en même temps qu'elle les perçoit, et classe simultanément ces derniers selon des arrangements et des procédures éminemment évolutives. Personne n'ayant instillé dans tel cerveau particulier les instructions auxquelles obéirait cette reconstruction ininterrompue du souvenir, ce cerveau peut être donné pour fonctionnant sans programme préétabli.

Reste que la théorie rosenfieldienne de la mémoire s'appuie largement sur le "darwinisme neuronal" de G. Edelman, et semble ainsi réintroduire au niveau de l'espèce, ou du genre, le finalisme qu'elle prétend avoir évacué à celui de l'individu : si l'on ne peut créditer aucune mémoire *particulière* d'un comportement visant à la réalisation précise d'un programme explicite, le fonctionnement de *toute* mémoire aurait pour objectif ultime la meilleure adaptation possible de l'individu concerné à son environnement. En somme, cette théorie ne serait pas étrangère à toute idée de finalisme, mais seulement à celle qui se trouve associée à la notion de programme d'ordinateur, et elle ne rejeterait explicitement cette dernière que pour mieux accueillir la finalité adaptatrice apparemment sous-jacente à la théorie darwinienne de l'évolution. Sans rouvrir ici un vieux débat, remarquons que cet ajustement dynamique de l'individu au milieu renvoie à la notion d'homéostasie d'un système. Or,

"depuis l'époque de Claude Bernard, il existe deux définitions de l'homéostasie (...): 1° l'homéostasie comme *fin*, ou état, plus précisément l'existence d'une certaine constance en dépit des changements (externes); 2° l'homéostasie comme *moyen*, c'est-à-dire les mécanismes de rétroaction négative qui servent à atténuer les répercussions d'un changement. L'ambiguïté de ce double usage, et par suite les applications très étendues de ce terme, souvent d'ailleurs bien vagues, ont obscurci sa commodité d'analogie précise et de principe explicatif. Il vaut mieux actuellement parler de *l'état stable* ou de la *constance* d'un système, état qui est en général maintenu grâce à des mécanismes de *rétroaction négative*." [P. Watzlawick, J. Helmick-Beavin et D. Jackson, 1972, p. 144].

De manière analogue, l'adaptation de l'individu à son environnement exprime aussi bien la finalité expresse de cet individu lors du fonctionnement de sa mémoire, que le caractère adéquat des moyens mis en oeuvre lors de ce même fonctionnement. Par-delà les inconvénients entraînés par une telle polysémie, cette dernière est donc en elle-même révélatrice du fait que la finalité présumée peut parfaitement se transformer en une pure et simple condition d'existence - au sens où être adapté veut simplement dire être vivant. Car nier une telle équivalence déboucherait immédiatement sur une circularité de raisonnement, clairement épinglée par C. Castoriadis :

' nous ne pouvons pas dire que le fonctionnement global du vivant vise la conservation de quelque chose de définissable. On ne peut pas affirmer qu'il vise la conservation de l'individu, car cela serait circulaire (le fonctionnement de l'individu vivant viserait la conservation de l'individu en tant qu'individu vivant, évidemment (...)) bref: on invoque la conservation, en négligeant le fait que cette conservation, si elle est quelque chose, est conservation d'un état qui ne serait définissable que par renvoi à la conservation.' [1978, p. 182].

Les développements précédents montrent que le modèle rosenfieldien de la mémoire ne perd rien de son sens ni de sa portée en faisant l'économie de l'hypothèse de comportement consistant en une finalité adaptatrice que I. Rosenfield a cru bon d'emprunter à G. Edelman. En définitive, ce modèle décrit tout simplement le fonctionnement *normal* de la mémoire chez un être vivant. La perspective de l'apprentissage associée à la conception de la mémoire défendue par I. Rosenfield se révèle donc aussi libre de toute connotation finaliste que celle du modèle d'auto-organisation proposé par H. Atlan: ni l'un ni l'autre ne fondent l'apprentissage sur une notion de programme préalable, et c'est là le premier point commun existant entre leurs théories respectives.

Ce point débouche immédiatement sur le suivant: s'ils refusent de considérer l'apprentissage en termes de programme préétabli à l'intérieur des systèmes qu'ils étudient, c'est parce que I. Rosenfield et H. Atlan s'accordent à attribuer un poids considérable au rôle de l'environnement de ces systèmes dans l'évolution de ceux-ci. S'agissant du processus de reconnaissance-perception que I. Rosenfield affirme être à l'oeuvre lors du fonctionnement de notre mémoire, nous avons vu toute l'importance du contexte actuel dans la remémoration du passé. Cette remémoration en acquiert inévitablement un aspect de reconstruction - pour ne pas dire d'invention, ce mot employé de manière un peu provocante par I. Rosenfield nous paraissant trop fort pour le contenu qu'il désigne: il n'y a tout simplement pas de remémoration possible sans contexte approprié, et un tel contexte imprime nécessairement sa marque au souvenir actuellement produit. Dans le sillage de la théorie edelmanienne, la remémoration rosenfieldienne consiste donc en une constante réévaluation du passé à la lumière du contexte présent. Or, que représente pour nous un tel contexte, sinon la partie pertinente de notre environnement lorsque nous sommes en situation de remémoration? Telle est bien la signification large qu'il convient d'associer à la notion de contexte si volontiers mise en avant par I. Rosenfield: celle qui rassemble les voies multiples - émotionnelles comme purement sensorielles - par lesquelles nous touche le milieu dans lequel nous sommes insérés au moment précis de notre remémoration.

C'est une logique analogue à celle ainsi décrite par I. Rosenfield à propos du fonctionnement de notre cerveau qui sous-tend le phénomène d'auto-organisation, analysé par H. Atlan dans le cadre des systèmes non construits en général - et non dans celui du seul cerveau, animal ou humain. En effet, c'est bien parce qu'un système de ce genre se révèle capable d'intégrer à son organisation propre le bruit, produit en son sein par des facteurs aléatoires en provenance de l'environnement, qu'il est auto-organisateur. Et pour un tel système, être doté de cette propriété auto-organisatrice ne signifie rien d'autre que pouvoir faire l'apprentissage de lui-même en voyant augmenter son degré de complexité dans la mesure exacte où diminue son degré de redondance. Mais réciproquement, sans bruit organisationnel, pas d'ambiguïté-autonomie ni de complexité accrue : pas d'apprentissage donc, sans environnement source de facteurs de bruit. Remarquons en outre qu'ici, comme dans la conception de I. Rosenfield, l'environnement n'est pas davantage programmé que le système pour faire réagir celui-ci - sous forme d'une remémoration ou d'un accroissement de complexité: de même que I. Rosenfield ne cesse de souligner que les stimuli en provenance de notre environnement ne sont pas d'emblée organisés - les cartes cérébrales ont au contraire pour tâche d'organiser ces stimuli en les catégorisant -, H. Atlan souligne le caractère aléatoire du bruit organisationnel, dont la survenue ne correspond pas à un programme préétabli dans l'environnement et destiné à organiser ou désorganiser le système. En définitive, pour ces deux auteurs l'apprentissage effectué par un système non construit, en particulier celui que réalise notre cerveau lors de la remémoration, résulte toujours d'une interaction non programmée entre un tel système et son environnement. C'est là le deuxième point commun entre leurs théories respectives.

L'une des implications majeures de cette position partagée à propos des conditions d'apprentissage de systèmes non construits (tel le cerveau) est d'ouvrir la possibilité d'analyser le phénomène de l'innovation. En effet, I. Rosenfield et H. Atlan font preuve tous deux d'une particulière attention aux problèmes soulevés par l'apparition de sens nouveaux dans leurs objets respectifs d'investigation. Pour le premier, un tel souci répond à une nécessité évidente: comme il l'affirme explicitement lui-même à plusieurs reprises, une théorie pertinente de la mémoire doit être capable d'expliquer notre adaptation dynamique à un monde sans cesse changeant. Une telle théorie est donc nécessairement celle d'un système évolutif, dont l'apprentissage se réalise par une permanente intégration à son mode normal de fonctionnement des aspects inattendus sous lesquels se présente à lui l'environnement. C'est bien pourquoi I. Rosenfield se déclare tellement séduit par le darwinisme neuronal de G. Edelman: en niant la réalité d'un déterminisme génétique strict dans l'embryogenèse du cerveau comme dans son fonctionnement ultérieur, le darwinisme neuronal installe *ipso facto* dans notre conception de la mémoire le degré de labilité dont celle-ci doit être pourvue afin d'être en mesure de figurer cet instrument au service de notre survie que I. Rosenfield voit en elle. Dans la perspective de H. Atlan, la même préoccupation fondamentale (expliquer l'apparition du nouveau) se présente sous un jour légèrement différent: traduite en

termes beaucoup plus abstraits, elle semble davantage destinée à répondre explicitement à un objectif proprement théorique qu'à une nécessité affichée de réalisme du discours.

Non que le concept d'auto-organisation de H. Atlan soit dénué de tout souci de réalisme - bien au contraire: avec un tel concept, cet auteur se propose bien d'analyser le fonctionnement de systèmes *réellement* non construits. Mais enfin, ce concept apparaît surtout voué à résoudre, dans la veine du formalisme développé par C. Shannon qui les a laissés en suspens,

“trois sortes de problèmes qu'on ne peut pourtant que difficilement éviter dès qu'on parle d'information . 1, Ceux qui sont liés à la création d'information: le deuxième théorème de Shannon, de la voie avec bruit, énonce explicitement que l'information transmise dans une voie ne peut pas se créer, puisqu'elle ne peut qu'être détruite par les effets du bruit, et au mieux, se conserver. 2, Ceux qui sont liés à la signification de l'information: la formule de Shannon ne permet de quantifier l'information moyenne par symbole d'un message qu'à la condition de négliger le sens éventuel de ce message. 3, Enfin ceux qui sont liés aux formes hiérarchiques d'organisation: dans la mesure où la formule de Shannon a pu servir de mesure d'organisation, elle ignore totalement les problèmes d'emboîtements de différents niveaux d'organisation, plus ou moins intégrés, les uns dans les autres.” (H. Atlan, 1979, p. 65).

Nous avons montré de manière suffisamment détaillée les solutions apportées par H. Atlan à ces trois problèmes liés pour qu'il soit inutile d'y revenir ici. Cependant, il est remarquable que I. Rosenfield, aussi concerné par les problèmes de création de sens et de signification de l'information que H. Atlan, bien qu'il ne se pose pas ces questions en termes de système hiérarchisé et malgré la légère différence d'approche mentionnée à l'instant, y réponde en invoquant sensiblement le même type de logique que ce dernier: dans la théorie rosenfeldienne de la mémoire comme dans celle de l'auto-organisation de H. Atlan, la croissance de l'information chez le sujet s'accompagne toujours de différenciation ou de complexification - jamais d'homogénéisation.

Sur ce point précis, il nous serait difficile d'être plus explicite que H. Atlan lui-même dans le passage suivant:

“Lorsqu'il s'agit d'apprentissage non dirigé, deux propriétés, conséquences du principe de complexité par le bruit, peuvent être reconnues La première est que le processus d'apprentissage peut être compris comme une création de patterns par diminution de redondance, où des spécifications de patterns bien particuliers en excluent d'autres. Ainsi, à la question: qu'est-ce qui augmente et qu'est-ce qui diminue dans l'apprentissage?, on peut répondre, d'après ce principe, que ce qui augmente, c'est la différenciation, la spécificité des patterns appris, et ceci implique donc une augmentation de la variété, de l'hétérogénéité; au contraire, ce qui diminue, c'est la redondance de l'ensemble du système, c'est le caractère non différencié (...). Un deuxième aspect du principe de complexité par le bruit dans les mécanismes d'apprentissage non dirigé consiste en ce que les patterns, une fois créés, sont comparés avec les nouveaux stimuli ou, plus exactement, sont projetés et appliqués sur eux. Dans la mesure où patterns et nouveaux stimuli peuvent coïncider, on dit qu'on 'reconnaît' de nouveaux patterns dans l'environnement. *Mais*

dans la mesure où ils sont vraiment nouveaux, cette coïncidence ne peut être qu'approximative. Il y a là une ambiguïté dans cette application, dans cette projection de ces patterns sur les nouveaux stimuli, et cette ambiguïté elle-même a alors un rôle positif dans la mesure où elle entraîne une action en retour sur les patterns eux-mêmes, c'est-à-dire une modification des patterns initiaux. Ceux-ci, modifiés, vont ensuite être projetés à nouveau sur les nouveaux stimuli, et ainsi de suite. On peut ainsi se représenter ces mécanismes d'apprentissage non dirigé par une espèce de va-et-vient entre des patterns qui sont créés, puis projetés sur des stimuli aléatoires, et ceux-ci, qui, dans la mesure où ils ne peuvent pas coïncider exactement avec les premiers, modifient alors la classe de patterns qui va servir de référence, et ainsi de suite. Autrement dit, tout se passe comme si notre appareil cognitif était une espèce d'appareil créateur, une fois de plus, d'ordre de plus en plus différencié, soit de complexité à partir du bruit." [1979, pp. 145-146].

Si nous avons tenu à citer quasi-intégralement ce passage de H. Atlan malgré sa relative longueur, c'est pour deux raisons liées. D'abord, parce qu'il porte explicitement sur le genre d'apprentissage effectué par notre système cognitif envisagé comme un exemple de système non construit et présentant, par suite, les propriétés auto-organisatrices dont est plus généralement crédité ce type de systèmes par H. Atlan. En somme, ce dernier s'exprime ici très précisément sur le même sujet que celui traité par I. Rosenfield. Ensuite et surtout, parce qu'il est particulièrement frappant de voir que, sur ce sujet identique, les propos de H. Atlan rejoignent *exactement* ceux tenus par I. Rosenfield: mises à part d'inévitables différences d'ordre purement lexical, il n'y aurait pas un seul mot à changer dans cette citation pour qu'elle paraisse provenir du texte de I. Rosenfield.

Pour achever de se convaincre de cette parfaite identité de vues entre nos deux auteurs, il suffit de considérer l'analyse de la reconnaissance et de l'acquisition de la parole chez l'enfant - un cas significatif de système auto-organisateur "fiable" au sens de H. Atlan - par I. Rosenfield. En s'appuyant sur un grand nombre de données expérimentales, ce dernier montre en effet que l'apprentissage à l'oeuvre chez l'enfant, lors de la reconnaissance et de l'acquisition de la parole, procède par généralisation à partir de différents indices visuels et parlés [1989, pp. 95 sq.]. En quelque sorte, l'enfant interprète ces indices en les catégorisant - en construisant des "patterns", dirait Atlan: il traduit sa perception des sonorités associées à certaines mimiques labiales en une combinaison qu'il s'efforce de reproduire lors de ses essais d'articulation. Pour le dire comme Rosenfield, l'enfant repère d'abord les "contours" de la parole porteuse de sens, et notamment ses contours prosodiques, bien qu'il puisse ne pas comprendre les mots isolés ni les phrases :
 "Certains jeunes enfants commencent, par exemple, par imiter les variations d'intonation et d'intensité de la parole de l'adulte. Il leur arrive d'en reproduire si bien les diverses inflexions que les adultes croient les entendre prononcer des phrases authentiques, alors qu'en fait les suites de syllabes ne forment pas de véritables mots.~ [op. cit., pp. 96-97].

Mélodie ascendante de l'interrogation et de la surprise, mélodie descendante de la déclaration, etc.: tels sont les contours prosodiques du son que l'enfant cherche d'abord à

reproduire avec son babillage, dans lequel les adultes ne reconnaissent qu'un "charabia" charriant parfois, tel le fleuve une pépite, un éclat de leur propre voix. Après ces mimiques intonatives, vient l'acquisition des syllabes et des segments phonétiques (consonnes et voyelles) composant les syllabes. Segments et syllabes apparaissent progressivement, au fur et à mesure que l'enfant apprend à maîtriser, en les synchronisant correctement, les composantes des mouvements articulatoires nécessaires à la production correcte des sons.

Le point important est ici que ce type d'apprentissage évoque fortement celui décrit plus haut par Atlan, et qui se compose en définitive de deux démarches complémentaires: une opération de *catégorisation* que l'on peut voir toujours à l'oeuvre dans la *synchronie* du processus de l'apprentissage, et une opération de *différenciation* qui scande la *diachronie* de ce même processus. Ainsi, les stades successifs de ce dernier portent sur des objets obtenus par une partition de plus en plus fine de l'ensemble des catégories correspondant à chaque stade - un ensemble lui-même qualitativement modifiable par le caractère toujours approximatif de la projection des catégories anciennes sur les nouveaux stimuli. En effet, écoutons une dernière fois I. Rosenfield résumer la logique de l'apprentissage de la parole par l'enfant : celui-ci

"commence donc par reconnaître des contours d'ensemble, mais lui fait défaut la synchronisation nécessaire à la séquence de mouvements, bien qu'il l'ait détectée. (...) Il ne saisit d'emblée ni le détail de cette organisation, ni l'ensemble de ses variations d'une émission à l'autre. Mais il est frappant de constater que les enfants établissent des catégories de mouvements, dont ils perçoivent l'importance dans la parole adulte, et qu'ils s'efforcent de reproduire les mots en ayant recours à ces catégories. Par la suite, un schéma articulatoire, une série de mouvements, permet de définir une catégorie pour un phonème spécifique ou une unité de son. Ainsi, la signification d'un jeu de mouvements évolue selon une séquence allant du général au spécifique (les enfants commencent par acquérir une capacité générale à imiter les intonations), tout comme cela se produit pour l'acquisition du sens. (En effet), les enfants acquièrent d'abord la partie sémantique commune à deux mots, et seulement plus tard la partie sémantique qui différencie les deux mots." [op. cit., pp, 98-99].

Les conceptions de l'apprentissage de H. Atlan et de I. Rosenfield se rejoignent donc sur trois points liés : 1) il s'agit d'un apprentissage non dirigé, au sens où il ne répond à aucun programme préétabli, ni dans le système considéré, ni dans l'environnement de ce système; 2) la cause efficiente de cet apprentissage réside en la rencontre aléatoire du système et de certains facteurs en provenance de son environnement; 3) le produit de ce processus d'apprentissage consiste en une succession de catégories toujours plus fines et différenciées, dont la définition même, voire le mode de construction, sont susceptibles d'être remis en cause à chaque étape de ce processus.

C'est très exactement dans pareille optique que nous ouvrirons plus loin la seconde partie de cet ouvrage par la présentation d'un modèle formel d'apprentissage non dirigé, avec création d'information, dans un système économique conçu comme une société de systèmes cognitifs

individuels. Mais pour l'instant, il convient de clore cette première partie avec un texte dans lequel P. Garrouste et J.-A. Héraud montrent d'autres approches de l'auto-organisation, et mettent en relief la double dimension de ce concept: l'une qui voit un système doté de propriétés auto-organisatrices utiliser l'information en provenance de son environnement pour *s'auto-organiser* en modifiant qualitativement l'agencement et le fonctionnement des différentes parties le constituant; l'autre qui voit ce même système apposer sa marque propre sur son environnement - être *auto-organisateur* vis-à-vis de celui-ci.

Notes

(1) La modularité fonctionnelle est une version moderne de la localisation, au sens où elle conçoit le cerveau comme étant composé d'unités fonctionnelles spécialisées appelées modules, pouvant être localisées anatomiquement ou non. Selon I. Rosenfield, ce point de vue est aujourd'hui très largement partagé par la plupart des psychologues et des spécialistes en neurosciences [1989, p. 83].

(2) Cf. I. Rosenfield : "En définitive, pour Dejerine, la perte de la lecture pouvait s'expliquer de deux façons: soit par la destruction des 'images visuelles des mots' - les images engrammées dont il prétend qu'elles sont reproduites quand on écrit; soit par une déconnexion entre les centres de la vision et ceux du langage. Dans ce dernier cas, l'information visuelle n'atteint jamais le 'centre de images verbales', où les mots acquièrent un sens linguistique s'ajoutant à la forme purement visuelle (et non linguistique) issue des centres visuels du cerveau. Dans la première situation - lorsque les images visuelles des mots sont détruites -, le tableau clinique se caractérise à la fois par l'agraphie (puisque'il n'y a aucune image à copier) et par l'alexie (pour la même raison, aucune image ne pouvant être confrontée à ce qui est lu); c'est ce que l'on observa dans le cas relaté en 1891 et lors des derniers jours d'Oscar C. C'est pourquoi l'agraphie se trouve toujours associée à une inaptitude à lire: ces mêmes images visuelles des mots sont également indispensables à la lecture. Mais l'inverse n'est pas exact: une alexie ne se trouve pas forcément associée à une agraphie. Dans ce cas, le problème est uniquement celui de la cécité verbale - le premier symptôme présenté par Oscar. Cela n'est pas dû à la destruction d'images mnésiques, mais à l'incapacité de les rappeler pour les confronter à ce qui est en train d'être lu; elles sont 'déconnectées' de l'information visuelle afférente." [1989, p. 54].

(3) Huhlings-Jackson cite l'exemple de l'une de ses patientes, qui était aphasique, et s'écria "au feu"! alors qu'un incendie s'était déclaré dans une salle voisine: "N'est-il pas absurde de croire que cette femme ne se souvenait que de l'expression *au feu*!" (cité par I. Rosenfield, 1989, p. 71). Ainsi que l'ajoute immédiatement I. Rosenfield: "Elle était capable de crier au feu en raison de l'émotion qu'elle éprouvait devant un immeuble en flammes - dans cette situation, le mot exprimait ce qu'elle ressentait et voyait. Mais, deux jours plus tard, elle était sans doute incapable de parler calmement du feu qui avait ravagé le bâtiment; la lésion cérébrale la privait de l'usage de la pensée dans une situation nouvelle." [Ibid.].

(4) Pour une analyse culturaliste des relations entre contexte et signification, voir E. T. Hall [1979, pp. 87-104, 117-128]. Hall rapporte notamment certaines expériences montrant que l'adaptation interne de l'être humain à son contexte culturel lui permet de corriger automatiquement les déformations ou les oublis dans un message, exactement sur le même mode que le patient de A. A. Low lisant spontanément des syllabes ou, plus généralement, des énoncés en y rétablissant un sens inexistant dans le document soumis à sa lecture [op. cit., pp. 117 sq.].

(5) Sur la multiplicité des formes de mémoire chez l'homme (auditive, visuelle, olfactive, tactile, gustative), et l'erreur commise par notre civilisation lorsqu'elle privilégie par trop, dans son système d'éducation, la facilité verbale aux dépens d'autres facultés importantes de l'esprit, voir l'ouvrage cité de E. T. Hall [pp. 167 sq.]. D'après Hall, A. R. Luria estimait que les "centres du cerveau où sont localisées des fonctions spécifiques telles que la colère, la peur, la faim, ainsi que les zones visuelles, auditives et motrices du cerveau [étaient des] stations de

traitement et d'interprétation, et non pas des 'dépôts de conservation'. Leur rôle est de classer, d'interpréter, de pomper ou de chasser les informations, mais non de les emmagasiner." [op. cit., pp. 190-191].

(6) Contrairement à la théorie émise en 1940 par Pauling, selon laquelle la présence d'une bactérie dans l'organisme déterminerait la *nature* de l'anticorps fabriqué, G. M. Edelman montra en 1969 que tous les animaux naissent avec un répertoire *complet* d'anticorps: au lieu de considérer l'organisme comme un système apprenant à combattre un germe spécifique par la fabrication d'un anticorps spécifique, il faut donc le concevoir comme doté d'emblée d'une molécule dont les éléments sont susceptibles de se modifier pour produire les millions de types d'anticorps différents nécessaires à la protection du corps face aux organismes étrangers. La présence de la bactérie ne détermine pas la nature de l'anticorps fabriqué - il existe déjà dans l'organisme - mais seulement sa *quantité*. Tout se passe comme si la bactérie ou le virus *sélectionnaient* l'anticorps approprié - qui est différent pour chaque individu. I. Rosenfield rappelle à ce propos l'observation de Darwin soulignant qu'il n'existe pas, dans le monde de la biologie, d'animal typique ou de plante typique, mais seulement des êtres singuliers dont les communes qualités sont des abstractions que nous inventons en négligeant les variations interindividuelles. Or c'est à partir des variations survenues dans la population que la sélection peut s'opérer: c'est la variation qui est réelle, et non la moyenne [op. cit., p. 156].

(7) Suggérée dès 1989, pareille extension n'est pleinement assumée par I. Rosenfield que dans un ouvrage ultérieur [1990], que nous avons peu sollicité pour l'instant car sa mention se justifie davantage dans notre contribution à la seconde partie de ce livre. Dans *L'invention de la mémoire*, I. Rosenfield évoquait à l'appui de l'extension de la théorie d'Edelman aux fonctions cérébrales supérieures (gouvernant le langage) l'existence de "Darwin II", un automate construit par G. M. Edelman et G. Reeke Jr. en conformité avec les principes du darwinisme neuronal. De fait, les performances de cet automate semblent aller bien au-delà de celles obtenues par l'approche informatique traditionnelle ou par le PDP: destiné à simuler l'activité cartographiante du cerveau, Darwin II est apparemment capable d'abstraire (d'organiser des stimuli en catégories) sans que ne lui soient préalablement fournies d'instructions spécifiques pour cela [cf. I. Rosenfield, 1989, pp. 171-176, 195-196].

(8) A notre connaissance, la première présentation française de son modèle d'auto-organisation fut donnée par H. Atlan en 1972, dans un ouvrage rassemblant une série de conférences prononcées l'année précédente en Israël. Le modèle fut repris ensuite dans une publication destinée à un plus large public [H. Atlan, 1979], puis amélioré ou complété sur certains points [cf., par exemple, H. Atlan, 1986a]. Pour une bibliographie relativement récente sur la question, voir H. Atlan [1986b, p. 367, note 2.30].

(9) Dans leur célèbre ouvrage, W. Weaver et C. E. Shannon [1949] l'appellent "théorème fondamental pour un canal discret avec bruit". Les éléments de Théorie de l'information contenus dans les quatre premiers chapitres de H. Atlan [1972] offrent une présentation très pédagogique de la théorie shannonienne. L'exposition et la discussion du "théorème de la voie avec bruit" font l'objet du chapitre IV de cet ouvrage [pp. 41-61].

[10] Il faut que le système S présente une redondance initiale suffisamment élevée pour que le rôle bivalent du bruit puisse exister. Imaginons en effet qu'il n'existe qu'une seule voie élémentaire de transmission de l'information entre deux sous-systèmes A et B, seuls éléments de S: l'éventuelle autonomie de B par rapport à A est alors synonyme de mauvais fonctionnement du système, et implique en pratique la destruction de celui-ci. Du point de vue du système, la quantité d'information de B en tant que ce dernier est indépendant de A perd toute signification dans une telle situation. Il doit donc exister un nombre minimal de sous-systèmes et de voies pour que l'indépendance, même totale, de l'un (par exemple B) par rapport à un autre (par exemple A), ne signifie pas disparition du système S. En d'autres termes, ce dernier doit présenter une redondance initiale suffisamment élevée pour que l'ambiguïté-autonomie ait un sens.

BIBLIOGRAPHIE

- ANCORI B. et al. [1987], "*Evolution économique et Théories de l'information*", ARI sur les Sciences de la Communication, BETA/CNRS, mars
- ATLAN H. [1972], "*L'organisation biologique et la théorie de l'information*", Hermann
- ATLAN H. [1979], "*Entre le cristal et la fumée. Essai sur l'organisation du vivant*", Seuil
- ATLAN H. [1986a], "Créativité biologique et auto-organisation du sens", in "*Cognition et Complexité*", Cahiers du CREA, 9, mars
- ATLAN H. [1986b], "*A tort et à raison. Intercritique de la Science et du mythe*", Seuil
- BATESON G. [1977], "*Vers une écologie de l'esprit, I'*", Seuil
- BRILLOUIN L. [1959], "*La science et la théorie de l'information*", Masson
- CASTORIADIS C. [1978], "*Les carrefours du labyrinthe*", Seuil
- HALL T. H. [1979], "*Au-delà de la culture*", Seuil
- KUHN T.S. [1962], "*The Structure of Scientific Revolutions*", The University of Chicago Press
- LEVY P. [1987], "Le paradigme du calcul", in STENGERS I. éd., "*D'une science à l'autre. Des concepts nomades*", Seuil
- MONOD J. [1970], "*Le hasard et la nécessité*", Seuil
- PERSONNAZ L., DREYFUS G. & GUYON I. [1988], "Les machines neuronales", *La Recherche*, Vol. 19, n°204, novembre
- ROSENFELD I. [1989], "*L'invention de la mémoire. Le cerveau, nouvelles données*", Eshel
- ROSENFELD I. [1990], "*La conscience. Une biologie du moi*", Eshel
- VON FOERSTER H. [1960], "On self-organizing systems and their environments", in Yovitz & Cameron éd., "*Self-organizing Systems*", Pergamon Press
- WATZLAWICK P., HELMICK-BEAVIN J. & JACKSON D. [1972], "*Une logique de la communication*", Seuil
- WEAVER W. & SHANNON C. [1949], "*The mathematical theory of communication*", University of Illinois

CHAPITRE 3

**AUTO-ORGANISATION ET CREATION
TECHNOLOGIQUE**

par

Pierre GARROUSTE et Jean-Alain HERAUD

Ce texte examine l'application du modèle de l'auto-organisation à l'économie des changements technologiques. Après une partie introductive (1.) qui présente le modèle général et évoque la place de celui-ci en économie, nous prenons le temps d'un retour sur ses versions originales, qui sont celles des biologistes et des chimistes (2.), pour pouvoir ensuite mieux analyser, classer et évaluer les contributions à l'économie de la technologie qui y renvoient - explicitement ou non - (3.). L'approche systémique de l'évolution technico-économique que proposent les contributions inspirées par l'auto-organisation apparaît utile dans la mesure où elle trace un programme de recherche en sciences économiques potentiellement très riche, mais ce programme ne saurait se conformer à l'exigence de réalisme sans introduire la dimension stratégique dans le modèle initial de l'auto-organisation hérité des sciences de la nature (4.).

1. Introduction à la notion d'auto-organisation

L'introduction de la référence à l'*auto-organisation* dans l'analyse des changements technologiques est récente: elle date de la fin des années 70. Il est sans doute trop tôt pour juger s'il s'agit d'un nouveau paradigme ou d'un nouveau programme de recherche scientifique - voire d'une simple métaphore pour exposer autrement un certain type d'analyses. A défaut de terminologie plus précise, nous parlerons ici de diffusion du "modèle" interdisciplinaire de l'auto-organisation. Celui-ci s'est construit progressivement sur plusieurs champs des sciences de la nature puis des sciences humaines (1), avant de donner lieu à des applications croissantes dans le domaine de l'*économie de la technologie*. Par ailleurs, certaines approches évolutionnistes du changement technique peuvent être réexaminées à la lumière de l'auto-organisation même si elles n'ont pas été formulées par les auteurs dans ces termes. Notre propos est de fournir des éléments d'évaluation des divers courants de pensée en économie de la technologie qui peuvent être rattachés à ce modèle interdisciplinaire - qu'ils affichent explicitement ou non une référence à l'auto-organisation.

Pour cela, il est indispensable de revenir tout d'abord sur les approches originelles - celles des biologistes et des chimistes principalement - et sur les applications que les économistes en ont faites, particulièrement en économie de l'information et dans les travaux s'inspirant de la science des systèmes. Les conceptions de base de l'auto-organisation en sciences de la nature sont à notre avis de trois types qui correspondent respectivement aux travaux d'Atlan, de Varela-Maturana et de Prigogine-Stengers (nous verrons que la terminologie varie également d'un auteur à l'autre et que d'autres concepts doivent être explicités, comme ceux d'*autonomie* ou d'*autopoïèse*, mais nous partons ici de l'hypothèse qu'il existe une notion générale commune à toutes ces approches: l'auto-organisation). Les diverses analyses du changement technique vu comme processus auto-organisé sont à rapprocher des trois versions du modèle général. Il est alors possible d'évaluer l'utilité du modèle - en faisant ressortir sa version la plus pertinente pour l'économie de la technologie. Inversement, on peut classer les contributions économiques faisant

référence au modèle selon le degré de pertinence de cette référence: ce ne sont pas forcément les analyses introduisant explicitement le modèle de l'auto-organisation qui apparaissent les plus convaincantes.

Dans cette partie, nous verrons sous quels aspects l'évolution de la théorie économique tend à converger vers le modèle général de l'auto-organisation (1.1), avant d'examiner les applications de ce modèle au domaine particulier du changement technologique (1.2).

1.1 L'émergence de l'auto-organisation en économie

Après s'être inspirée longtemps des modèles des sciences physiques, l'économie tend, par divers chemins, à défricher des champs théoriques plus proches non seulement de la biologie (cf. par exemple R. Passet [1979]), mais aussi et surtout de la systémique (B. Walliser [1988]).

B. Walliser propose la périodisation suivante:

- phase *mécaniciste* (importance accordée à la notion d'équilibration), jusqu'aux années 40;
- phase *cybernétique* (notion de régulation), qui s'affirme principalement après guerre;
- phase *systémique* (notion d'auto-organisation), commencée dans la période récente.

La distinction entre cybernétique et systémique est importante: outre les relations de *feed-back* (régulations) que possèdent les deux types de modèles, un modèle systémique présente la propriété de se *reconstruire de lui-même* à certaines occasions au moins.

On peut replacer les propriétés des systèmes auto-organisés par rapport aux propriétés habituellement attribuées au monde vivant, telles qu'elles sont rappelées par R. Ayres[1988]:

- persistance
 - réplique
 - modification de l'environnement
 - capacité de se modifier soi-même,
- et constater que seule la dernière propriété distingue réellement les modèles d'auto-organisation de ceux de la cybernétique.

Pour caractériser un système auto-organisé (ou un processus auto-organisateur), on peut reprendre l'expression de B. Walliser [1988, p.254] - qui nous semble particulièrement bien adaptée à la problématique économique:

"capacité d'un système à faire émerger des processus locaux une structure globale, qui s'avère ainsi ni imposée autoritairement par une instance supérieure ni élaborée délibérément par les éléments de base".

Parmi les thèmes théoriques illustrant la notion générale d'auto-organisation, l'auteur s'intéresse plus spécifiquement à la représentation de l'économie en termes de jeux dynamiques "entre acteurs dotés de capacités limitées, mais participant à un réseau dense d'interactions matérielles et informatives" (ibid.).

Ce type d'approche amène à s'intéresser aux diverses formes de *rationalité* des agents (au lieu de retenir l'a priori traditionnel concernant *la* rationalité économique) ainsi qu'ont tenté de la faire des auteurs comme H. Simon et J.G. March. On cherche à expliciter les *représentations* que se font les agents de leur environnement (notons que le degré maximum de connaissance est celui postulé dans les modèles à anticipations rationnelles). L'accent est mis sur diverses formes de "frottements" dans les interactions - ce qui porte atteinte au rôle des prix dans la coordination des actions des agents et fait émerger d'autres signaux pour la réalisation de cette fonction (les quantités dans les modèles d'équilibre avec rationnement). Les transactions sont médiatisées par des *institutions* plus variées que les seuls marchés (organisations, réseaux, etc...): le modèle de l'auto-organisation s'affirme naturellement à partir du moment où l'on ne se contente plus de considérer les institutions comme des données, mais que l'on cherche au contraire à comprendre leur *genèse*.

La multiplicité - et/ou la non-optimalité - des équilibres spontanés possibles pose la question théorique du choix de l'un d'eux à travers des processus évolutifs. U. Witt [1985] exprime l'idée que la coordination des activités individuelles est un processus évolutif d'auto-organisation aboutissant à des formes de *marchés coordonnés mais de manière non intentionnelle*. Aucun acteur individuel, à l'intérieur ou hors du système, ne joue un rôle d'organisateur complet et délibéré. Des *structures* comme les marchés ou d'autres formes organisationnelles sont à la fois sujets et *objets* de la création. J. Lesourne [1985] donne comme exemples de systèmes auto-organisés, c'est-à-dire *capables d'acquérir des propriétés nouvelles en modifiant eux-mêmes leur organisation* : l'entreprise qui réussit à structurer la concurrence (la réponse de la firme à son environnement consiste à agir sur cet environnement, engendrant éventuellement de nombreuses itérations), le marché qui donne naissance à des marchés nouveaux, le système de l'emploi qui engendre les syndicats (un marché suscitant une organisation).

La création de structures nouvelles par les agents eux-mêmes est la marque d'un processus d'auto-organisation. L'évolution structurelle n'est pas pilotée par une instance hiérarchique (ni voulue par une instance exogène), de même qu'elle ne saurait être parfaitement anticipée par un acteur quelconque du système. Ce qui nous semble essentiel à souligner dans ce phénomène de réseau qu'est l'auto-organisation, c'est l'*interférence des niveaux*: la réorganisation structurelle qui matérialise l'acquisition des propriétés nouvelles est *à la fois* celle de l'agent et celle du système (cf. J.-A. Héraud [1988a]). La propriété d'auto-organisation d'un système est fondée sur la capacité des agents à modifier à la fois leur milieu interne et leur

environnement (2). En ce qui concerne les sous-systèmes que forment les agents (typiquement: les firmes), on peut appliquer le qualificatif "auto-organisé" à ceux qui se révèlent capables de se restructurer en fonction de certaines circonstances environnementales et qualifier d'"auto-organisateur" ceux qui, de surcroît, ont une stratégie de modification de leur environnement. Mais il est difficile de faire une distinction trop tranchée entre le système global et les sous-systèmes lorsqu'on parle d'auto-organisation car il s'agit, fondamentalement, d'un processus *global* d'interaction créatrice. C'est particulièrement le cas en sciences sociales - donc en économie - où une forme particulière d'interférence des niveaux d'organisation caractérise les processus évolutifs: les agents étant des êtres conscients et volontaires, *chacun porte en lui une image du réseau global*, cette image étant à la fois une *représentation* et un *projet*. (le texte suivant de B. Ancori illustrera en particulier cette question). L'imperfection de l'information se traduit ici par le fait que les agents n'ont pas la même image du système global; et c'est bien entendu cette propriété qui permet l'évolution du système - lequel, sinon, ne saurait être que stationnaire. Dans un tel système, la notion d'information prend un sens fort: "in-former" signifiant se former soi-même (apprentissage) et former les autres (stratégie).

Comme on peut le constater, le modèle de l'auto-organisation apporte à l'économie la dimension de la création informationnelle: à travers ce type d'approche, le théoricien cherche à décrire le processus endogène par lequel le système économique se construit lui-même - autrement dit, comment les agents, à l'occasion de leurs interactions habituelles, créent de nouvelles règles du jeu. (Signalons qu'une approche possible, bien que limitée à un aspect de la question, est celle des *jeux répétés*; mais il s'agit là de tout un domaine de recherche que nous ne pouvons aborder dans le cadre de ce texte). Lorsque l'information nouvelle implique le système technique, on se situe un domaine particulier, celui de l'économie du changement technologique, que nous allons maintenant aborder.

1.2 L'approche du changement technologique par le modèle de l'auto-organisation

Il nous semble que trois critères doivent être mobilisés pour mener à bien notre tentative d'évaluation des approches du changement technologique s'inspirant des modèles de l'auto-organisation.

a) Un des premiers critères discriminants est le positionnement par rapport aux *travaux fondateurs des non-économistes en matière d'auto-organisation*:

- On peut distinguer parmi les contributions des économistes celles qui se rattachent à une conception identifiée (ou identifiable) des travaux sur l'auto-organisation (G. Silverberg, G. Dosi et L. Orsenigo; R. Nelson et S. Winter), celles qui parviennent explicitement à des convergences en termes de résultats ou de méthodes avec ces travaux (W.B. Arthur), et celles

qui partent des résultats de ces travaux (D. Foray et P. Garrouste).

- Dans la mesure où les économistes se réfèrent aux modèles fondateurs comme ceux d'Atlan, de Varela ou de Prigogine, il faut tenir compte de notables différences entre ces modèles eux-mêmes.

b) Les différents niveaux auxquels s'expriment les processus auto-organiseurs permettent de distinguer les analyses qui considèrent la *technologie* comme un système auto-organisé (P. P. Saviotti), celles qui définissent le *processus de compétition technologique* comme étant auto-organisé (W.B. Arthur), celles qui localisent l'auto-organisation dans les *relations entre changements technologiques et structures industrielles* (Dosi, Orsenigo, Silverberg), celles qui envisagent l'auto-organisation comme liée au comportement de formes productives (Foray, Garrouste), celles pour qui l'auto-organisation caractérise un *système économique* (R. Ayres), ou celles enfin qui l'intègrent à plusieurs niveaux de leur analyse (D. Sahal).

c) Le dernier critère discriminant que nous évoquerons concerne un mécanisme particulier du processus d'auto-organisation: il s'agit de la nature de l'*intervention de l'environnement* sur les caractéristiques propres au système considéré. On distinguera les analyses où l'environnement joue un rôle perturbateur essentiel dans le déroulement du processus auto-organisateur (la sélection chez Nelson et Winter; les "petits événements" chez W.B. Arthur), et celles où l'environnement est plutôt considéré comme un facteur "résiduel", n'intervenant que pour renforcer le système de cohérence interne au système (Foray, Garrouste).

Afin de donner des éléments d'*évaluation externe* des divers "courants" que nous distinguons dans l'approche auto-organisationnelle du changement technologique (c'est-à-dire pour définir ce qui permettra de les situer par rapport aux conceptions "standard" de l'auto-organisation), il nous semble intéressant, dans un premier temps (2.), de présenter les approches originelles de l'auto-organisation avant d'étudier (3.) les travaux des économistes du changement technologique qui se réclament - ou que l'on peut rattacher à - des approches en termes d'auto-organisation. Notre évaluation de ces travaux se fera donc, d'une part, en les situant par rapport aux conceptions génériques, en les regroupant par classes (tout en insistant sur leurs spécificités), et d'autre part en nous efforçant de montrer en quoi ces approches proposent une analyse renouvelée de la création technologique.

2. Les conceptions originelles de l'auto-organisation

Nous nous contenterons ici de rappeler les approches de H. Atlan, de H. Maturana et F. Varela, de I. Prigogine et I. Stengers. On peut avancer deux justifications de ce choix: à quelques différences près quant à leur définition, ces approches correspondent aux "trois voies"

par lesquelles l'idée d'auto-organisation a fait son chemin dans les années soixante-dix, selon H. Atlan [1986]; elles constituent aussi de bons modèles de référence pour la classification des analyses du changement technologique. Nous ne pouvons mettre en évidence ici la généalogie de l'analyse en termes d'auto-organisation. Nous nous contenterons de décrire les proximités de pensée et de conception.

2.1 L'auto-organisation chez H. Atlan

Pour H. Atlan, la nécessité d'introduire une réflexion en termes d'auto-organisation ne constitue pas une rupture. On part du fait que les organismes vivants sont caractérisés par des performances résultant de principes spécifiques qui sont "postulés en continuité avec les autres domaines de la cybernétique, les mieux connus, ceux qui s'appliquent aux automates artificiels" (H. Atlan [1972], p.242). Les conséquences de cette position sont d'importance: d'une part, "la spécificité des organismes vivants est rattachée à des principes d'organisation plutôt qu'à des propriétés vitales irréductibles" et d'autre part, lorsqu'on a mis en évidence ces principes, "rien ne devrait empêcher de les appliquer à des automates artificiels dont les performances deviendraient alors égales à celles des organismes vivants" (ibid.). Les débats actuels portant sur l'orientation méthodologique des sciences cognitives montrent le caractère problématique d'une telle approche (H.L. Dreyfus [1979]; F. Flores et T. Winograd [1989]; F. Varela [1989b] et le travail de I. Rosenfield analysé par B. Ancori dans le texte précédent).

Le modèle d'Atlan ayant été présenté et commenté en détail précédemment, nous nous contenterons d'en résumer les points essentiels, ceux qui le distinguent nettement des autres modèles. En mobilisant la théorie de l'information, H. Atlan montre que la variation de la quantité d'information de certains systèmes se traduit par une tendance à la *diminution de sa redondance* et à l'*augmentation de sa complexité* quand une perturbation extérieure se manifeste. Le rôle de l'environnement est central dans ce modèle, mais paradoxalement sa contribution est qualitativement modeste puisqu'il se réduit à une source de bruit.

L'être auto-organisé évolue selon une logique purement interne, même si l'occasion (le prétexte, l'incitation) est environnementale. Le processus d'auto-organisation présente également un terme programmé: le stock de redondance initiale ne peut pourvoir indéfiniment à l'augmentation de complexité nécessaire pour répondre aux facteurs aléatoires issus de l'environnement. Lorsque ce stock est épuisé, l'être meurt, en quelque sorte fossilisé - car la propriété du vivant réside précisément dans la capacité d'adaptation permanente.

On voit que ce concept d'apprentissage structurel proposé par H. Atlan fonde un modèle d'auto-organisation typiquement adapté à la biologie, et plus précisément à l'évolution des *individus*. Nous reviendrons à la fin de cette partie sur les implications d'un modèle si fortement entropique lorsqu'on envisage de l'appliquer à la technologie, à l'économie ou d'une manière

générale au domaine des sciences sociales. On ne retrouve pas dans les modèles suivants une dimension entropique aussi centrale.

2.2 L'auto-organisation chez H. Maturana et F. Varela: autopoïèse et autonomie

Pour F. Varela, il est nécessaire de penser les systèmes auto-organisés en mobilisant des méthodes différant des méthodes classiques. Ces systèmes présentent des caractéristiques spécifiques: ils possèdent des processus internes interconnectés qui permettent de stabiliser leurs trajectoires dans un espace de phases; cette qualité définit leur identité en rendant possible leur distinction d'un espace substrat. La représentation théorique de l'auto-organisation du vivant fait appel chez Maturana et Varela au concept d'autopoïèse (on pourrait traduire par "auto-crédation").

"Un système autopoïétique est organisé (défini comme une unité) comme un réseau de processus de production, transformation et destruction de composants qui: (1) à travers leurs interactions et transformations régénèrent continuellement et réalisent le réseau des processus (relations) qui les ont produits; et (2) le constituent (le système) comme une unité concrète dans l'espace dans lequel ils existent en spécifiant le domaine topologique de sa réalisation comme réseau" (F. Varela [1979], p.13).

Cette définition suppose l'acceptation d'une distinction préalable entre les notions de *structure* et d'*organisation* d'un système:

"Le mot structure concerne les relations entre les parties aussi bien que l'identité des parties qui constituent un tout" (H. Maturana [1974], p.177). Par contre, l'organisation concerne les relations qui définissent un système comme unité (définissant par là ses propriétés) sans faire référence à la nature des composants, qui peuvent être tels ou tels pourvu qu'ils satisfassent ces relations.

Bien que la terminologie employée pose problème (3), la différence ainsi faite entre modifications altérant ou non l'*identité du tout* est très pertinente. Le propre d'un système auto-organisé étant de se recréer lui-même en permanence, il est en effet essentiel de lui conférer la possibilité théorique de se modifier sans se trahir. Les systèmes *autopoïétiques* ont la capacité, contrairement aux systèmes *allopôïétiques* (systèmes créés par une entité extérieure), de produire les conditions du maintien de leur identité, c'est-à-dire de leur organisation. Le maintien de l'invariance de l'"organisation" d'un système autopoïétique suppose des possibilités de modifications "structurelles": celles-ci lui permettent d'absorber les perturbations. Dès l'instant où un tel système est dans l'incapacité de promouvoir des changements de structure qui soient de nature à préserver l'invariance de son organisation, il disparaît comme système autopoïétique. (Ce qu'il devient peut être un nouveau système autopoïétique, mais en tout état de cause ce ne sera pas le même).

F. Varela, pour sa part, distingue soigneusement le concept de système *autopoïétique* de

celui de système *autonome*. Le second est plus général et c'est celui que nous devrions retenir, si l'on suit l'auteur sur ce point, pour une utilisation en dehors du champ de la biologie.

Ce qui caractérise les systèmes *autonomes*, c'est qu'ils sont "*organisationnellement clos*", ce qui signifie que leur organisation est caractérisée par des processus tels que:

" (1) les processus sont reliés en un réseau, de façon qu'ils dépendent récursivement les uns des autres dans la génération et la réalisation des processus eux-mêmes, et (2) ils constituent le système comme unité identifiable dans l'espace (domaine) dans lequel les processus existent." (F. Varela [1979], p. 55).

En fait, la distinction essentielle entre les deux concepts repose sur le fait que les relations qui caractérisent l'autopoïèse sont des relations de *production de composants*:

"... l'idée d'autopoïèse est, par définition, limitée aux relations de production de quelque sorte, et réfère à des frontières topologiques. Ces deux conditions sont clairement non satisfaites pour les autres systèmes autonomes" (ibid., p.54).

Ici, F. Varela pense en particulier aux systèmes *sociaux*. La restriction de l'autopoïèse aux systèmes vivants renvoie au fait que les interactions sociales ne peuvent être décrites, selon cet auteur, en termes de production de composants. Cette position quelque peu matérialiste est certainement critiquable (elle se situe entre autres à l'opposé de toute une école de pensée en sciences sociales, à savoir le *constructivisme* (4)). Elle justifie en tout cas la réticence du biologiste à exporter le paradigme de l'autopoïèse en dehors du domaine de l'organisation du vivant.

Le concept d'autonomie (et son contraire, l'*hétéronomie*) mérite quelques commentaires supplémentaires, ne serait-ce qu'en raison de cette plus grande généralité que nous venons d'évoquer.

En ce qui concerne la caractérisation du système: celui qui est hétéronome se présente comme une boîte d'entrée-sortie (paradigme de la commande) alors que l'autonome possède un comportement propre (clôture opérationnelle);

En ce qui concerne la relation à l'environnement: dans le cas d'un système hétéronome, elle passe par la représentation du monde extérieur et on a affaire à une logique de correspondance; dans le cas d'un système autonome, la relation passe par un monde émergent de significations qui est inséparable de la clôture opérationnelle et on a affaire à une logique de cohérence (l'environnement est en quelque sorte internalisé).

2.3 L'auto-organisation chez I. Prigogine et I. Stengers

L'ambition de Prigogine et Stengers est de caractériser les exigences véhiculées par une nouvelle vision de la science. Pour penser l'évolution, les points suivants sont mis en avant:

- l'irréversibilité, c'est-à-dire "*la brisure de symétrie entre l'avant et l'après*" (I. Prigogine et I. Stengers [1988], p.46);
- l'événement comme porteur d'un *sens*;
- la *cohérence* que font émerger les événements, qui sont ainsi susceptibles "de transformer le sens de l'évolution qu'ils scandent" (ibid., p. 47).

En s'attachant à l'étude des processus loin de l'équilibre, ils montrent que leur irréversibilité est source de cohérences. L'apparition de cette activité cohérente de la matière - les "*structures dissipatives*" - nous impose un nouveau regard, une nouvelle manière de nous situer par rapport au système que nous définissons et manipulons. Alors qu'à l'équilibre, et près de l'équilibre, le comportement du système est, pour des temps suffisamment longs, entièrement déterminé par les conditions aux limites, nous devons désormais lui reconnaître une certaine autonomie qui permet de parler des structures loin de l'équilibre comme de phénomènes d'"auto-organisation". (Ibid., p.59). Un système auto-organisé est alors caractérisé par sa sensibilité, son instabilité et les points de bifurcation que son évolution manifeste.

Ce qui spécifie cette approche est le rôle éminent dévolu aux *fluctuations* (5). L'aléa est une des dimensions du fonctionnement du système, mais il ne caractérise pas la relation de celui-ci à son environnement (contrairement à la conception d'Atlan et aux systèmes hétéronomes de Varela):

"C'est l'activité intrinsèque du système qui détermine comment nous devons décrire son rapport à l'environnement, qui engendre donc le type d'intelligibilité qui sera pertinente pour comprendre ses histoires possibles." (Ibid. p.60).

Les modèles physiques de Prigogine présentent par rapport aux modèles biologiques - particulièrement celui d'Atlan - la particularité d'échapper à une certaine fatalité: tant que le système peut puiser de la néguentropie dans l'environnement, il continue à "fonctionner". A l'opposé, l'organisme biologique est promis à une mort certaine: une fois épuisées ses réserves de redondances, le premier choc en provenance de l'environnement entraînera sa destruction. En conclusion de cette partie, nous commenterons cet aspect des modèles d'auto-organisation dans la perspective des applications en sciences sociales.

Comme nous le verrons plus loin, l'optique d'Atlan rejoint des approches en termes d'épuisement de trajectoires, que l'on retrouve par exemple dans les travaux de Nelson et Winter, ou plus explicitement dans ceux de Silverberg, Dosi et Orsenigo: l'environnement crée dans un premier temps les pressions qui génèrent la constitution d'un paradigme, lequel, sous l'effet de ces mêmes contraintes, va ensuite s'épuiser le long d'une trajectoire pour laisser la place à un nouveau paradigme. Mais l'évolution des technologies (comme celle des espèces dans les sciences de la nature) n'obéit pas nécessairement aux mêmes règles que le déroulement de la vie

des individus. Les systèmes technologiques peuvent générer leur propre dépassement par une forme de créativité intrinsèque plus proche du modèle de Maturana-Varela. De même, l'apparition et la disparition d'industries au sein de l'ensemble des activités productives en évolution s'interprètent mieux en termes de système autonome ou de structures dissipatives qu'en termes d'auto-organisation consommant des redondances.

3. L'analyse des changements technologiques en termes d'auto-organisation

Dans cette partie, nous analyserons plusieurs des principales contributions à l'économie du changement technologique en nous efforçant de mettre en évidence les mécanismes auto-organisateurs qu'elles proposent (ou présupposent). L'évaluation de ces travaux pose un problème de localisation de la critique. En effet, comme nous l'avons signalé plus haut (1.2), ces approches diffèrent quant aux objectifs qu'elles s'assignent. Ainsi, il nous semble nécessaire de distinguer les critiques internes que l'on peut adresser à ces différentes approches des critiques externes qui seront centrées sur leur capacité à expliquer le processus de création technique.

Notre grille de lecture de la littérature consacrée aux changements technologiques découle d'un essai de classification selon le niveau de l'analyse (en particulier: la technologie, l'industrie ou l'ensemble du système socio-économique) et selon le modèle d'auto-organisation de référence. C'est ainsi que nous sommes amenés à distinguer cinq types de contributions, que nous relierons aux problématiques suivantes:

- la dynamique d'une industrie: R. Nelson et S. Winter; G. Dosi et al. (modèle d'Atlan)
- le processus de compétition technologique: W.B. Arthur (modèle de Prigogine)
- le système autonome d'une industrie: D. Foray et P. Garrouste (théorie des systèmes autonomes de Maturana et Varela)
- le système de la technologie: P.P. Saviotti (une combinaison originale de modèles, dont ceux d'Atlan et de Varela)
- l'évolution globale: D. Sahal (un modèle spécifique).

Dans chacun des cas nous procéderons successivement à une présentation puis à une évaluation de la contribution.

3.1 La dynamique d'une industrie comme processus d'auto-organisation

G. Dosi et L. Orsenigo ont cherché à analyser les relations existant entre les *structures industrielles* et l'*innovation technologique*, en mettant en évidence le fait que:

"l'évolution de l'environnement industriel implique, de la part des agents de décision, un comportement

stratégique orienté vers la genèse de leurs propres "mutations" (par exemple les innovations technologiques et organisationnelles), comportement dont résulte un processus d'apprentissage, au cours duquel les décideurs apprennent à maîtriser leur environnement à leur avantage." (G. Dosi, L. Orsenigo [1988], p.14).

Pour ces auteurs, une *structure industrielle* peut être décrite en termes de répartition des entreprises selon la taille, les capacités technologiques, le type d'organisation et la stratégie. L'*innovation* concerne les modifications des processus de production et des produits fabriqués par chaque entreprise.

Cette conception des rapports entre les caractéristiques structurelles du système industriel et le comportement des entreprises n'est pas en soi une nouveauté, dans la mesure où certains des précurseurs dans le domaine de l'économie industrielle ont développé un schéma séquentiel où la notion d'effet de feed-back permettait d'établir un bouclage entre les structures de marché et le comportement stratégique des firmes. L'idée selon laquelle les agents créent ainsi leur propre information plutôt que de l'attendre de l'environnement est simple mais importante (notons qu'elle renvoie aux modalités d'endogénéisation de l'incertitude qu'étudie G. Lambert dans le présent ouvrage à propos de la décision d'investissement).

Ce qui fait la nouveauté des travaux de Dosi et Orsenigo, suivant en cela ceux de Nelson et Winter, provient de ce que:

- la notion de "structure industrielle" remplace celle beaucoup plus restrictive de "structure de marché";
- la "technologie" y est définie de manière plus riche.

(i) Dosi et Orsenigo estiment que les économistes ont trop axé leur attention sur des caractéristiques comme la taille des entreprises ou la concentration des marchés - au point de négliger les autres facteurs. Ils complètent la caractérisation des industries et des marchés par la prise en compte d'*asymétries technologiques* entre firmes, de différences de *gammes d'offres technologiques*, de diversité de comportement vis-à-vis de l'innovation (*innovateur* contre *imitateur*, par exemple), etc...

(ii) Alors que traditionnellement on définit la technologie comme un ensemble d'informations relatives à certaines combinaisons d'inputs, G. Dosi, dans une optique moins réductrice et davantage axée sur le processus d'évolution de la technologie, a proposé les concepts de *paradigme* et de *trajectoire* technologiques:

Un paradigme technologique est défini comme une "configuration (*pattern*) de solutions à des problèmes technoéconomiques sélectionnés et basés sur des principes dérivés des sciences de la nature, conjointement à des règles spécifiques permettant d'acquérir de nouvelles connaissances et de les sauvegarder, si possible, contre leur diffusion rapide chez les concurrents" (G. Dosi [1988], p. 1127).

Une trajectoire technologique est décrite comme "l'activité du processus technologique le long des arbitrages technologiques et économiques définis par un paradigme" (Ibid., p.1128).

Cette approche permet d'envisager la dynamique d'une industrie comme:

"un processus d'auto-organisation, dans le contexte duquel la technologie, les structures et les modes de comportement interviennent pour établir des trajectoires évolutives relativement ordonnées pour chaque secteur industriel ou pour l'économie dans son ensemble, tandis que la morphologie du système exerce à son tour un effet de rétroaction sur les capacités, incitations et comportements des différents agents de décision".

(G. Dosi et L.Orsenigo [1988], p.24).

Le moteur de ce processus d'évolution est localisé à la fois dans l'apprentissage et l'innovation, et dans la concurrence et la sélection.

Les travaux de Nelson et Winter, comme ceux de Dosi, Orsenigo et Silverberg, s'intègrent dans la mouvance du modèle d'Atlan. Il est d'ailleurs intéressant de noter que les premiers, bien qu'ils ne se réclament pas explicitement du courant de l'auto-organisation, défendent une conception peut-être plus conforme encore aux caractéristiques de ce modèle de référence.

La notion de *routine* est essentielle à la compréhension du modèle théorique de Nelson et Winter et c'est à travers elle que l'on peut mettre en évidence le rapport avec les modèles d'auto-organisation. Cette notion qu'a développée S. Winter dans ses travaux sur la firme permet de rendre compte des règles de décisions - en distinguant justement les grandes décisions (comportant des éléments de nouveauté radicale dans les questions posées par l'environnement et/ou les réponses de la firme) de celles qui relèvent du choix "routinier". L'analyse des changements technologiques se fonde sur cette notion et sur l'existence affirmée de *régimes* et de *trajectoires* technologiques: sur la base d'un régime technologique une trajectoire se déploie et s'épuise. Le rôle des routines et de leur évolution est alors de permettre la mise en oeuvre des potentialités que porte en lui un régime technologique. Il est important de noter que Nelson et Winter identifient les routines à des *gènes*. L'innovation apparaît alors comme une mutation réussie - le critère de réussite étant la *sélection* par l'environnement.

On peut considérer qu'une routine est ce qui permet de transformer, compte tenu des contraintes de l'environnement, la *redondance* incluse dans un régime donné en *complexité* que traduit l'existence d'une trajectoire déterminée. Quand ce processus s'avère condamné du fait des limites propres à ce régime, on constate un épuisement de la trajectoire. On est ici très proche des analyses de H. Atlan.

De même, les travaux de Dosi, Orsenigo et Silverberg expliquent la création technologique comme rendue nécessaire par un phénomène d'épuisement des retombées lié à l'évolution le long d'une trajectoire. Pour une part, ces auteurs ne font que prolonger l'analyse

des précédents en faisant apparaître nettement la dimension de l'auto-organisation. Dans leur approche, l'accent est également mis sur le phénomène de *variété* microéconomique, vu comme le moteur de l'évolution (au lieu de la source externe de bruit du modèle d'Atlan), mais l'idée d'épuisement d'un certain stock de ressources internes évolutives (d'adaptation) rattache clairement l'approche au modèle de référence biologique.

Une telle théorie ne rend pas compte du processus de création lui-même dans la mesure où il reste nécessaire de recourir, compte tenu de sa définition même, au caractère exogène de la technologie, ainsi que le signalent M. Amendola et J.-L. Gaffard [1988]. Le renouvellement de la perspective théorique dans l'étude de la dynamique industrielle proposé par G. Dosi est important, mais il laisse subsister un certain nombre de problèmes:

(i) En premier lieu, il existe une indétermination quant au processus de constitution d'un nouveau paradigme technologique. Celui-ci naît d'une évolution des connaissances scientifiques et techniques dont le processus n'est pas décrit. De même, la nature de la transformation autorisant le passage d'un paradigme technologique à l'autre n'est pas explicitée (6).

(ii) En second lieu, la spécification des caractéristiques qui assurent la "fermeture" d'une industrie est absente des travaux de G. Dosi. Ainsi, on ne peut connaître le processus qui permet le maintien des trajectoires évolutionnaires dans les limites au-delà desquelles le processus d'explosion serait engagé. Si le paradigme technologique est l'instrument de la stabilité et de la non-divergence des trajectoires, le problème se déplace dans la mesure où il est alors nécessaire de spécifier ce qui produit la stabilité d'un paradigme. Pour échapper à cette exigence, il serait nécessaire de disposer d'une définition de l'industrie qui d'une part détermine les trajectoires viables et d'autre part assure leur existence.

Au total, on peut conclure avec M. Amendola et J.-L. Gaffard que les analyses "évolutionnistes", quel que soit leur mérite par ailleurs (elles renouvellent l'analyse économique en montrant le caractère peu pertinent des préceptes de la théorie économique standard en ce qui concerne l'étude des changements technologiques), souffrent d'une lacune théorique pour ce qui concerne le processus de *création* de nouveauté. Avec les outils analytiques hérités de Nelson et Winter, on dispose d'une théorie de l'évolution comme processus de sélection, mais on ne sait pas rendre compte de façon *endogène* de la *création technologique*.

3.2 La compétition technologique comme processus d'auto-organisation

W. B. Arthur considère que l'étude des changements technologiques doit se faire en analysant le *processus* par lequel une technologie *devient* prépondérante, plutôt que les conditions dans lesquelles une technologie supérieure s'impose (*technological competition*). Ce processus de compétition technologique est censé posséder quatre propriétés:

- *inefficience* globale compatible avec des choix individuels rationnels;

- *irréversibilité* des choix;
- *imprédictibilité* (la connaissance de paramètres tels que les conditions de l'offre ou de la demande est insuffisante pour prévoir le partage ultérieur du marché);
- *non-ergodicité* (ruptures brutales possibles comme conséquence d'un enchaînement de "petits événements").

W. B. Arthur [1988] relie aussi le processus de compétition entre technologies à l'existence de "rendements croissants d'adoption" d'une technologie. (D'autres contributions pourraient être citées à propos des effets d'auto-renforcement ou de rendements croissants d'adoption, en particulier P. David [1986] et R. Cowan [1988]; nous ne les présentons pas afin de ne pas alourdir ce texte).

Ces rendements peuvent, selon Arthur, avoir plusieurs sources, mais cinq d'entre elles lui paraissent particulièrement importantes:

- (i) le *learning by using* (N. Rosenberg [1982]): plus une technologie est pratiquée, plus on en apprend à son propos; (nous tenons à souligner ici que le processus auto-entretenu de développement et d'amélioration qui s'ensuit joue particulièrement à travers des interactions informationnelles client/fournisseur, équipementier/utilisateur, etc...);
- (ii) les *externalités de réseau* (M. Katz et C. Shapiro [1985]): l'adoption d'une technologie est renforcée par divers avantages liés au fait d'appartenir à son réseau d'utilisateurs;
- (iii) les *économies d'échelle dans la production*: une fois qu'une technologie est incorporée dans un produit, le coût de production de ce dernier diminue avec le nombre d'unités produites: c'est un phénomène classique d'indivisibilité (coûts fixes) qui explique que la technologie devienne plus attractive quand le niveau d'adoption s'accroît;
- (iv) les *rendements croissants d'information* : une technologie plus largement adoptée est mieux connue et mieux comprise (on peut aussi parler d'un effet d'image);
- (v) l'*interrelation technologique* (M. Frankel [1955]): au fur et à mesure qu'une technologie est adoptée, un certain nombre de sous-technologies et de produits deviennent partie intégrante de son infrastructure.

Sur la base de ces propriétés, W.B. Arthur suggère que la compétition technologique s'apparente à un processus de Polya (processus que l'on retrouve d'ailleurs chez F. Varela quand il caractérise le comportement propre d'un système autonome). Rappelons le modèle d'un tel processus: considérons une série de tirages aléatoires avec remises dans une urne contenant des boules de deux couleurs en proportions identiques au départ; à chaque tirage, on rajoute dans l'urne une boule supplémentaire de la même couleur que celle qui a été tirée; lorsque le nombre de tirages tend vers l'infini, la proportion de boules d'un type (donc également la probabilité de tirer une boule de ce type) tend vers une certaine limite et on montre que cette limite est une variable aléatoire uniformément répartie entre 0 et 1. Il est intéressant de constater, premièrement, que cette limite est quelconque (intuitivement, on aurait pu penser que le

processus converge obligatoirement vers 0 ou 1) et, deuxièmement, que le résultat final de chaque expérience est fortement influencé par les résultats des premiers tirages.

Appliquant ce processus à la compétition technologique, W.B. Arthur montre que:

- cette compétition ne se résoud pas nécessairement par l'adoption de la meilleure technologie au regard de critères instantanés de sélection;
- il est difficile, voire impossible, de prévoir quelle technologie sera adoptée à long terme;
- des "petits événements" sont susceptibles de faire basculer les choix (ici, nous pouvons faire référence à un paradigme interdisciplinaire qui n'est pas sans relation avec l'auto-organisation: la percolation (7)).

Pour W.B. Arthur, la compétition se solde par un effet de "*lock-in*" (verrouillage) quand une technologie devient prépondérante, donc quand le processus de Polya s'est stabilisé. Ce processus de compétition technologique est qualifié par l'auteur d'*auto-organisé* au sens de I. Prigogine et I. Stengers [1979 et 1988] et rend compte d'une caractéristique importante de la compétition entre technologies - que l'on peut identifier à un *processus chaotique*.

Pour évaluer la capacité de ce modèle de compétition à rendre compte de la *création* technologique, il faut considérer exactement ce qu'on appelle une *technologie*. Dans ses travaux, W.B. Arthur, à la suite de N. Rosenberg, considère qu'une technologie, à sa naissance, s'apparente plutôt à une potentialité. En effet, ce n'est qu'au cours de sa diffusion qu'elle se constitue concrètement: elle est l'objet d'une création continue tout au long de son existence et n'est en aucun cas un objet fini à son origine. Le modèle d'Arthur est donc bien, en un sens, un modèle de création technologique, mais on ne peut pas caractériser le phénomène par lequel une nouvelle technologie est introduite dans le jeu. On ne génère pas non plus ainsi les ruptures éventuelles du processus historique. En fait, l'espace au sein duquel se déploie la compétition technologique n'est pas spécifié: on ne connaît pas les conditions d'application du modèle concernant les technologies qui sont en jeu.

3.3 L'industrie comme système autonome

Les travaux sur l'industrie de D. Foray et P. Garrouste [1989] s'attachent à analyser les relations entre changements technologiques et modification des formes productives. Utilisant explicitement la théorie des systèmes autonomes de F. Varela, cette approche met en évidence:

- en ce qui concerne le concept d'*industrie*: les conditions d'émergence, au sein du système productif, de formes productives stables au moins pour une certaine période historique;
- en ce qui concerne la nature et la fonction des *changements technologiques*: l'existence d'un ensemble de contraintes liées au maintien de cette stabilité.

L'analyse de l'*autonomisation* de formes productives passe par la recherche de cohérences sur trois plans (D. Foray, 1987):

- (i) celui des *fonctions industrielles* (cela correspond à la manière dont une activité économique parvient à conférer aux besoins exprimés par la société une forme physique et marchande déterminée);
- (ii) celui de la *technologie* (ce qui permet de mettre en évidence les principes techniques qui orientent les changements technologiques);
- (iii) celui de la *structure du capital productif*.

Ces plans sont en correspondance avec les trois ensembles de caractéristiques de S. Metcalfe et P. P. Saviotti [1984] sur lesquels nous reviendrons en (3.4).

Cette analyse met l'accent sur les changements technologiques dont certaines orientations, à des moments déterminés de l'évolution, vont favoriser la constitution de formes objectives autonomes (des "industries"), puis piloter leur évolution en tant qu'entité désormais bien individualisées. La morphologie du système productif apparaît comme le résultat d'un processus historique - au terme duquel certaines activités de production ont revêtu des caractéristiques leur permettant de se développer de manière autonome, c'est-à-dire de créer des industries particulières. Celles-ci ne sont cependant pas considérées comme immuables: elles peuvent disparaître lorsque les conditions qui leur confèrent leur identité sont remises en cause. Ces conditions concernent, pour une activité donnée:

- les directions spécifiques d'évolution des techniques concernées;
- les fonctions industrielles remplies et la capacité d'organiser l'évolution des produits;
- le dimensionnement de la production et la captation des innovations génériques.

P. Garrouste [1984] montre sur l'exemple de l'industrie de la forge qu'il est possible de rendre compte de la dynamique d'un ensemble de procédés de production à partir de la caractérisation de relations spécifiques entre:

- (i) l'organisation du fonctionnement de l'ensemble des procédés techniques (caractéristique de la *rigidité* d'une filière technique)
- (ii) le différentiel de rendement technique de ces procédés (caractéristique de la *cohérence* de la filière)
- (iii) la *complexité* de l'ensemble formé par ces procédés.

Dans cette optique, une filière technique est définie comme un système autonome. On trouvera une analyse d'industrie (la fonderie) selon le même type d'approche dans D. Foray [1987].

L'approche de l'*industrie* en termes d'*autonomie* s'inscrit explicitement dans le courant initié par H. Maturana et F. Varela. Elle définit les conditions d'émergence et de maintien de formes productives stables compte tenu des orientations des changements technologiques. Elle permet d'exprimer l'éventail des variations technologiques qu'une industrie est susceptible de

supporter sans disparaître. Mais il est impossible de la solliciter pour expliquer le processus par lequel une technologie, définie sur la base de principes techniques radicalement nouveaux, se met en place. Elle ne peut que déterminer les conséquences de son introduction, au niveau de la structuration du système de production. Elle apparaît aussi, à ce titre, comme une approche plus normative que positive.

Cette conception qui reste encore programmatique peut cependant intégrer les phénomènes de rupture technologique en développant les bases d'une taxonomie technologique qui permette de spécifier les directions dans l'évolution des techniques (voir D. Foray et A. Grübler [1990] pour une application à l'industrie de la fonte).

3.4 La technologie comme système auto-organisé

P.P. Saviotti [1986] propose de dépasser les approches évolutionnistes en utilisant les résultats de la théorie des systèmes. Selon cet auteur:

"la meilleure façon de comprendre la relation entre changement technologique et développement économique consiste à séparer analytiquement ces deux groupes de phénomènes afin de mieux comprendre leurs interactions. Ainsi la séparabilité analytique du changement technologique ne doit en aucun cas être considérée comme s'inscrivant dans le cadre d'une théorie exogène du changement technologique, mais comme permettant de mieux comprendre l'interaction entre le changement technologique et le développement économique".

(op. cit. p.774).

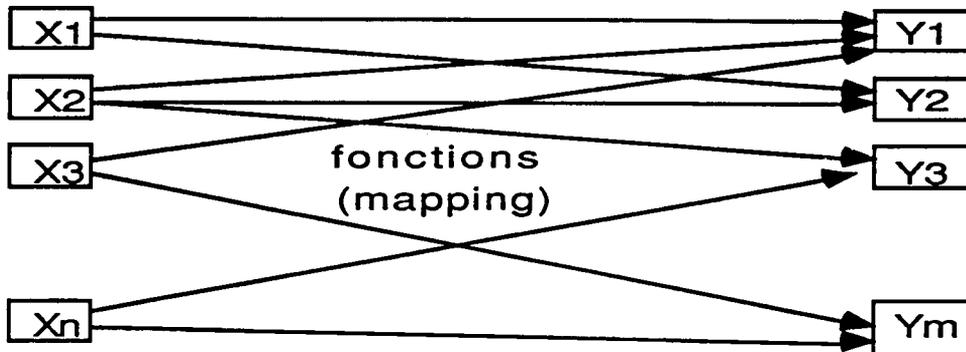
Cet article prolonge l'analyse de S. Metcalfe, P.P. Saviotti [1984], laquelle s'inscrivait dans la lignée des travaux de K.J. Lancaster [1966] et Z. Griliches [1971]. Metcalfe et Saviotti établissent une distinction entre trois ensembles reliés de caractéristiques qui seront appelées caractéristiques *techniques, de procédés et de services*.

"Un produit pourrait être décrit comme une combinaison de deux ensembles de caractéristiques, appelées respectivement caractéristiques techniques (Xi) et de services (Yi), et par une configuration de fonctions ("pattern of mapping") reliant les deux ensembles de caractéristiques mentionnés plus haut."

(S. Metcalfe, P.P. Saviotti [1984], p. 142).

Caractéristiques techniques

Caractéristiques de services



Cette représentation permet de définir les dimensions possibles d'un changement technologique, à savoir:

- un changement dans les valeurs absolues des X_i
- une modification des poids relatifs des X_i
- un changement dans la configuration des fonctions qui mettent en correspondance (X_i) et (Y_j)
- un changement dans le poids relatif des Y_j
- un changement des valeurs absolues des Y_j

Pour Metcalfe et Saviotti, cette approche permet d'améliorer la conceptualisation des processus de substitution technologique et de spécialisation ainsi que des phénomènes franchement qualitatifs comme ceux touchant à l'émergence de produits nouveaux (en fait, elle rend possible une représentation des relations entre environnement, innovations de produits et innovations de procédés). Elle aboutit aussi à la redéfinition des concepts de *régime* et de *trajectoire* technologiques, le premier étant considéré comme "une liste donnée de caractéristiques techniques X_i ", le second comme "une structure (pattern) d'amélioration des différents X_i au cours du temps." (op. cit., p.145).

Cette perspective est intéressante dans la mesure où elle permet une représentation explicite du changement technologique dans laquelle les caractéristiques techniques offrent une

représentation "de l'intérieur de la boîte noire". La technologie y apparaît comme un système complexe conforme à la représentation de H.A. Simon [1981]: possédant un "environnement interne", un "environnement externe" et un ensemble de buts. Les *caractéristiques techniques* d'une technologie en décrivent la structure interne et les *caractéristiques de services* définissent l'interface avec l'environnement efficace de la technologie (la *niche* dans la terminologie des éthologues).

Le problème est alors de savoir comment une technologie s'adapte à son environnement. Saviotti utilise ici la notion d'*homéostasie* pour montrer qu'un système ouvert qui se trouve dans un état stable tend à préserver cet état, en compensant les perturbations consécutives aux modifications de son environnement, ou à générer un nouvel état stable. Ainsi, tant que l'ampleur des modifications de l'environnement est inférieure à une limite donnée, l'évolution des variables internes est faible, sinon nulle. Au-delà de cette limite, le système est menacé de destruction.

P. P. Saviotti explique ainsi, dans le cadre du changement technologique, l'émergence de "guideposts", schèmes dominants susceptibles d'orienter des trajectoires technologiques: les apports de Nelson, Winter et Dosi apparaissent englobés dans l'approche de P.P. Saviotti [1986].

L'analyse du processus par lequel un système s'auto-régule revêt dans cette optique un caractère essentiel. Suivant les travaux de W.R. Ashby [1951] sur la variété requise pour le contrôle d'un système, Saviotti considère qu'une technologie s'adapte d'une part en *augmentant sa propre variété* et d'autre part en *réduisant celle de son environnement*. Ce processus opère grâce à l'existence de variables adaptatives qui permettent de protéger les variables essentielles du système en gérant les relations du système avec son environnement.

"Les changements au niveau des variables non essentielles, qui améliorent l'adaptation du système à son environnement, sont nécessaires pour maintenir la stabilité des variables essentielles et ainsi le régime technologique." (P.P. Saviotti [1986], p.781).

Ainsi, une technologie, en tant que système complexe, cherche à augmenter son rapport de variété avec son environnement en "internalisant des variables adaptatives et en réduisant la variété de l'environnement externe". (Ibid. p.783).

L'analyse de P.P. Saviotti (que l'on peut dissocier, en ce qui concerne notre propos, de celle qu'il mène en compagnie de S. Metcalfe), paraît critiquable sur le plan de sa capacité à expliquer le contenu des technologies nouvelles. En effet, on peut considérer que Saviotti inscrit ses travaux d'une part dans la logique des "approches en termes d'épuisement", et d'autre part dans celle de l'autonomie: il évalue la capacité d'une technologie à évoluer en modifiant à la fois son environnement et ses variables non essentielles. Il considère alors que l'émergence du nouveau en matière de technologie peut s'exprimer comme le résultat d'une difficulté à survivre

du système. A ce titre il ne peut, pas plus que les approches qu'il englobe ou "synthétise", rendre compte du contenu effectif de la nouveauté technologique. Il se limite à déterminer les conditions d'une rupture technologique et à la localiser dans le temps.

Malgré sa généralité - qui limite quelque peu sa pertinence - le modèle semble autoriser des développements intéressants à condition de préciser les relations entre ensembles de caractéristiques qui définissent une technologie. Il est probable que les applications seront plus convaincantes sur champs sectoriels très précis que sur des entités comme "la chimie" ou "les technologies de l'information". Ce qui attire notre attention en tout état de cause dans l'approche de Saviotti est que l'on y suppose la technologie capable d'imposer des formes dans le système socio-économique, et non pas seulement de s'adapter aux contraintes de son environnement.

3.5 L'évolution comme processus global d'auto-organisation

Pour D. Sahal, le processus de l'évolution technologique ne saurait être compris en faisant abstraction du fonctionnement global des sociétés. Cependant, "la logique de l'innovation technologique n'est pas entièrement réductible à des variables de nature socio-économique" (D. SAHAL [1981], p.14). Il faut tenir compte de certaines variables physiques, spécifiques du point de vue technologique, et de variables de nature purement économique. Cette conception permet de considérer que l'analyse économique peut prendre comme objet le phénomène du changement technologique - à condition que les résultats de cette appropriation restent compatibles avec ceux qu'obtiennent les autres types d'analyses propres à l'économie.

Cette approche, chez Sahal, est liée à une certaine conception de la réalité: une réalité qui serait comme "gouvernée" par la terminologie employée pour la décrire (8). Ceci permet à l'auteur de présenter une analyse *endogène* des changements technologiques et de critiquer les conceptions traditionnelles de la technologie, pour introduire une conception évolutionniste de l'activité innovatrice basée sur les notions de "guideposts" et "d'avenues" technologiques. Celle-ci est ancrée dans une approche systémiste de la technologie et s'appuie sur l'existence de configurations émergentes du point de vue de la conception des machines, qui déterminent l'éventail des évolutions techniques possibles.

De ce point de vue, les travaux de D. Sahal fournissent une approche intéressante et originale de la dynamique de l'auto-organisation. La notion proposée de *learning by scaling* correspond à l'image de la *croissance d'un organisme par stades successifs*: la croissance plus ou moins homothétique (à structure identique) est impossible au-delà de certains seuils en raisons de lois naturelles fondamentales (par exemple les non-linéarités dont l'archétype géométrique est le rapport surface/volume). Dans ce modèle d'auto-organisation, ce n'est pas un bruit exogène qui impose les restructurations, mais diverses nécessités logiques apparaissant

sur le chemin de la croissance. Le moteur premier reste la croissance (cette manière d'articuler l'extension des activités et le changement qualitatif n'est pas sans rappeler la dynamique du progrès technique et la division du travail d'A. Smith).

Bien que les apports de Nelson, Winter et Dosi s'apparentent à l'analyse des changements technologiques de Sahal, notamment à travers les concepts de paradigme et de trajectoire technologiques, il existe une différence notable dans la mesure où Sahal cherche à fonder théoriquement l'apparition de la nouveauté radicale (ses "guidepost" technologiques). Pour lui, c'est l'évolution elle-même qui est un *processus auto-organisé*.

La caractéristique essentielle des travaux de Sahal est qu'ils mettent en évidence le rôle déterminant de la *conception des machines*. On rentre ainsi dans le domaine de la réalité proprement technique des changements technologiques. Ce faisant, il nous semble qu'un fondement plus solide est conféré aux notions de *régime* et de *paradigme* en déterminant leur possibilité d'émergence au niveau technique. En ce sens on peut considérer que l'approche de Sahal *complète* les approches évaluées précédemment.

La question que l'on peut poser à une telle approche a trait à la logique d'évolution de la conception des machines, ce qui doit se traduire non plus par une analyse de la technologie comme système, par opposition à la vision traditionnelle, mais par la mise en évidence des relations entre artefacts. A défaut, on peut douter de la capacité d'une telle approche à rendre compte non pas seulement de la forme que prend le processus de création technologique mais aussi du contenu des nouveaux artefacts.

Nous concluerons cette partie en soulignant que les travaux qui rendent compte des changements technologiques en termes d'auto-organisation montrent tous la nécessité d'analyser la technologie comme un processus d'évolution et non pas comme un ensemble d'informations techniques. Ce processus - et en particulier ses points de rupture - viennent ainsi au centre de l'analyse; il est qualifié d'auto-organisateur dans la mesure où il se présente comme un bouclage récurrent interne au système étudié, bouclage qui prend différentes formes selon la manière dont on définit le système supposé pertinent pour l'analyse du changement technologique - et c'est précisément sur ce point que diffèrent les approches.

- Pour W.B. Arthur, le système pertinent est celui de la *compétition technologique*, au sein duquel un phénomène d'auto-renforcement crée des conditions de rupture par effet de *lock-in*. L'existence de cet effet amène à relativiser fortement l'idée d'espace concurrentiel, en le rendant tout à fait contingent.

- Pour G. Dosi *et al.*, C'est le mouvement de *déploiement le long d'une trajectoire technologique* qui constitue l'objet pertinent de l'analyse des changements technologiques, donc une forme d'auto-organisation qui est contrainte par l'épuisement d'une capacité de redondance. Cette capacité du système à réagir (à apprendre) dans le cadre d'un couplage avec l'environnement se situe au coeur de l'analyse.
- L'approche de P.P. Saviotti, au contraire, met l'accent sur la capacité de la technologie à imprimer sa marque sur l'environnement. Ce dernier est conçu non plus comme le lieu de production de perturbations auxquelles le système répond en se réorganisant, mais comme un espace malléable qu'il est possible de structurer.

Cette dernière approche nous paraît particulièrement pertinente. Elle correspond plutôt au modèle d'auto-organisation de F. Varela, où l'environnement du système est d'une certaine manière "sélectionné" par le système autant qu'il le sélectionne:

"Si l'on considère l'environnement du système également comme un système structurellement plastique, alors le système et son environnement auront une histoire imbriquée de transformations structurelles, chacun sélectionnant les trajectoires de l'autre" (F. Varela [1979], p. 33).

Cette manière d'envisager le mouvement d'auto-organisation technologique correspond bien, à notre avis, au cadre dans lequel il faut traiter les stratégies technologiques - lesquelles sont au coeur du processus d'innovation.

4. Auto-organisation et stratégie d'innovation

Si le modèle de l'auto-organisation prend une place croissante sur l'ensemble du champ de l'économie, c'est particulièrement vrai en économie du changement technique comme nous espérons l'avoir montré ci-dessus. Mais il faut en outre souligner un fait important: la technologie est de plus en plus au centre des *stratégies* industrielles. Le progrès technique étant un apprentissage (sauf à caricaturer ce processus en exogénéisant totalement l'invention), on retrouve bien les deux dimensions de l'innovation comme phénomène d'auto-organisation:

- *auto-information*, car l'apprentissage effectif suppose la matérialisation de l'information nouvelle dans la firme (le "développement" d'un processus, d'un produit ou d'un marché nouveau correspond à une véritable mutation organique de la firme);
- *mise en forme de l'environnement* (stratégie), car l'innovation n'est réussie que lorsque la firme a imposé son nouveau standard dans une partie du système global - son environnement pertinent.

N. Rosenberg [1982] a montré que l'innovation se constitue très souvent au niveau de l'interaction des agents: le *learning by using* se distingue du *learning by doing* de K. Arrow en ce que l'apprentissage est autant le fait de l'utilisateur que du producteur, et la capacité pour le

producteur d'assimiler cette information en retour est un facteur crucial de progrès. E. Zuscovitch [1985] a montré à propos de la diffusion intersectorielle des matériaux synthétiques que l'imitation, lors de la phase de diffusion d'une idée nouvelle, constitue un véritable processus d'apprentissage; l'innovation se présente alors comme un phénomène progressif, de caractère méso plus que microéconomique. On pourrait donner encore bien d'autres exemples concrets à l'appui de la thèse selon laquelle le changement technique est plutôt un phénomène de réseau: une *auto-organisation du système technico-économique* (9).

Comme on le voit, introduire (ou redécouvrir) le principe d'auto-organisation en sciences économiques revient à approfondir la réflexion sur les modes de coordination des agents. Entre les modèles extrêmes (et forcément réducteurs) du marché et de l'organisation, doit s'inscrire une nouvelle approche, qui élabore le contenu possible d'un concept d'apprentissage des systèmes, qui mette l'accent sur l'émergence des formes institutionnelles et des conventions (y compris bien sûr la technologie) en les considérant non pas comme une donnée mais comme un acquis progressif du système.

L'interface des champs économique et technologique est certainement le lieu crucial du développement capitaliste. Nous savons depuis A. Smith que le ressort du progrès réside dans l'enchaînement d'un processus économique (accumulation du capital et extension des marchés) sur un processus organisationnel et technologique (division du travail et rationalisation de la production). La *firme* apparaît au cœur de ce système évolutif et sa fonction est double: en mettant en œuvre la division du travail elle s'auto-organise *pour elle-même*, mais à travers la création technologique elle contribue à l'organisation du système global, elle *in-forme* son environnement.

Notre lecture de la littérature consacrée à l'auto-organisation, ainsi que l'ensemble des réflexions d'où est issu le présent ouvrage, nous a convaincu qu'il n'était pas logiquement cohérent, au moins en sciences sociales, de chercher à décrire l'auto-organisation d'une entité sans faire référence au processus d'évolution co-déterminé dans l'environnement de l'entité. Dans le cas du sous-système que constitue la firme, il est par exemple peu satisfaisant de chercher à décrire son processus d'auto-organisation de manière indépendante de celui de son environnement. D'une part, on sait bien que, de plus en plus, certaines firmes contribuent massivement à infléchir l'évolution de leur *industrie*. D'autre part, il faut toujours garder à l'esprit que, fondamentalement, l'environnement pertinent d'un agent est la *perception* qu'il en a. Ce n'est que par facilité opératoire que l'on construit, dans les modèles standard, une réalité extérieure définie de manière indépendante.

Cette conception de la firme comme entité auto-organisée et auto-organisatrice correspond à une théorie particulière de l'innovation. On ne peut pas se contenter de décrire le

changement technologique comme la résultante d'un processus de sélection/diffusion, car ce serait adopter une vision darwinienne pure où la source de bruit (la variation des gènes ou des routines) est posée comme extérieure à la problématique. C'est là, nous semble-t-il que réside la principale limite à l'exploitation de paradigmes évolutionnistes issus de la biologie. La réalité économique est autant lamarkienne que darwinienne. Elle dépasse même cette distinction: non seulement les modifications utiles sont tout de suite inscrites dans les routines, mais les agents font fréquemment des conjectures sur l'évolution du système et intègrent ces représentations dans leur stratégie, ce qui contribue en retour à former cette même évolution. De plus, un certain nombre d'entre eux savent que leur action n'est pas de poids nul sur l'avenir du système global, ou au moins celui de leur environnement pertinent. C'est bien de cela qu'il s'agit lorsqu'on parle d'action *stratégique* - au vrai sens du terme. Or, c'est cette dimension stratégique de l'auto-organisation technico-économique qui fait la différence avec les modèles utilisables en sciences de la nature. Bien sûr, un tel modèle d'auto-organisation doit être extrêmement difficile à formaliser, ce qui explique que l'on soit encore loin de disposer d'outils satisfaisants en économie du changement technologique. Il ne suffit pas d'entrer dans la "boîte noire de la technologie", il faut en plus intégrer les "boîtes noires stratégiques" de nombreux décideurs. L'auto-organisation propre à l'économie du changement technologique est donc, pour nous résumer, non seulement tissée dans l'interface micro/macro, mais elle résulte aussi d'un apprentissage *stratégique*, ce qui la différencie totalement des modèles d'évolution systémique en sciences de la nature.

Notes

(1) Voir en particulier les Actes du Colloque de Cerisy (P. Dumouchel, J.-P. Dupuy [1983]) et le numéro spécial d'Economie Appliquée (cité sous J. Lesourne [1985]).

(2) En fait, les aller-retour sont incessants entre agent et système. Pour ne prendre qu'un exemple (J.-A. Héraud [1988b]) : l'accroissement de l'instabilité environnementale en matière d'énergie a amené les firmes à adopter un nouveau comportement qui est celui de la polyénergie; en retour, l'existence d'équipements polyvalents a entraîné une innovation institutionnelle: le développement des marchés spots de l'énergie. Ce nouvel environnement marqué par la variété et l'incertitude a par ailleurs incité certains agents qui n'y trouvent pas leur intérêt à rechercher des formes de contrats particuliers pour échapper à la forme du marché. On retrouve ici deux ingrédients des systèmes auto-organisés: (i) des effets en retour - non pas programmés une fois pour toutes dans le système, mais reposant sur un apprentissage conjoint des agents et du système; (ii) le développement de formes nouvelles d'interactions entre agents, en dehors du cadre institutionnel habituel.

(3) Le problème terminologique réside dans l'usage habituel (à la suite d'auteurs comme Piaget) des mots "système" et "structure" qui pousserait à les utiliser à la place, respectivement, de "structure" et "organisation" dans l'acception de Maturana.

(4) Pour une récente mise au point, voir : P. Watzlawick (ed.), "L'invention de la réalité. Contributions au constructivisme", Seuil, 1988.

(5) Ceci est à rapprocher de l'inspiration que Prigogine et Stengers trouvent dans la théorie des catastrophes de R.Thom. Cependant, là où ce dernier cherche à réintroduire une vision déterministe en élaborant une méthode qui permette de répertorier toutes les catastrophes possibles, les premiers affirment une conception de la réalité où le radicalement nouveau peut être pensé comme résultant du mouvement propre d'un système.

(6) L'indétermination quant à l'existence ou non d'une possibilité de concurrence entre deux paradigmes peut être rapprochée de la confusion entretenue entre la notion de paradigme introduite par T. Kuhn et celle de programme de recherche scientifique due à I. Lakatos.

(7) Un des premiers articles de synthèse sur ce sujet est celui de P. Gille de Genne: "La percolation: un concept unificateur", *La Recherche*, novembre 1976.

(8) Une telle conception de la réalité peut être rapprochée de celle que développe J. Hamburger [1984], et qu'il relie au concept de césure : "discontinuité qui interdit au chercheur d'unifier totalement les résultats qu'il obtient sur le même objet à des échelles et avec des méthodes différentes." (p.40).

(9) Cette observation renvoie à une question: celle de l'autonomie de la sphère technique. R. Ayres [1988] arrive

à la conclusion que, s'il y a autonomie, cela ne peut concerner que l'ensemble (économie + technosystème). Ajoutons que, déjà privés ainsi de déterminisme technologique, nous risquons de ne pas avoir non plus de déterminisme technico-économique strict: les systèmes complexes possèdent de multiples possibilités d'auto-organisation à chaque étape de développement; tout au plus pourrait-on anticiper les "attracteurs" de cette évolution "chaotique", c'est-à-dire décrire des scénarios d'évolution sans possibilité de trancher (J. Lesourne dit à propos du processus d'auto-organisation qu'il sélectionne sa trajectoire d'équilibre dans un ensemble vaste de futurs possibles). Pour reprendre la classification de B. Walliser que nous avons citée en entrée, les modèles systémiques se distinguent des constructions traditionnelles des économistes en remplaçant l'équilibre unique par un ensemble d'équilibres possibles définis en termes d'anticipations autoréalisatrices. La notion de cohérence des visions des agents devient centrale plutôt que celle d'optimum, ce qui exclut le déterminisme absolu (économique comme technologique) et implique une analyse en termes de scénarios possibles ("futuribles") sur le champ global technico-économique.

BIBLIOGRAPHIE

- AMENDOLA L., GAFFARD J.L. [1988], *"La dynamique économique de l'innovation"*, Economica.
- ARTHUR W.B. [1988], "Competing technologies: an overview", in G. Dosi et al., *"Technical change and economic theory"*, Pinter.
- ATLAN H. [1972], *"L'organisation biologique et la théorie de l'information"*, Hermann.
- ATLAN H. [1979], *"Entre le cristal et la fumée - essai sur l'organisation du vivant"*, Seuil.
- ATLAN H. [1986], "Créativité biologique et auto-crédation du sens", in *"Cognition et complexité"*, Cahiers du C.R.E.A., n°9, pp. 145-189.
- AYRES R. [1988], *"Self-Organization in Biology and Economics"*, International Institute for Applied System Analysis, Laxenburg (Austria), janvier.
- COWAN R. [1988], *"Nuclear Power Reactors: a study in technology lock in"*, New York University (non publié).
- DOSI G. [1988], "Sources, procedures, and microeconomic effects of innovation", *Journal of economic Literature*, vol. XXVI, pp.1120-1171.
- DOSI G., ORSENIGO L. [1988], "Structure industrielle et évolution technologique, in A. Heertje (ed.), *"Innovation, technology and finance"*, Blackwell.

DAVID P. [1986], "Understanding the economics of QWERTY: the necessity of history", in PARKER (ed.) "*Economic history and the modern economist*", Basil Blackwell.

DREYFUS H.L. [1979], "*What computers can't do: a critique of artificial reason*", Harper and Row.

DUMOUCHEL P., DUPUY J.P. [1983] (eds.), "*L'auto-organisation. De la physique au politique*", Actes du Colloque de Cerisy, Seuil.

FLORES F., WINOGRAD T. [1989], "*L'intelligence artificielle en question*", P.U.F.

FORAY D. [1987], "*Innovations technologiques et dynamique industrielle: l'exemple de la fonderie*", Presses Universitaires de Lyon.

FORAY D., GARROUSTE P. [1989], "*Dynamique du changement technologique et stabilité des formes productives*", Journées européennes de l'AFSE, Strasbourg.

FORAY D., GRÜBLER A. [1990], "Morphological Analysis, Diffusion and Lock-Out of Technologies: Ferrous Casting in France and in the FRG", *Research Policy*, vol.19, n°6, pp. 535,555.

FRANKEL M. [1955]. "Obsolescence and technological change in a maturing economy", *American Economic Review*, vol. 45, pp.1296-131

GARROUSTE P. [1984], "*Filières techniques et économie industrielle: l'exemple de la forge*", Presses Universitaires de Lyon.

GRILICHES Z. [1971], "*Price indexes and quality change*", Harvard University Press.

HAMBURGER J. [1984], "*La raison et la passion, réflexions sur les limites de la connaissance*", Seuil.

HERAUD J.A. [1988a], "Routine et innovation dans un système économique auto-organisé", *Fundamenta Scientiae*, Vol 9, 2/3.

HERAUD J.A. [1988b], "La polyénergie: un exemple de transmission des contraintes de flexibilité", in P.COHENDET et P.LLERENA (éds.), "*Flexibilité, information et décision*", Economica.

KATZ M., SHAPIRO C. [1985], "Network externalities, competition and compatibility", *American Economic Review*, vol.75, pp. 424-440.

LANCASTER K. [1966], "A new approach to consumer theory", *Journal of Political Economy*, vol. 74, pp. 132-157.

LESOURNE J. [1985], "A la recherche d'une théorie de l'auto-organisation dans les sciences sociales", *Economie Appliquée*, XXXVII, 3/4, pp. 559-567.

MATURANA H. [1974], "Stratégies cognitives", in E. Morin, M. Piatel-Pamarini (eds.), "*L'unité de l'homme; 2. Le cerveau humain*", Seuil.

METCALFE S., SAVIOTTI P.P. [1984], "A theoretical approach to the construction of technological output indicators", *Research Policy*, 13, pp. 141-151.

NELSON R., WINTER S. [1982], "*An Evolutionary Theory of Economic Change*", Harvard University Press.

PASSET R. [1979], "*L'économie et le vivant*", Payot.

PRIGOGINE I., STENGERS I. [1979], "*La nouvelle alliance, métamorphose de la science*", Gallimard.

ROSENBERG N. [1982], "*Inside the Black Box: Technology and Economics*", Cambridge University Press.

SAHAL D. [1981], "*Patterns of technological innovation*", Addison-Wesley.

SAHAL D. [1983], "Invention, Innovation and Economic Evolution", *Technological Forecasting and Social Change*, N°23, pp. 213-235.

S AVIOTTI P. P. [1986], "Systems theory and technological change", *Futures*, pp. 773-786.

SILVERBERG G., DOSI G., ORSENIGO L. [1988], "Innovation, Diversity and Diffusion: A Self-Organization Model", *The Economic Journal*, 98, décembre, pp. 1032-1054.

VARELA F.J. [1979], "*Principles of biological autonomy*", North-Holland.

VARELA F.J. [1989a], "*Autonomie et connaissance: essai sur le vivant*", Seuil.

VARELA F.J. [1989b], "*Connaître les sciences cognitives, tendances et perspectives*", Seuil.

WALLISER B. [1988], "Systémique et économie", *Revue Internationale de Systémique*, Vol. 2, N°3, pp. 245-260.

WITT U. [1985], "Coordination of Individual Economic Activities as an Evolving Process of Self-Organization", *Economie Appliquée*, XXXVII, 3/4, pp. 569-595.

ZUSCOVITCH E. [1984], "La diffusion intersectorielle des matériaux synthétiques: évolution et bilan économique" in P. COHENDET (ed.), "*La chimie en Europe*", Economica.

Ainsi que nous l'avons montré dans la première partie de cet ouvrage, les conceptions de l'apprentissage de H. Atlan et de I. Rosenfield se rejoignent sur trois points liés: 1) il s'agit d'un apprentissage non dirigé, au sens où il ne répond à aucun programme préétabli, ni dans le système considéré, ni dans l'environnement de ce système; 2) la cause efficiente de cet apprentissage réside en la rencontre aléatoire d'un système et de certains facteurs en provenance de son environnement; 3) si le produit instantané de l'apprentissage consiste en une catégorisation des stimuli perçus, son déroulement temporel voit se succéder des catégories toujours plus fines et différenciées, les règles de construction de celles-ci étant en outre susceptibles d'être modifiées à chaque étape de ce processus. C'est sur de telles bases que la présente contribution propose un modèle formel d'apprentissage, avec création d'information au niveau global d'un système économique défini comme une société de systèmes cognitifs individuels, pour lesquels le problème du choix se pose dès lors de manière relativement inédite (1).

L'Economie informationnelle ainsi entendue incorpore une conception de la communication sociale esquissée naguère par M. Serres:

"Imaginons, dessiné dans un espace de représentation, un diagramme en réseau. Il est formé, pour un instant donné (car nous verrons amplement qu'il représente un état quelconque d'une situation mobile), d'une pluralité de points (sommets) reliés entre eux par une pluralité de ramifications (chemins) (...) Par définition, aucun point n'est privilégié par rapport à un autre, aucun n'est univoquement subordonné à tel ou tel; ils ont chacun leur puissance propre (éventuellement variable au cours du temps) (...) Et par conséquent, quoique certains puissent être identiques entre eux, ils sont, en général, tous différents. De même pour les chemins, qui, respectivement, transportent des flux de déterminations différents, et variables dans le temps." [1968, p. 11].

Si la formalisation de l'économie en termes de graphes ou de réseaux n'est guère nouvelle dans la littérature propre à notre discipline, celle-ci se contente souvent de représenter les agents comme des sommets dénués de structure interne - comme de simples points qui sont autant de "boîtes noires" pour la théorie. Quant aux chemins ou aux arcs du graphe, ils se bornent traditionnellement à visualiser des transmissions d'informations effectuées entre agents sur un mode identique à celui d'un transfert de substance, ou à dessiner le parcours offert a priori à des trajectoires d'agents accédant séquentiellement à des partitions de plus en plus fines d'un espace fini d'états du monde possibles - ainsi dans les modèles de marchés contingents à la Arrow-Debreu (2). S'agissant de résoudre certains types de problèmes, l'utilité et la pertinence de telles représentations ne saurait être niée. Envisagée sous l'angle de sa contribution à la décision efficace de l'agent individuel, ou des effets de l'asymétrie de sa répartition parmi les agents sur les paiements d'un jeu, ou encore de sa distribution optimale au sein d'un système macroéconomique, la substance informationnelle n'a pas réellement à recevoir de contenu sémantique précis, car c'est surtout sa mesure qui importe: cette mesure suffit pour que la théorie puisse étudier les conditions d'optimalité de son acquisition ou de sa diffusion. Dans cette problématique calculatoire, l'espace des représentations possibles des agents est par définition toujours *donné*, ou modifié selon une procédure d'apprentissage bayésien dont le contenu reste en lui-même *exogène* à la théorie (3), de sorte que l'analyse est celle d'une recherche d'optimalité individuelle ou collective dans le cadre d'un espace *fermé*.

De manière complémentaire à la précédente, une seconde direction de recherches se propose depuis peu d'*ouvrir* pareil espace, en étudiant la problématique de l'évolution économique à l'aide du concept d'information et de ceux qui lui sont associés - notamment celui de complexité [cf. R. Passet, 1979, 1980, 1985, 1987]. Un aspect de ce programme consiste explicitement en l'étude de la dynamique historique des représentations possibles d'agents pour lesquels, suggère A. P. Kirman [1986], il ne s'agit pas tant de maximiser que, précisément, d'apprendre. Il est donc impossible de se contenter ici d'un formalisme en termes de graphes dont les sommets seraient dénués de structure interne, ce qui équivaut à dire que les agents doivent être représentés comme étant eux-mêmes des systèmes. Corollairement, suivant la recommandation très générale de G. Bateson [1977, 1980, 1984], ce type d'approche requiert de concevoir l'information dans le cadre d'une épistémologie de la *forme* (où elle obéit à des lois de nature *combinatoire*) plutôt que de la *substance* (où elle relève de lois *additives*).

C'est dans cette seconde perspective que se situe la présente contribution. Pour nous, l'espace de représentation évoqué par M. Serres est celui d'un système auto-organisateur dont chaque état est représenté par un graphe dont les sommets figurent ces sous-systèmes que sont les agents en tant que sujets connaissant individuels. Quant à la pluralité de ramifications ou de chemins mentionnée par ce même auteur, elle possède ici une double nature: celle d'un ensemble de communications *possibles* entre les agents, étant donné l'état du système; celle d'un ensemble de communications *réalisées* faisant passer le système de chaque état donné à un état ultérieur, et modifiant ainsi dans la durée du processus d'apprentissage la configuration même qui les avait rendues possibles. C'est dans ce cadre global que nous proposons une théorie de l'apprentissage intégrant simultanément l'incertitude dans laquelle les agents prennent leurs décisions, et l'irréversibilité des conséquences de celles-ci (4).

Nécessairement liées, l'irréversibilité des décisions et l'incertitude dans laquelle elles sont prises traduisent deux aspects complémentaires de l'écoulement du temps, et chacune contribue ainsi à installer chez les agents le sentiment de la durée: à chaque instant, la première leur rappelle le poids du passé par leur incapacité à modifier les conséquences non désirées de décisions qu'ils peuvent regretter d'avoir prises, tandis que la seconde les assure à l'avance qu'il en ira vraisemblablement de même à l'avenir. Ici comme là, ce qui éveille d'éventuels regrets parmi les agents provient d'un seul et même phénomène: leur capacité à engendrer des sens nouveaux en créant de l'information. Appuyant sa cohérence externe sur la théorie rosenfieldienne de la mémoire, notre modèle, comme le paradigme de l'auto-organisation de H. Atlan, vise à expliquer l'émergence du nouveau.

Nous effectuerons cependant une double modification de champ par rapport à ces deux types de contributions: notre perspective sera macroscopique autant que microscopique, de sorte que nous envisagerons un domaine plus étendu que celui de I. Rosenfield qui se cantonne au plan individuel. Par ailleurs, nous analyserons l'émergence de sens nouveaux parmi une population dont l'élément générique est l'*homo computens*, et restreindrons ainsi

le champ par rapport à celui de H. Atlan, qui considère les systèmes non construits en général.

Pour ce faire, nous diviserons notre texte en deux parties. La première vise à définir formellement l'espace des agents de telle sorte que ces derniers puissent évoluer sur un mode rosenfieldo-atlanien, donc à présenter un modèle d'apprentissage créateur de sens nouveaux (1). La seconde partie esquisse, dans le cadre analytique ainsi tracé, la problématique du choix individuel (2).

1 Apprentissage et création d'information

Malgré une conception de l'information plus satisfaisante que celle des modèles considérant cette dernière comme une substance entièrement détachée du sujet connaissant, une première formalisation d'un état et de la dynamique du système révélera vite ses limites, en se montrant incapable de concevoir la signification et la création de l'information (1.1). Nous serons ainsi amené à déplacer notre niveau d'analyse en situant l'information dans le monde de la forme, ainsi que nous y invite G. Bateson - un monde dont le caractère combinatoire des lois débouche sur une solution acceptable de ce double problème (1.2).

1.1 Première approche d'un état du système et de ses transformations

Notre objet d'analyse consiste ici en un système Z, composé d'éléments qui sont des agents économiques individuels doués de mémoires "rosenfieldiennes", et d'une structure consistant en l'ensemble des communications virtuelles, ou réellement établies entre ces agents. Un état de Z est formalisé par la donnée d'une matrice (m,n), dont les vecteurs-lignes représentent les agents A_i , indicés $i = 1, \dots, m$, et les vecteurs-colonnes, les catégories C_j , indicées $j = 1, \dots, n$, construites par ces agents. Soit M l'ensemble des agents, dont chacun est supposé différent de tous les autres (c'est-à-dire singulier), et S l'ensemble des catégories.

En accord avec la théorie de I. Rosenfield [1989], l'on serait tenté d'affirmer sans autre précision que chaque agent construit de telles catégories par abstraction à partir des stimuli perçus par lui. En réalité, le fait que ces agents mettent en oeuvre leurs fonctions cérébrales supérieures lors de communications mobilisant la parole et le langage implique de mieux détailler le contenu du processus de catégorisation qui se trouve ainsi en jeu. Car ce contenu est indissociable de la conscience qu'ont les agents d'eux-mêmes, ainsi que d'un type de mémoire propre à l'être humain. Précisons rapidement ce point, développé par I. Rosenfield [1990]. Selon celui-ci, la notion de "je" est plus abstraite que celle de "il", car la conscience que chaque sujet humain a de sa propre action résulte d'une abstraction, et non d'une observation directe. Ainsi,

"Tandis que je cours, je peux (...) abstraire d'un certain nombre d'indices (la sensation du mouvement de mes jambes, à laquelle s'ajoute le changement de décor autour de moi) la pensée que 'je suis en train de courir'. La conscience que j'ai de ma propre action n'est pas le fruit d'une observation directe; c'est une abstraction. Elle peut d'ailleurs découler en partie des observations que j'ai réalisées sur d'autres personnes en train de courir."
[I. Rosenfield, 1990, pp. 11-12].

Ce type d'abstraction implique que les catégories créées par notre cerveau sous forme de "concepts prélinguistiques" proviennent du développement chez l'homme d'aires cérébrales pouvant caractériser les activités cérébrales elle-mêmes [ibid., p. 74]. La conscience auto-réflexive naît alors de la "réintroduction" entre catégories perceptives et activité de catégorisation de sa propre activité par notre cerveau [ibid., p. 76]. La catégorisation ainsi effectuée par l'homme se situe donc à un niveau d'abstraction supérieur à celui impliqué par le type de catégorisation dont est capable l'animal en général. Corollairement, à la mémoire à court terme dont se trouve pourvu ce dernier s'ajoute, chez l'homme, une mémoire à long terme, supportée par l'existence du langage [ibid., pp. 82 sq.].

Bien qu'elles haussent d'un cran le niveau logique auquel nous devons situer les catégories proprement langagières, pareilles précisions ne modifient en rien l'interprétation globale de notre fonctionnement cérébral comme opérant des catégorisations: elles indiquent simplement que la structure de la mémoire humaine comporte un niveau logique de plus que celle d'autres espèces animales. Au-delà des seules catégories perceptives immédiates que créent les membres de chacune de ces dernières, et de leur aspect fugitif lié à la notion de mémoire à court terme, l'homme crée des concepts prélinguistiques comme autant de précurseurs des catégories du langage qui viennent supporter chez lui la constitution d'une mémoire à long terme. La structure même de la conception rosenfieldienne du fonctionnement de la mémoire ne se trouve nullement altérée par son extension aux fonctions cérébrales gouvernant la conscience et le langage. D'un point de vue purement fonctionnel, il est donc possible de nous en tenir ici au terme général de "catégorisation" pour décrire un processus dont l'activité cérébrale propre à l'homme, si extraordinaire soit-elle, ne constitue qu'une modalité particulière [cf. I. Rosenfield, 1990, p. 107].

Retenons alors simplement qu'à toute date considérée chaque agent "connaît" un certain nombre de catégories. Il est vraisemblable que ce nombre est fini. Soit $S_i(t)$, l'ensemble des catégories ainsi construites par l'agent A_i à la date t . $S(t)$ représente alors l'union de tous les ensembles de ce type à la même date - union que seul l'observateur du système (identifié au niveau englobant celui des sous-systèmes individuels) peut connaître exhaustivement. Insistons sur le fait que $S(t)$ est construit *à partir* des $S_i(t)$, non l'inverse. Quel que soit t , il est ainsi clair que chaque élément C_j de S est, ou n'est pas, élément d'un ensemble S_i donné, et il s'ensuit qu'à cette date, tout coefficient a_{ij} de la matrice (n,m) , désormais notée $[a_{ij}](t)$, peut prendre, respectivement, les valeurs 0 ou 1. Ainsi, $S_i(t)$ n'est rien d'autre que l'ensemble des catégories C_j telles que $a_{ij}(t) = 1$.

Chaque état du système Z apparaît donc défini par trois caractéristiques : 1) le cardinal m de l'ensemble M d'agents A_i ; 2) le cardinal n de l'ensemble S de catégories C_j ; 3) la répartition des coefficients $a_{ij} = 1$ dans la matrice $[a_{ij}]$. Ce système étant composé d'éléments qui sont autant de sous-systèmes cognitifs, Z représente la mémoire globale de l'économie.

A chaque date considérée, le *volume* de cette mémoire est défini par les caractéristiques 1) et 2), et sa *structure*, par la caractéristique 3). Notre problème immédiat est de préciser suffisamment ces caractéristiques pour que *Z* ait un caractère de système *organisé*, sans que de telles précisions ne viennent restreindre par trop la généralité du modèle.

Convenons ainsi de fixer la première caractéristique, en posant $m = \text{constante}$. Dans ces conditions, si nous supposons que nous ne savons strictement rien de la troisième caractéristique à la date t , la quantité d'information contenue dans *Z*, notée $H(t)$, devrait être considérée comme étant maximale. En effet, une telle méconnaissance signifierait que, hormis leur existence et leur nombre, nous ignorons tout des m agents éléments de *Z*. Il nous faudrait alors considérer ces éléments comme étant dénués de toute structure interne, n'intervenant pas au titre de véritables sous-systèmes dans l'analyse de *Z*, mais seulement comme autant de points sans dimension. De sorte que la caractéristique 2) ne pourrait être prise en compte dans la mesure de la quantité totale d'information contenue dans *Z*, car, rappelons-le, $S(t)$ est défini à partir des $S_i(t)$, et non l'inverse. La complexité du système serait alors mesurée par sa variété au sens de Ashby: $H(t) = \log_2 m$, de sorte que la redondance de *Z* à la date t , notée $R(t)$, serait nulle. En fait, tant que nous n'avons aucune information relative à la répartition des a_{ij} non nuls dans $[a_{ij}](t)$, le système ne peut présenter à nos yeux aucune propriété auto-organisatrice - *Z* n'est même pas un système organisé, notion dont nous savons qu'elle suppose l'existence de *relations* entre les éléments du système considéré.

Réciproquement, nous donner une répartition des a_{ij} dans $[a_{ij}](t)$ telle que tous les agents soient parfaitement identiques ($a_{ij} = 1, \forall i, \forall j$) détruirait également l'organisation du système. *Z* dégénérerait en une réplique de m sous-systèmes strictement semblables, dont on peut facilement admettre qu'ils n'auraient alors absolument rien à se communiquer l'un à l'autre: en pareil cas, $H(t)=0$ et $R(t)=1$. A notre mutisme concernant les communications *possibles* entre agents lorsque $H(t) = \log_2 m$ et $R(t) = 0$ - un mutisme de droit qui ne serait levé de fait qu'au prix du parfait arbitraire avec lequel nous déciderions que tels agents *établissent* une communication à la date t - succéderait ainsi notre totale incapacité à analyser des communications *réellement établies* entre agents - toutes les communications étant également possibles entre eux, mais toute communication établie ayant une productivité nulle en termes d'apprentissage lorsque $H(t) = 0$ et $R(t) = 1$. Dans la mesure où nous nous proposons précisément d'analyser les conditions de possibilités de l'établissement de telles communications entre les agents, ainsi que les effets sur *Z* de ces communications établies, nous devons donc éviter ces deux cas extrêmes dans notre représentation de $Z(t)$, en introduisant une certaine répartition des a_{ij} dans $[a_{ij}](t)$: il faut que notre description de l'état initial de *Z* imprime une structure à celui-ci, mais, répétons-le, cette structure doit être suffisamment pauvre pour que notre modèle demeure le plus général possible.

Suivant les développements contenus dans notre texte précédent, les effets de la communication concernent aussi bien la transmission de l'information dans une voie élémentaire établie entre deux ou plusieurs agents à l'intérieur du système, que celle qui est

transmise à l'observateur par le système considéré comme un tout. Construisons une telle analyse en commençant par nous interroger sur les conditions de possibilités de l'établissement d'une communication entre deux agents quelconques à une date donnée, car nous allons voir qu'une telle démarche débouche immédiatement sur une meilleure spécification de Z , qui en acquiert statut de système organisé.

Il est clair en effet que deux agents quelconques ne peuvent établir de communication qu'à la condition expresse de disposer d'un minimum de langage commun. Formellement, nous conviendrons qu'à la date t , une communication n'est possible (sa probabilité a priori n'est pas nulle) entre les agents A_i et A_k que si $S_i(t)$ et $S_k(t)$ contiennent *au moins un* élément commun - le minimum évoqué à l'instant (5). Pareille convention formelle n'a rien d'*ad hoc*, comme l'on serait tenté de le croire au premier abord, mais s'appuie bien sur la nécessité de donner dans notre description un minimum de structure à l'état initial de Z (6). Sans modifier autrement notre représentation de Z pour l'instant, supposons qu'à la date $t = t_0$ une telle communication s'établisse effectivement entre les deux agents A_i et A_k , et uniquement entre ces deux agents. Selon ce qui précède, cette seule supposition implique que nous disposions désormais d'une information, même partielle, sur la distribution des a_{ij} dans $[a_{ij}](t_0)$. Cette information supplémentaire est réduite à sa plus simple expression, puisqu'elle consiste simplement à poser $S_i(t_0) \cap S_k(t_0) \neq \emptyset$. Mais elle suffit à réduire la complexité de Z , et à nous assurer que le système va changer d'état, et passer de la date $t = t_0$ à la date $t = t_1$. En effet, n'imposer aucune autre contrainte aux agents A_i et A_k revient à supposer que la catégorie rendant possible la communication entre ces agents est la seule qui leur soit alors commune, de sorte qu'au moment où s'établit la communication entre eux, chacun de ces agents représente pour l'autre un environnement qui est source de bruit perturbateur - donc d'information.

S'il connaît un début d'organisation, le système Z ainsi décrit n'est cependant pas un système auto-organisateur pour autant. Bien au contraire, c'est sa *redondance* qui augmente du fait de la communication établie entre A_i et A_k , et non sa complexité: aucune information n'est ici *créée* dans le système, qui voit seulement *s'échanger* de l'information entre deux de ses sous-systèmes. Ce faisant, le volume de la mémoire globale de l'économie ne varie nullement: seule sa structure se modifie dans le sens d'une homogénéité accrue. En réalité, s'en tenir à une telle représentation implique de ne pouvoir penser qu'une *transmission* d'information à l'intérieur du système. Si elle modifie la répartition de la quantité totale d'information contenue dans Z - en accroissant simultanément les cardinaux I et K des ensembles S_i et S_k - chacun des agents ayant transmis à l'autre au moins une catégorie que celui-ci ne détenait pas auparavant -, une telle transmission laisse inchangé le cardinal n de l'ensemble S . Illustrons ce point par un exemple simple. Imaginons que Z ne comporte que deux sous-systèmes: les agents A_i et A_k , qui établissent une communication alors que la matrice $[a_{ij}](t_0)$ se présente ainsi:

	C_1	C_2	C_3	C_4	C_5
A_i	1	0	1	0	1
A_k	0	1	1	1	0

Supposons que l'agent A_i communique à l'agent A_k la catégorie C_5 , et que ce dernier lui communique simultanément la catégorie C_2 . Dans ces conditions, à s'en tenir au niveau d'analyse prenant en compte les seuls ensembles de type S_j , cette communication se présente comme une simple transmission d'information entre les agents: l'agent A_i transmet l'élément C_5 de l'ensemble $S_i(t_0)$ à l'agent A_k qui lui transmet l'élément C_2 de l'ensemble $S_k(t_0)$. Abstraction faite de tout phénomène d'oubli (7), la matrice $[a_{ij}](t_1)$ a donc la forme suivante:

	C_1	C_2	C_3	C_4	C_5
A_i	1	1	1	0	1
A_k	0	1	1	1	1

Il est évident qu'ainsi analysée, la communication établie entre les agents A_i et A_k ne crée aucune information au niveau de Z , et qu'aucun effet de complexité ne peut apparaître. Chaque période élémentaire correspondant à une telle transmission d'information entre les agents A_i et A_k , il est facile d'admettre qu'au bout d'un nombre fini de périodes (ici, deux seulement) A_i et A_k seront devenus absolument identiques, au sens où $S_i \equiv S_k$. Plus généralement, en supposant que ce soit sur un tel mode que s'établissent des communications entre un nombre quelconque d'agents, il en irait de même pour l'ensemble de ces agents: à une certaine date $t = T$, nous aurions ainsi $S_i(T) \equiv S_k(T)$, $\forall i, \forall k$.

Incapable de rendre compte d'une création d'information au niveau global qui est celui du système Z lui-même, la formalisation adoptée jusqu'à présent présente une seconde faiblesse, tout aussi importante que la précédente: elle ne peut traiter de manière satisfaisante la question de la *signification* de l'information. En effet, considérons les catégories C_j définies précédemment comme étant les éléments de $S(t)$, c'est-à-dire les éléments de l'union de tous les $S_j(t)$. Chacune de ces catégories est implicitement posée comme parfaitement indépendante de toutes les autres dans $S(t)$, et il en va de même au niveau individuel: dans chaque ensemble $S_j(t)$, les catégories C_j constituant ses éléments (les C_j telles que $a_{jj}(t) = 1$) sont purement et simplement *juxtaposées*. Un tel traitement analytique suppose que, lors de la transmission d'information entre A_i et A_k , les catégories échangées présentent la même signification pour l'agent émetteur que pour l'agent récepteur. En d'autres termes, la signification de l'information véhiculée par une catégorie donnée (ou par un ensemble de telles catégories) apparaît ici totalement *indépendante* de l'ensemble dont cette catégorie (ou cet ensemble de catégories) est, ou devient, élément. En un sens, cela revient à dire que l'information est une substance dénuée de signification, ce qui est une manière de nier les spécificités individuelles des agents.

Car le savoir instantané de chaque agent concret entrant en communication avec un autre agent - cette mémoire que tous deux mobilisent alors sur un mode rosenfeldien - n'est pas formé d'une juxtaposition de catégories dont chacune aurait pour lui la même signification que pour tout autre agent, mais se présente au contraire comme une *totalité*: l'information est une forme organisée selon des processus opératoires spécifiques à l'agent considéré, et se trouve dotée d'une unité organique portant la marque propre de cet agent.

Sans que la signification de l'information ne se dilue pour autant dans un relativisme intégral qui nierait la possibilité même d'établissement d'une communication entre les agents, notre représentation d'un état de Z doit tenir compte de cet aspect, en montrant explicitement qu'une catégorie donnée C_j n'a jamais exactement la même signification selon qu'elle participe du savoir de l'un ou l'autre agent (8). La formalisation adoptée jusqu'ici présente l'avantage de ne pas considérer l'information comme étant une substance totalement détachée du sujet connaissant: au contraire, elle fait partie intégrante de la définition même de celui-ci, qui apparaît dans le modèle sous la forme explicite de la mémoire dont il est porteur à tout instant. Mais cette formalisation laisse en suspens la double question de la signification de l'information et de sa création au niveau global: en réduisant la communication à une simple transmission d'information, c'est encore au monde de la substance qu'elle rapporte cette dernière.

Traiter convenablement les questions liées de la signification de l'information et de sa création au niveau du système Z lui-même impose de déplacer le niveau de notre analyse en situant l'information dans le monde de la *forme*, et non dans celui de la *substance*. Loin d'opérer sur le mode *additif* selon lequel s'effectuait la transmission d'information décrite à l'instant, dans le monde de la forme les lois sont de nature plutôt *combinatoire*.

1.2 Création et signification de l'information: seconde approche d'un état du système et de ses transformations

C'est précisément le caractère combinatoire des lois à l'oeuvre dans le monde de la forme qui va nous permettre, dans le cadre d'une description légèrement modifiée d'un état de Z , et en parfait accord avec les conceptions de I. Rosenfield comme avec celles de H. Atlan, de préciser la *signification* de l'information tout en expliquant la *création* de cette dernière au niveau du système considéré dans son ensemble: ainsi que nous allons le voir, l'accroissement du volume global de la mémoire de l'économie est indissociable de l'attribution de significations différentes à une information donnée par des agents différents.

Il existe sans doute divers moyens de formaliser la notion de signification de l'information, mais l'un des plus simples est probablement le suivant. Considérons les agents A_i et A_k , et les ensembles S_i et S_k correspondants. Au lieu de nous en tenir comme précédemment à S_i pour représenter le savoir de l'agent A_i , considérons à cette fin *l'ensemble des parties* de S_i , noté $P(S_i)$. Il est immédiat que $P(S_i)$ traduit parfaitement l'unité organique du savoir de A_i , évoquée plus haut: dans $P(S_i)$, chaque sous-ensemble de S_i apparaît *combiné* à tous les autres sous-ensemble de S_i , ce qui rend bien compte de l'aspect *structuré* du savoir de l'agent A_i . Mais d'un autre côté, cette représentation se révèle apte à exprimer le caractère *spécifique*, variable avec chaque agent, d'une telle structuration, car chaque catégorie C_j élément d'un ensemble S_i apparaît sous deux aspects différents dans $P(S_i)$: en tant que singleton, mais aussi en tant qu'élément des autres sous-ensembles de $P(S_i)$ dans lesquels elle est combinée aux autres éléments de S_i .

Il est clair que la première modalité d'apparition de cette catégorie C_j dans $P(S_j)$ traduit la capacité de celle-ci à être transmise d'un agent à un autre, c'est-à-dire la *communicabilité* de la signification de l'information véhiculée par elle, alors que sa seconde modalité d'apparition dans $P(S_j)$ rend compte du caractère *relatif* de cette signification, en modulant cette dernière selon l'ensemble $P(S_j)$ précis dans lequel vient s'inscrire l'élément C_j considéré.

Afin de mieux faire apparaître cette remarquable ambivalence, considérons l'intersection des ensembles S_i et S_k , notée S_c , et supposons, pour plus de simplicité, que cette intersection ne comporte qu'un élément. Soit C_c cet élément. Appelons $C(P_{csi})$ et $C(P_{csk})$ les ensembles complémentaires de $P(S_c)$ par rapport, respectivement, à $P(S_i)$ et $P(S_k)$. La catégorie C_c est donc à la fois élément de $P(S_i)$ et de $P(S_k)$, et sa signification présente, dans chacun de ces deux ensembles, le double aspect évoqué à l'instant. En effet, cette signification apparaît bien, d'une part, comme étant *partagée* par les agents A_i et A_k , puisque l'élément C_c figure au titre du singleton $\{C_c\}$ dans $P(S_c)$. La signification associée à la catégorie C_c se présente ainsi comme étant *communicable* entre A_i et A_k . Mais d'autre part, la catégorie C_c apparaît également dans certains éléments de $C(P_{csi})$, ainsi que dans certains éléments de $C(P_{csk})$. Or la définition de ces deux derniers ensembles implique que leur intersection soit vide. Il s'ensuit que la signification de la catégorie C_c - connue à la fois de A_i et de A_k - devient différente pour chacun de ces agents, en apparaissant respectivement dans $C(P_{csi})$ et dans $C(P_{csk})$: ces derniers étant entièrement différents, une telle catégorie intervient ici à titre d'élément combiné à d'autres selon des sous-ensembles de parties spécifiques de chaque agent considéré. En ce sens, la signification associée à C_c apparaît *particulière* à chacun d'eux.

Un état du système Z reste donc associé à la donnée d'une matrice de type $[a_{ij}]$, mais sa description doit désormais tenir pour significatif le niveau qui considère les agents A_i du point de vue des ensembles de parties $P(S_j)$ qu'ils sont en mesure de concevoir, plutôt que celui qui s'en tiendrait à considérer les ensembles S_j eux-mêmes. Ce n'est qu'en plaçant l'analyse à un tel niveau que la notion de signification de l'information acquiert une pertinence pour les agents, éléments transmettant l'information par les voies élémentaires de communication établies au sein d'un tel système. Mais nous allons voir que c'est également à ce niveau qu'il est possible de concevoir la *création* d'information par le système, c'est-à-dire de donner une pertinence à la notion de signification de l'information transmise par le système à l'observateur que nous sommes.

Avant de donner une mesure possible de la variation de la quantité d'information contenue dans le système, explicitons sur un plan purement qualitatif ce processus de création d'information. Selon ce qui précède, considérons à la date $t = t_0$ deux agents A_i et A_k , désormais respectivement caractérisés par $P(S_i)(t_0)$ et $P(S_k)(t_0)$. Convenons de l'existence d'une possibilité de communication entre ces deux agents en notant $S_c(t_0) \neq \emptyset$ - le cardinal de $S_c(t_0)$ n'ayant pas à être spécifié davantage pour l'instant. Lors d'une communication réellement établie entre ces agents, la réception par l'agent A_i de tout élément appartenant à $C(P_{csk})(t_0)$ donne lieu à une recombinaison qui transforme $P(S_i)(t_0)$ en $P(S_i)(t_1)$, et le

même type de phénomène se produit simultanément en ce qui concerne l'agent A_k , qui recombine tout élément de $C(P_{csi}) (t_0)$, reçu par lui de A_i lors de leur communication de telle sorte que $P(S_k) (t_0)$ devienne $P(S_k) (t_1)$. Au niveau du système Z , ces deux opérations simultanées exercent un double effet, dont l'un augmente la redondance du système, alors que l'autre accroît sa complexité. Le signe de la variation nette de la quantité d'information contenue dans Z dépend ainsi de la force respective de ces deux effets.

Reprenons davantage en détail chacun de ces derniers, et d'abord ce que nous pouvons appeler *l'effet de redondance*. Dans notre modèle, l'existence d'une tel effet se marque par l'augmentation du cardinal de $P(S_c)$ de la date t_0 à la date t_1 : en l'absence de phénomène d'oubli, et vu le caractère inaliénable de l'information, l'ensemble $P(S_c)(t_1)$ contient nécessairement davantage d'éléments que $P(S_c)(t_0)$. Nous retrouvons ici ce que H. Atlan appelle *ambiguïté destructrice* dans un système auto-organisateur considéré du point de vue de l'observateur. Au niveau d'une voie élémentaire, ce type d'ambiguïté signifie qu'une certaine quantité d'information est bien transmise entre les sous-systèmes reliés par cette voie, équivalant à l'exercice d'une contrainte du sous-système émetteur sur le sous-système récepteur - ici, une contrainte réciproque entre A_i et A_k . Une telle contrainte implique un gain d'information pour chaque sous-système concerné, mais non pour l'observateur. Néanmoins, il faut bien voir que dans notre modèle cet effet de redondance n'est pas totalement associé à une simple *transmission* d'information entre les agents A_i et A_k : il résulte en partie d'une *recombinaison* de l'information par ces agents, opération qui se trouve également à l'origine d'une véritable création d'information représentant ce que nous pouvons appeler *l'effet de complexité*.

Pour bien comprendre ce point, reprenons l'exemple donné plus haut lorsque nous supposons que Z ne comportait que deux sous-systèmes, les agents A_i et A_k "échangeant" les catégories C_5 et C_2 . Mais à présent, considérons celles-ci comme étant, respectivement le *singleton* $\{C_5\}$ élément de $P(S_i)(t_0)$, et le *singleton* $\{C_2\}$ élément de $P(S_k)(t_0)$, et non plus comme étant des éléments de $S_i(t_0)$ et $S_k(t_0)$. Nous pouvons ainsi mettre en évidence une création d'information au niveau de Z , car c'est ici du processus de *recombinaison* de $P(S_i)$ et de $P(S_k)$, effectué par les agents A_i et A_k entre les dates t_0 et t_1 , que résultent simultanément un effet de complexité et la majeure partie de l'effet de redondance évoqué ci-dessus.

Montrons-le d'abord en ce qui concerne l'effet de redondance. Nous avons dit que celui-ci se mesurait par l'augmentation du cardinal de $P(S_c)$ de la date t_0 à la date t_1 . De fait, $P(S_c)(t_0)$ ne contenait que deux éléments, alors que $P(S_c)(t_1)$ en contient huit. Mais parmi ces huit éléments, seuls deux ont, à proprement parler, été *transmis* par la communication établie entre A_i et A_k à la date t_0 : $\{C_2\}$ et $\{C_5\}$. Il n'en va qu'apparemment de même pour certains autres. Ainsi, l'élément $\{C_2, C_3\}$ *semble* avoir été transmis par A_k à A_i lors de leur communication: cet élément appartenait à $P(S_k)(t_0)$ et non à $P(S_i)(t_0)$, alors qu'il est évidemment élément de $P(S_i)(t_1)$ puisqu'il est élément de $P(S_c)(t_1)$.

En réalité, il n'y a là nulle transmission: dans ce dernier ensemble, $\{C_2, C_3\}$ a été obtenu par la *combinaison* de $\{C_3\}$, qui était déjà élément de $P(S_i)(t_0)$, et de $\{C_2\}$ qui, lui, a été réellement transmis par A_k à A_i entre t_0 et t_1 . Il en va bien sûr de même pour $\{C_3, C_5\}$, maintenant élément de $P(S_k)(t_1)$ alors qu'il ne l'était pas de $P(S_k)(t_0)$ et qu'il figurait dans $P(S_i)(t_0)$: loin d'avoir été transmis par A_i à A_k , cet élément a été combiné par ce dernier à partir de l'élément $\{C_3\}$ de $P(S)(t_0)$, et de $\{C_5\}$, communiqué à A_k par A_i entre t_0 et t_1 . Mais surtout, l'ensemble $P(S_c)(t_1)$ contient deux éléments qui n'ont manifestement pu faire l'objet d'*aucune* transmission entre A_i et A_k , puisqu'ils n'appartenaient ni à $P(S_i)(t_0)$, ni à $P(S_k)(t_0)$: $\{C_2, C_5\}$ et $\{C_2, C_3, C_5\}$. Ces derniers résultent d'un pur et simple processus combinatoire effectué par les agents.

En d'autres termes, ces deux derniers éléments apparaissent comme résultant d'une *création nette* d'information au niveau du système Z lui-même. Bien qu'ils aient pour particularité de signifier à la fois une augmentation de redondance et une augmentation de complexité au niveau de Z , ces deux éléments ne sont pas les seuls à témoigner d'une création d'information dans Z : en t_1 , la complexité du système s'est également accrue du fait de l'apparition au niveau de chaque ensemble de type $P(S_i)$ d'un certain nombre de parties totalement inédites, et spécifiques du savoir de l'agent considéré - ainsi $\{C_1, C_2, C_3\}$ pour l'agent A_i , ou $\{C_3, C_4, C_5\}$ pour l'agent A_k .

Sur la base de ce qui précède, redéfinissons un état quelconque du système Z . Un tel état apparaît désormais défini par la donnée d'une matrice (m, z) , dont les vecteurs-lignes représentent comme précédemment les agents A_i , indicés $i = 1, \dots, m$, mais dont les vecteurs-colonnes représentent à présent les éléments de *l'union de tous les ensembles de parties des ensembles de catégories* formées par tous les agents, sur un mode rosenfeldien, à la date considérée. Soit Γ cet ensemble, dont les éléments sont indicés $x = 1, \dots, z$. Calculons la dimension de l'ensemble $\Gamma(t) = \cup P(S_i)(t)$. Il est clair que $z(t)$, cardinal de $\Gamma(t)$, est égal à la somme de tous les cardinaux des $P(S_i)(t)$, diminuée du nombre d'éléments de ces $P(S_i)(t)$ qui sont communs à deux ou plusieurs d'entre eux. Notons $\mu_i(t)$ le cardinal de l'ensemble des parties $P(S_i)(t)$, et $\Omega(t)$ la somme de ces cardinaux: $\Omega(t) = \sum \mu_i(t)$. Soit $\Pi(t)$ l'ensemble de toutes les parties $P(S_i)(t)$, $\forall i$, communes à au moins deux agents A_i et A_k à la date t . Le cardinal de $\Pi(t)$, noté $r(t)$, représente un indice de la redondance globale de Z à cette même date. Il vient donc:

$$z(t) = \Omega(t) - r(t)$$

Appliquons ce résultat au cas particulier étudié ci-dessus, où les valeurs significatives étaient les suivantes:

$$\begin{aligned} \mu_i(t_0) &= \mu_k(t_0) = 8, \text{ et } r(t_0) = 2. \text{ D'où } z(t_0) = 14 \\ \mu_i(t_1) &= \mu_k(t_1) = 16, \text{ et } r(t_1) = 8. \text{ D'où } z(t_1) = 24 \end{aligned}$$

En notant $\Delta z = z(t_1) - z(t_0)$, il apparaît donc qu'entre t_0 et t_1 la dimension de l'ensemble Γ s'est accrue de $\Delta z = 10$ unités. Il s'agit bien ici de la différence *nette* existant entre les effets de complexité et de redondance évoqués plus haut. Redondance et différenciation des savoirs individuels croissant ainsi simultanément, la complexité nette de l'ensemble demeure

nécessairement positive ou nulle à toutes les dates considérées. Il n'y a là rien d'étonnant, car cela provient tout simplement de la propriété mathématique selon laquelle l'union de deux ensembles différents contient toujours strictement leur intersection - union et intersection devenant égales en cas d'ensembles identiques. C'est la raison pour laquelle le résultat précédent est immédiatement généralisable à un système composé d'un nombre quelconque de sous-systèmes. Dès lors que les messages échangés par deux ou plusieurs agents entre t et $t+1$ véhiculent un certain degré de nouveauté pour au moins l'un d'entre eux, la communication établie se traduit par un accroissement simultané de la redondance des savoirs de l'ensemble de ces agents, et du degré de différenciation du savoir de tout agent récepteur d'une information nouvelle. Cette différenciation croissante est signalée par l'augmentation de z entre t et $t+1$. Le nombre m de sous-systèmes de Z étant donné, z représente un indice acceptable de la complexité de ce dernier.

Dans notre modèle de sous-systèmes en interaction, chaque état du système global est défini par la superposition de tous les états correspondants des sous-systèmes. A chaque recombinaison de la morphologie et du volume de l'information associés à l'état de l'un de ces derniers, le système dans son ensemble change d'état. C'est ce changement qui implique le passage de la date t à la date $t+1$, et non l'inverse: *ce n'est pas le système qui se déplace dans un cadre temporel préalable, mais le cadre temporel qui est progressivement dessiné par l'évolution du système. Ce dernier constitue donc une horloge proprement économique.*

Bien qu'il rappelle à la fois celui de I. Rosenfield et celui de H. Atlan, un tel modèle présente apparemment certaines différences avec chacun de ces derniers. Avant d'explorer la problématique du choix rationnel d'un agent immergé dans un système qui évolue en se complexifiant sur le mode analysé à l'instant, précisons la portée de telles différences.

Sur la base de l'exemple étudié plus haut, nous avons quasiment rejoint la commune position de I. Rosenfield et H. Atlan quant au rôle créateur de l'ambiguïté dans un système d'information. De fait, s'agissant du cerveau mettant en oeuvre des processus opératoires visant à catégoriser les stimuli *a priori* ambigus, présents dans le contexte actuel du sujet individuel, ou créant des patterns en réponse à des stimuli aléatoires en provenance de son environnement, notre précédent texte avait montré ces deux auteurs mettre l'accent sur un apprentissage non dirigé, effectué par un système d'information à l'occasion de son interaction avec son environnement, et consistant en un processus de différenciation interne synonyme de création de sens nouveaux. A quelques nuances près, il en va de même dans notre modèle : de t_0 à t_1 , l'accroissement de la quantité totale d'information contenue dans le système tient au fait que deux sous-systèmes de Z se sont différenciés bien qu'ils se soient simultanément homogénéisés lors de la communication établie entre eux. C'est bien l'interaction entre agents au cours de leur communication, chacun représentant l'environnement de l'autre, qui est ici occasion de création de sens totalement inédits - au niveau de l'ensemble du système Z comme à celui de chacun des agents. Née d'une différence initiale entre ces derniers, une telle création reconduit plus ou moins celle-ci tout en accroissant la complexité du système global .

Les nuances mentionnées plus haut apparaissent ainsi immédiatement. Là où I. Rosenfield envisage la remémoration individuelle comme étant un processus créateur occasionné par la perception de stimuli non organisés présents dans le contexte du sujet (dont la mémoire invente précisément en organisant ces stimuli par leur catégorisation), nous présentons un apprentissage individuel tenant à la réception par chaque agent de stimuli déjà *organisés*: les catégories (ou les combinaisons de catégories) communiquées par un autre agent lui tenant lieu de contexte. Le même type de nuance sépare apparemment notre modèle de celui de H. Atlan : loin de lier le processus auto-organisateur à un bruit créé dans le système par des facteurs *totalelement aléatoires* en provenance de l'environnement de ce dernier, notre représentation suspend la possibilité même de ce processus à l'existence préalable de catégories communes aux agents concernés.

Qui plus est, au lieu d'envisager un système s'auto-organisant par la destruction continue d'une quantité *donnée* de redondance, corrélative d'un accroissement continu de complexité, notre modèle traite d'un système où redondance et complexité semblent croître *simultanément*. De sorte qu'au contraire du modèle d'auto-organisation présenté par H. Atlan, il ne paraît guère possible d'envisager ici la dégénérescence, puis la mort, du système, en termes d'un épuisement progressif de sa redondance initiale, le conduisant à ne plus être fiable après une certaine date (t_m), et de ne plus pouvoir assurer finalement aucune fonction vitale lorsque l'ambiguïté destructrice l'a définitivement emporté sur l'ambiguïté-autonomie. Au contraire, le système Z semble capable de trouver indéfiniment en lui-même de nouvelles ressources, en créant continuellement en son sein la redondance nécessaire à la possibilité d'établissement de communications productrices d'une ambiguïté-autonomie toujours croissante. Nous allons revenir sur l'interprétation qu'il est possible de donner à ce dernier point, mais auparavant montrons que les nuances ainsi soulignées tiennent en grande partie à la double modification de champ que nous avons effectuée par rapport aux modèles respectifs de I. Rosenfield et H. Atlan, et non à l'existence d'une différence intrinsèque de logique entre ces modèles et celui qui est présenté ici.

En étendant notre investigation au plan macroscopique autant que microscopique, nous avons envisagé un domaine plus étendu que celui de I. Rosenfield, puisque celui-ci s'intéresse principalement au cerveau, donc à la mémoire individuelle; mais simultanément, nous avons ainsi restreint, par rapport à cet auteur, la définition de ce qui doit être considéré comme étant le contexte du sujet individuel: alors que Rosenfield désigne par ce terme l'ensemble de l'environnement du sujet, y compris l'environnement naturel de celui-ci, nous restreignons le contexte de l'*homo cogitans* à son environnement social, principalement perçu par lui sur un mode symbolique. A l'évidence, les stimuli provenant d'un tel contexte sont d'emblée organisés par le langage, et sont reçus selon ce type d'organisation par chaque agent engagé dans une communication avec un autre agent. Il en va de même, pour ce qui ne nous sépare qu'apparemment sur ce point de H. Atlan. Il est clair que la nature *totalelement aléatoire* des facteurs de bruit organisationnel envisagés par ce dernier est à rapporter entièrement au degré de généralité avec lequel il présente le "noyau dur" de son modèle.

Au niveau d'abstraction auquel se situe le discours de cet auteur dans cette partie précise de son développement, une telle conception est doublement justifiée: elle place d'emblée le modèle hors de tout finalisme connotant inéluctablement la notion de programme, et lui confère simultanément une plasticité permettant d'envisager une très large gamme d'applications possibles. Reste que chaque application de ce type demande de préciser soigneusement la nature de ces facteurs de perturbation, dont le caractère purement aléatoire se trouve *ipso facto* estompé. Nous avons d'ailleurs vu H. Atlan lui-même apporter une telle précision en parlant de *l'approximative coïncidence* existant entre les stimuli nouveaux et les patterns anciens que le sujet projette sur eux afin de les reconnaître: s'il est vrai que ce caractère approximatif est alors source de création de nouveaux patterns de la part du sujet concerné, il est non moins vrai que ce même caractère implique simultanément la persistance d'une coïncidence partielle entre patterns anciens et stimuli nouveaux - faute de quoi, le sujet ne pourrait rien reconnaître du tout dans ces derniers.

Nous n'avons rien voulu exprimer d'autre en exigeant ici des agents qu'ils connaissent au moins une catégorie commune pour pouvoir établir une communication à un moment donné. L'exemple traité plus haut montre que c'est bien parce que $P(S_i)(t_0)$ et $P(S_k)(t_0)$ contiennent tous deux le singleton $\{C_3\}$ que les agents A_i et A_k peuvent "échanger" les catégories C_5 et C_2 lors de leur communication. Concevons par exemple celle-ci comme un échange de deux morphèmes, dont l'un contiendrait la combinaison de catégories " C_3 et C_5 " (émise par A_i), et l'autre la combinaison " C_3 et C_2 " (émise par A_k). Chaque stimulus nouvellement perçu par chacun de ces deux agents coïncide *approximativement* avec un pattern déjà connu de lui : ce qu'il y reconnaît lui permet de percevoir, et le reste lui permet d'apprendre.

La capacité auto-organisatrice manifestée par le système Z, comme celle qu'analyse H. Atlan s'agissant des systèmes non construits en général, s'articule donc bien autour de l'existence simultanée, en chaque état de ce système, d'un certain degré de redondance et d'un certain degré de variété entre sous-systèmes. De fait, à chaque date considérée, la redondance est ce qui donne à Z un caractère de système *organisé*, en installant un lien potentiel entre les agents qui en figurent les éléments, et la variété présentée par ces derniers confère au système son caractère *auto-organisateur*: la complexification de Z résulte de l'établissement *effectif* de tels liens lors de communications entre agents. En dernière analyse, le moteur d'une telle évolution consiste donc exclusivement en le fait que *ce qui est émis au cours d'une communication a toujours une signification différente de ce qui est reçu*: la spécificité individuelle du savoir organisé de chacun, synonyme de différences dans les significations associées par des agents différents à une information donnée, constitue le fondement même de la croissance de complexité de l'ensemble du système. Tout en marquant le profond accord de notre conception avec la position darwinienne rappelée par I. Rosenfield, selon laquelle la biologie ne connaît ni animal ni plante typiques mais seulement des êtres singuliers, cette dernière proposition nous permet de mieux comprendre l'apparente différence subsistant entre notre modèle et celui de H. Atlan. Car celle-ci provient également en partie d'une modification de champ, et non d'une différence intrinsèque de logique.

En effet, si H. Atlan perçoit les systèmes auto-organiseurs comme étant destinés à cesser de l'être, et à finir par disparaître sous les effets répétés du même type de facteurs ayant suscité auparavant leur augmentation de complexité, c'est parce que l'objet d'étude privilégié de ce biologiste consiste naturellement, et explicitement, en l'être vivant *individuel* - fût-il inséré dans un environnement. Ce dernier n'intervient que marginalement dans ce que nous avons appelé le "noyau dur" de son modèle - et nous avons vu qu'il y avait de bonnes raisons pour cela. Rien n'est dit, en particulier, de *l'évolution de l'environnement lui-même*. Tout se passe comme si une force immuable perturbait de manière aveugle un système organique qui, après avoir pu un temps organiser à son profit l'énergie véhiculée par ces perturbations, finirait par succomber sous leur impact. Créatrice puis destructrice de complexité, l'interaction se situe ici *entre* un système organisé et son environnement monolithique, bien qu'évidemment les effets de cette interaction soient sensibles au sein même du système.

Il en va différemment dans notre modèle, qui analyse l'évolution de cet autre type de système non construit que représente un système économique, à travers les effets de processus symboliques mis en oeuvre par ses agents lors de leurs communications. Ici, l'interaction se situe *au sein même* du système - bien que ses effets soient sensibles au niveau de l'ensemble de ce dernier, où elle paraît continuellement créatrice de complexité et de redondance. En termes atlaniens, les niveaux d'organisation respectifs auxquels se situe l'observateur ne sont pas les mêmes dans l'un et l'autre modèle, de sorte que le même type de grille de lecture conduit à des observations différentes: contrairement à celui de H. Atlan, notre modèle montre explicitement la *co-évolution* des deux parties en interaction, et celle-ci consiste en l'accroissement net de complexité résultant d'un mouvement simultané d'homogénéisation et de différenciation. Reste que ce n'est là qu'une explication partielle: encore faut-il préciser le *contenu* de l'augmentation continue de redondance du système Z dans notre modèle. Terminons cette première partie par quelques indications sur ce point.

Cette augmentation continue de redondance paraît associée aux notions, elles-mêmes liées, d'*oubli* et de *convention*. S'agissant de cette dernière, son lien conceptuel avec notre représentation de la redondance, et surtout avec la croissance de celle-ci dans le modèle, est pratiquement immédiat: une convention étant globalement définie comme une régularité qui a sa source dans les interactions sociales, mais qui se présente aux acteurs sous une forme objectivée [J.-P. Dupuy, 1989], il est facile d'admettre que ces interactions sociales apparaissant dans notre modèle comme autant de communications établies donnent lieu à certaines régularités représentables sous forme de savoir partagé par les agents (les acteurs). Nous pouvons ainsi formaliser tous les degrés d'extension concevables de conventions données, depuis celles qui voient s'accorder deux agents seulement (par exemple, deux coéquipiers habitués à travailler ensemble), jusqu'à celles qui rassemblent tous les agents (telle la monnaie): la première situation serait rendue formellement par la donnée de certaines catégories C_j simultanément connues des deux agents en cause et d'eux seulement, et la seconde par une catégorie C_0 - voir la note (6) *infra*. Entre ces deux extrêmes, le poids relatif d'une convention donnée se mesure au nombre d'agents connaissant simultanément la ou les catégorie(s) C_j pertinente(s), rapporté à celui de l'ensemble des agents.

L'accroissement continu de la redondance globale de Z prend donc clairement la signification d'une prégnance croissante de certaines conventions parmi des agents dont le nombre total est fixé. C'est ici que s'effectue le lien avec la notion d'oubli, plus précisément avec l'un des aspects de cette notion. Car de nombreux auteurs soulignent l'importance extrême du mécanisme de l'oubli dans le fonctionnement de la mémoire. Ainsi G. Bateson - après Samuel Butler disant que, mieux un organisme "connaît" quelque chose, moins il est conscient de cette connaissance - affirme l'existence d'un processus par lequel la connaissance (ou l'habitude d'action, de perception ou de pensée) s'enfonce à des niveaux de plus en plus profonds de l'esprit [1977, p. 146]. Pareil processus de "refoulement" des aspects parfaitement maîtrisés du savoir (des aspects généraux qui restent toujours valables) permettrait à l'organisme de concentrer son attention sur la pragmatique impliquée par telle ou telle situation [ibid., p. 153]. De même pour E.T. Hall [1979] qui précise:

"En fait, selon Powers, le système nerveux humain est ainsi fait que les modèles qui commandent la conduite et la perception ne remontent à la conscience que lorsqu'il y a changement de programme. C'est pourquoi les codes les plus importants qui commandent notre existence fonctionnent au-dessous du niveau des contrôles conscients, et ne se prêtent généralement pas à l'analyse." [op. cit., p. 49].

Qu'ils apparaissent sous une forme objectivée aux acteurs dont les interactions sont cependant à leur origine, ou qu'ils aillent tellement de soi que personne ne s'interroge plus sur leurs fondements voire que chacun oublie leur existence, codes de comportements et autres conventions se laissent donc formaliser par notre modèle en termes de redondance: l'accroissement continu de celle-ci paraît susceptible de déboucher sur la production régulière de catégories de type C_0 qui, en advenant, quittent le niveau logique des catégories de type C_j dont elles sont issues, pour venir s'installer à un niveau à la fois supérieur et davantage enfoui - celui qui détermine *la manière même* dont de nouvelles catégories de type C_j seront construites.

2 La problématique du choix dans un système complexe

2.1 Position du problème

Au niveau du système Z comme à celui des sous-systèmes en interaction, l'évolution complexifie toujours davantage le cadre dans lequel se pose à chaque agent individuel le problème du choix. Puisque la rationalité d'un choix s'apprécie toujours par rapport à une stratégie dont la forme dépend elle-même de la perspective analytique dans laquelle les unités de décision sont définies et ont à se déterminer, l'optique évolutionniste qui est la nôtre pèse sur la manière même dont il convient d'envisager un tel problème. En quels termes celui-ci se pose-t-il dans notre modèle? Les unités de décision sont ici des agents individuels A_i , éléments figurés du système Z évoluant qualitativement sur le mode analysé ci-dessus. Dans un tel cadre, les aspects directement pertinents du problème du choix rationnel d'un agent donné nous semblent pouvoir être articulés autour de la notion de *participation optimale* de celui-ci à l'évolution d'ensemble du système. Précisons: la rationalité de l'agent se marque ici dans sa stratégie d'émission-réception d'information lors de communications qu'il établit, ou peut souhaiter établir (ou ne pas établir), avec d'autres agents, à l'exclusion de toute autre

considération. Ainsi, nous ne tiendrons aucun compte de coûts ou de ressources liés matériellement à de telles opérations - contraintes spatiales venant éventuellement peser sur la circulation de l'information, inégalités de ressources monétaires qui différencieraient les agents quant à leurs possibilités d'accès à l'information, etc. Bref, dans ce qui suit nous considérons des agents *a priori* identiques en tous points, sauf en ce qui concerne leurs dotations initiales en information, apparaissant dans le modèle sous la forme de leurs $P(S_i)(t_0)$ respectifs.

Ainsi circonscrit, le problème du choix rationnel peut être défini de la manière suivante: pour l'agent A_i , participer optimalement à l'évolution d'ensemble de Z consiste à minimiser, de période en période, la probabilité de diminution du rapport μ_i/z . A chaque date considérée, ce rapport revêt en effet une double signification pour cet agent: d'une part, il mesure sa part actuelle dans la complexité globale du système; d'autre part, il exprime l'ouverture de l'éventail des choix qui lui sont actuellement offerts concernant ses communications futures. En effet, l'établissement d'une communication entre deux agents A_i et A_k n'étant possible que s'il existe une intersection non vide entre leurs savoirs respectifs, chaque élément de $P(S_i)$ représente pour A_i une communication potentielle avec au moins un autre agent. Le rapport μ_i/z mesure ainsi de manière fidèle la position instantanée de A_i par rapport à l'ensemble du système, en résumant à tout moment le passé de leurs évolutions respectives qu'il exprime simultanément en rapportant l'une à l'autre, et en indexant les possibilités d'évolutions à venir de l'agent considéré: il traduit bien l'idée de participation dynamique de l'agent individuel à l'évolution de l'ensemble du système, idée corrélative de celle d'un présent conçu comme une véritable *tension* entre le passé et l'avenir (9).

Cela dit, la rationalité du choix individuel s'apprécie naturellement à l'aune d'une certaine optimalité - celle que l'agent lui-même estime mettre en oeuvre dans sa stratégie d'émission-réception de l'information, lorsqu'il la juge optimale: d'une manière ou d'une autre, le choix de l'agent porte sur les *quantités* d'information respectivement émises et reçues par lui lors des communications qu'il est amené à établir avec d'autres agents. A chaque date considérée, la stratégie de participation optimale de A_i à l'évolution de Z est donc susceptible de se traduire en termes quantitatifs: il s'agit pour lui d'agir de telle manière que, dans l'évolution globale de μ_i/z , la partie dépendant directement de ses propres communications successives évolue de manière à minimiser la probabilité de diminution d'un tel rapport. Un tel objectif paraît analytiquement d'autant plus raisonnable qu'il témoigne d'une sorte de réalisme de l'agent quant aux limites de ses capacités d'action sur l'évolution globale du système: à chaque date considérée, dans l'évolution globale de μ_i/z la partie directement liée à la communication établie entre l'agent A_i et l'agent A_k se réduit en effet à l'évolution du rapport $\mu_i/[\mu_i+\mu_k-\mu_c]$, où μ_c est le cardinal de l'intersection $P(S_c)$ de $P(S_i)$ et $P(S_k)$. De sorte que la définition proposée à l'instant s'agissant de l'objectif de l'agent A_i incorpore l'idée que ce dernier sait pertinemment n'avoir de prise sur l'évolution globale du système que par ses possibilités d'agir sur celles de μ_i et de μ_k - donc, aussi, sur celle de μ_c . Lors de chaque communication, il s'agit donc pour lui de maintenir au moins constant, voire de faire croître de période en période, le rapport $\mu_i/[\mu_i+\mu_k-\mu_c]$.

A priori, pareil objectif semble aussi bien pouvoir être atteint en jouant sur le numérateur que sur le dénominateur de ce rapport. Pour notre agent A_i , ces deux tactiques alternatives consisteraient respectivement à maximiser l'accroissement de $P(S_i)$ pour toute variation donnée de $P(S_k)$, ou à minimiser l'accroissement de $P(S_k)$ pour toute variation donnée de $P(S_i)$, lors de sa communication avec A_k . Plus généralement, pour qui viserait à participer de manière optimale à l'évolution globale de Z , il conviendrait d'en dire le moins possible, lors de chaque communication, tout en essayant d'en apprendre le plus possible à cette occasion - bref, de maximiser une sorte de profit informationnel afin de minimiser la probabilité de voir diminuer sa part d'un "marché" sur lequel s'échangeraient des "biens" informationnels. Néanmoins, une telle manière de poser le problème néglige la dimension évolutionniste de celui-ci, et attribue à l'agent un individualisme strict qui n'est pas nécessairement de mise ici. Dans la mesure où chacun sait qu'il n'est pas élément d'un système dénué de mémoire qui redistribuerait de manière totalement aléatoire les identités respectives des agents communiquant à chaque période, sa participation optimale à l'évolution de ce système exige qu'il ne borne pas son horizon temporel à une seule période.

Dans ces conditions, l'objectif poursuivi par A_i peut parfaitement inciter celui-ci à émettre *le plus* d'information possible lors d'une période donnée, s'il pense que l'établissement de sa communication avec l'agent A_k se reproduira au moins encore une fois: plus l'agent A_k recevrait d'information de sa part lors de la première période, plus le savoir issu de la recombinaison de son savoir antérieur avec l'information nouvelle serait étendu, et plus l'agent A_i pourrait ultérieurement tirer d'enseignements de sa communication avec un agent A_k ainsi rendu auparavant plus riche en information par lui. Comme dans la théorie des jeux coopératifs, il peut donc être de mauvaise stratégie pour l'agent d'être guidé par un strict individualisme lorsque l'enchaînement des périodes vient lier ses intérêts à ceux d'autres agents. Nous allons précisément clore ce texte en montrant que la communication établie entre deux agents peut avoir pour effet de renforcer la probabilité de sa réitération, ce qui nous permettra de déterminer l'évolution la plus probable à court terme du système global, et de repenser la stratégie de l'agent individuel à la lumière de ce type d'évolution.

2.2 Evolution globale et choix individuel

Définissons tout d'abord l'état initial de Z , de manière que celui-ci soit suffisamment spécifié pour que sa redondance $R(t_0)$ ne soit pas nulle, sans qu'une telle spécification ne vienne par trop limiter la portée de l'analyse de son évolution: contentons-nous d'affirmer le caractère organisé de $Z(t_0)$, et de donner une spécification minimale à la forme d'organisation que connaît celui-ci. Convenons ainsi que la quantité totale d'information contenue dans Z est initialement répartie parmi les agents de telle sorte que lui soit associé un graphe connexe. Chaque agent est alors en mesure d'établir une communication avec au moins un autre agent, avec lequel il partage initialement un certain savoir dont il peut mobiliser tout ou partie dans le langage utilisé par lui lors de leur communication. Spécifions davantage le système Z en formulant deux hypothèses relatives à la forme de son organisation : supposons que le graphe connexe associé à son état initial est tel que : 1) il n'existe pas de chemin direct entre plus de deux agents; 2) lorsqu'un tel chemin existe entre deux agents, il est unique.

Ces deux hypothèses sont moins restrictives qu'il n'y paraît. En fait, elles reviennent à donner au système l'organisation initiale la plus ténue qui soit en minimisant la redondance globale alors présentée par celui-ci: puisque nous savons que cette redondance croît avec le nombre de communications établies, et que par définition aucune communication n'a encore été établie à la date initiale, pareilles hypothèses sont parfaitement justifiées à cette date. Déterminons plus précisément $r(t_0)$ qui représente un indice de $R(t_0)$, en faisant désormais abstraction de l'ensemble vide (10). La seconde hypothèse mentionnée ci-dessus signifie que les ensembles de type $P(S_i)(t_0)$ de chaque couple d'agents quelconques contiennent au maximum un élément commun, alors que la première interdit à cet élément commun d'être le même pour tous les $P(S_i)(t_0)$. Le graphe étant néanmoins connexe, il en résulte que chaque $P(S_i)(t_0)$ contient au moins deux éléments. Supposons en outre que la probabilité non nulle, pour chaque agent, d'établir une communication avec au moins un autre agent en t_0 est, à cette date, *égale pour tous les agents*.

Nous reviendrons plus loin sur la mesure qu'il convient de donner à cette probabilité. Remarquons pour l'instant que l'état initial de Z est alors associable à un graphe fermé, dont les sommets sont des agents munis d'ensembles $P(S_i)(t_0)$ comportant exactement deux éléments: en t_0 , chaque agent est potentiellement lié à deux autres agents selon une chaîne fermée.

Dans ces conditions, $r(t_0) = m$. Puisque $\Omega(t_0) = \sum \mu_i(t_0)$ est alors égal à $2m$, la complexité initiale minimale du système, indiquée par $z(t_0) = \Omega(t_0) - r(t_0)$, est aussi égale à m : redondance et complexité initiales de Z sont égales, et elles sont également partagées parmi les agents. Il suit que chacun d'eux contribue identiquement à définir l'état initial de Z . Formellement, $\forall i, [\mu_i / z](t_0) = 2/m$. Il est vrai que les spécificités individuelles des agents semblent alors se réduire à des différences interindividuelles dont la somme est nulle au niveau global, puisque ce qui différencie chaque agent d'un autre est aussi ce qui le rapproche d'un troisième. Néanmoins, la spécificité de chacun continue ici d'être marquée par le fait qu'aucun $P(S_i)(t_0)$ n'est identique à aucun autre.

Considérons alors deux agents *donnés* A_i et A_k . Quelle que soit la date, il est vraisemblable que leur probabilité d'entrer actuellement en communication l'un avec l'autre est une fonction croissante de ce que nous pourrions appeler un *indice de la ressemblance* existant alors entre eux. Un tel indice varierait entre deux extrêmes, tous deux exclus du modèle: à l'un, ces agents se ressembleraient au point d'être parfaitement identiques, et à l'autre, ils seraient totalement dissemblables. Dans ces conditions, notons r_{ik} l'indice de ressemblance entre les agents A_i et A_k considérés. Le contenu sémantique ainsi indicé semble pouvoir être naturellement quantifié par $r_{ik} = [\mu_c - 1] / [\mu_i + \mu_k - \mu_c]$, qui évoque bien un degré de ressemblance entre A_i et A_k en rapportant leur savoir partagé (μ_c) à l'étendue totale de leurs savoirs ($[\mu_i + \mu_k - \mu_c]$). Les ensembles $P(S_i)$ et $P(S_k)$ n'étant pas vides, le dénominateur de ce rapport est toujours strictement positif, de sorte que r_{ik} est toujours défini. Plus précisément, il apparaît que $0 \leq r_{ik} < 1$: si dissemblables que soient les agents A_i et A_k , l'intersection $P(S_c)$ des ensembles $P(S_i)$ et $P(S_k)$ contient toujours au moins l'ensemble vide, de sorte que r_{ik} est toujours positif ou nul; si ressemblants soient-ils, ces mêmes agents ne sont jamais identiques, de sorte que r_{ik} est toujours strictement inférieur à l'unité.

Néanmoins, cette quantification apparemment naturelle de r_{ik} se révèle en fait mal adaptée à notre problème. Nous voulons définir ici une fonction croissante entre ressemblance et probabilité afin d'étudier les *variations* de probabilité d'établissement d'une communication entre agents donnés d'une date à l'autre: les ressemblances entre ces derniers variant en fonction des quantités d'information échangées lors de leurs communications, nous pourrions ainsi lier leurs stratégies de communication à leur probabilité d'établir ultérieurement une autre communication. D'un point de vue purement formel, une telle étude serait rendue particulièrement lourde par l'adoption d'un indice de ressemblance exprimé en termes de cardinaux d'ensembles de parties d'ensembles! Essayons de définir une expression quantitative plus simple de l'indice de ressemblance entre les agents A_i et A_k , en déterminant les conditions dans lesquelles il serait possible de raisonner directement sur les S_i des agents, et non pas sur les $P(S_i)$ correspondants. En fait, de telles conditions se réduisent à une hypothèse concernant la mesure de l'information contenue dans un message.

Pour bien comprendre ce point, revenons à l'exemple dans lequel l'agent A_i "échangeait" avec A_k le singleton $\{C_5\}$ de $P(S_i)(t_0)$ contre le singleton $\{C_2\}$ de $P(S_k)(t_0)$. Discutant plus loin de l'analogie existant entre notre modèle et celui de H. Atlan, nous avons évoqué cet échange sur un mode légèrement différent, en parlant alors de morphèmes contenant respectivement les éléments $\{C_3, C_5\}$ de $P(S_i)(t_0)$ et $\{C_3, C_2\}$ de $P(S_k)(t_0)$. Ce faisant, nous avons considéré comme identiques deux messages, dont l'un comportait $\{C_5\}$ et l'autre $\{C_3, C_5\}$ (respectivement, $\{C_2\}$ et $\{C_3, C_2\}$), au sens où ils auraient contenu la même quantité d'information pour leur récepteur (respectivement A_k et A_i). Un tel traitement apparaît justifié dans la mesure où ce qui est nouveau pour A_k recevant $\{C_3, C_5\}$ est exclusivement véhiculé par $\{C_5\}$: pour cet agent *qui connaît déjà* $\{C_3\}$, la combinaison "C₃ et C₅" ne contient en elle-même pas plus d'information que la catégorie élémentaire "C₅". De même pour A_i , indifférent à recevoir "C₃ et C₂" plutôt que "C₂", puisqu'il *connaît déjà* "C₃". En d'autres termes, nous avons négligé la valeur informative d'un *effet de renforcement* de catégories existant dans un $P(S_i)(t_0)$ quelconque, éventuellement provoqué par la réception de l'une d'elles lors d'une communication établie entre les dates t_0 et t_1 .

Il importe de bien marquer la signification et la portée d'une telle hypothèse. Celle-ci *ne consiste pas* à supposer qu'une combinaison de catégories rosenfieldiennes contient exactement la même quantité d'information que la somme arithmétique des quantités d'information contenues dans les catégories élémentaires la composant. Bien entendu, une telle supposition serait parfaitement contradictoire avec la spécificité de la signification associée par chaque agent à une catégorie donnée: on ne peut affirmer simultanément, d'une part, que la catégorie "C₃" n'a pas la même signification dans "C₃ et C₅" et "C₃ et C₂", et d'autre part, que la combinaison de catégories élémentaires ne diffère en rien de leur simple juxtaposition. Pour reprendre notre exemple, si A_k est indifférent à recevoir "C₃ et C₅" ou "C₅", c'est *uniquement parce qu'il connaît déjà* "C₃" au moment de la réception du message, *et non* parce que "C₃ et C₅" ne contiendrait pour *aucun* agent d'autre information que la somme arithmétique de celles contenues dans les catégories élémentaires "C₃" et "C₅". Il en va bien sûr de même pour A_i , "C₃ et C₂", et "C₂".

C'est donc bien un effet de renforcement qui est ainsi négligé, et non une conception globale de l'information qui se verrait remise en cause. Reste que pareille hypothèse est très restrictive, car elle implique qu'en l'absence de renouvellement du lexique $S(t_0)$ à partir duquel est composé le volume global de la mémoire initiale de Z , il existe nécessairement une date T à laquelle chaque $P(S_i)$ contient tous les éléments qu'il était possible de construire à partir des catégories existant initialement en un nombre fini n . En d'autres termes, dès lors que $m = \text{constante}$, cette hypothèse signifie que l'évolution de Z conduit nécessairement ce système vers un état final dans lequel sa complexité est nulle car sa redondance est maximale: dans cet état de mort entropique du système, $z(T) = 0$ du fait que $r(T) = \Omega(T)$, et tous les agents sont identiques (11). Il est néanmoins nécessaire de la formuler si nous voulons obtenir ici une expression quantitative simple de r_{ik} . En effet, ce sont bien les catégories ou les combinaisons de catégories figurant dans leurs savoirs respectifs envisagés comme des totalités organiques qu'échangent les agents lors de leurs communications, c'est-à-dire des messages dont chacun revêt toujours des significations différentes pour l'émetteur et le récepteur de l'information, et non les catégories platement juxtaposées de la matrice $[a_{ij}]$ - faute de quoi, aucune information nouvelle ne pourrait jamais apparaître au niveau de Z . Mais, sous l'hypothèse d'absence d'effet de renforcement, les parties pertinentes des $P(S_i)(t)$ respectifs évolueront nécessairement, d'une période à l'autre, dans le même sens que les $S_i(t)$ correspondants.

Car la mesure de la quantité totale d'information détenue par l'agent A_i à la date t est alors strictement liée par une fonction croissante à $I(t)$: $\mu_i(t) = 2^{I(t)}$. Ainsi, même si A_i , recevant " C_3 et C_2 " lors de sa communication avec A_k à la date t_0 , recompose son savoir de telle sorte qu'un élément de type $\{C_3^2, C_2\}$ figure dans $P(S_i)(t_1)$, notre modèle considère un tel élément comme parfaitement équivalent, du point de vue de sa valeur informative pour cet agent, à l'élément $\{C_3, C_2\}$. Cette hypothèse va ainsi nous permettre de définir un nouvel indice de ressemblance des agents A_i et A_k , aussi significatif que le précédent puisqu'il rapporte de manière analogue à celui-ci le savoir partagé des agents à l'étendue totale de leurs savoirs, mais beaucoup plus facile à manipuler.

Soit $r_{ik} = C / [I+K-C]$ ce nouvel indice, où, rappelons-le, C est le cardinal de l'intersection des ensembles S_i et S_k , de cardinaux respectifs I et K . Les ensembles S_i et S_k n'étant pas vides, le dénominateur de ce rapport est toujours strictement positif, de sorte que r_{ik} est toujours défini. De manière plus précise, $0 \leq r_{ik} < 1$: deux agents dont les savoirs respectifs ne se recouperaient sur aucun point seraient totalement dissemblables - de sorte que r_{ik} serait nul; inversement, deux agents dont les savoirs seraient strictement identiques ne constitueraient en fait qu'un seul et même agent - de sorte que $r_{ik} = r_{ii} = r_{kk} = 1$. En notant $p_{ik}(t)$ la probabilité d'établissement d'une communication entre les agents A_i et A_k à la date t , nous pouvons dès lors identifier purement et simplement degré de ressemblance et probabilité: $\forall i, \forall k, \forall t, p_{ik}(t) = r_{ik}(t)$. L'avantage principal d'une telle identification tient à la simplicité du traitement formel qu'elle autorise, sans obérer pour autant la portée des résultats ainsi obtenus quant aux liens existant entre quantité d'information échangée lors d'une communication et probabilité de réitération de celle-ci (12).

Revenons en effet à notre hypothèse d'une égalité de toutes les probabilités initiales des agents d'établir une communication avec au moins un autre agent. Cette hypothèse signifie que lorsque deux agents se ressemblent, ils se ressemblent toujours exactement autant que deux autres agents qui se ressemblent. Alliée à la connexité du graphe associé à $\Gamma(t_0)$, elle implique donc que la probabilité initiale d'établissement d'une communication entre tous les agents d'un groupe donné varie uniquement avec le nombre d'agents de ce groupe selon une règle de probabilités composées si chaque agent ressemble à un autre, et que cette même probabilité devient nulle dès lors qu'un élément de ce groupe ne ressemble à aucun autre. Dans ce dernier cas, le sous-graphe ayant pour sommets les agents membres du groupe considéré n'est pas connexe, et

$$\forall i, \forall k, \forall l, p_{ik}(t_0) = 0 \text{ ou } p_{kl}(t_0) = 0 \Rightarrow p_{ikl}(t_0) = 0$$

Par contre, lorsque la chaîne des ressemblances n'est nulle part brisée (le sous-graphe ayant pour sommets les agents membres du groupe considéré est connexe),

$$\forall i, \forall k, \forall l, p_{ik}(t_0) \neq 0 \text{ et } p_{kl}(t_0) \neq 0 \Rightarrow p_{ik}(t_0) = p_{kl}(t_0)$$

et

$$p_{ikl}(t_0) = p_{ik}(t_0) \cdot p_{kl}(t_0) = p_{ik}^2(t_0) = p_{kl}^2(t_0)$$

Chaque $p_{ik}(t_0)$ donnée étant strictement inférieure à l'unité, la probabilité initiale d'établissement de communications entre agents d'un tel groupe décroît donc rapidement avec la taille de celui-ci, sans devenir jamais nulle pour autant. C'est ainsi que notre hypothèse de connexité du graphe associé à $\Gamma(t_0)$ implique que la probabilité initiale d'établissement d'une communication entre tous les agents est très faible - d'autant plus faible que m est grand, puisqu'elle est égale à $p_{ik}^{m-1}(t_0)$ - , mais non nulle: $\forall m$, elle reste donc supérieure à celles d'établissements de communications au sein de groupes pourtant plus restreints, mais où l'un des agents ne ressemble à aucun autre. Au total, les groupes établissant le plus probablement une communication à la date t_0 sont ceux qui comportent exactement deux agents.

Supposons donc ici qu'à la date initiale un certain nombre de communications s'établissent entre des paires d'agents A_i et A_k donnés, et, en relation avec ces communications établies, voyons comment évolue la distribution des probabilités de communication de la date t_0 à la date t_1 . Remarquons immédiatement que nous devons définir l'état initial de Z comme suffisamment complexe pour que la mort du système ne se produise pas avant un nombre suffisamment élevé de périodes. Pour cela, il suffit de supposer que chaque $P(S_i)(t_0)$ contient un assez grand nombre d'éléments qui lui soient spécifiques, ce nombre, noté $\beta(t_0)$, étant identique pour tous les agents. Ainsi $z(t_0) = m(\beta+1)$ est d'autant plus supérieur à $r(t_0) = m$, que β est grand, sans que rien ne change par ailleurs dans la description de $\Gamma(t_0)$, toujours associé à un graphe connexe, spécifié comme ci-

dessus. En particulier, si nous ne tenons pas compte de \emptyset , chaque agent continue de contribuer identiquement à définir l'état initial de Z: $\forall i, [\mu_i / z](t_0) = [2 + \beta] / m[\beta + 1]$.

Cela dit, supposons qu'au cours de leur communication, les agents A_i et A_k échangent des messages comportant D informations nouvelles, au sens où $D = E + F$, en notant E le cardinal du sous-ensemble des catégories élémentaires C_j appartenant à $C(P_{csi})(t_0)$ qui figurent dans le message émis par A_i et reçu par A_k , et F le cardinal du sous-ensemble des catégories élémentaires C_j appartenant à $C(P_{csk})(t_0)$ qui figurent dans le message émis par A_k et reçu par A_i . Il est alors immédiat que

$$[1] \quad p_{ik}(t_1) = r_{ik}(t_1) = [C+D] / [I+E+K+F-(C+D)] = [C+D] / [I+K-C]$$

de sorte que

$$[2] \quad \Delta p_{ik} = p_{ik}(t_1) - p_{ik}(t_0) = D / [I+K-C]$$

Et, puisque $0 < D < I+K-2C$, il vient

$$0 \leq \Delta p_{ik} \leq 1 - C / [I+K-C]$$

C'est-à-dire

$$[3] \quad 0 \leq \Delta p_{ik} \leq 1 - p_{ik}(t_0)$$

Ces résultats permettent d'établir trois propositions: 1) L'équation [2] montre que la probabilité d'établissement d'une communication entre A_i et A_k à la date t_1 est supérieure à ce qu'elle était à la date t_0 , si, et seulement si, la communication établie de t_0 à t_1 entre ces deux agents est synonyme de nouveauté pour au moins l'un d'entre eux: $[I+K-C]$ étant toujours positif, et D étant positif dès lors que E ou F n'est pas nul, Δp_{ik} est nécessairement positif. 2) Les conditions initiales étant fixées, cet accroissement est d'autant plus important que la quantité d'information échangée entre les agents est importante: pour I, K et C donnés, Δp_{ik} est d'autant plus élevé que D est grand. A la limite, si chaque agent communique à l'autre l'intégralité du savoir spécifiquement détenu par lui cet accroissement est maximal. L'équation [3] montre qu'il est alors égal à $1 - p_{ik}(t_0)$. Dans ce cas, $D = D_{\max} = I+E-2C$, et l'équation [1] montre que $p_{ik}(t_1)$ est alors égal à l'unité: il est clair que si chaque agent communique à l'autre la totalité du savoir qui les différenciait au début de la communication, l'issue de celle-ci voit les agents devenus parfaitement identiques, quelles qu'aient pu être leurs différences initiales. 3) Le savoir initial global des agents étant donné, un accroissement Δp_{ik} donné est obtenu avec d'autant moins d'information nouvelle que la fraction de ce savoir global partagée par les agents, donc aussi la probabilité initiale de communication entre eux, sont élevés: l'équation [1] montre que, avec I et K fixés, plus C est grand, plus $p_{ik}(t_0)$ est grand, et plus la quantité D correspondant à une variation donnée $[\Delta p_{ik}]_0$ est petite.

L'ensemble de ces trois propositions signifie donc que *toute communication synonyme d'information nouvelle pour au moins un participant favorise, en s'établissant, les conditions de sa réitération*. Bien qu'ils aient été obtenus dans le cadre d'une communication dyadique et sur une seule période, ces résultats paraissent généralisables à des communications établies entre un nombre quelconque d'agents donnés et, sous certaines conditions, à plusieurs périodes.

Entre un nombre quelconque d'agents donnés: fût-elle très faible, la probabilité d'établissement d'une communication entre un nombre quelconque d'agents n'est jamais nulle tant que la chaîne des ressemblances n'est nulle part brisée entre eux, et, lors d'une communication de ce genre, l'information émise par chaque agent est, par définition, reçue par tous. De sorte que, si cette communication s'établit, la logique cumulative décrite à l'instant se trouve nécessairement à l'oeuvre. La faiblesse relative de la probabilité initiale de communication au sein de groupes composés d'au moins trois agents, par rapport à celle d'une communication dyadique, peut alors se trouver compensée, voire plus que compensée, par la quantité supérieure d'informations nouvelles circulant dans un cadre ainsi élargi, et ceci au terme d'un nombre d'autant plus réduit de périodes que, la population globale étant faiblement différenciée, les p_{ik} sont proches de l'unité.

Mais ces résultats peuvent également être étendus à un nombre quelconque de périodes, tant que le caractère cumulatif du processus analysé plus haut l'emporte sur l'effet de *chevauchements* apparaissant éventuellement entre savoirs respectifs d'agents engagés dans des groupes différents de communication. Afin de préciser ce que seraient de tels chevauchements, imaginons qu'à une date $t \neq t_0$ les probabilités de communications des agents A_i , A_k , A_u et A_v soient distribuées de la manière suivante

$$p_{iu} \ll p_{iv} \ll p_{ku} \ll p_{kv} \ll p_{ik} < p_{uv}$$

La structure de communication la plus probable est alors celle qui rassemble A_i et A_k d'une part, et A_u et A_v d'autre part. Mais il se peut que les messages effectivement échangés entre ces agents lors de ces deux communications simultanées et indépendantes soient tels qu'à la date $t+1$, nous ayons

$$p_{ik} \ll p_{ku} \ll p_{iv} \ll p_{uv} \ll p_{kv} < p_{iu}$$

Il suffit pour cela que, de t à $t+1$

$$\Delta p_{ik} < \Delta p_{uv} \ll \Delta p_{kv} < \Delta p_{iu}$$

c'est-à-dire que, l'effet de chevauchement étant ici plus fort que l'effet cumulatif, la structure de communication la plus probable est devenue celle qui rassemble A_k et A_v d'une part, A_i et A_u d'autre part. Par hypothèse exclue de t_0 , où tous les p_{ik} sont soit égaux, soit nuls, une telle situation n'est pas impossible en t_1 . Sauf si les stratégies d'agents sont très fortement différenciées, elle est néanmoins d'autant moins probable à cette date que les $\mu_i(t_0)$ sont

relativement petits. Par contre, les possibilités de chevauchements sont d'autant plus grandes au fur et à mesure que le volume de Z croît lors des périodes ultérieures. Seuls de tels chevauchements pourraient venir contrecarrer l'évolution que nous voyons se dessiner au bout d'une seule période. Les conditions de l'état initial de Z - beaucoup de complexité pour peu de redondance - rendent ces chevauchements entre savoirs peu vraisemblables.

De tels phénomènes n'ont cependant rien d'impossible, mais l'analyse de leurs conditions d'apparition nécessiterait de spécifier le modèle bien plus précisément que nous ne l'avons fait ici. En effet, notre objectif consistait simplement à montrer que l'une des trajectoires *possibles* du système voyait se poursuivre uniformément l'évolution ici analysée sur une période. Cette trajectoire est telle qu'à partir d'un état partagé de manière homogène entre l'ensemble des agents, il est plus probable que le système évolue vers des états constitués d'*amas différenciés* de sous-systèmes, que son évolution ne tende à reconduire l'homogénéité initiale.

A l'intérieur de chacun de ces amas tendant ainsi à se former, l'*homogénéisation* progressive des savoirs des agents est synonyme d'effacement de leurs singularités individuelles, tandis qu'*entre les amas* le processus est de *différenciation* des savoirs et non d'homogénéisation: une seule et même évolution, conduisant à l'accroissement du volume global de l'économie, tend simultanément à agréger localement et à désagréger globalement cette mémoire. Cette évolution est *la plus probable* pour le système Z, et elle entraîne principalement deux conséquences

Au niveau macroéconomique tout d'abord, une telle évolution paraît susceptible d'abrégé la durée de vie du système. De fait, la probabilité d'établir une communication entre deux agents appartenant à des amas différents ne cessant de décroître, et chacun communiquant ainsi de plus en plus probablement dans une aire purement locale avec d'autres qui lui ressemblent de plus en plus, le système voit se tarir plus rapidement la diversité permettant à sa complexité d'être toujours supérieure à sa redondance. Au lieu d'arriver à la parfaite homogénéité des savoirs au niveau de l'ensemble de Z, impliquée par l'hypothèse d'absence d'effet significatif de renforcement, il est ainsi possible que la structure du système se fige dans un état intermédiaire, consistant en une partition quasi parfaite en sous-ensembles presque entièrement différents, eux-mêmes composés de sous-systèmes parfaitement identiques. Pour cela, il suffirait qu'à force d'être activées à l'intérieur de sous-ensembles voisins d'agents, plutôt qu'entre ces sous-ensembles, les catégories demeurées communes aux éléments de ceux-ci voient la communicabilité de leurs significations respectives s'effacer totalement devant la spécificité qu'elles acquièrent en s'intégrant dans des savoirs différents. Dans ces conditions, ce qu'il resterait de diversité en Z serait enfermé dans une juxtaposition de compartiments pratiquement étanches, de sorte que cette diversité résiduelle serait pour ainsi dire inutilisable par le système: ce dernier ne pourrait aller jusqu'au bout de la logique du processus venant globalement l'homogénéiser, et devrait se contenter d'une homogénéité locale des savoirs laissant inexploitée une réserve de diversité dont l'épuisement aurait prolongé sa trajectoire.

A quelques nuances près dont nous nous sommes expliqué, ce modèle permet de lier les enseignements de I. Rosenfield à ceux de H. Atlan dont il retrouve, en élargissant certains et en restreignant d'autres, les principaux résultats. Dans le champ propre de ce système non construit particulier constitué par la communication entre agents d'une économie décentralisée, comme dans le cadre général tracé par H. Atlan, l'accroissement de complexité a pour cause efficiente la survenue d'un bruit organisationnel. Mais ici, ce bruit est celui que représente, pour chaque sous-système, la mise en communication avec un autre sous-système associant des significations différentes à une information pourtant communicable. Dans notre modèle en outre, le jeu de la diversité du système engendre d'emblée de la redondance simultanément à la complexité: loin de s'opposer selon une logique de vases communicants, redondance et complexité croissent ensemble tout le long de la trajectoire du système, et si celui-ci finit par mourir, ce n'est pas qu'il ne dispose plus de suffisamment de redondance pour retourner (en l'organisant à son profit) l'impact du bruit organisationnel, mais plutôt qu'il étouffe progressivement sous le poids d'une redondance venant inexorablement recouvrir un accroissement net de complexité restant positif jusqu'au moment exact de la mort du système - où cet accroissement net devient nul.

Ainsi que nous venons de le voir, cette mort peut être précipitée par la formation d'amas différenciés au sein du système - un processus ressortissant à la logique même de l'établissement de la communication entre éléments de ce dernier. "Qui se ressemble s'assemble": vieux dicton assurément, dont la profonde vérité trouve ici à s'exprimer en un langage inattendu, et surtout se voit prolongée, par ce langage ensuite retraduit, par "qui s'assemble se ressemblera le plus probablement de plus en plus". Tout au long de la trajectoire du système, cette logique cumulative tend à immobiliser le système avant son terme naturel d'entropie maximale. Mais d'autres forces sont susceptibles de contrecarrer une telle logique, notamment les phénomènes de chevauchement dont nous avons parlé. Par ailleurs, il est possible que le système proprement économique ait trouvé un moyen d'enrayer ce processus de formation d'amas avec l'émergence de la *monnaie*. Car cette dernière fournit à chaque agent la possibilité de s'évader à tout moment de l'amas dans lequel il serait autrement cantonné, en lui permettant d'acheter toute information offerte à la vente par un agent de n'importe quel amas, et d'offrir à la vente toute information susceptible d'intéresser un agent de n'importe quel amas: en ce sens, la monnaie constitue pour le système l'ultime moyen de maintenir son caractère organisé, de même qu'elle constitue pour l'agent individuel la dernière communication efficace possible. Comme celle d'autres conventions, l'émergence de la monnaie signifierait ainsi qu'un processus d'apprentissage d'un type logique supérieur à celui qui consiste à combiner des catégories sur un mode donné est à l'oeuvre dans le système: avec la monnaie, c'est la manière même de communiquer, et de combiner les catégories reçues lors de la communication, qui change. Enfin, d'autres forces encore peuvent, elles, faire reculer indéfiniment l'épuisement de diversité signant la fin du système: le renouvellement du lexique initial, dont seul le caractère fini rendait ici inéluctable la survenue d'une parfaite homogénéisation du savoir, mais aussi les effets de renforcement négligés par notre modèle, et qui n'en sont pas moins partout à l'oeuvre dans ce que nous appelons le langage.

Mais cette évolution du système vers une constellation d'amas entraîne également d'importantes conséquences sur le plan purement microéconomique. Elle imprime notamment sa marque à la nature des différentes stratégies possibles de l'agent visant à participer de manière optimale à une telle évolution. Sans prétendre aller au-delà d'une première exploration de ce problème, déterminons, à partir des conséquences des différentes stratégies mises en oeuvre par un agent au cours de la communication, les stratégies qui se révèlent pour lui les plus pertinentes. Bien entendu, l'analyse menée ici en termes de communications dyadiques peut être étendue à des groupes comportant un nombre quelconque d'agents, dans la mesure où chacun de ces derniers considère l'ensemble des autres comme un seul et même agent.

Dans ces conditions, les stratégies pertinentes s'ordonnent autour de deux comportements contrastés, alternativement observés par l'agent A_i au cours de la communication, et renvoyant chacun à une interprétation du concept de valeur d'option - incitant respectivement l'agent à attendre passivement l'information supplémentaire, et à adopter une attitude plus active en injectant un surcroît d'information dans le système. En effet, la première stratégie possible de l'agent consiste à pratiquer une rétention de l'information, quel que soit le comportement de l'autre agent impliqué dans la communication. L'équation [2] ci-dessus, et son interprétation dans la première des trois propositions qui suivent, montrent bien que la conséquence d'une telle stratégie est de rendre minimale la probabilité de réitération de la communication actuellement établie: pour A_i , quelle que soit la quantité d'information émise par A_k , indiquée par F , pratiquer une rétention d'information est minimiser E , donc minimiser D pour tout niveau donné de F . A la limite, $\Delta p_{ik} = F / [I+K-C]$, lorsque $E = 0$. Une telle attitude apparaît rationnelle à chaque fois qu'il s'agit de prendre une position d'attente, ou de disponibilité envers l'éventuel établissement d'une communication estimée *a priori* plus fructueuse. Elle correspond donc à une stratégie optimale de l'agent A_i dans deux types de situations: 1) lorsque l'agent A_k est perçu par A_i comme pratiquant lui-même une rétention d'information, c'est-à-dire quand F anticipé est nul; alors D anticipé par A_i est nul, et la probabilité de réitération de la communication reste identique à la probabilité de celle-ci avant qu'elle ne s'établisse; 2) lorsque l'agent A_k est perçu par A_i comme étant peu susceptible d'émettre en sa direction, suite à la réception d'une quantité importante d'information, un message riche en potentialités nouvelles de combinaisons de catégories pour A_i ; c'est alors $\mu_k(t_0)$ anticipé par A_i qui est petit, de sorte que A_i refuse d'accroître la probabilité de réitération d'une communication potentiellement très pauvre en information pour lui.

Parfaitement rationnelle dans les deux cas mentionnés à l'instant, la rétention de l'information consiste donc à différer la redondance afin de rester ouvert à la complexité. Une telle stratégie évoque le cadre classique du concept de valeur d'option, dans lequel l'agent reporte les décisions irréversibles dans l'attente d'une information supplémentaire, produite de manière exogène au modèle. Il est évident qu'un tel comportement ne présente aucun caractère rationnel dans le long terme: à moins que tous les agents ne se comportent de même, auquel cas la structure du système Z demeurerait parfaitement inerte, ce n'est pas en participant petitement à des communications pauvres à toutes les dates que l'agent A_i va

participer de manière optimale à l'évolution du système! Dès lors qu'à une date au moins un seul couple d'agents suivrait la stratégie inverse en manifestant une commune propension à émettre un maximum d'information lors de la communication, l'objectif de A_i serait manqué. Sur le long terme, il lui faut donc préférer l'autre stratégie possible, consistant précisément à émettre le maximum d'information lors de sa communication avec un autre agent A_k . Au lieu de différer la redondance pour rester ouvert à la complexité, il s'agit ici à la fois de créer de la redondance et de la complexité entre lui et A_k , et de créer tous deux de la complexité pour le reste du système. Il est clair que cette stratégie, correspondant à une valeur d'option "active", est la plus rationnelle dans les deux cas inverses à ceux mentionnés auparavant, c'est-à-dire lorsque : 1) A_i anticipe une valeur élevée de F lors de la communication présente; 2) lorsqu'il anticipe $\mu_k(t_0)$ élevé, de sorte que la recombinaison de $P(S_k)(t_0)$ avec l'information correspondant à une valeur élevée de E se traduira par $P(S_k)(t_1)$ donnant lieu à une valeur anticipée élevée de F lors de la communication établie entre A_i et A_k à la date t_1 - une communication dont la probabilité anticipée est ainsi supérieure à celle établie entre ces agents à la date t_0 .

Naturellement, la rapide présentation des stratégies optimales des agents, proposée à l'instant, est des plus caricaturales: il est bien évident que la rétention d'information en laquelle consiste la première de ces stratégies ne peut jamais être totale, puisque les agents, étant différents, ne peuvent pas ne pas communiquer (échanger une quantité nulle d'information) dès lors qu'ils sont mis en présence au sein d'une partie commune de la structure de communication de l'ensemble du système. De même, l'émission réciproquement maximale figurant la seconde stratégie ne peut, ni ne doit, être réellement maximale lors de la communication, puisque cela signifierait qu'à l'issue de celle-ci les agents, étant devenus parfaitement identiques, ne sauraient que se dire lors de la communication certaine qui suivrait, chacun étant désormais parfaitement transparent à l'autre. Comme souvent, tout ici est dans le dosage, la stratégie optimale consistant à trouver dans chaque période, l'horizon temporel des agents étant donné, l'arbitrage le plus adapté aux circonstances - c'est-à-dire à ce que chacun peut percevoir de l'état actuel du système, et anticiper de ses évolutions futures possibles.

Pour aussi caricaturale qu'elle soit, cette présentation esquisse cependant à grands traits ce que pourrait être une stratégie individuelle optimale au sein d'un système évolutionniste liant, par la durée que trace son mouvement même, les intérêts réciproques de agents. En particulier, une telle présentation débouche directement sur la nécessité d'une communication efficace au sein de groupes stables, rationnellement animés d'un souci de flexibilité dynamique dès lors qu'ils sont insérés dans un environnement incertain synonyme de régime d'informations perturbées, ainsi que le montrent P. Llerena et P. Cohendet dans le texte suivant.

Notes

(1) Ce modèle représente la version actuelle d'une théorie de la communication sociale que nous développons depuis longtemps [cf. B. Ancori, 1979, 1983, 1987a, 1987b], amorcée sans que nous ne connaissions les très importants travaux de H. Atlan, et sensiblement revivifiée aujourd'hui par notre récente lecture de I. Rosenfield. S'il atteint son objectif, ce modèle constituera donc en lui-même une illustration du processus qu'il prétend décrire.

(2) Pour une introduction à l'utilisation de telles structures d'information dans ce type de marchés, voir D.M. Kreps [1990, pp. 217-224]. Notamment utilisée par C. Henry [1974a, 1974b] dans le cadre de la théorie de la valeur d'option appliquée à l'investissement irréversible, pareille formalisation d'une information croissante ressortit à la théorie de l'information plus généralement développée par J. Kampe de Fériet et B. Forte contre l'approche probabiliste de l'information [cf. M. Willinger, 1988, pp. 306 sq.].

(3) Cependant, une "logique épistémique" reliée à la théorie des jeux essaie depuis peu de mieux endogénéiser ce type d'apprentissage à la théorie économique, en voulant notamment s'appliquer à la genèse des institutions. Voir, sur ce point, B. Walliser [1989, 1990].

(4) Il est clair en effet qu'incertitude et irréversibilité n'ont de portée économique qu'ensemble: en l'absence d'incertitude, l'irréversibilité de ses décisions ne porterait nullement à conséquence pour l'agent, qui se situerait d'emblée sur une trajectoire optimale dont toutes les conditions de réalisation lui seraient connues à la date initiale; réciproquement, sans l'irréversibilité pesant sur les conséquences de ses décisions, cet agent serait parfaitement indifférent à l'incertitude dans laquelle il prend celles-ci : par hypothèse, il pourrait en permanence, et sans coût, reformuler ses plans à la lumière d'informations supplémentaires sur les états du monde, le système des prix ou son propre système de préférences. Or il se trouve que, loin de prendre en une seule fois la suite de décisions optimales dont il n'aurait plus qu'à enregistrer les résultats parfaitement conformes à ses prévisions, ou de transformer gratuitement le passé en présent, l'agent économique prend séquentiellement des décisions dont les conséquences successives ponctuent une suite irréversible de périodes, toutes marquées de l'incertitude affectant alors son environnement pertinent.

(5) Rien n'assure qu'à chaque date, tout agent puisse communiquer effectivement avec tout autre. Au contraire, I. Rosenfield [1989] ne cesse d'insister sur la singularité individuelle du cerveau de chacun. Transposée au niveau des fonctions cérébrales supérieures régissant la conscience et le langage, et radicalisée à l'extrême à un tel niveau, cette singularité pourrait déboucher sur un système Z tel que $\forall t, i \neq k \Rightarrow S_i(t) \cap S_k(t) = \emptyset$. Un cas très particulier de cette absence totale de spécification de Z serait celui où $[a_{ij}](t)$ serait une matrice unité. Puisqu'une telle matrice est symétrique, chaque $S_i(t)$ ne contiendrait qu'un seul élément, de sorte qu'il n'existerait aucun langage commun entre agents: chacun d'eux ne connaissant qu'une seule catégorie, différente de celle connue de tout autre agent, nous aurions ainsi la représentation d'un solipsisme généralisé dans Z .

(6) Evidemment, la manière la plus immédiate de satisfaire à l'exigence précédente consisterait à introduire la catégorie "monnaie" comme élément de $S(t)$. Formellement, la monnaie serait ainsi la catégorie C_0 , élément de l'ensemble S , telle que $\forall t, \forall i, a_{i0}(t) = 1$. Une telle catégorie associerait un graphe K_m -complet à toute matrice $[a_{ij}](t)$: quelle que soit la date considérée, l'ensemble des sommets de ce graphe étant l'ensemble M des agents, et l'ensemble de ses arcs étant celui des communications possibles entre ces agents, il existerait au moins un chemin entre deux sous-ensembles quelconques de sommets (le graphe est connexe), et ce chemin serait direct entre deux sommets quelconques du graphe (qui est donc complet). Quant à la signification d'une telle monnaie, elle serait ici double. Au niveau des voies élémentaires de transmission à l'intérieur de Z , elle accorderait les deux dimensions antinomiques de l'agent, dont l'être social s'oppose à la singularité : aussi singulier soit-il, tout agent donné pourrait toujours établir au moins une communication avec tout autre agent par l'intermédiaire de la monnaie. Du point de vue de la voie établie entre l'observateur et Z , l'introduction de la monnaie diminuerait le degré de complexité de ce dernier en lui conférant statut de véritable système organisé. Ainsi, le seul fait de considérer l'un des éléments C_j de S comme étant commun à tous les S_i associerait une redondance non nulle à l'état correspondant du système. Néanmoins nous n'utiliserons pas une telle formalisation, car il est clair que le modèle n'exige pas qu'il existe une telle monnaie pour se prêter à la représentation d'états non solipsistes de Z . Cette existence est une condition suffisante, mais nullement nécessaire, pour que la communication soit possible entre deux agents donnés: cette possibilité existe dès lors que ces deux agents connaissent simultanément une catégorie C_j quelconque. Simplement, cette catégorie

n'étant pas la monnaie (formalisée par C_0), la communication ne peut être tenue pour possible entre deux éléments *quelconques* du système Z. Dans la dernière partie de ce texte, nous suggérerons que la monnaie *émerge* comme un effet possible de l'évolution du système Z, plutôt que d'être donnée d'emblée dans la description de celui-ci.

(7) Raisonnant ici sur une seule période, nous n'avons guère à prendre en compte le phénomène de l'oubli. Nous suggérerons plus loin une interprétation que peut recevoir ce phénomène, fondamental pour toute théorie de la mémoire, dans les termes de notre modèle.

(8) Cette observation apparaît encore plus fondée si l'on considère le type de mémoire associée au langage. Selon E.T. Hall en effet, "quel que soit le point de départ utilisé, les symboles ont toujours une composante collective et une composante individuelle. Jamais deux personnes n'utilisent le même mot exactement de la même façon, et plus le symbole est abstrait, plus il est vraisemblable que le facteur individuel ait de l'importance." [1979, p. 29, n. 2].

(9) Pareille conception est tout à fait fidèle à la pensée de I. Rosenfield [1990], qui écrit: "Toute image consciente, toute pensée, est une transformation du passé et du présent, une création qui n'est ni l'un ni l'autre." [op. cit. p., 19]. Cette idée revient souvent sous sa plume [cf. pp. 27, 40, 51-52, 56-57, 78, 110-111]. Nous ne faisons que lui ajouter ici la dimension du devenir, conformément à la perspective finaliste qui est traditionnellement celle de l'*homo oeconomicus*.

(10) Nous pouvons en effet admettre que, de manière analogue mais non identique aux conventions formalisées par des catégories de type C_0 , le contenu sémantique correspondant à l'ensemble vide ne se situe pas au même niveau logique que celui des autres catégories C_j , mais à un niveau à la fois supérieur et davantage enfoui: celui où s'opère la conscience de soi du sujet connaissant. Voir sur ce point, J. A. Miller [1966] et G. Rosolato [1985, pp. 133-165].

(11) Lever cette hypothèse consisterait à introduire ici des combinaisons de catégories de type " C_3^2 et C_5 " ou " C_3^2 et C_2 ", c'est-à-dire à coefficienter davantage toute catégorie déjà connue de l'agent et à nouveau reçue par lui lors d'une communication. Pareil traitement formel rendrait bien compte du processus biologique de renforcement des jonctions synaptiques de notre cerveau après la naissance, décrit par G. Edelman et I. Rosenfield [1989, p. 166], et, sur un plan sociologico-politique, correspondrait à la notion de "valeur": de manière générale, la présence d'une "valeur" socio-politique (liberté, solidarité, patriotisme, etc.) serait d'autant plus marquée chez un sujet connaissant A_i que les catégories C_j correspondantes auraient des exposants élevés dans son $P(S_i)$ à chaque date considérée. Au prix de quelques sophistications supplémentaires, notre modèle est donc parfaitement capable de décrire une évolution incorporant certains contenus historiques, et évitant au système Z la perspective quelque peu mélancolique de la mort entropique évoquée ici.

(12) Sur l'exemple traité plus haut, il est facile de vérifier qu'entre t_0 et t_1 , l'indice r_{ik} serait passé de $1/14$ à $7/24$ si nous l'avions quantifié par $[\mu_c - 1] / [\mu_i + \mu_k - \mu_c]$, alors que cet indice passe de $1/5$ à $3/5$ entre ces deux dates s'il est quantifié par $C / [I + K - C]$. L'ampleur de la variation de r_{ik} est évidemment différente selon la quantification adoptée, mais seul le *signe* de cette variation nous intéresse, et l'hypothèse adoptée ici est sans effet sur celui-ci.

BIBLIOGRAPHIE

ANCORI B. [1979], *"Pouvoir et information : d'un instant à l'autre"*, Thèse de Doctorat en Sciences de l'information, Université des Sciences Juridiques, Politiques, Sociales et de Technologie, Strasbourg

ANCORI B. [1983], "Communication, information et pouvoir", Séminaire Interdisciplinaire du Collège de France, in Lichnerowicz A., Perroux F. et Gadoffre, *"Information et communication"*, Maloine, pp.59-84.

ANCORI B. [1987a], "Génération d'information dans une économie complexe", in Ancori B. et al. : *"Evolution économique et Théories de l'information"*, A.R.I. sur les Sciences de la Communication, B.E.T.A./C.N.R.S., mars, pp. 194-209

ANCORI B. [1987b], "L'émergence d'un consensus monétaire dans le cadre d'une Economie informationnelle", Communication au deuxième *Colloque de l'Association Charles Gide pour l'Etude de la Pensée Economique*, Nice, septembre

BATESON G. [1977], *"Vers une écologie de l'esprit, I"*, Seuil

BATESON G. [1980], *"Vers une écologie de l'esprit, II"*, Seuil

BATESON G. [1984], *"La nature et la pensée"*, Seuil

DUPUY J.-P. [1989], "Convention et Common knowledge", *Revue Economique*, Vol. 40, n°2, mars, pp. 361-400

HALL E.T. [1979], *"Au-delà de la culture"*, Seuil

HENRY C. [1974a], "Option values in the economics of irreplaceable assets", *Review of Economic Studies*, Symposium on the Economics of Exhaustible Resources, pp. 89-104

HENRY C. [1974b], "Investment decisions under uncertainty: the 'irreversibility effect'", *American Economic Review*, Vol. 64, december, pp. 1006-1012

KIRMAN A.P. [1986], "Organisation et Communication dans les marchés", *Economie Appliquée*, XXXVII, n°3, pp. 597-609

KREPS D.M. [1990], *"A course in microeconomic theory"*, Harvester Wheatsheaf

MILLER J.A. [1966], "La Suture. (Eléments de la logique du signifiant)", *Cahiers pour l'analyse*, 1, janvier-février, pp. 37-49

PASSET R. [1979], *"L'économie et le vivant"*, Payot

PASSET R. [1980], "La thermodynamique d'un monde vivant. Des structures dissipatives à l'économie", *Futuribles*, décembre, pp. 3-25

PASSET R. [1985], "L'Economie: des choses mortes au vivant", *Encyclopaedia Universalis*, Symposium, pp. 831-841

PASSET R. [1987], "Les systèmes complexes comme grille de lecture de l'évolution économique", Communication au séminaire *"Linguistique et Informatique"*, Université des Sciences Humaines de Strasbourg, 16 février

ROSENFELD I. [1989], *"L'invention de la mémoire. Le cerveau, nouvelles donnes"*, Eshel

ROSENFELD I. [1990], *"La conscience, une biologie du moi"*, Eshel

ROSOLATO G. [1985], *"Eléments de l'interprétation"*, Gallimard

SERRES M. [1968], *"Hermès I. La Communication"*, Minuit

WALLISER B. [1989], "Logique épistémique et Théorie des Jeux", ENPC-CERAS

WALLISER B. [1990], "Théorie des Jeux et genèse des institutions", *Recherches Economiques de Louvain*, mars

WILLINGER M. [1988], *"Information et fondements de la rationalité dans l'incertain: trois essais sur le statut de l'information dans la théorie de la décision en incertain"*, Thèse de Doctorat de Sciences Economiques, Université Louis Pasteur (Strasbourg I)

CHAPITRE 5

**NATURE DE L'INFORMATION, COÛTS DE TRANSACTION ET
ORGANISATION DE L'ENTREPRISE**

par

Patrick COHENDET et Patrick LLERENA

Le rôle de l'organisation dans la théorie économique est celui d'une éviction. Dès l'origine de la constitution de la science économique, le "marché" a occupé la place de la catégorie fondatrice. Or, depuis plusieurs années, un certain nombre de travaux issus d'origines très différentes (analyse des conséquences des marchés incomplets et de la concurrence imparfaite, théorie des incitations en présence d'asymétrie d'information, théorie des contrats, etc ...) tendent à replacer l'organisation au centre des réflexions de la théorie économique. L'objectif de cet article est d'apporter un nouvel éclairage à la reconnaissance de l'importance de la notion d'organisation en tant que catégorie fondatrice de la science économique, en analysant de manière approfondie les rapports entre la firme et son environnement, et de façon plus précise en montrant la dépendance de la structure de la firme par rapport à la nature des informations issues de son environnement. Dans cet esprit, deux types extrêmes d'environnement économique (l'environnement économique d'une firme étant caractérisé à un moment donné par un certain état de la demande, de la concurrence et de la technologie) peuvent être distingués.

Premièrement, un environnement économique caractérisé par un régime d'information "stable", au sens où les informations issues de l'environnement peuvent être appréhendées sous la forme de fréquence d'apparition ou de loi de probabilité connues par les agents économiques. On peut montrer dans ce cadre (qui est celui de la théorie traditionnelle) qu'il existe une indépendance au sens large entre l'organisation de l'entreprise et l'évolution de l'environnement. Même si dans une telle situation, selon la variabilité de l'environnement, une firme peut effectivement se doter d'une certaine souplesse d'adaptation à travers une organisation "flexible" (firme multiproduits, recherche d'économie de variété, mise en parallèle de processus de production, etc ...), cette forme d'adaptation de l'entreprise est déterminable une fois pour toutes grâce à l'hypothèse du caractère prévisible de l'information. Il est alors aisé dans un tel contexte d'isoler, d'un côté, ce qui relève des mécanismes économiques de marché (détermination des quantités à produire en fonction des prix, détermination des inputs, etc ...) et de l'autre, de la gestion proprement dite de l'organisation dont la structure est donnée.

Deuxièmement, à l'autre extrême, on peut considérer un environnement caractérisé par un régime d'information fondamentalement perturbé, c'est-à-dire un contexte où les variables de l'environnement relèvent d'une incertitude au sens pur. Dans ce cadre, les individus ne peuvent plus anticiper l'évolution des variables de leur environnement. Ils sont contraints d'effectuer un processus d'apprentissage de l'information, de sorte que l'hypothèse d'indépendance entre l'organisation et l'évolution de l'environnement ne tient plus. L'organisation devient une variable endogène au décideur, en même temps que la frontière entre la firme et le marché est remise perpétuellement en cause. L'impossibilité qui en résulte de séparer l'analyse interne des organisations et l'analyse des mécanismes cognitifs individuels peut être interprétée selon les travaux fondamentaux de H.Simon sur le sujet. Mais de redoutables

problèmes théoriques demeurent, notamment ceux que pose la prise en compte de dispositifs cognitifs collectifs comme réponse appropriée à un environnement d'information perturbé pour atteindre une certaine efficacité minimale.

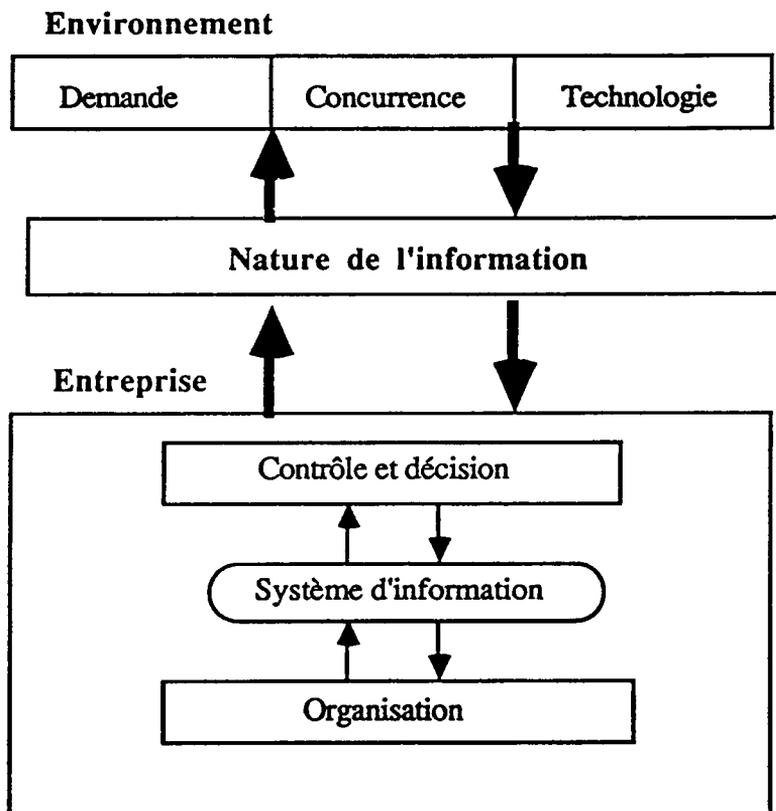
Avant d'aborder ces problèmes, on étudiera tout d'abord dans une première partie (**L'évolution des systèmes de production : d'un régime d'information stable à un régime d'information perturbée**), les raisons pour lesquelles on peut légitimement émettre l'hypothèse que les organisations industrielles sont amenées à affronter des environnements de plus en plus instables. Dans une seconde partie (**La recherche de flexibilité: une stratégie structurante**), on montrera pourquoi face à un environnement caractérisé par un régime d'information de plus en plus perturbée, la recherche de flexibilité (au sens dynamique de ce terme) apparaît être un objectif de plus en plus fondamental, autour duquel se réordonnent les autres objectifs de l'entreprise, et à partir duquel doivent aussi se reconstruire les méthodes et les critères d'évaluation. Une troisième partie s'attachera à la question de l'**organisation** et de sa structuration face à un environnement d'information perturbé (pour atteindre, en particulier, une certaine efficacité minimale). Dans ce contexte, on montrera le rôle des procédures d'intégration (**Le rôle des procédures d'intégration comme facteur clé de fonctionnement des systèmes de production**). Finalement, nous tenterons une proposition: l'utilisation de la théorie des coûts de transaction comme fondement théorique des procédures d'intégration interne des entreprises (**Les coûts de transaction comme mode d'évaluation de l'intégration**)(1). Cette réflexion nous amènera à tracer quelques pistes de recherches pour l'avenir.

1. L'évolution des systèmes de production: d'un régime d'information stable à un régime d'information perturbée

Depuis le début de l'ère industrielle, la conception des systèmes de production n'a jamais cessé d'évoluer vers des niveaux de complexité de plus en plus élevés. De manière schématique, si l'on admet qu'un système de production à un moment donné se caractérise par la donnée d'un certain type d'organisation, pour répondre à un certain type d'environnement économique (défini par la donnée de l'état de la demande, de la concurrence et de la technologie) selon un schéma de référence classique (cf. Figure 1). Il est possible de distinguer trois stades successifs dans l'évolution des modèles de production : le modèle de "standardisation", puis le modèle de "variété", puis enfin le modèle de "réactivité". L'interprétation de ces trois modèles successifs, qui est décrite ci-dessous, fait apparaître deux phénomènes essentiels : le premier est la "montée en puissance" de la notion de flexibilité, concept multidimensionnel qui devient progressivement l'un des objectifs majeurs des organisations au fur et à mesure que celles-ci expérimentent des niveaux de complexité élevés ; et le second est la **profonde rupture qui se produit dans la transition entre le modèle de variété et le modèle de réactivité, en raison d'un**

changement radical dans la manière dont circule le flux d'informations entre l'entreprise et son environnement. Jusqu'au modèle de réactivité, le système d'information pouvait, en effet, être considéré comme stable et répétitif; à partir du modèle de réactivité, il devient essentiellement perturbé. Or, on montrera que c'est à cause de cette dernière caractéristique que la conception de l'organisation de l'entreprise doit être nécessairement profondément repensée dans l'optique d'une organisation globale "intégrée".

Figure1: Le modèle de référence



1.1 Le modèle de standardisation ou le monde de la certitude

Durant une longue période, le principe économique dominant, autour duquel se sont structurés les différents modèles économiques et les représentations de la firme, a été celui établi par Adam Smith [1776], affirmant l'existence d'un **cercle vertueux** entre extension du marché - accroissement de la productivité - baisse des prix - extension des marchés etc... Le modèle économique de l'entreprise correspondant le mieux à ce principe fondamental est très certainement le modèle de **standardisation** (dont l'archétype est le modèle dit "taylorien-fordien") que l'on peut décrire de la manière suivante :

L'environnement économique du modèle de standardisation est défini par une demande mono-caractéristique (ou à caractéristiques réduites) en croissance régulière, par un régime de concurrence fondé sur les prix (via les quantités produites) et par une technologie axée sur le développement du machinisme accélérant la division du travail. A cet environnement économique correspond une **organisation** de l'entreprise (organisation pyramidale, avec séparation des fonctions, cloisement des services, etc...) caractérisée par le respect (à l'échelle microscopique) du principe fondamental d'extension du marché - division du travail, à travers la recherche de rendements d'échelle. Un aspect remarquable est que les normes de gestion, qui se mettent progressivement en place, viennent épouser cette même logique. Ainsi la division du travail permet à chaque étape de son développement de dégager une architecture précise des tâches. Or chaque tâche correspond à une activité génératrice de coûts, activité que l'on attribue soit à un homme, soit à une machine, soit à un couple homme-machine. Alors, il est logique dans ce cadre de centrer les indicateurs de performance sur le travail direct, et de ranger ces indicateurs dans la classe des mesures de la productivité des facteurs de production. A partir de cette notion de travail direct, vont s'organiser les principes d'évaluation de la comptabilité analytique, en se fondant sur l'hypothèse que pour chaque sous-ensemble de production, le coût de travail direct est une part importante du coût total : cette hypothèse implique tout d'abord que c'est dans le coût du travail direct que résident les principaux "gisements" de productivité (les économies de coûts potentielles) ; cette hypothèse fournit aussi des clefs de répartition cohérentes pour l'affectation des coûts indirects (les coûts indirects sont affectés à une activité au prorata des coûts directs induits par cette activité et une affectation erronée d'un coût indirect n'aura donc qu'une importance limitée en raison de sa part relativement faible dans le coût global).

Dans le cadre du modèle de standardisation, la cohérence du système d'évaluation comptable permet non seulement l'évaluation des performances (ex post), elle permet aussi de fonder les prises de décisions d'investissement (ex ante). En effet, pour les investissements, le système comptable est en mesure de dégager des comptes prévisionnels, et notamment des cash-flows anticipés permettant le calcul des critères financiers classiques (délai de retour, valeur actualisée nette, taux interne de rentabilité, etc...). Ces critères sont d'autant plus pertinents que le caractère séquentiel des différentes phases productives, séparées les unes des autres, implique

en première approximation une propriété d'additivité des coûts relatifs aux différentes phases productives. Il est ainsi possible de cerner la contribution relative de chaque phase productive au coût total de production, et donc de mesurer l'intérêt de remplacer un type d'équipement par un autre. Enfin, la possibilité de mesurer assez précisément l'intérêt d'investir dans une machine particulière est facilitée par le fait que les machines sont, dans un tel modèle, dédiées à un produit donné (qui, lui même, a un cycle de vie généralement long).

De plus, cette cohérence globale n'est pas seulement interne, elle est aussi externe. Ce sont les mêmes critères financiers qui fondent les méthodes de valorisation des entreprises par leurs partenaires (banques, actionnaires, acheteurs potentiels, etc...). Ainsi, on "voit" l'entreprise de la même façon que l'on soit un observateur interne ou un observateur externe. Si une telle cohérence est possible et envisageable, cela provient d'une forme de "consensus" sur le devenir de l'entreprise, sur la puissance du mécanisme sous-jacent d'extension du marché - division du travail. Cette confiance diffuse implique une certaine convergence des anticipations, l'émergence d'une "quasi-certitude" et par conséquent une fiabilité et une homogénéité des méthodes d'évaluation.

Le modèle de standardisation se caractérise donc par une remarquable cohérence entre les principes économiques qui le fondent et les normes de gestion qui définissent la conduite des entreprises. Or, l'origine de cette cohérence doit être mise en lumière, car elle constitue le fondement même du fonctionnement du modèle. Cette cohérence repose, en effet, sur la perception par l'entreprise d'un système d'information stable en provenance de l'environnement économique (cf. Figure 1). Stable est ici pris dans le sens de réparable par des lois de probabilités ou des fréquences d'apparition connues du décideur. Qu'il s'agisse des prix, des stratégies des concurrents, de l'état de la technologie, le fonctionnement du modèle de standardisation suppose la réception par l'entreprise d'informations répétitives et prévisibles qui autorisent une caractéristique fondamentale, à savoir que **l'organisation de la firme peut être alors conçue indépendamment de l'état de l'environnement**. Cette propriété (qui est l'une des hypothèses de base de la théorie économique néo-classique) délimite une frontière claire entre l'univers économique (le marché) et celui de la gestion (l'organisation de la firme). Les rôles sont non seulement clairement distribués, ils sont aussi étroitement complémentaires.

1.2 Le modèle de variété ou le monde du divers

Les efforts réalisés dès 1920 par General Motors, pour diversifier sa gamme de production en pratiquant la concentration par intégration technique, économique et financière pour acquérir une taille suffisante permettant la mobilisation de ressources financières à affecter au développement de nouveaux modèles, ouvrent une deuxième période : celle que l'on peut qualifier de **variété programmée** qui s'est étendue au moins jusqu'au début des années soixante. La transition progressive d'une demande homogène vers une demande variée,

incertaine et exigeante sur la qualité des produits, apparaît comme une tendance irréversible de l'environnement économique. Les utilisateurs réclament ainsi non seulement des caractéristiques personnalisées, mais aussi exigent de plus en plus que des "services" soient davantage incorporés dans les produits (délais de livraison courts, service après-vente, maintenance, fiabilité, etc ...). Naturellement, cette exigence en matière de services est devenue l'un des facteurs décisifs de concurrence entre entreprises. Par ailleurs, le développement technologique des machines à commande numérique et des centres d'usinage, en permettant aux machines d'effectuer différents types d'opérations tout en réduisant les difficultés de reconversion des lignes, a considérablement facilité la transition vers un modèle de variété. Sur le plan théorique, le modèle de variété repose sur la notion de "economies of scope" (économies de champ, ou de variété) qui apparaissent lorsque le coût de combiner deux (ou plusieurs) lignes de production, dans une unité de production, est inférieur à celui de les produire séparément.

L'ensemble des caractéristiques qui viennent d'être évoquées montre de notables différences entre le modèle de variété et le modèle de standardisation. Toutefois, sur un point tout à fait fondamental, les deux modèles coïncident toujours : à savoir qu'il n'y a, dans aucun des modèles, de remise en cause de la logique d'indépendance entre l'organisation de la firme et l'état de l'environnement. Et le maintien de cette logique provient de ce que le flux d'information, issu de l'environnement, peut être considéré comme stable, prévisible, répétitif, rendant les décisions programmables. Il n'y a pas de remise en cause radicale de la cohérence entre les principes économiques et les normes de gestion. Par ailleurs, dans le modèle de variété, on assiste à la même séparation entre les services fonctionnels et les services de production. Naturellement, un certain nombre d'adaptations sont nécessaires pour permettre à la structure d'organisation, dans le modèle de variété, de rester économiquement viable. Ainsi, le concept de "différenciation retardée", qui signifie une recherche de standardisation des composants à l'amont pour permettre une plus grande diversité à l'aval, constitue l'une des formes d'adaptation requise pour assurer la viabilité du système.

Sur le plan théorique, il est remarquable de rappeler que l'on peut démontrer que, lorsque les économies de variété sont suffisamment importantes, un phénomène d'économie d'échelle peut être obtenu même si sur chacune des lignes de production, considérée individuellement, des rendements constants voire décroissants sont constatés (W.J. Baumol, J.C. Panzar, R.D. Willig, 1982). Ces résultats montrent qu'il n'y a pas nécessairement incompatibilité entre la variété des produits et les économies d'échelle. En d'autres termes, le nouveau modèle de variété possède la capacité d'"exprimer" les avantages du modèle de standardisation.

1.3 Le modèle de réactivité ou le monde de l'incertain

A partir des années 60, un certain nombre d'entreprises développent une stratégie d'organisation basée sur la "réactivité", en privilégiant la réduction des délais de réponse de l'entreprise aux sollicitations extérieures. Cette stratégie correspond, d'une certaine manière, à un troisième stade de l'état de la concurrence : après la concurrence par les prix et quantités correspondant au modèle de standardisation, puis le stade de la différenciation dans le régime de variété, la concurrence se déplace sur les délais de réaction (tout en essayant naturellement de continuer à répondre à une demande de plus en plus variée). Ce déplacement de la concurrence vers la notion de délai de réaction correspond à de nouvelles exigences, apparues au niveau de la demande, où les utilisateurs réclament non seulement des caractéristiques personnalisées mais exigent de plus en plus que des services rapides et fiables soient davantage incorporés dans les produits (délais de livraison courts, services après vente, etc...). Par ailleurs, les progrès de la technologie, notamment à travers les développements de la microélectronique, de l'informatique industrielle et des logiciels associés (CAO, CFAO, etc...) ont donné aux entreprises les moyens techniques de fonctionner dans des délais plus courts.

Pour s'adapter à un environnement économique de ce type, les entreprises se trouvent contraintes de bouleverser leurs règles de comportement et leurs méthodes d'organisation. Mais, surtout le caractère essentiellement **déstabilisateur** de la recherche permanent d'une réduction des délais de réaction, introduit la nécessité d'une **étroite dépendance entre l'organisation de l'entreprise et l'environnement**. Il s'agit là d'une rupture capitale avec le système ancien, et ce aussi bien sur le plan théorique que sur le plan pratique. Faisant face à une demande évolutive et de plus en plus imprévisible de produits variés à durée de vie très courte, l'entreprise ne peut plus considérer son organisation comme une donnée immuable. **Confrontée à la réception d'informations perturbées**, l'entreprise doit adapter continuellement son organisation à l'évolution de l'environnement. La distinction claire qui existait dans le modèle de standardisation entre le marché et l'organisation ne tient plus. L'un des aspects remarquables de cette rupture dans les modèles de production est qu'elle coïncide avec une profonde rupture au niveau théorique. En effet, dans la théorie microéconomique traditionnelle, l'une des hypothèses fondamentales est celle d'**indépendance** entre l'organisation de la firme et l'environnement. Cette hypothèse assurait une stricte répartition thématique entre l'économie (analysant les relations du marché) et la gestion (analysant les organisations). La disparition de l'hypothèse d'indépendance ouvre ainsi non seulement la voie à un renouvellement de la théorie économique mais repose fondamentalement le problème de la séparation entre économie et gestion. La dépendance entre organisation et environnement peut être conçue de deux manières: soit l'organisation est capable de s'adapter rapidement aux changements de l'environnement, soit elle est capable de maîtriser, d'influencer cet environnement.

Par ailleurs, les principes économiques régissant les conditions de fonctionnement et de réussite des entreprises se trouvent découplés de normes de gestion utilisées et héritées des modèles du passé. L'une des conséquences perceptibles a trait, par exemple, à l'évaluation du coût d'un produit : l'ordre des séquences d'opérations n'étant plus immuable, la somme des coûts de production de chaque phase productive n'est plus appropriée - même en première approximation - pour appréhender les coûts du processus dans son ensemble. La propriété d'additivité, utilisée pour le calcul du coût d'un produit ou d'une activité, disparaît. En effet, elle nécessite une indépendance totale ("intersection vide") des éléments objets de la sommation. Or, les nouveaux systèmes de production se caractérisent par l'existence de fortes interdépendances entre les phases de production. Il devient ainsi de plus en plus difficile de mesurer la contribution relative au coût globale de l'entreprise d'un atelier particulier, voire d'un type de machine particulier. En revanche, dans la mesure où l'agencement des opérations est susceptible de varier dans le temps, ce sont les coûts des interactions entre opérations, et plus généralement ceux liés aux choix d'un ordonnancement, qui prennent de plus en plus d'importance; et précisément ces interactions révèlent des coûts en termes d'optimisation des temps de fonctionnement des équipements, de stocks, de flux d'énergie et d'information, de contrôle, etc... donnant notamment lieu à des coûts cachés précédemment évoqués. En d'autres termes, à la notion de processus séquentiel de phases distinctes se substitue la notion de processus en réseau, et à la recherche prioritaire de temps unitaire minimale se substitue celle de temps de reconversion minimal des machines. Dans la même optique, l'impossibilité d'évaluer la durée de vie d'un équipement et le nombre de processus dans lesquels il sera employé pose de graves problèmes pour juger de sa rentabilité. Pour prendre un exemple dans l'automobile:

"étant donné l'importance de l'investissement mis en jeu, c'est aujourd'hui à l'occasion du démarrage d'un nouveau modèle qu'un atelier robotisé d'assemblage-soudure est mis en place. Progressivement, en fonction de l'arrivée des nouveaux modèles dans l'usine où elle est intégrée, l'engagement de la chaîne flexible dépend de moins en moins de la seule courbe de vie du premier modèle, en s'identifiant à l'engagement de l'usine, au même titre que les installations de peinture ou les chaînes de montage. Si l'on compare une telle chaîne flexible avec les moyens spécifiques automatiques, la durée d'étude n'est plus la durée de vie du premier produit, mais la durée de la chaîne flexible, que l'on espère beaucoup plus longue" (J.Bultel, 1983)

Si la perception d'une rupture profonde par rapport à l'ancien modèle paraît manifeste, il reste d'une part à en interpréter l'impact exact sur le comportement des entreprises notamment en termes de recherche de flexibilité et d'autre part à en retrouver les principes directeurs en termes organisationnels et notamment en explicitant les processus d'intégrations que ces évolutions suscitent. Il sera alors possible de proposer une voie alternative pour l'évaluation des coûts dans un système intégré de production.

2. La recherche de flexibilité: une stratégie structurante

Dans ce paragraphe, nous chercherons à montrer que la nature de la flexibilité des systèmes de production s'est elle-même transformée au cours du temps pour mieux répondre aux exigences de l'environnement. Par ailleurs, cette recherche va structurer l'organisation des entreprises. Avant de développer de manière plus précise cette hypothèse, et dans la mesure où le concept de flexibilité est un concept à plusieurs dimensions dont la méconnaissance provoque de trop fréquentes confusions, on s'efforcera dans un premier temps de définir les différents sens de cette notion, et d'indiquer les outils adéquats pour son analyse.

Si, au sens le plus large, la flexibilité caractérise la possibilité pour une organisation de s'adapter à son environnement, on peut admettre en première analyse qu'il existe deux catégories distinctes d'environnements économiques conduisant à deux formes conceptuellement différentes d'adaptation ou de flexibilité: d'un côté, un environnement économique dont les caractéristiques d'évolution sont connues (au moins par la connaissance des lois de probabilité et des fréquences d'apparition des variables qui caractérisent cet environnement), d'un autre côté, un environnement dont les caractéristiques sont fondamentalement incertaines (c'est-à-dire non appréhendables en termes de probabilité, et ne relevant que d'un processus d'apprentissage). Le premier type d'environnement suppose une adaptation en termes de **flexibilité statique**, le second en termes de **flexibilité dynamique** [cf aussi P. Cohendet, P. Llerena, 1989b].

2.1 Flexibilité statique

La flexibilité statique relève de l'existence, à un instant donné, d'un ensemble plus ou moins vaste d'opportunités. Elle est une réponse à une qualité particulière d'environnement : les situations risquées, c'est-à-dire les situations où la variabilité de l'environnement est parfaitement appréhendée par une distribution de probabilités.

Ainsi, la flexibilité statique d'un système de production se caractérise nécessairement par la présence de surcapacités spécifiques à chaque processus ou résultant de l'entretien d'une offre simultanée de plusieurs produits. Dans les branches, services ou activités industrielles confrontées à la saisonnalité, cette surcapacité permet d'éviter des coûts de stockages excessifs. Dans le même ordre d'idée, certaines entreprises n'hésitent pas à "doubler" leurs lignes de machines existantes par des machines obsolètes, pour répondre à une éventuelle "surchauffe" de la demande. De même, lorsque le bon déroulement de la production est susceptible d'être perturbé par des incidents incertains mais de probabilité identifiable, il est intéressant d'installer des équipements de secours (par exemple, on s'équipera d'un générateur de secours pour parer d'éventuelles coupures de courant).

Ce type de comportement s'apparente au choix d'un portefeuille de produits et/ou de processus. Dans ce cas, le critère d'espérance-variance de la théorie des portefeuilles financiers peut être utilisé. Cette forme de flexibilité demeure en grande partie compatible avec le modèle "taylorien-fordien" puisqu'il s'agit d'une simple juxtaposition de processus, répondant chacun à l'ancienne logique de production. Dans ce cadre, il est clair que le choix de la flexibilité statique génère habituellement des **surcoûts** pour l'entreprise et qu'il pose le classique dilemme flexibilité-productivité, contrairement au cas de la flexibilité dynamique qui est analysée ci-dessous.

L'apparition de ce dilemme montre que, par rapport à l'objectif majeur de l'entreprise qui reste celui de la productivité, la flexibilité ne joue qu'un rôle auxiliaire qui se mesure par des surcoûts, sans que l'organisation fondamentale de l'entreprise ne soit affectée. Les méthodes et les critères d'évaluation peuvent au mieux être aménagés pour tenir compte de ces éventuels "surcoûts". Ils n'ont nul besoin d'être reconçus. La flexibilité statique est de toute façon **programmable** et ceci quelles que soient les formes concrètes qu'elle revêt : capacités excédentaires, main d'oeuvre supplémentaire ou stocks dans le modèle de standardisation ; mise en parallèle de processus de production ou recours à la polyvalence dans le modèle de variété.

Dans de tels schémas concrets, on peut remarquer que, de même que la flexibilité, d'autres objectifs comme celui de la **qualité** ont le statut d'objectifs "annexes" qui viennent se greffer sur l'objectif fondamental de productivité. En effet, la fiabilité d'une machine ou d'un composant est considérée comme une donnée externe au processus. Et, pour améliorer la qualité du fonctionnement d'un atelier ou d'un secteur homogène, il suffit de multiplier les redondances jusqu'à obtenir le niveau de fiabilité souhaité. Il n'existe aucune incitation particulière pour privilégier de manière interne l'amélioration de la qualité des fonds productifs.

2.2 Flexibilité dynamique

La flexibilité dynamique est une **capacité de réagir continûment, dans le temps, aux variations de l'environnement**. Elle constitue une réponse à des variations qui ne sont pas régies par des lois immuables. Au contraire, le décideur doit faire face à une incertitude dont la maîtrise relève d'un processus temporel d'apprentissage de l'information. Il en résulte la nécessité pour l'entrepreneur de transférer, d'une période à l'autre, un "**portefeuille d'actions optimales**", lui permettant à la fois de préserver le maximum de réponses possibles aux modifications perçues de sa propre situation, et d'assurer une réaction dans les meilleurs délais par rapport à la vitesse d'évolution des paramètres de l'environnement. La stratégie d'une entreprise qui consiste à développer et maintenir en activité un processus de production, aujourd'hui non rentable, simplement dans l'espoir qu'il s'agisse d'un processus efficace dans l'avenir, correspond à la recherche d'une telle flexibilité.

Pour une entreprise, cette flexibilité dynamique peut prendre deux formes: soit une minimisation du **décal de réaction** par rapport à une variation de l'environnement (flexibilité de réponse), soit une capacité de maîtriser, de modifier et de "concevoir" l'environnement (flexibilité d'initiative). Le délai de réaction dépend principalement de la maîtrise des temps propres aux différentes opérations de production et à leur enchainement. Il présente au moins deux aspects:

- D'une part, à **court terme**, ce délai de réaction, indicateur du degré de flexibilité dynamique, peut s'estimer par le **temps de production** des produits. Dans le cas d'une ligne de fabrication ne produisant qu'un bien unique, le temps de production correspond à la durée de passage de la pièce entre l'instant de son entrée et celui de sa sortie de la ligne. Dans le cas d'une chaîne de fabrication produisant une gamme de produits donnée, les temps de production incluent les temps morts liés aux changements d'outils.

- Par ailleurs, à **moyen et long termes**, nous devons retenir le **temps d'adaptation** qui traduit une capacité à renouveler les produits et à transformer, en conséquence, les processus de production et leur organisation. Il s'agit, entre autres, du cycle technologique du produit. Ce temps d'adaptation marque également la capacité d'une intégration aisée des évolutions technologiques.

La flexibilité dynamique sous la forme d'une flexibilité d'initiative a pour enjeu "l'élargissement progressif de la gamme des solutions et des problèmes productifs " (J.L. Gaffard, 1990, p.363). Cet enjeu implique une modification profonde de la manière de concevoir les stratégies et leur évaluation. Elles seront alors forcément "forward looking". Le point capital est que l'on peut émettre l'hypothèse que l'une des principales différences qui caractérisent la flexibilité dynamique (par rapport à la flexibilité statique) est qu'elle constitue un objectif dominant autour duquel s'articulent les autres objectifs de l'entreprise. Si cette hypothèse est vérifiée, elle implique naturellement le besoin de reconcevoir en profondeur les critères et les méthodes d'évaluation de la production. Quelques arguments peuvent cependant être avancés, dès à présent, pour étayer cette hypothèse. Ainsi, alors que dans l'optique taylorienne, l'objectif primordial de productivité est visé par une réduction des temps unitaires d'exécution, une augmentation des séries de lancement et un accroissement des stocks (du fait, en particulier, d'une prise en compte peu soutenue des problèmes de fiabilité des équipements), en inversant le raisonnement, si l'impératif devient la réduction des temps de reconversion pour des produits variés (afin de satisfaire l'objectif de flexibilité dynamique) un véritable nouveau "cercle vertueux" peut se mettre en oeuvre avec une diminution des séries économiques de lancement et donc des stocks. On obtient alors une plus grande fluidité et une plus grande souplesse du processus (sans renoncer toutefois à satisfaire, mais de manière annexe cette fois-ci l'objectif de productivité). De même, on peut montrer que l'objectif de flexibilité dynamique entraîne une vision très différente de l'objectif de qualité. Dans une stratégie de recherche de flexibilité dynamique, la qualité, la fiabilité et la capabilité des équipements deviennent des préoccupations

centrales, incontournables. En effet, l'un des moyens pour atteindre l'objectif de flexibilité consiste en une réduction des temps d'attente des produits, etc... c'est à dire en une suppression des redondances "inutiles" et des aléas perturbants. Par conséquent, pour que le système demeure viable, et surtout pour qu'il puisse fournir au marché le bon produit au bon moment et sans défaut, c'est à dire pour qu'il corresponde aux critères du nouvel environnement, il faut agir directement sur les différentes composantes de la qualité: par un contrôle "on line" à chaque étape de la production, par chaque intervenant dans le processus, et par la mise en place de structures organisationnelles, de dispositifs de communication et de concertation adéquats. Pour reprendre le cas cité ci-dessus, il ne s'agit plus de rendre le système plus fiable étant donné le degré de défaillance des composants, mais de le rendre plus fiable en améliorant systématiquement et sans relâche la fiabilité des composants. L'objectif de qualité est ainsi "endogénéisé" (cf. P. Cohendet, P. Llerena, 1987b).

La recherche de flexibilité dynamique apparait ainsi comme un objectif fondamental autour duquel se réordonnent les autres objectifs de l'entreprise. Derrière ce phénomène, on peut considérer que l'on assiste à une rupture de l'enchaînement extension du marché - division du travail, qui avait fondé le développement industriel sur une longue période. La nécessité nouvelle de décloisonner les fonctions, de privilégier une vision globale des ateliers plutôt qu'une spécialisation des tâches individuelles, la conception de nouveaux outils de production davantage orientés vers la réalisation d'opérations rapides, fiables et variées que vers la seule recherche de gain de productivité, sont autant de manifestations qui plaident pour centrer la réflexion autour de la flexibilité dynamique (2). C'est, en particulier, autour de cet objectif qu'il convient d'entreprendre un réexamen des méthodes d'organisation et des critères d'évaluation. Or, cette recherche de flexibilité dynamique s'insère dans un environnement caractérisé par une information perturbée. Il nous sera donc impossible de faire l'économie d'une réflexion sur la nature de la rationalité des agents. De plus, la recherche de flexibilité va être la source d'une reconception de l'organisation des entreprises, notamment par ce que nous appellerons des "processus d'intégration".

3. Une interprétation de l'évolution des systèmes de production et du rôle de l'intégration

L'évolution des systèmes de production du modèle de standardisation vers le modèle de variété, puis vers le modèle de réactivité, montre une capacité croissante pour les firmes d'opérer à des niveaux de complexité de plus en plus élevés. Néanmoins, la logique de cette évolution, et les forces qui sont à la base de cette dynamique n'apparaissent pas toujours de manière évidente. L'hypothèse de travail, qui est soutenue ici est, d'une part, que c'est la recherche d'une flexibilité sans cesse accrue qui est à la base de la dynamique de l'évolution des systèmes de production mais que, d'autre part, cette évolution n'a été rendue

possible que grâce à la mise en oeuvre de processus d'intégration qui ont assuré la viabilité économique des nouvelles formes d'organisation qui étaient expérimentées. Ayant défini ci-dessus les deux principales formes de flexibilité, flexibilité statique et flexibilité dynamique, il est possible d'en étudier directement les conséquences pour l'organisation de l'entreprise.

La conséquence majeure de la recherche de flexibilité dynamique est la nécessité d'introduire dans l'organisation de l'entreprise **des procédures d'intégration pour rendre économiquement viable le nouveau système**. En effet, la recherche de flexibilité dynamique induit une intégration du système de production de manière à réduire le délai de réaction ou à maîtriser les évolutions de l'environnement. L'intégration, ainsi conçue, apparaît comme une condition permissive de fonctionnement du système de réactivité. Néanmoins, dans la mesure où l'intégration apparaît (comme la flexibilité) être un concept multidimensionnel, un examen de la définition précise de ce concept est maintenant nécessaire. Cet examen sera mené à la lumière des trois modèles de production exposés précédemment.

3.1 Le rôle de l'intégration comme facteur-clé de fonctionnement des systèmes de production

L'hypothèse de travail que l'on soutiendra ici est que l'intégration correspond à un ensemble de procédures, variables selon les impératifs de la production, permettant de rendre économiquement viables des organisations cherchant à satisfaire les objectifs qu'elles se sont fixés. Il est ainsi remarquable de noter que, selon le modèle que l'on considère, l'intégration correspond à des procédures différentes dépendant de l'objectif économique que l'on cherche à atteindre. Ainsi :

- Dans le cadre du **modèle de standardisation**, où l'objectif est d'atteindre la productivité maximale à travers la recherche d'économie d'échelle, la procédure qui a permis de rendre viable le fonctionnement du système, est une procédure d'"**intégration**" consistant à rassembler, au sein de sections spécialisées de la production, des machines conventionnelles agencées en travées (on considèrera ainsi la section des tours, des fraiseuses, des aloseuses, etc ...). Cette forme d'intégration "spatiale" revient à rassembler, dans un lieu donné, des moyens de production auparavant éclatés dans l'espace.

- Dans le cadre du **modèle de variété**, les procédures d'intégration s'articulent autour de la stratégie de différenciation retardée. On "intègre" le plus possible, à l'amont par exemple, en standardisant les composants pour obtenir le maximum de variété à l'aval. Des méthodes telles que la TGAO correspondent aux objectifs d'intégration du modèle de variété. L'intégration consiste ainsi dans ce modèle, essentiellement en une simplification "physique" de la nature des produits fabriqués, pour faire apparaître entre les différentes gammes et différents modèles le plus d'éléments communs possibles pour rendre viable le régime de variété.

- Dans le **modèle de réactivité**, les procédures d'intégration visent à assurer une meilleure coordination entre les différents éléments de la production. Des moyens tels que les réseaux locaux ou des méthodes telles que le juste-à-temps illustrent bien ce type de méthodes

d'intégration requises pour permettre d'obtenir le plus court délai de réaction possible vis-à-vis de l'environnement. L'intégration dans le modèle de réactivité est essentiellement celle de flux "informationnels" (dans le cas des réseaux locaux) ou de flux physiques de production (dans le cas de juste-à-temps).

Ce regard "historique" sur l'intégration permet de mieux situer le concept et de ne pas conférer trop vite à celui-ci le sens exclusif (physiologique) du dictionnaire "coordination des activités de plusieurs organes, en vue d'un fonctionnement harmonieux, réalisée par divers centres nerveux" (Larousse, 1981). Même si cette définition capte une partie importante du concept tel qu'on l'utilise en production, elle ne rend pas compte de différents aspects de l'intégration. Assimiler, en effet, l'intégration à la volonté de rendre cohérent quelque chose qui est dans un tout ne permettrait pas, par exemple, de distinguer entre le modèle fordien et le modèle de réactivité, puisque le modèle fordien était parfaitement cohérent (par rapport aux objectifs qu'il s'était fixés et, compte tenu de l'environnement économique de l'époque).

Pour préciser la notion d'intégration, il faut prendre aussi en considération les propriétés suivantes :

- La forme concrète que revêt la procédure d'intégration est fondamentalement dépendante des objectifs poursuivis par la firme (intégration davantage spatiale dans le modèle de standardisation, intégration davantage physique dans le modèle de variété, intégration informationnelle dans le modèle de réactivité). Toutefois, ces procédures ne sont pas exclusivement "dédiées" à un modèle particulier. Par exemple, dans le modèle de réactivité, à côté des méthodes d'intégration des flux physiques et informationnels, on peut tout-à-fait concevoir des mises en oeuvre de stratégie d'intégration spatiale (pour minimiser l'espace occupé par le système productif) ou d'intégration physique (pour réduire la complexité des pièces).

- Les procédures d'intégration correspondent toujours à la mise en place d'une certaine forme de rigidité dans l'organisation de l'entreprise. Cette rigidité assure les principes structurants autour desquels s'organise l'entreprise, en fonction des objectifs prioritaires de celle-ci. Mais l'introduction d'une forme de rigidité induit aussi naturellement une certaine perte de flexibilité. De manière plus précise, on peut vérifier qu'une procédure d'intégration correspond toujours à l'abandon d'une certaine forme de flexibilité, au profit d'une autre forme de flexibilité. **L'intégration est donc un principe d'allocation.** Par exemple, si l'on considère la procédure d'intégration par différenciation retardée, on abandonne, à l'amont, une certaine forme de flexibilité dans la conception des composants pour privilégier une flexibilité plus grande à l'aval ; de même, si l'on considère la procédure d'intégration par le recours au juste-à-temps, on améliore la réactivité (flexibilité dynamique) d'un poste en l'"intégrant" aux postes amont, et donc en lui faisant perdre un certain nombre de degrés de liberté, etc ... Naturellement, dans l'allocation entre deux formes de flexibilité, que réalise la procédure d'intégration, la forme de flexibilité qui est privilégiée correspond à l'objectif prioritaire de l'entreprise. Cette situation d'arbitrage provient de ce que la flexibilité n'est pas "gratuite", et

qu'une entreprise, à un moment donné, ne peut pas disposer de toutes les formes de flexibilité (statique : capacités excédentaires, stocks, polyvalence ; ou dynamique : réactivité, capacité de disposer de plusieurs options, etc ...), mais qu'il lui faut choisir en quelque sorte, aussi bien au niveau de chaque élément de production qu'au niveau global, la "composition" optimale de flexibilité en fonction des objectifs et des considérations de coûts. Et ce sont les procédures d'intégration qui "opèrent" le choix de l'allocation, tout en structurant l'organisation de l'entreprise, selon le modèle désigné.

- Les procédures d'intégration peuvent aussi être considérées comme des moyens de réduire la complexité de l'entreprise. Cette propriété est une autre manière de considérer la propriété précédente, car il existe une relation directe entre flexibilité et complexité. Introduire dans un système une flexibilité supplémentaire revient, en effet, à augmenter les interactions entre les composants d'un système et entre ses composants et l'environnement. En d'autres termes, il en résulte une plus grande complexité. L'augmentation de la complexité peut résulter aussi bien d'un nombre d'éléments plus élevés dans le système (par exemple, une juxtaposition de processus pour la versatilité) que d'une multiplication des interactions entre les éléments et/ou de l'existence d'interactions qui ne sont plus immuables mais fonctions des circonstances (flexibilité dynamique).

Nous plaçons ce concept de complexité au centre du mécanisme d'évolution du processus de production, car il fonctionne comme un révélateur du changement de la logique de production. Dans le modèle de standardisation, par exemple, le raisonnement se fait à complexité constante. Dans ce cadre, les investissements consistent en de simples substitutions, ne bouleversant pas les principes d'organisation. En revanche, le modèle de réactivité implique une étroite dépendance entre l'organisation de l'entreprise et l'évolution de l'environnement. Le niveau de complexité devient alors une variable endogène qu'il faut maîtriser. Dans le cas d'un accroissement de la flexibilité du processus, nous avons une connexion plus étroite de celui-ci avec son environnement. Pour l'assurer, cela nécessite un volume accru d'informations à traiter et, par conséquent, une complexité plus importante du système. L'exemple d'une flexibilité acquise par une mise en parallèle de processus est, à cet égard, symptomatique. Le nombre de composants ainsi que des interconnexions augmente, et donc, la complexité du système aussi. Il découle de cet accroissement de la complexité, au moins deux conséquences : la première prend la forme d'un théorème intuitif (un "folk-theorem"), la seconde celle d'un constat économique.

Le "folk-theorem" établit une relation entre l'accroissement de la complexité et celui de la stabilité d'un système. "L'argument intuitif ... est qu'une plus grande complexité génère un réseau plus dense d'interconnexions entre les composants du système et, de ce réseau plus dense, résulte un système capable d'absorber des perturbations déstabilisantes "(J. Casti, 1986).

Le constat économique est plus simple à établir - qui dit complexité accrue, dit aussi une gestion du système plus difficile, un volume d'informations plus important, et donc une

augmentation du coût de gestion du système. Ces coûts croissants, avec le degré de complexité, constituent une caractéristique déterminante pour la recherche de la flexibilité. Les entreprises vont alors chercher à réduire ces coûts pour une **maîtrise de la complexité**. Pour cela, elles ont recours à un certain nombre de principes opératoires qui prennent principalement la forme de **processus d'intégrations de natures différentes**.

Il est possible, à ce stade, de distinguer plus précisément dans la nature de la complexité, et donc, d'effectuer une distinction dans la nature des procédures d'intégration associées. En effet, la complexité est un concept qui recouvre :

- d'une part, l'idée qu'un objet (ou un système) est complexe parce qu'il est composé de nombreux éléments. On peut parler ici de "complication". Dans ce cas, en l'absence de tout caractère incertain, les procédures d'intégration, du moment qu'elles sont justifiées économiquement, seront toujours souhaitables, quelque soit le modèle de référence envisagé. Ainsi, par exemple, l'intégration (technologique) de pièces, où l'on réduit le nombre de pièces composant un objet donné afin de réduire le coût de l'assemblage. Ce phénomène est particulièrement fréquent dans le cas des substitutions de matériaux (cf. P. Cohendet, P. Llerena, 1987a).

- d'autre part, l'idée que la complexité est née de la nécessité d'introduire, volontairement ou involontairement, des redondances dans un système (mise en parallèle d'équipements, par exemple), pour assurer (par une flexibilité statique) l'adaptation (passive) à l'environnement risqué. Le type d'intégration qui est lié à cette forme de complexité (relative à une incertitude dont on connaît les lois de probabilités) s'applique principalement (mais pas exclusivement) au modèle de variété ; la **procédure d'intégration des phases productives** répond bien à ce contexte. Elle consiste à réduire la complexité, soit en faisant disparaître certaines phases de production, soit par l'exécution en une seule et même opération de ce que jusqu'alors exigeaient deux ou plusieurs opérations consécutives. Le mécanisme et l'intérêt de cette procédure d'intégration peuvent s'analyser à partir d'un modèle cybernétique très simplifié (Cf. P. Cohendet, P. Llerena, 1987a)

- enfin, il existe une autre acception de la complexité, qui s'applique à des systèmes présentant outre les divers éléments et les différents niveaux hiérarchiques qui les composent, des flux dynamiques d'information (non régulière). La réduction de la complexité (à l'aide de procédures d'intégration) passe par la régulation de ces flux d'information. C'est dans cette optique que se situent, par exemple, les procédures d'intégration par la mise en oeuvre de réseaux locaux. C'est naturellement dans le cadre de ce modèle de réactivité que le recours à la réduction de complexité par intégration informationnelle apparaît la plus appropriée.

Les réflexions précédentes ont montré la richesse des procédures d'intégration qui s'offrent aux entreprises, pour atteindre les objectifs qu'elles se sont fixés. A titre d'illustration, le tableau suivant résume les principaux cas concrets d'application que l'on rencontre en pratique.

Tableau 1: Exemples de procédures d'intégration.

Propriétés			
Cas d'intégration	Nature de l'intégration	Objectif recherché	Modèle de référence le plus approprié
Constitution de sections homogènes.	Intégration spatiale.	Economies d'échelle.	Modèle de standardisation.
Intégration de phases.	Intégration physique (technologique).	Economies sur les facteurs de production.	Tous les modèles.
Intégration de pièces.	Intégration physique (technologique).	Economies sur les facteurs de production.	Tous les modèles.
Différenciation retardée.	Intégration physique (organisationnelle).	Accroissement de la variété aval.	Modèle de variété.
TGAO.	Intégration physique (technologique).	Accroissement de la variété aval.	Modèle de variété.
Intégration de fonction.	Intégration organisationnelle (exemple : CFAO, découloisonnement).	Meilleure réactivité.	Modèle de réactivité.
Juste-à-temps .	Intégration organisationnelle.	Meilleure réactivité.	Modèle de réactivité.
Réseaux locaux.	Intégration informationnelle et technologique.	Meilleure réactivité.	Modèle de réactivité.
Réduction de l'espace productif.	Intégration spatiale.	Economies sur les facteurs de production.	Tous les modèles.

3.2 Le choix stratégique de l'intégration dans le modèle de réactivité

Ayant défini et situé la notion d'intégration dans une perspective historique, il est maintenant possible d'interpréter les changements en cours que nous observons dans le modèle dit de "réactivité". Ces changements en cours incitent à un changement radical de perspective dans la conception de l'organisation de la firme. Pour le montrer et pour situer dans ce contexte la nature précise et l'intérêt d'une politique d'intégration, on reviendra, au préalable, sur la caractéristique majeure qui définit le nouveau modèle de production : la nécessité d'opérer dans un régime d'informations perturbées, impliquant en particulier des prises de décisions pour l'entreprise qui sont non programmables.

3.2.1 Les modifications de l'organisation en régime d'information perturbée

La prise en compte d'un temps porteur d'informations nouvelles et perturbée a pour conséquence qu'un décideur donné ne peut plus appréhender son environnement par le biais d'une appréciation objective des phénomènes, mais bien au contraire par une appréciation subjective de ceux-ci. Il y a donc glissement, pour le décideur, d'une rationalité parfaite vers une rationalité contextuelle, au sens donné à cette expression dans le présent ouvrage. Le fondement logique des décisions est donc transformé en postulant explicitement une limitation du champ d'investigation du décideur: la complexité de l'environnement et la présence d'une incertitude irréductible (issue de la nature de la concurrence) contraignent le domaine des possibles envisageables par lui à être plus réduit. Ces limites portent aussi bien sur l'étendue des potentialités prises en compte que sur la détermination des gains qui peuvent en résulter, c'est-à-dire de leur mode d'évaluation.

Il est également apparu que les différences d'horizon (temporel et hiérarchique) deviennent primordiales. Car, dès que l'on accepte le principe d'une rationalité contextuelle, les horizons temporel et hiérarchique des décideurs et des organisations deviennent déterminants. En effet, le champ d'investigation sur lequel va s'appliquer le processus de recherche des solutions va dépendre de ces horizons. Or, l'un des moyens pour pallier, au moins partiellement, les inconvénients de cette limitation du champ d'investigation des micro-décideurs est d'enrichir les tâches. En d'autres termes, on cherchera à réduire le nombre de niveaux hiérarchiques dans l'entreprise, et à intégrer les fonctions. Grossièrement, on assiste à un affaiblissement de la hiérarchisation des entreprises. Mais ce constat, bien qu'indiquant une voie de recherche possible, ne nous dit rien sur l'une des questions qui devient alors centrale : **quel est le découpage le plus approprié, le plus pertinent dans l'entreprise, pour améliorer la recherche de solutions, c'est-à-dire la prise de décision ?** Et, de plus, même si ce découpage existe, il ne résoudrait en rien l'incontournable absence de cohérence entre les différents acteurs du processus de décision. Or, sur ce point précis, on peut émettre l'hypothèse que bien qu'il soit utile de réduire les effets de cette absence (par un élargissement des champs

d'investigation), la cohérence interne, prise dans le sens strict de rationalité parfaite, n'est en rien nécessaire. En effet, les divergences et les conflits vont être les moteurs d'une négociation entre les différents acteurs, vont être le signe d'une richesse d'analyse bénéfique pour la conduite d'un processus d'apprentissage court et efficace. Il y a dans cette recherche de "consensus" ou de "compromis" un aspect pédagogique non négligeable pour comprendre le bon déroulement d'un processus de décision en rationalité limitée. Ces réflexions rejoignent celles exprimées par O. Favereau [1989] pour qui "une organisation hiérarchique n'est pas une forme stable dans les systèmes sociaux empiriques". En effet, une organisation hiérarchique possède une structure décomposable, c'est-à-dire qu'elle peut être subdivisée en sous-partie, structurée par un lien de causalité. Si les interrelations entre les membres de l'organisation sont représentées par une matrice, nous avons une matrice triangulaire du type :

$$\begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix}$$

Une telle organisation a une tendance "naturelle" à évoluer vers une structure parfaitement décomposable, où les sous-parties s'autonomisent, deviennent totalement indépendantes de l'environnement et de l'organisation globale ($a_2 = 0$, la matrice devient diagonale). Cette structure perd alors les avantages de la structure décomposable puisqu'il n'y a pas de coordination souple entre les sous-systèmes. L'une des raisons de cette autonomisation est la multiplication des règles limitatives, au-delà des règles strictement nécessaires ("constitutives") pour l'existence de l'organisation. Or, cet excès de règles limitatives va avoir pour principal effet une indépendance croissante du sous-système vis-à-vis du système dans son ensemble, et du système par rapport à son environnement. Le cas du développement d'un marché interne du travail est symptomatique de cette évolution. Celle-ci est renforcée par l'existence d'une "tendance de tout ensemble humain à s'autonomiser par rapport au monde extérieur, en s'organisant de plus en plus autour de considérations et de motivations internes" (O.Favereau, 1989).

La question qui est alors posée est celle de l'évolution du système vers une nouvelle structure décomposable, voire indécomposable, où il existe "à nouveau" des liens organiques entre les sous-systèmes assurant, en partie, la coordination de l'ensemble et permettant de dépasser l'inertie inhérente aux systèmes parfaitement décomposables. L'apparition des "cercles de qualité", l'accent mis sur la nécessité d'intégrer les fonctions dans l'entreprise, etc ..., participe de cette évolution vers un système d'entreprise "moins décomposable".

3.2.2. L'importance du flux informationnel

Puisque la rupture fondamentale entre les modèles de production précédents et le nouveau modèle de réactivité est issue du changement dans la nature de l'information (de l'information "stable" à l'information "perturbée"), on peut logiquement en déduire que la

nouvelle organisation qui est à concevoir doit refléter fortement ce changement dans la nature de l'information. De manière plus précise, l'organisation dans les modèles de production à information stable était structurée autour du flux matière, selon un principe séquentiel, et un système de décision correspondant à une hiérarchie immuable, puisque chaque décision pouvait être "programmée" à l'avance. Il n'y a pas de place dans de tels modèles pour des initiatives locales, inutiles et perturbatrices, pas d'intérêt non plus pour que se développent des processus d'apprentissage locaux, puisque tout peut être programmé à l'échelle globale, puis décentralisé.

Dans le cas d'un régime d'information perturbée, il n'est plus possible de programmer les décisions de manière globale, puisqu'il n'y a plus de support informationnel suffisamment répétitif pour fonder ces décisions. L'intensité et la nature du flux matière deviennent fortement incertaines. Il est donc nécessaire de structurer l'entreprise de manière à se comporter de façon la plus efficace possible face à un flux d'information perturbée. Or, dans ce contexte, **l'hypothèse que nous soutenons est que c'est la mise en oeuvre de processus locaux d'apprentissage qui constitue la réponse organisationnelle la plus appropriée pour rendre viable une organisation.** A partir de cette hypothèse de travail, trois questions principales se posent : comment se définissent ces "processus locaux d'apprentissage" ? comment "intégrer" l'ensemble de ces processus dans un ensemble cohérent ? comment passer en pratique d'un ancien modèle de production au nouveau, si l'on doit modifier radicalement l'organisation ?

Les processus locaux d'apprentissage constituent les éléments les plus invariants de l'entreprise dans un régime d'information perturbée. Ils se définissent non seulement à partir de l'information perturbée provenant de l'environnement de la firme mais aussi de l'horizon temporel et décisionnel de chaque groupe de décideurs. Ainsi, il existe une double détermination des limites du support de l'apprentissage: l'horizon et la nature de l'information nécessaires à la prise de décision. Ces processus d'apprentissage reposent sur la notion d'"équipe" locale capable, à partir de la réception d'informations extérieures, de satisfaire un service ou de résoudre un problème en disposant des caractéristiques suivantes :

- Capacité d'identifier le problème.
- Capacité d'assimiler l'information extérieure, de comparer cette information avec les situations précédemment expérimentées pour élaborer une décision.
- Capacité de transformer le signal reçu en une production (matérielle ou immatérielle).
- Capacité d'identifier et de communiquer avec les autres éléments de l'entreprise impliqués dans la production considérée.
- Capacité de préparer la décision de production.
- Capacité de contrôler la production et d'assurer la maintenance des équipements.
- Capacité de faire un bilan de la production réalisée et d'enregistrer l'information.

De telles caractéristiques impliquent que les processus d'apprentissage en question ne se résument pas à une "routine" a priori automatisable, puisque des conditions sur l'accumulation d'expérience sont requises; ils ne se résument pas non plus à la définition classique d'un atelier dont le découpage est imposé par la hiérarchie et qui est dépourvu généralement des capacités de dialogue avec les autres éléments de production. L'autre différence avec l'atelier classique est que le fonctionnement de telles "équipes" ne suppose pas une stricte division des tâches à l'intérieur de l'équipe, ni d'une hiérarchie stricte à l'intérieur de l'équipe. Il peut y avoir, par exemple, des "postes tournants" pour aboutir à une forme de savoir collectif, y compris pour la fonction de contrôle de l'équipe. Toutefois, il est clair que la définition de telles équipes ne peut pas complètement s'éloigner des formes traditionnelles d'organisations locales. Ainsi, si l'on admet le découpage en plusieurs niveaux de régulation du système d'information de l'entreprise, il est probable que chaque équipe que l'on considère se situera le plus généralement en majorité sur un seul niveau, car il est difficile d'imaginer un apprentissage local jouant sur l'ensemble des niveaux. En revanche, ce qui caractérise ici le processus local d'apprentissage, c'est qu'il doit incorporer des connaissances sur l'ensemble de l'entreprise, et donc posséder une connection avec les niveaux de régulation supérieurs et inférieurs.

Nous retrouvons ainsi en partie les résultats de M. Aoki [1986] dans le cas particulier de la gestion d'un processus de production soumis à un aléas probabilisable. L'analyse de M. Aoki, qui procède à une comparaison de structures stylisées d'informations d'une entreprise japonaise dite "J-firm" et d'une entreprise (américaine) dite "A-firm" peut se résumer de la manière suivante:

Tableau 2 : Comparaison des structures informationnelles.

Structure d'information horizontale

- Coordinations horizontales d'ateliers semi-autonomes.
- Connaissances incomplètes du processus global de production.
- Réponse rapide aux aléas et apprentissages de la solution optimale.

Structure d'information verticale

- Connaissance complète de l'ensemble des technologies.
- Imperfections dans la perception des aléas.
- Retard dans l'application des solutions.

Paramètres de la comparaison

- Retard d'adaptation aux événements.
- Degré d'imprécision de la perception des aléas par la direction.
- Capacité et vitesse d'apprentissage des ateliers
- Degré d'imperfection de la coordination horizontale entre ateliers.

Le tableau 2 explicite ces deux structures. La première, se caractérise par une coordination horizontale. Cette coordination horizontale permet de procéder à un ajustement fin aux modifications des exigences du marché, de l'aval vers l'amont, dans le cadre de plans de production plus généraux conçus par le niveau hiérarchique supérieur. Une telle structure d'information se fonde sur des équipes de travail semi-autonomes qui ont la capacité et le know-how pour réagir aux aléas du marché et/ou du processus de production. Pour y parvenir, une certaine connaissance du processus de production est nécessaire et elle est acquise par un "learning by doing", rendu possible grâce à une rotation dans les tâches à exécuter par les différents opérateurs. La principale qualité de cette structure est de permettre une réponse rapide aux aléas et une grande souplesse dans l'organisation des productions. L'archétype d'une telle structure est le système Kanban, où l'information circule et la coordination se fait de l'aval vers l'amont. Cette coordination horizontale s'oppose à une organisation verticale, au contrôle hiérarchique important, héritée de la tradition "taylorienne". Pour cette structure informationnelle, il existe une connaissance parfaite du processus de production et une conception centralisée des décisions, et donc des réactions aux aléas. L'efficacité de ces réactions dépendra de la qualité de la perception des aléas et du délai d'application des solutions.

Pour procéder à une comparaison des structures informationnelles, nous disposons notamment de quatre paramètres :

- délai d'adaptation aux événements ;
- degré de précision de la perception des aléas par la direction ;
- capacité et vitesse d'apprentissage des ateliers ;
- efficacité de la coordination horizontale entre ateliers.

S'il existe d'importants délais de réaction, une imprécision dans la perception des aléas, un savoir-faire important des ateliers et une bonne coordination horizontale, alors une structure horizontale est plus appropriée. Or, un accroissement de l'instabilité implique aussi bien des délais de réaction allongés, une plus grande imprécision et une coordination horizontale plus difficile. Inversement, si l'on tente une évaluation grossière des effets d'une intégration telle que les réseaux locaux industriels, il est probable que les délais de réaction soient plus courts, la précision plus importante et la coordination horizontale améliorée.

En d'autres termes, la comparaison des structures d'informations ne permet pas de mettre en évidence la supériorité intrinsèque d'une organisation sur l'autre. En revanche, l'organisation "à la japonaise", d'après M. Aoki, semble plus adaptée aux situations où l'environnement est instable car une coordination horizontale (ou intégrée) permet de faire l'économie d'une cohérence et d'une efficacité du contrôle hiérarchique accrue, rendue nécessaire par l'environnement. Dans le cas de l'analyse menée par M. Aoki, il est intéressant de noter l'importance de l'apprentissage. Pour que la structure horizontale soit supérieure, il faut notamment que non seulement les connaissances de départ soient suffisamment élevées mais aussi que l'assimilation des informations nouvelles soit rapide. Ce résultat confirme notre

hypothèse que, face à une information incertaine, mais surtout face à une information perturbée, le rôle des processus d'apprentissage devient primordiale. Par conséquent, s'il doit y avoir des formes d'intégration permettant de maîtriser ce type de complexité, elles doivent se fonder sur ces mêmes processus. La question du mode d'organisation ne doit cependant pas être détachée de celle de l'évaluation et notamment de l'évaluation des coûts de production. Dans une dernière partie, nous nous attacherons à définir une piste de recherche pour un mode original d'estimation de ces coûts.

4. Les coûts de transaction comme mode d'évaluation de l'intégration (3)

Face à une information perturbée, nous avons déjà remarqué l'importance croissante des interactions entre les composantes de l'entreprise. Les processus d'intégration constituent des principes opératoires pour maîtriser la complexité des systèmes à la recherche d'une flexibilité accrue. Nous allons maintenant proposer, à l'aide de la théorie des coûts de transaction, des fondements économiques à la décision d'intégrer. Les hypothèses développées le seront avec prudence dans la mesure où la théorie des coûts de transaction a été développée dans le cadre des relations entre la firme et le marché, et non pour déterminer le fonctionnement interne des organisations. Notre idée se résume en une application de concepts d'économie industrielle à la gestion industrielle

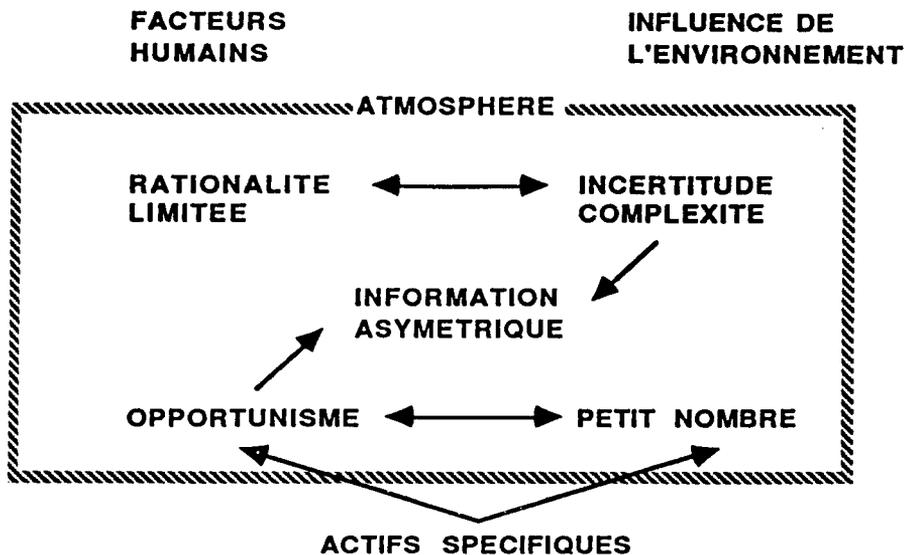
4.1 Les principes de l'économie des coûts de transaction

A la suite des travaux de R. Coase [1937] et d'O. Williamson [1981], le "paradigme" transactionnel s'est développé, à partir de l'idée selon laquelle la production de biens et de services résulte de l'enchaînement d'activités élémentaires, en recherchant "les raisons pour lesquelles la partition de cet ensemble d'activités se modifie et les motifs qui conduisent la firme à intégrer (ou à abandonner) une activité particulière" (P. Joffre - G. Koenig, 1985). La notion de coût de transaction a précisément pour objet d'expliquer le fait que, dans les organisations, des dispositifs de type hiérarchique se sont maintenus en remplaçant le système de prix comme instrument d'allocation des ressources. C'est ainsi, par exemple, que peut se justifier le phénomène "d'intégration verticale", bien connu des économistes. Des coûts de transaction trop importants entre deux éléments distincts d'un ensemble économique peuvent ainsi justifier le recours à l'intégration plutôt que laisser fonctionner les ensembles de manière décentralisée. Une transaction se définit alors comme une interaction entre deux composantes du système économique, entre deux activités élémentaires. En d'autres termes, l'économie des coûts de transaction constitue une théorie (parmi d'autres) des limites d'une firme ou de manière équivalente du degré de décentralisation d'un système industriel. On peut dès à présent remarquer la similitude de problématique avec la définition de l'organisation interne de l'entreprise.

Pour O. Williamson [1981], l'économie des coûts de transaction est favorisée dans le contexte suivant :

- **Un régime d'incertitude et de complexité.** Or, en régime d'information perturbée, nous avons un accroissement aussi bien de l'incertitude interne de l'entreprise (qualité, délais entre les différentes étapes du processus de production par exemple) et externe (exigences du client,...) que de la complexité de l'environnement et de l'organisation de la firme.
- **Un comportement de rationalité limitée** de la part des agents économiques. Nous retrouvons une des caractéristiques d'une situation en régime d'information perturbée.
- **Des risques d'opportunisme**, c'est-à-dire des risques liés à la tendance qu'ont les agents économiques à agir dans leur propre intérêt, au détriment de leurs partenaires. Ce type de comportement est d'autant plus fréquent que l'incertitude s'accompagne d'une asymétrie de l'information et que le nombre d'agents concernés et de partenaires possibles de la transaction est limité. Ce nombre dépend en particulier d'une quatrième caractéristique du contexte.
- **Des actifs spécifiques** (un actif est spécifique lorsque sa productivité dépend d'une articulation harmonieuse de son usage avec celui d'un autre facteur de production déterminé ou lorsqu'il s'avère que, pour certains agents, sa valeur d'usage est supérieure à sa valeur d'échange). On peut ainsi proposer plusieurs mesures de la spécificité d'un actif. Selon E. Anderson [1982], il s'agit du volume requis de connaissances spécialisées ; selon K. Monteverde et D. Teece [1982], il s'agit des efforts consentis pour développer un composant; selon Walker et Weber [1984], il s'agit du nombre de fournisseurs potentiels. A l'intérieur d'une entreprise, nous avons bien une situation similaire: à chaque fonction, à chaque étape du processus correspond une spécificité qui est non substituable (ou difficilement), d'où des risques d'opportunisme et surtout d'asymétrie d'information importants. En fait un actif spécifique est celui qui présente peu de substituts. La figure 2 représente l'interaction entre ces différents facteurs des coûts de transaction.

Figure 2: facteurs de coûts de transaction (d'après Charreaux et alii, 1987)



Les caractéristiques de l'économie des coûts de transaction qui viennent d'être rappelées se retrouvent dans leur quasi-intégralité lorsqu'on observe les mutations qui s'imposent à la firme dans le contexte nouveau d'informations perturbées. Il est donc légitime de tenter d'appliquer ces principes à l'examen des relations entre composantes de production dont la tendance est d'être de plus en plus décentralisée avec les risques d'incohérence que l'on a évoqués dans l'introduction.

4.2 Coûts de transaction et organisation interne de la firme

L'application des coûts de transaction à la firme a fait l'objet de tentatives pionnières par R.S. Kaplan et, plus récemment, par E. Vollman, W. Berry et D. Whybark [1988]. Les coûts de transaction seront très souvent des coûts de structure. En effet, selon ces auteurs, les coûts (réels) de transaction sont à l'origine d'un accroissement continu des coûts de structure (l'un des "générateurs de coûts" les plus importants dans les recherches sur la refonte de la comptabilité industrielle apparaît ainsi bien être la notion de transaction: cf les travaux du CAM-I, et P. Lorino, 1989). Les principaux coûts de transaction sont alors les suivants:

i. **Les coûts de transaction logistiques** qui incluent entre deux composantes de production, les coûts de commandes, d'exécution et de confirmation du flux de matière, allant d'un point à un autre. (En revanche, la livraison elle-même, qui correspond au transfert physique de la commande et qui implique des activités de stockage et de transport, doit être rattachée aux activités de production, et donc être bien distinguées du coût de transaction). A un niveau plus général, entre l'entreprise et ses partenaires, le coût de transaction logistique intègre les ressources utilisées à la conception et au suivi du "contrat" portant transfert des "droits de propriété". "Ces ressources comprennent notamment le travail nécessaire à la recherche d'un compromis, les efforts de standardisation et de certification des liens échangés, les honoraires des conseils juridiques chargés de s'assurer du caractère légal des dispositions envisagées et les impôts attachés à certains types de transactions. A la différence des coûts de livraison, les coûts de transaction sont intimement liés à la méthode retenue pour l'allocation des ressources" (P. Joffre, G. Koenig [1985], p.135).

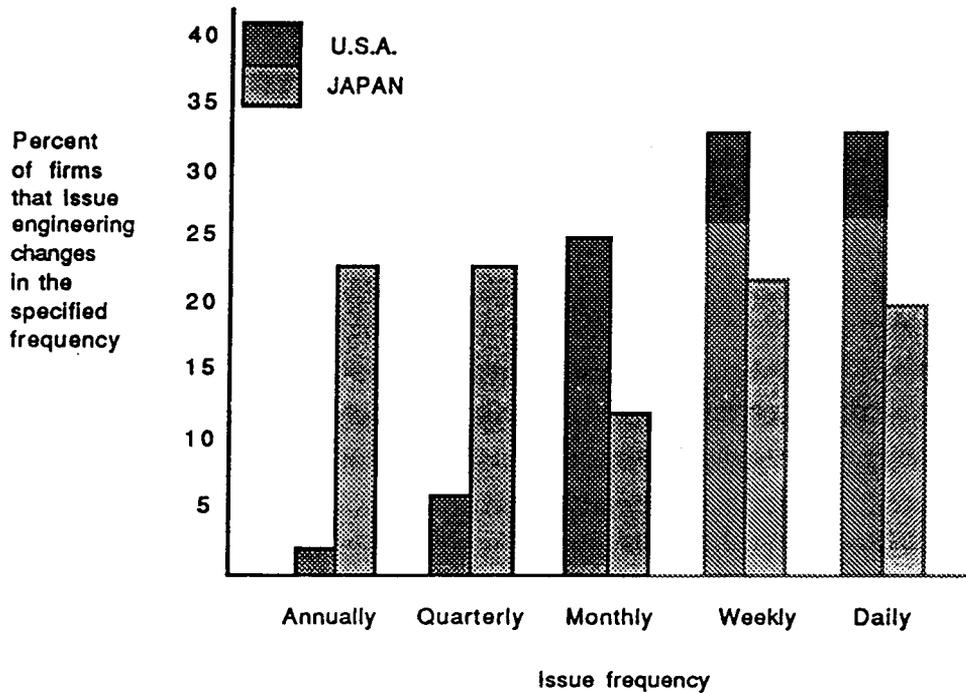
ii. **Les coûts "d'équilibrage"** qui sont relatifs à l'adéquation des charges par rapport aux ressources de la production. Ils sont largement corrélés aux coûts de logistique. Ces coûts incluent le contrôle de la production, l'activité d'achat, les coûts d'ordonnancement, de prévision, etc ... selon E. Vollman et alii. (1988), ces coûts peuvent atteindre jusqu'à 10 à 20 % du total des coûts de structures.

iii. **Les coûts liés aux activités de gestion de la qualité.** Ces coûts ne se limitent pas au simple contrôle de la qualité. Ils incorporent les coûts de transaction liées à l'identification et à la communication de spécifications, à la certification que les autres transactions se déroulent normalement, à l'enregistrement des données pour les bilans de contrôle, à la rédaction des procédures liées au contrôle, etc ... Les coûts d'infrastructure, de gestion de la qualité en font partie (cercles de qualité par exemple).

iv. **Les coûts de transaction relatives aux modifications de production** (changement affectant la définition technique des produits et des équipements, par exemple). Ces changements portent sur l'organisation globale du processus de production, sur la politique d'achats et les spécifications techniques. Ces coûts sont d'après E. Vollman et alii. [1988] les plus importants. Par exemple, un tel changement nécessite la réunion des responsables du contrôle de production, des lignes de production, de l'ingénieur de production, des responsables des achats, etc ... Il doit, de plus, être approuvé, planifié et son introduction doit être pilotée. Toutes ces étapes sont par définition coûteuses en ressources et notamment en temps. La figure 3 représente une comparaison entre les USA et le Japon de la fréquence de telles transformations. La fréquence d'une transaction apparaît ainsi clairement comme un élément non négligeable de son coût.

Ces quatre formes de coûts de transaction nous semblent couvrir la majorité des cas. Seule une application systématique sur le terrain pourrait permettre une meilleure définition.

Figure 3 : Fréquences des modifications du processus aux USA et au Japon
(Source : E. Vollman et alii., 1988, p.248).



Ainsi, nous parvenons à une première proposition que nous chercherons à montrer: **l'intégration peut s'interpréter comme un mode de réduction des coûts de transaction.** En effet, l'intégration va permettre une diminution du nombre des transactions génératrices des coûts ("cost driver"). Plus généralement, les procédures d'intégration peuvent être considérées comme des moyens de réduire la complexité de l'entreprise. Cette propriété est liée à la relation existante entre flexibilité et complexité (cf 3). Introduire dans un système une flexibilité supplémentaire revient à augmenter les interactions entre les composantes du système et entre ce système et son environnement. En d'autres termes, le nombre de transactions croît. Ainsi, la notion de complexité pourra être approchée par le nombre de transactions générées par les procédures liées à la gestion du système.

Le Tableau 1 ci-dessus donne quelques exemples de procédures d'intégration. Ces procédures s'insèrent dans un contexte stratégique très spécifique. Si l'on devait exprimer une évolution globale de l'organisation interne de la production, elle s'énoncerait ainsi : à la hiérarchie ancienne, facteur de rigidités empêchant les réactions locales aux événements, est substitué un principe de coordination plus décentralisée (réduisant les niveaux hiérarchisés mais ne supprimant pas les hiérarchies, au contraire même, cf. P. Zarifian, [1990]). Cette coordination doit s'avérer capable de maîtriser les coûts de transaction qu'elle génère et repose donc sur des mécanismes d'intégration et sur la redéfinition des interfaces entre composantes.

A titre d'exemples, prenons deux procédures d'intégration : l'intégration de phases productives et le juste-à-temps (JAT) et examinons leurs effets sur les coûts de transactions.

Exemple 1:

L'intégration de phases productives correspond à une réduction de la "complication" du système. Elle consiste soit à faire disparaître certaines phases de production, soit à exécuter en une seule et même opération ce qui jusqu'alors exigeait plusieurs opérations consécutives (P. Cohendet - P. Llerena, 1987a). Ce type d'intégration permet de compenser l'introduction de redondance dans les systèmes de production, nécessaires pour assurer une flexibilité statique face à un environnement risqué.

L'intégration de phases productives agit directement sur les coûts de transaction. Elle élimine par définition les coûts logistiques et d'équilibrage. Son effet sur les coûts liés à la gestion de la qualité est a priori indéterminé et il est vraisemblable que les coûts de transactions relatives aux modifications de production augmentent. Ainsi, lors d'une intégration de phases, il s'agit d'estimer l'effet net pour le comparer aux coûts d'investissement nécessaire.

Exemple 2:

Le juste-à-temps (JAT) correspond à une maîtrise des flux matières et informationnels en éliminant toutes redondances. Cette procédure d'intégration réduit fortement les coûts de structure en favorisant un processus délocalisé de décision, de gestion des aléas et de la qualité. Le juste-à-temps permet de réduire énormément les coûts de transaction logistique, en partie par une correspondance terme à terme des besoins et des flux matières. Par conséquent, la logistique est "intégrée" au flux des matières. Par ailleurs, d'après E. Vollman et alii. [1988], la complexité de la planification des besoins en composants se trouve réduit de 75 à 90 %. Cette réduction implique donc une forte diminution des coûts de transaction d'équilibrage. Les procédures de gestion de la qualité sont grandement simplifiées par le JAT, notamment par un lien étroit entre deux postes successifs de production. Ainsi, le contrôle de qualité est plus rapide et la réaction à une erreur immédiate. Finalement, la figure 3 montre également une réduction de la fréquence des modifications des processus de production et donc des coûts de transaction qu'elles

gènèrent. Cette vision extrêmement positive du JAT doit cependant être relativisée. En effet une telle organisation est souvent incompatible avec des changements majeurs de production, surtout s'ils sont mal anticipés.

Par ces deux cas de procédure d'intégration, nous avons tenté d'illustrer la pertinence de l'utilisation de la théorie des coûts de transaction comme fondement économique de l'intégration. Elle permet, notamment, de dégager un mode d'évaluation de l'intégration, cohérent avec les développements récents de la comptabilité industrielle fondés sur les notions d'"activités" et de "générateur de coûts". Dans ce cas, les transactions deviennent la source principale des coûts de structure. Nous avons par ailleurs pu procéder à un essai de cette méthodologie (cf. P. Cohendet et alii, 1990).

4.3 Vers un nouveau modèle économique de l'organisation de la firme

Nous avons tenté de montrer que la recherche de flexibilité en régime d'information perturbée nécessite un accroissement de la complexité interne des entreprises. Les processus d'intégration constituent alors aussi bien un mode d'allocation de la flexibilité qu'un principe opératoire pour gérer, voire réduire le degré de complexité de l'entreprise. Or, le fondement économique de l'intégration semble devoir se trouver dans la notion de coût de transaction. Il nous est maintenant possible d'entrevoir un "nouveau" modèle d'organisation de l'entreprise dans un régime d'information perturbée. L'intégration devient alors un mode de structuration adéquat de l'entreprise en sous-ensemble cohérent, permettant une réduction aussi importante que possible des coûts de transaction. Schématiquement, le "modèle" proposé reposerait sur les principes suivants, exprimés sous forme de propositions logiques:

1°) Les coûts totaux de l'entreprise sont constitués de deux arguments principaux:

- les coûts de production proprement dits, au sens classique du terme, dépendant des quantités produits (impliquant des frais de main d'oeuvre, capital, etc...). Ces coûts, en première analyse, ne dépendent pas de la structure et de l'organisation de l'entreprise.

- les coûts de gestion liés à la complexité de l'entreprise (c'est ici que les hypothèses divergent avec la représentation classique qui ne considère à ce niveau que des coûts fixes).

2°) La complexité de l'entreprise est dépendante du nombre de produits fabriqués (variété statique) et des délais au sens large de production de ces produits (réactivité dynamique).

3°) Les coûts de gestion liés à la complexité sont dépendant du nombre de transactions (unités physiques). Ces transactions, en première analyse, sont décomposables en transactions de nature logistique, en transactions liées à la qualité, en transactions liées à la reconception de produits et en transactions d'équilibrage. Le nombre de transactions est lui-même dépendant de l'organisation de l'entreprise.

4°) Une procédure d'intégration visera à réduire les coûts de complexité de l'entreprise à travers la diminution du nombre de transactions.

5°) Par définition, les procédures d'intégration peuvent donc revêtir des modalités très différentes, et en pratique on peut effectivement constater la grande diversité de ces politiques (ex: TGAO qui minimise les coûts de transaction de reconception, le KANBAN qui minimise les coûts de transaction logistique, etc...)

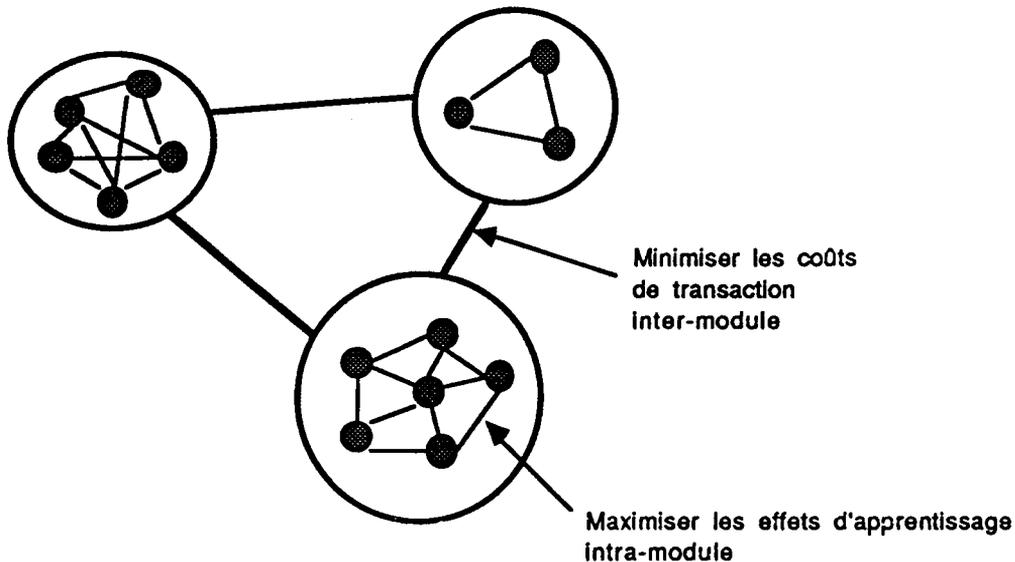
6°) Il existe une architecture adéquate de l'entreprise conçue comme limite de l'ensemble des procédures d'intégration, qui correspond pour une configuration industrielle donnée (type de produits et quantités de produits à produire) à un coût de transaction minimum.

7°) Cette architecture repose sur la définition de sous-ensembles de production (qui peuvent se confondre avec des îlots, des ateliers, des équipes ou toutes autres formes de sous-ensembles d'entreprise) bénéficiant **d'économies d'apprentissage**. En effet, en régime d'information perturbée, c'est l'existence des processus d'apprentissage qui assure la viabilité du système dans son ensemble. Les lieux de ces processus d'apprentissage constituent le découpage recherché de l'entreprise. Ils représentent les invariants par rapport aux évolutions de l'environnement, et peuvent donc être les unités élémentaires de l'organisation de l'entreprise (cf. figure 4).

8°) Ces économies d'apprentissage naissent de la capacité d'équipes de production au sein de l'entreprise d'absorber les chocs externes en assurant la meilleure réactivité interne grâce à leur expérience commune (car à l'intérieur de ces équipes les membres peuvent se remplacer, s'entraider mutuellement, etc...). Il s'ensuit des **gains de temps immédiats** si l'apprentissage pré-existait, ou des effets potentiels positifs à attendre à terme si l'apprentissage démarre. Ces effets d'apprentissage permettent d'affaiblir la relation entre l'augmentation du nombre de transactions et l'augmentation des coûts de complexité. Ces sous-systèmes internalisent donc certains coûts de transaction.

9°) L'entreprise cherchera donc l'architecture qui lui permettra de minimiser des coûts de complexité. Elle recherchera pour cela **une organisation qui minimise les coûts de transaction entre les sous-systèmes, tout en maximisant les effets d'apprentissage à l'intérieur de ceux-ci.**

Figure 4: Structure organisationnelle fondée sur les effets d'apprentissage



Ainsi l'entreprise cherchera d'une part à internaliser les coûts de certaines transactions lorsque celles-ci favorisent les processus d'apprentissage, par la constitution de sous-ensembles, et d'autre part à réduire, voire minimiser les coûts de transaction résultant de la coordination de ces différents processus. Nous voyons ainsi émerger un modèle de l'entreprise centré autour des notions d'apprentissage et de transaction. L'unité de base de l'organisation y devient le processus local d'apprentissage.

10°) La maximisation des effets d'apprentissage ne signifie pas que le découpage de l'entreprise est figé une fois pour toute. Selon l'évolution de l'environnement et la nécessité d'une mise en cohérence de l'ensemble des processus d'apprentissage, l'entreprise devra décider la création ou la destruction de nouveau lieux d'apprentissage (en regroupant par exemple des sous-systèmes dans des cercles de qualité; par de la formation; par des actions incitatives,...)

Ces propositions demeurent pour la plupart à démontrer aussi bien d'un point de vue analytique qu'empirique. Elles constituent cependant d'intéressantes pistes de recherche.

L'objet de ce texte est de montrer le rôle de la nature de l'information dans la définition des comportements et de l'organisation des entreprises. De cette analyse résulte au moins un

constat: les modes d'évaluation et d'organisation doivent nécessairement s'articuler autour de processus d'apprentissage dès lors qu'il existe une information perturbée. Ce constat ne fait en fait qu'ouvrir de nombreuses interrogations et évoquer certaines pistes de recherches futures.

Notes

- (1) Ce paragraphe résulte d'une recherche menée en collaboration avec l'ADEPA, le CESIP et l'IFRESI; cf. P. COHENDET, P. LLERENA [1990].
- (2) Cette analyse de l'évolution n'est pas uniformément partagée. Cf. R. BOYER, G. SCHMEDER [1990].
- (3) Nous tenons à remercier A. BURETH pour sa contribution à cette partie de la recherche.

BIBLIOGRAPHIE

ANDERSON E. [1982] : "The sales person as outside agent or employer : a transaction cost analysis", Wharton School of Pennsylvania, 1982.

AOKI M. [1986]: "Horizontal versus vertical information structure of the firm", *American Economic Review*, Vol. 76, n°5, pp. 971-983.

BAUMOL W.J., PANZAR J.C, WILLIG R.D. [1982]: "*Contestable markets and the theory of industry structure*", Harcourt Brace Jovanovitch, 1982.

BOYER R., SCHMEDER G. [1990]: "Division du travail, changement technologie et croissance: un retour à Adam Smith", *Revue Française d'Economie*, Vol.V, n°1, hiver 1990, pp. 125-194.

BULTEL J. [1983]: "Flexibilité de production et rentabilité des investissements. L'exemple de la robotisation de l'assemblage solaire ou soudage par points", *Revue d'Economie Industrielle*, 4ème trim., 1983.

CASTI J. [1986]: "Manufacturing as a system determined science", I.I.A.S.A., Laxenbourg, 1986.

CHARREAUX G. et alii [1987] : "De nouvelles théories pour gérer l'entreprise", Economica - Gestion, Paris, 1987.

COASE R.H.[1937] "The nature of the firm", traduit en français *Revue Française d'Economie*, Vol. II n°1, hiver, pp.133-157.

COHENDET P. et alii [1990]: "Intégration, coûts de transaction et organisation par modules", contribution du BETA à "*L'entreprise face à l'intégration*", ADEPA, Phase 2, 1990.

COHENDET P., LLERENA P. [1987a]: "Flexibilités, complexité et intégrations dans les processus de production", in B. ANCORI et alii, "*Evolution économique et théorie de l'information*", Rapport pour l'ARI Sciences de la Communication, ATP Droit et Economie de l'Information et de la Communication, Mars 1987.

COHENDET P., LLERENA P. [1987b]: "Productique et nouvelle représentation des processus de production", in P. COHENDET, P. LLERENA, P. PECQUET [1987], "*La productique: concepts, méthodes, mise en oeuvre*", Economica, 1987, pp.15-38.

COHENDET P., LLERENA P. [1989a] : "*Flexibilité, information et décision*", Economica, Paris, 1989.

COHENDET P., LLERENA P. [1989b] : "Flexibilités, risque et incertitude dans la théorie de la firme : un survey", in P. Cohendet, P. Llerena [1989a].

FAVEREAU O. [1989] : "Valeur d'option et flexibilité : de la rationalité substantielle à la rationalité procédurale", in P. Cohendet, P. Llerena [1989a].

GAFFARD J.L. [1990] : "*Economie industrielle et de l'innovation*", Dalloz

JOFFRE P., KOENIG G. [1985] : "*Stratégie d'entreprise*", Economica, 1985.

LORINO P. [1989] : "L'économiste et le manager", La découverte, 1989.

MONTEVERDE K., TEECE D. [1982] : "Supplier switching costs and quasivertical integration", *Journal of Law and Economics*, 1982.

SMITH A. [1776] : "*Recherches sur la nature et les causes de la richesse des nations*", 1776, édition partielle en français, Gallimard, 1976.

VOLLMAN E., BERRY W., WHYBARK D. [1988] : "*Manufacturing planning and control system*", Irwin, Homewood, 1988.

WACKER, WEBER [1984] : "A transaction cost approach to make or buy decisions", *Administrative Science Quarterly*, 1984.

WILLIAMSON O. [1981] : "The modern corporation, origins, evolution, attributes", *Journal of Economic Literature*, Vol. XIX december, pp.1537-1568.

ZARIFIAN P. [1990] : "Hiérarchisation, réseaux et systèmes de décision dans les grandes entreprises industrielles", à paraître dans "Réseaux".

CHAPITRE 6

**APPRENTISSAGE INFORMATIONNEL
ET RATIONALITE ECONOMIQUE
DES CRITERES D'INVESTISSEMENT**

par

Gilles LAMBERT

L'impact d'une procédure d'investissement sur la firme n'est pas neutre comme le considère l'analyse financière classique des investissements. Parallèlement aux flux financiers générés par le projet d'investissement, d'autres types d'effets peuvent être appréhendés par le calcul économique afin d'enrichir l'analyse d'impact de l'investissement sur l'entreprise.

Il s'agit, d'une part, d'analyser le rôle joué par la décision d'investissement sur la **capacité d'adaptation** à court et moyen termes de la firme aux prises avec des incertitudes sur le prix des intrants, sur le niveau et la nature de la demande et/ou sur la technologie. Une façon d'y parvenir consiste à prendre en considération le degré d'irréversibilité de la décision en l'intégrant au calcul de la rentabilité de l'investissement, ceci ayant pour effet de mettre en avant les stratégies flexibles (C. Henry [1974], A. Richard [1982], P. Llerena [1986], G. Lambert [1988]). Il est important, d'autre part, de ne pas négliger la **dimension organisationnelle** de l'investissement pouvant générer des effets de structure qui dépassent le cadre strict classiquement retenu pour analyser le projet. Nous ne pouvons négliger le fait qu'un investissement est très souvent générateur d'apprentissages technologique et organisationnel, dont les effets peuvent rétroagir en modifiant de façon significative la rentabilité attendue d'un projet. Certes, ce second aspect est parfois difficilement tangible au moment du calcul de rentabilité permettant de guider le choix des décideurs, et nous nous efforcerons de rendre compte des limites de la méthode quantitative que nous proposerons.

Cette volonté affichée par les théoriciens et les praticiens d'ouvrir le champ d'investigation des critères de choix en matière d'investissement répond à un double objectif. Tout d'abord, elle permet, selon nous, de réhabiliter les critères d'investissement en les rendant mieux adaptés aux prises de décision dans un environnement économique incertain conduisant les décideurs à rechercher des solutions flexibles. Elle fait ensuite écho, en ce qui concerne le calcul de rentabilité au niveau de la firme, à l'évolution récente de l'analyse économique de projets publics dont l'ambition est d'évaluer toutes les formes d'impact d'un projet d'investissement au plan social (A.K. Dasgupta, D.W. Pearce [1972], E.J. Mishan [1988], J. Benard [1989]). Ces impacts peuvent être directs ou indirects (c'est-à-dire sortant du cadre du projet proprement dit), tangibles ou intangibles (c'est-à-dire n'ayant pas de valeur monétaire issue d'un marché). L'évaluation des projets d'investissement public sur la base d'une analyse coûts-avantages intègre des éléments tels que l'impact du projet sur la productivité du secteur ou de l'activité industrielle en générale, ou encore la mesure économique d'externalités négatives engendrées par le projet comme, par exemple, les nuisances environnementales (bruit, pollution,...). La méthode proposée par les économistes pour analyser l'impact sur la productivité repose sur le modèle "input-output" qui retrace la diffusion intersectorielle du progrès technique en se basant sur une information macroéconomique provenant du tableau Entrées-Sorties. L'une des difficultés inhérente à l'utilisation de cette méthode provient du fait

La compréhension de cet article n'exige pas la lecture des modélisations en italiques. Celles-ci sont conçues comme support théorique aux arguments que nous développons, et n'apparaissent que dans un but pédagogique.

que les produits vendus par un secteur aux autres secteurs de l'économie incorporent la même quantité de R&D. La mesure des effets de nature environnementale fait appel pour sa part à deux techniques d'évaluation distinctes. La première consiste en une évaluation directe des dommages environnementaux générés par le projet (santé, habitations, corrosion des matériaux,...) et suppose par conséquent l'existence de nomenclatures fiables et homogènes en la matière. La seconde technique d'évaluation mesure indirectement ces effets en observant la dépréciation qu'ont subi, ou que pourraient subir, des biens et sites touchés par la pollution (valeur des immeubles par exemple). L'évaluation économique consiste alors à internaliser ces externalités, et la valeur de la dépréciation est supposée égale au coût de la pollution engendrée par le projet. De tels développements pourraient avoir lieu au niveau du calcul économique de la firme dans la mesure où le progrès technique issu d'un processus d'apprentissage comme les problèmes de pollution vont rétroagir sur la rentabilité escomptée de l'investissement.

Les travaux de H. Simon [1976-1978] en ce qui concerne la rationalité des preneurs de décision constituent un point d'appui non négligeable afin de mener à bien ce projet. Il convient toutefois de préciser la nature exacte de cet emprunt à la théorie économique. Car, en effet, deux sens distincts, ne recouvrant pas les mêmes incidences théoriques, peuvent être donnés au concept de rationalité simonien. Le premier, le plus couramment répandu, ne retient de la **rationalité limitée** que le caractère fortement restreint des capacités de calcul du décideur. Cette constatation conduit à analyser le comportement du preneur de décision, non pas en maximisateur, mais en bon statisticien, se basant sur une information partielle. Malgré l'importance que revêt cette correction comportementale du point de vue scientifique, elle ne nous paraît pas fondamentale en vue de l'élaboration d'une règle de décision pouvant aider à la prise de décision dans l'entreprise. Le second apport simonien oppose les caractères **substantiel et procédural** de la rationalité du décideur. L'interprétation que nous pouvons faire de la rationalité procédurale, pour laquelle la façon dont a été élaborée la décision importe autant que le résultat financier direct attendu, révèle des dimensions jusque là négligées de l'investissement et nous conduit à corriger la manière d'aborder l'évaluation d'un projet pour la firme. L'introduction du concept de **valeur d'option** (C. Henry [1974 a,b]) dans les critères classiques de choix d'investissement, constitue un premier pas significatif en ce sens. Le concept de valeur d'option est apparu de façon intuitive dans la littérature économique (B. Weisbrob [1964]) dans les domaines de décision où l'irréversibilité est la plus lourde de conséquence: la décision environnementale qui porte sur des actifs irremplaçables (une forêt par exemple). Elle correspond dans ce contexte à la valeur monétaire qu'un agent est prêt à payer pour acquérir une information supplémentaire en vue de préserver l'actif irremplaçable. Cette définition est proche dans sa démarche du "test de compensation" d'Hicks-Kaldor (J.H. Hicks [1940], N. Kaldor [1969]) permettant de donner une valeur économique à des effets intangibles. Ce test largement utilisé dans les analyses coûts-avantages mesure en effet les bénéfices et pertes intangibles d'un projet par la volonté de payer des agents économiques subissant l'influence du projet. Parallèlement aux développements théoriques qui ont eu lieu autour du concept de valeur d'option dans un cadre décisionnel statique (R. Schmalensee [1972], C.J. Cicchetti, A.M. Freeman [1974], P. Bohm [1975], R.J. Anderson [1981], R.C. Bishop [1982], V.K. Smith [1983]), des cas d'applications pratiques ont eu lieu dans le domaine des décisions

environnementales (A.M.III Freeman [1979], V.K. Smith - W.H Desvougues - A. Fisher [1983], C.F. Runge [1984]).

Dans le cadre plus restreint de la firme, le concept de valeur d'option permet, d'une part, de mettre en évidence la focalisation du décideur en situation d'incertitude sur les décisions les plus irréversibles. L'argument peut être poussé plus loin et conduire à des sélections dans les différents états du monde qui affectent un projet d'investissement (G. Lambert [1988]). Ce processus de sélection des perspectives de gains attachés à un projet va dans le sens d'un plus grand poids donné aux coûts des décisions irréversibles dans le calcul économique. Ce comportement permet d'expliquer des phénomènes de report d'investissement en vue de réaliser une décision d'investissement sur une meilleure base informationnelle (contexte d'information croissante). C'est le cadre classique des modèles de valeur d'option appliqués à la décision d'investissement (B.S. Bernanke [1983], P. Cohendet [1983], P. Llerena [1986]). La valeur d'option met, par ailleurs, en évidence l'aptitude de certaines procédures d'investissement à générer des phénomènes d'apprentissage (G. Lambert [1987]). Ces procédures sont très souvent les plus lourdes financièrement, et donc les plus irréversibles pour l'entreprise. La valeur d'option permet alors de prendre en compte des processus d'apprentissage au niveau du calcul économique (projet d'investissement générateur de progrès technique par exemple). Ces effets d'apprentissage sont issus de la procédure d'investissement elle-même, et échappent à toute analyse classique en termes de dépenses/résultats basées sur le calcul des cash-flows actualisés.

Le point commun entre ces différentes évolutions des critères de choix en matière d'investissement réside dans la part qui est faite à l'information dans le processus décisionnel, ainsi que dans la distinction de nature de celle-ci (exogène ou endogène). Ainsi, dans un cadre procédural de décision, l'information est considérée comme étant un bien productif au même titre que les hommes ou les machines. Nous verrons, en effet, dans une première partie consacrée à la prise en compte des effets irréversibles de la décision d'investissement, que l'information sur les variables de l'environnement, pouvant influencer la rentabilité du projet d'investissement, est une variable stratégique. La seule perspective de son acquisition peut justifier le report d'un projet d'investissement (1). Dans ce contexte, l'information de nature exogène peut faire l'objet d'une décision d'investissement au sens propre lorsque, dans le but d'accélérer son processus d'acquisition, l'industriel envisage de faire par exemple appel à un audit. Enfin, dans une seconde partie (2), nous évoquerons en quoi l'investissement peut être lui-même générateur d'informations (concept d'information endogène) technologique et organisationnelle pour la firme, et ainsi par des effets de structure, rétroagir sur le niveau de rentabilité réel du projet.

1. Flexibilité d'adaptation de la firme et analyse bayésienne de la décision d'investissement

Il existe des techniques très simples permettant de prendre en compte dans le calcul d'investissement le degré d'irréversibilité engendré par la décision d'investissement. Elles permettent de relier directement dans l'évolution financière d'un projet l'exigence de flexibilité

avec la rentabilité globale escomptée. C'est l'objectif de cette section. Nous évoquerons dans une section ultérieure une attitude "plus active" vis-à-vis de l'information en analysant les possibilités d'achat d'informations stratégiques préalablement à la décision d'investissement.

1.1. Critères séquentiels a posteriori avec irréversibilité et croissance de l'information : une analyse alternative du risque

L'analyse financière d'un projet d'investissement limite le diagnostic à l'aspect "dépenses-recettes" engendrées par celui-ci. Elle néglige par conséquent le fait que la décision d'investissement peut modifier les capacités d'adaptation de la firme. Le principe d'indépendance entre les plans d'investissement et de financement qui est au fondement des modèles d'évaluation financière des projets d'investissement (valeur actualisée nette, taux interne de rentabilité, etc ...) conduit à scinder le problème des ressources de financement avec celui de la rentabilité des projets. Cette relation d'indépendance fut démontrée par F. Modigliani - M. Miller [1958] dans le cadre d'hypothèses de marché parfait (absence notamment de faillites et de taux d'emprunt pour les entreprises). Cette rupture dans la chaîne causale des décisions, par ailleurs totalement justifiée quand il s'agit de comparer plusieurs projets ayant les mêmes sources de financement, a eu pour conséquence de négliger l'incidence inverse, c'est-à-dire celle de la décision d'investissement sur la trésorerie de l'entreprise pendant toute la durée de vie du projet. Ceci explique selon nous pour une large part l'insuccès des critères de décision actuariels inadaptés à la prise de décision lorsque les variables de l'environnement économique contraignent la firme à une adaptation continue.

Si l'on considère comme stable l'environnement économique de la firme, les problèmes de trésorerie se posent au début du processus de décision, au moment du choix des sources de financement. Les propositions comptables reviennent à ne choisir que les solutions qui respectent un équilibre financier entre capitaux propres et endettement (A. Charrue [1980]). Elles se restreignent à l'examen du bilan annuel comptable. L'exigence qui consiste à respecter cet équilibre sur une période plus longue réside dans le choix d'un taux d'actualisation des bénéfices, reflet de la rentabilité minimale exigée d'un investissement (notion de coût du capital).

Dans un contexte d'instabilité de l'environnement économique, la trésorerie constitue une source perpétuelle de financement d'investissement, et donc, d'adaptation de la firme aux incertitudes affectant sa performance économique. R.A. Thompson & J. Thuesen [1985] font une proposition intéressante dans cette direction. Sur la base d'un critère séquentiel de décision basé sur la notion de valeur actuelle nette, les auteurs lient l'impact comptable engendré par la décision d'investissement au ratio d'autonomie financière. L'astuce théorique consiste à ne pas raisonner sur des cash-flows au niveau de la valeur actuelle nette, mais sur les changements intervenus dans les fonds propres (notion plus globale), et ceci pour chaque période de l'investissement. Toutefois, si l'idée est originale, le raisonnement tenu par les auteurs repose sur la notion d'utilité et suppose donc connu le degré d'aversion face au risque du décideur, c'est-à-dire la déformation psychologique des gains monétaires par le décideur.

1.1.1 Expression du critère

L'analyse bayésienne nous permet, au sein d'un critère unique de décision, de rétablir cette relation entre les exigences de trésorerie en vue d'une capacité d'adaptation de la firme et de la décision d'investissement en permettant :

- d'une part, d'introduire, dans un contexte de croissance de l'information, une hypothèse d'irréversibilité de la décision. Cette hypothèse s'oppose à la myopie du décideur dans les critères de choix d'investissement (K. Arrow [1970] P.Cohendet [1983]) qui considère l'investissement productif comme parfaitement fluide, c'est-à-dire à tout moment liquidable à sa valeur comptable.

- d'autre part, de comparer financièrement à chaque décision d'investissement la solution flexible qui consiste à reporter l'investissement et donc à bénéficier d'un surplus informationnel sur l'incertitude au moment du choix.

Plusieurs formalisations (C. Henry [1974a], A. Richard [1982], G. Lambert [1988]) intégrant ces deux aspects ont été proposées sur la base du critère de valeur actuelle nette. Elles décrivent un critère de choix en matière d'investissement directement applicable par l'entreprise, et aboutissent à une synthèse entre les objectifs de liquidité permettant à l'entreprise de réitérer ces choix et l'objectif de rentabilité globale de l'investissement.

Face au critère de prudence extrême, représenté par le délai de récupération et ses dérivés (temps de retour, ROI, etc ...), l'utilisation d'une valeur actualisée nette permet de rendre compte de la rentabilité globale d'un investissement sur l'ensemble de sa durée. Le critère que nous proposons débouche sur un processus de décision des projets d'investissement plus fin qui évite le va-et-vient observé dans la pratique (suivant l'état de la conjoncture) entre un critère de sélection exclusivement tourné vers le retour des liquidités (délai de récupération) et un critère de pure maximisation des bénéfices (valeur actuelle nette).

Formulation du critère: Supposons qu'une entreprise, pour satisfaire sa demande en $t=1$, envisage de réaliser un investissement de capacité en $t=0$. Le niveau de demande n'est pas certain sur toute la durée de vie du capital productif, car il y a une possibilité que l'entreprise perde des parts de marché (existence d'une concurrence extérieure). Une décision susceptible de modifier le niveau futur de la demande devant avoir lieu en $t=k-1$, elle n'entrera en vigueur qu'en $t=k$ ($k \in [1, 10]$), date à laquelle l'industriel anticipe deux états du monde possibles :

-l'état 1 avec une probabilité de réalisation Π : existence d'une concurrence extérieure forte en $t=k, \dots, 10$, et qui diminue en partie la demande à la firme,

-l'état 2 avec une probabilité de réalisation $1 - \Pi$: la concurrence extérieure faible ne diminue pas la demande à la firme .

L'investissement dans la nouvelle capacité en $t=0$ constitue un choix irréversible , et l'industriel devra subir des coûts de surcapacité dans le cas où l'état 1 se réalise . Le capital productif qui fait l'objet de la décision d'investissement est divisible ex-ante, mais ne peut être décapitalisé ex-post . La concurrence est supposée se maintenir jusqu'à la fin de l'horizon et entraîne une baisse significative des cash-flows . La structure d'information est supposée parfaite. Ceci se traduit dans le critère par la connaissance en $k-1$ de l'état du monde qui va prévaloir

en k .

Le problème pour l'entrepreneur consiste donc à prendre une décision d'investissement alors que l'information sur les conséquences de ce choix n'arrive qu'ultérieurement. Plusieurs stratégies sont donc possibles, chacune d'elles donnant lieu à un résultat espéré qu'il conviendra de comparer sur la base du critère de la valeur actuelle nette retenue par le décideur.

Une décision irréversible (notée d) consiste à réaliser, quelque soit l'état du monde futur, l'investissement de capacité (noté I) dans sa totalité en début de période. Cette décision permet d'améliorer les cash-flows de premières périodes en prenant le risque de subir des coûts de surproduction sur la fin de l'horizon.

La valeur actualisée nette de cette décision irréversible s'écrit :

$$VAN_d = \sum_{t=1}^{k-1} \alpha^t \overline{C_t} + ((\pi) \sum_{t=k}^n \alpha^t \overline{C_t^1} + (1-\pi) \sum_{t=k}^n \alpha^t \overline{C_t^2}) - I$$

où n : durée de vie du capital,
 C_t : cash-flow de la période t ,
 C_t^i : cash-flow de la période t dans l'état i
 I : montant initial de l'investissement,
 α^t : facteur d'actualisation.

L'effet d'une baisse de la demande en $t = k$ entraîne des coûts de surcapacité qui se traduisent par :

$$\overline{C_t^1} < \overline{C_t^2}$$

Le décideur neutre vis-à-vis du risque réalisera la totalité de l'investissement I si :

$$VAN(d) > 0$$

L'entrepreneur peut opter pour une décision flexible (notée d) qui consiste à différer l'investissement. S'il est sensible au gain d'information (supposée ici parfaite) sur les états du monde qu'il peut obtenir de l'attente, il comparera la rentabilité d'une telle stratégie par rapport à la stratégie irréversible. Il investira donc en $t = 0$ dans la capacité juste nécessaire pour assurer le niveau de demande qui est certain. Il réalisera l'investissement de complément en $k-1$ dans le cas où c'est l'état 2 qui prévaut. Une propriété de divisibilité du capital productif lui permet de différer une partie de l'investissement.

La valeur actualisée nette s'écrit alors :

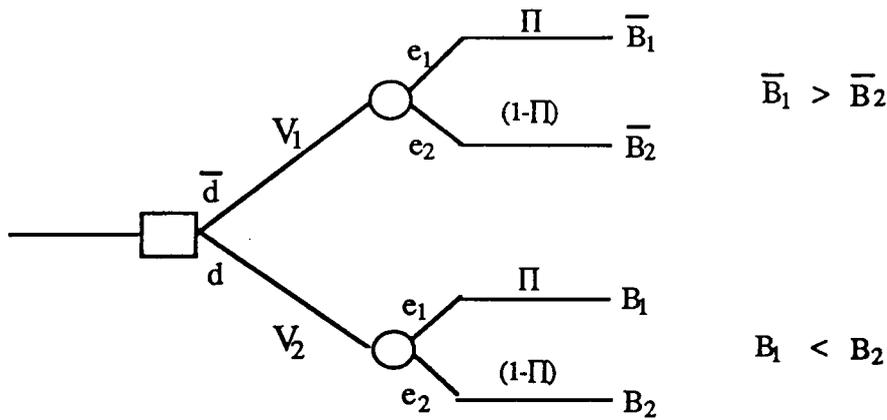
$$VAN_d = \sum_{t=1}^{k-1} \alpha^t C_t + [(\pi) \sum_{t=k}^n \alpha^t C_t^1 + (1-\pi) (\sum_{t=k}^n \alpha^t C_t^2 - \alpha^{k-1} I_{k-1})] - I'$$

où I' : montant de l'investissement réalisé en $t = 0$ et correspondant au niveau de demande certaine.

I_{k-1} : montant de l'investissement différé.

Supposons qu'un phénomène d'économies d'échelle existe en ce qui concerne le dimensionnement du capital productif. Il se traduit par l'inégalité suivante : $I < I' + I_{k-1}$.

Le choix du décideur consiste donc à comparer les solutions irréversible et flexible d'investissement conformément à l'arbre de décision ci-dessous :



Avec:

$$\begin{aligned} \bar{B}_1 &= \sum_{t=k}^n \alpha^t \bar{C}_t^1 & \bar{B}_2 &= \sum_{t=k}^n \alpha^t \bar{C}_t^2 \\ B_1 &= \sum_{t=k}^n \alpha^t C_t^1 & B_2 &= \sum_{t=k}^n \alpha^t C_t^2 \end{aligned}$$

et

$$VAN_{\bar{d}} = v_1 \quad VAN_d = v_2$$

La stratégie flexible de report d'investissement permet d'éviter un état de surcapacité relatif à la demande du marché mais ceci au prix d'un manque à gagner dans les premières périodes tel que l'on vérifie les inégalités suivantes:

$$\forall t \in [1, k-1] \quad , \quad \bar{C}_t > C_t$$

$$\forall t \in [k, n] \quad , \quad C_t^1 > \bar{C}_t^1 \\ C_t^2 \leq \bar{C}_t^2$$

Un degré d'aversion au risque supplémentaire peut être obtenu avec l'introduction de l'effet irréversibilité dans le critère de la valeur actuelle nette. Ce pas supplémentaire vers une plus grande prudence du décideur peut être franchi avec le concept de **valeur d'option** tel qu'il est défini par C. Henry [1974 b] et B.S. Bernanke[1983]. Ce concept équivaut au regret que le décideur pourrait avoir s'il choisissait la stratégie irréversible. Il se note :

$$V = \Pi (B_1 - \bar{B}_1) + (1-\Pi) (B_2 - \bar{B}_2)$$

Le critère de décision s'obtient alors en comparant cette valeur d'option avec les gains relatifs certains de la solution irréversible dans les premières périodes. La solution flexible sera adoptée si :

$$V - \sum_{t=1}^{k-1} \alpha^t (\bar{C}_t - C_t) + I - I' - \alpha^{K-1} I_{K-1} > 0$$

Un degré supplémentaire de sensibilité face au risque peut être obtenu dans le critère de sélection des projets. La solution consiste à déplacer le poids des croyances du décideur (mesurées en termes de probabilités subjectives) en faveur de la flexibilité. L'idée consiste à observer qu'en cas de défiance vis-à-vis d'un acte irréversible, le décideur va pondérer plus fortement la probabilité de l'état défavorable. Si

β et γ caractérisent les coefficients de déformation des probabilités initiales π et $1-\pi$, la valeur maximale de la perte encourue dans la seconde période s'écrit :

$$V' = \beta \cdot \Pi (B_1 - \bar{B}_1) + \gamma (1-\Pi) (B_2 - \bar{B}_2)$$

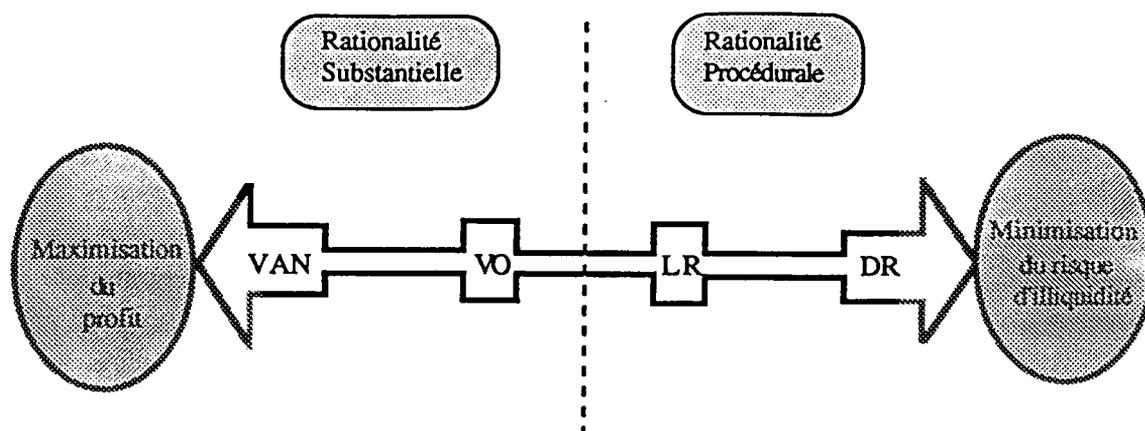
$$\text{avec } \beta \cdot \pi + \gamma \cdot (1-\pi) \leq 1 ; \beta > 1 ; \gamma < 1$$

Le décideur, supposé neutre vis-à-vis du risque, comparera cette perte espérée maximale aux bénéfices certains des premières périodes $[0, k-1]$ dégagés par la décision d par rapport à la décision d . Il n'investira pas tant que :

$$V' - \sum_{t=1}^{k-1} \alpha^t (\bar{C}_t - C_t) + I - I' - \alpha^{k-1} I_{k-1} > 0$$

Ce dernier critère que nous proposons (appelé critère "Liquidité-Rentabilité" et noté L-R; voir G. Lambert [1988]) met en avant une focalisation du décideur sur les perspectives extrêmes lors d'une prise de décision dans un environnement instable. La rationalité d'un tel critère se rapproche de celle contenue dans le délai de récupération dans la mesure où elle peut aller jusqu'à l'usage partiel de l'information disponible en vue d'accroître la prudence du décideur. C'est le cas lorsque l'agent applique un degré de confiance nul à l'état du monde favorable ($\gamma = 0$). Le calcul économique qu'il met en oeuvre ne prend alors en considération que l'état 1 défavorable à la décision irréversible d'investissement. Le critère opère donc un glissement dans la rationalité relativement aux critères de la valeur actuelle et de valeur d'option puisque l'agent ne se comporte plus en pur maximisateur. Alors qu'avec la valeur d'option le critère de décision reste inscrit dans une logique de rationalité substantielle principalement tournée vers les résultats financiers du choix, le critère LR débouche sur une rationalité de type procédurale proche de celle définie par (H.A. Simon [1978]). Dans son comportement l'individu, explique H.A. Simon, ne cherche pas à optimiser toutes ses décisions car le processus du choix est limité. L'agent ne maximise pas mais il choisit la procédure qui le satisfait. H.A. Simon [1978] écrit : "*yet as economic analysis acquires a broader concern with dynamics of choice under uncertainty, it will become more and more essential to consider choice processes*".

L'interprétation de la rationalité procédurale présente dans le critère que nous proposons à la place des critères financiers classiques signifie que l'objectif rationnel de maximisation du bénéfice cède le pas sur la préoccupation de l'agent à maintenir intact son espace de choix possibles futurs. La rationalité procédurale est donc ici entendue au sens où l'agent est sensible à la contrainte que les choix présents font peser sur les choix futurs. On peut représenter ceci à l'aide de la figure suivante :



La notion de temps séquentiel présente dans le critère classique de valeur actuelle nette est remplacée ici par un temps de la mémoire davantage préoccupé par l'incidence des décisions présentes sur l'ensemble de choix des périodes futures. L'investissement n'est plus indépendant de la capacité d'adaptation de la firme comme le considère la théorie financière, il l'influence, mieux, il la détermine. Ce critère de décision d'inspiration bayésienne rompt avec les résultats classiques de l'analyse financière dans le sens où il montre qu'il est parfois opportun de rejeter un projet d'investissement, malgré une rentabilité escomptée strictement positive ($VAN > 0$). Avec ce critère, l'investissement n'est plus considéré comme un placement financier parfaitement liquide, mais comme un acte pouvant diminuer l'ensemble des choix futurs possibles. Ce résultat constitue une alternative théorique non négligeable à l'argumentaire communément utilisé pour décrire la décision de désinvestissement dans l'entreprise ; argumentaire qui repose sur le principe de la cession d'actifs supposant l'existence de marchés de l'occasion pour les équipements productifs (P. Masse[1968], A. Quintart & R. Zisswiller [1982]).

1.1.2 Une représentation probabiliste de l'aversion au risque

Partant du cadre de la théorie bayésienne construite autour d'une notion simple d'espérance mathématique de gains, et en y ajoutant une hypothèse de déformation subjective des croyances, notre critère apporte une réponse à un dilemme souvent présent dans les procédures de choix d'investissement dans l'entreprise qui oppose deux objectifs stratégiques. Les aménagements apportés à la notion de valeur actualisée nette ont, tout d'abord, consisté à introduire les conséquences de l'irréversibilité de la décision d'investissement et, dans une seconde étape, à appliquer des correctifs sur la structure initiale d'information du décideur.

Ces deux étapes s'inscrivent dans la logique de dépassement du concept d'utilité espérée ayant lieu dans la théorie de la décision en incertain. En effet, ce critère jusqu'à présent fédérateur de toutes descriptions décisionnelles en économie comme en gestion, axiomatisé par J. Von Neuman & O. Morgenstern [1947] en ce qui concerne des probabilités objectives et un peu plus tard par L.J. Savage [1954] pour des probabilités subjectives, souffre de critiques internes remettant en cause la cohérence même de son édifice conceptuel, mais aussi d'une inadaptation aux formes réelles que prend la décision en entreprise. Pour s'en tenir à l'essentiel, le paradoxe de M. Allais [1953] remet en cause l'hypothèse d'indépendance entre les croyances

(caractérisées par des probabilités) et les préférences présentent dans la théorie de l'utilité espérée. Le paradoxe montre que le degré d'aversion au risque du décideur est croissant au voisinage de la certitude de gains. Par ailleurs, le cadre statique de cette théorie empêche toutes incidences des phénomènes de croissance de l'information et d'irréversibilité des choix sur le processus décisionnel de l'agent. Enfin, du point de vue pratique de la décision d'investissement en entreprise le fait de caractériser l'aversion au risque par une déformation psychologique des gains monétaires suppose que celle-ci soit connue et invariante dans le temps.

Concernant la prise en compte du caractère irréversible de la décision dans la détermination de la rentabilité d'un projet, plusieurs propositions théoriques méritent d'être mentionnées. La théorie alternative de la décision proposée par G.L.S. Shackle [1949] inspirée de la notion de "surprise potentielle", issue de sa théorie des probabilités, suppose que le décideur, lors d'une prise de décision, se focalise sur les alternatives extrêmes caractérisées par la meilleure et la plus mauvaise. Cherchant à combler le vide qui entoure à ce propos la théorie de l'utilité espérée, les théories du regret et du désappointement élaborées par D.Bell [1982, 1985] intègrent au moment du choix les conséquences et les manques à gagner provenant de ce choix. Ainsi, tout en maintenant le critère d'utilité espérée dans le premier cas, est pris en compte le coût d'opportunité sous forme d'utilité alternative à celle obtenue avec la décision d'investissement. La technique consiste dans la théorie du désappointement à tenir compte du regret potentiel que peut ressentir ex post l'agent.

Le second aspect présent dans notre critère lié à l'introduction de correctifs sur la structure initiale d'information est fondamental car il permet, dans un cadre décisionnel de **neutralité vis-à-vis du risque**, de réintroduire un degré de prudence dans le critère de décision. Les préférences du décideur sur la distribution des probabilités introduit un degré de prudence face au risque qui ne nécessite pas l'utilisation du concept d'utilité plus délicat à définir. Dans notre critère, le décideur ne déforme pas les valeurs monétaires (que reflète normalement la courbure de la fonction d'utilité), mais il va traduire son attitude face au risque en **priviliégiant une distribution de probabilités des états du monde**. L'intérêt théorique d'une telle démarche réside dans le fait que le critère qui en découle ne fige pas le comportement du décideur, comme dans le critère d'utilité espérée, puisque son aversion face au risque, qui se traduit par une déformation des probabilités, peut évoluer en fonction de son appréciation de l'incertitude. Le décideur change l'appréciation qu'il a des scénarii d'évolution future au fur et à mesure que lui parvient de l'information nouvelle, plutôt que, comme le suppose la théorie classique, de déformer les gains monétaires qui lui sont attachés. Dans un article éclairant à ce sujet, W. Fellner [1961] montre que les distortions des probabilités subjectives, que peuvent introduire les agents dans certains cas avant de réellement prendre leur décision, reflètent leur réaction à l'incertitude.

Les critères de décision proposés par M. Allais [1986], et de façon plus prononcée encore par M.J. Machina [1983], font une large place à la distribution des probabilités dans l'attitude face au risque du décideur. Car un des problèmes lié à l'utilisation du critère d'utilité espérée réside dans le fait que l'aversion au risque est confondue avec la notion d'utilité

marginale de la monnaie, ou encore d'attitude vis-à-vis de la richesse (utilité marginale croissante ou décroissante). M. Allais [1984] les dissocie en proposant un critère où, d'une part l'utilité marginale de la monnaie se traduit par la courbure de la fonction d'utilité supposée invariante, et d'autre part l'aversion au risque liée aux perspectives de gains est une fonction des probabilités d'autant importante que la perspective de gains est faible (Expected Utility with Rank Dependent Probabilities). M.J. Machina introduit la notion de **préférence lisse sur les distributions de probabilités**, ces dernières ayant de fortes incidences sur le degré d'aversion face au risque du décideur dans la mesure où l'utilité pour cet auteur est une fonction qui dépend de la valeur monétaire des conséquences mais aussi de la distribution des probabilités affectées à ces conséquences. L'une des limites du modèle proposé par M.J. Machina pour une application en économie provient du fait que les probabilités sont de nature objective, cette démarche présente un intérêt théorique et opérationnel non négligeable dans la mesure où l'attitude face au risque peut varier sans que l'appréciation de la valeur d'une même perspective de gain change (notion d'utilité invariante), il suffit que se modifient les croyances du décideur (reflétés par les probabilités). De même dans le critère que nous proposons, le degré de prudence du décideur est accru au voisinage de l'irréversibilité, et cela malgré l'hypothèse de neutralité vis-à-vis du risque au sens du ratio d'Arrow & Pratt qui se caractérise par des fonctions d'utilité linéaires.

Des propositions théoriques nouvelles sont plus radicales avec la notion d'utilité et ne rendent compte du phénomène d'aversion au risque qu'à travers le système de pondérations des agents. G. Shafer [1976] dans sa "théorie de l'évidence" met en avant des règles de comportement basées exclusivement sur des fonctions de croyance qui dépendent des gains monétaires. L'idée est intéressante mais ne donne pas lieu à un critère clair de décision. M.E. Yaari [1987] dans sa théorie duale du choix en situation de risque abandonne lui aussi la notion d'utilité et fait jouer aux probabilités le rôle tenu par l'utilité pour exprimer le degré d'aversion ou de préférence pour le risque du décideur (principe de dualité). Le critère de décision correspond alors à la maximisation des valeurs monétaires pondérées par une fonction des probabilités qui aura pour effet dans le cas d'un décideur ayant de l'aversion au risque d'accentuer le poids des perspectives extrêmes de gains (voir pour un "survey" de ces théories M. Willinger [1990]).

Toutefois, si avec ces approches (dont le lieu commun consiste à faire jouer un rôle central aux probabilités dans la caractérisation du risque) un pas important a été franchi permettant de mieux rendre compte des mécanismes décisionnels en incertitude, il reste à construire un cadre conceptuel suffisamment général pour contenir des sources différentes d'aversion aux risques. La prudence du décideur provient-elle de sa richesse initiale, de la variance des bénéfices espérés, du caractère irréversible de certaines décisions, ou de tout à la fois ?

La présence explicite de l'irréversibilité dans les critères de choix enrichit considérablement l'analyse financière des projets et évite d'avoir recours à des critères exagérément tournés vers le court terme. Toutefois, nous pourrions franchir un pas supplémentaire dans l'évolution des critères de décision. En effet, le comportement du décideur décrit dans le critère précédent se situe dans un cadre bayésien *a posteriori*, c'est-à-dire dans

un contexte où la décision doit être prise, connaissant la croissance de l'information au cours des périodes futures. Nous pourrions modifier ce cadre décisionnel en supposant que l'agent ait la possibilité d'accélérer ce processus d'acquisition de l'information en ayant recours à un conseil d'expert ou des études de marché par exemple. Le problème qui apparaît pour le gestionnaire revient à se prononcer sur le prix limite de cet investissement informationnel préalable à la décision d'investissement proprement dite.

1.2. Critère séquentiel avec recherche d'information préalable

Le comportement du décideur décrit jusqu'ici consistait à attendre que la "Nature" (c'est à dire l'environnement économique de la firme) révèle les informations, afin de corriger la perception de l'incertitude et, progressivement, l'éliminer du problème de décision. L'apport essentiel consistait à introduire à la fois l'irréversibilité et la croissance de l'information supposée constante dans les critères classiques de décision. Toutefois, relativement à une attente passive de l'information, on peut se demander ce qu'il advient de la décision d'investissement lorsque le décideur a une politique de recherche active de l'information.

Le recours à un audit informationnel, en vue de diminuer la part d'incertitude dans les paramètres de décision de l'entreprise se réfère à un cadre décisionnel bayésien de type "pre posteriori", c'est-à-dire celui de l'examen de la valeur de l'information supplémentaire (voir par exemple J.P. Védrine [1985], M. Willinger [1988]). Ce cadre peut être étudié selon deux perspectives. Il peut, tout d'abord, s'envisager comme un moyen de corriger la structure d'information du décideur, et ainsi, corriger ses croyances relatives à chaque état du monde. Mais l'évaluation des gains engendrés par une telle solution pose le problème de la connaissance a priori des résultats de l'expertise et conduit ainsi à remplacer, dans le critère de décision, les probabilités de l'agent par celles de l'expert. Si, techniquement, le cadre de la théorie bayésienne nous permet de résoudre le problème, le calcul à la période courante d'un tel critère exigerait la prescience du décideur industriel. La conséquence de ceci, au niveau du calcul économique, est que l'évaluation de la valeur des informations supplémentaires achetées à des audits nécessite de la part de l'acheteur une connaissance précise du contenu de cette information future.

La seconde perspective plus fructueuse pour le calcul économique consiste à examiner le cas où l'audit permet à l'industriel de connaître, dans la période courante, l'état du monde qui va prévaloir dans les périodes futures, et ceci avec un degré de fiabilité propre à la société qui fournit le service informationnel. L'audit ne corrige plus la distribution du décideur comme précédemment, mais il va tout simplement lui révéler l'état du monde avec un degré de fiabilité supposé connu.

Si, de plus en plus d'entreprises font appel à ce type de service afin de décider d'un investissement (déléguant ainsi à un tiers la recherche de l'information relative à l'environnement économique qui conditionnera la rentabilité du projet) la question de la valeur de l'information supplémentaire et du prix limite que ces entreprises sont prêtes à payer devient essentielle. En

effet, la consultation d'expert est une pratique qui se développe surtout dans les secteurs d'activités sujets à un environnement économique dynamique (réduction de la durée de vie des produits, etc ...), et parfois instable. Ce recours à un organisme externe à l'entreprise est d'autant plus vital pour les entreprises dont la taille ne permet pas la réalisation par ses services d'études de marché propres, comme c'est souvent le cas pour les PME. Mais l'évaluation d'une recherche d'information supplémentaire concerne aussi bien les cas de consultations externes que ceux où le décideur réalise lui-même une étude de marché ou un marché-test qui l'éclairera sur l'état de la demande. Le problème qui apparaît alors, précédant le calcul de la rentabilité d'un projet d'investissement, est celui de l'évaluation du recours à l'expert en tant que **surplus informationnel** pour le problème de décision. La réponse à ce problème nous conduit, par voie de conséquence, à rechercher les paramètres qui influencent la propension du décideur à acheter de l'information supplémentaire.

Formalisation du critère : L'entrepreneur tente ici d'évaluer la valeur des services de l'expert afin d'accepter ou de rejeter l'achat d'un supplément d'information. Il affecte aux services de l'expert un degré de fiabilité qu'il détermine sur la base de prévisions réalisées par celui-ci dans le passé. Nous supposons, pour simplifier, que cette fiabilité de l'information, fournie par l'expert, est identique quel que soit l'état du monde, c'est-à-dire est indépendante des états de l'environnement. Ainsi, pour tout message I_j fourni, indiquant l'état du monde e_j , nous avons :

$$\Pi(I_1 / e_1) = \Pi(I_2 / e_2) = \alpha \quad (\text{bonne prévision})$$

$$\Pi(I_1 / e_2) = \Pi(I_2 / e_1) = 1 - \alpha \quad (\text{mauvaise prévision})$$

La probabilité que l'expert indique au décideur l'état e_1 résulte de ses propres probabilités subjectives sur chaque état (elles constituent l'unique base d'appréciation pour le décideur au Il s'agit en fait d'évaluer un service informationnel avant d'en connaître le contenu exact.

Le théorème des probabilités totales nous donne la probabilité du décideur d'obtenir le message I_1 de l'expert.

Elle s'écrit :

$$\Pi(I_1) = \Pi(I_1 / e_1) \cdot \Pi(e_1) + \Pi(I_1 / e_2) \cdot \Pi(e_2)$$

Précisons que $\Pi(I_1)$ est la probabilité subjective pour le décideur de recevoir le message I_1 de la part de l'audit. Elle dépend à la fois des croyances du décideur vis-à-vis de la loi d'occurrence des états du monde, ainsi que de la marge d'erreur qu'il prête à l'information reçue. Pour l'audit, cette probabilité sera sans doute différente, mais elle n'est pas connue ex-ante du décideur et ne peut donc fonder aucune base d'appréciation dans l'estimation de la rentabilité espérée.

Les probabilités de l'entrepreneur relatives à chaque état du monde sont complémentaires conformément au modèle précédent :

$$\Pi(e_1) = \Pi \quad \text{et} \quad \Pi(e_2) = 1 - \Pi$$

Par ailleurs, le théorème de Bayes nous permet d'écrire les probabilités de chaque état conditionnellement à l'information reçue au début de l'horizon :

$$\Pi(e_1 / I_1) = \frac{\Pi(e_1) \cdot \Pi(I_1 / e_1)}{\Pi(I_1)} = \frac{\Pi \cdot \alpha}{1 - \alpha - \Pi + 2\alpha\Pi}$$

L'entreprise fait face au problème de décision suivant. Tout d'abord, elle peut avoir recours à l'audit et se fier entièrement à son information pour la décision d'investissement. Elle peut aussi poursuivre un processus bayésien classique d'acquisition de l'information, révélée par l'environnement. L'ensemble des cas de figure possibles sont représentés sur l'arbre de décision ci-après.

Les B_j correspondent aux valeurs actualisées des cash-flows dans l'intervalle $[k, n]$ pour chacune des possibilités.

Ils se définissent selon la décision retenue, ainsi que l'état du monde de la façon suivante :

$$VAN_d(e_1) = \sum_{t=1}^{k-1} \alpha^t C_t + B_1 - I'$$

$$VAN_d(e_2) = \sum_{t=1}^{k-1} \alpha^t C_t + B_2 - \alpha^{t-1} I_{k-1} - I'$$

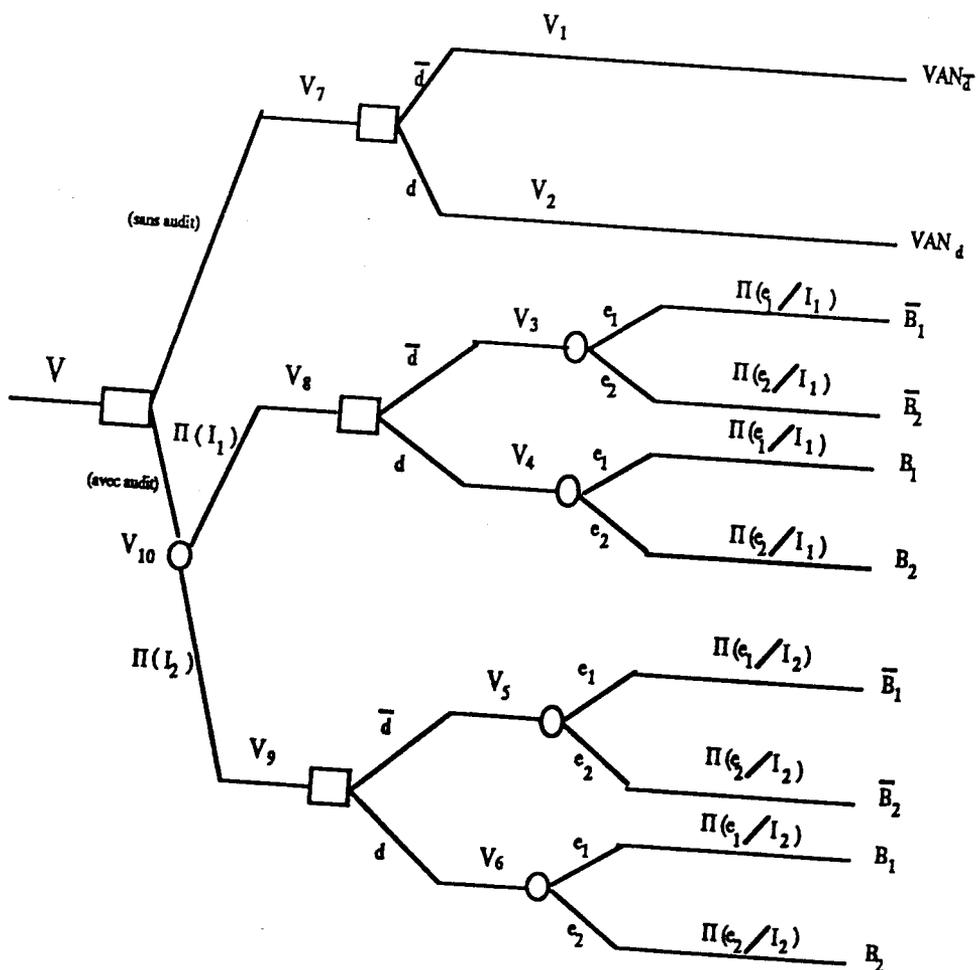
$$VAN_{\bar{d}}(e_1) = \sum_{t=1}^{k-1} \alpha^t \bar{C}_t + \bar{B}_1 - I$$

$$VAN_{\bar{d}}(e_2) = \sum_{t=1}^{k-1} \alpha^t \bar{C}_t + \bar{B}_2 - I$$

Les valeurs sur chaque branche de l'arbre ci-après sont les suivantes :

$$V_1 = VAN(\bar{d}) \quad V_2 = VAN(d) \quad V_7 = \max(V_1, V_2)$$

$$V_3 = \Pi(e_1/I_1) \bar{B}_1 + \Pi(e_2/I_1) \bar{B}_2 + \sum_{t=1}^{k-1} \alpha^t \bar{C}_t - I$$



$$V_4 = \Pi(e_1 / I_1) B_1 + \Pi(e_2 / I_1) B_2 + \sum_{t=1}^{k-1} \alpha^t C_t - I'$$

$$V_8 = \max(V_3, V_4)$$

$$V_5 = \Pi(e_1 / I_2) \bar{B}_1 + \Pi(e_2 / I_2) \bar{B}_2 + \sum_{t=1}^{k-1} \alpha^t \bar{C}_t - I$$

$$V_6 = \Pi(e_1 / I_2) B_1 + \Pi(e_2 / I_2) B_2 + \sum_{t=1}^{k-1} \alpha^t C_t - I'$$

$$V_9 = \max(V_5, V_6)$$

$$V_{10} = P(I_1) \cdot V_8 + P(I_2) \cdot V_9$$

$$\text{enfin } V = \max(V_7, V_{10}).$$

Le prix limite que le décideur est prêt à payer pour le supplément informationnel se déduit en comparant V_7 et V_{10} . Si le recours à l'expert permet d'espérer un gain substantiel, le prix limite n'est rien d'autre que la différence entre V_{10} et V_7 . Le prix limite p se définit alors de la manière suivante :

$$p = \max(0, V_{10} - V_7).$$

Dans la solution que nous proposons pour résoudre ce problème, les hypothèses du critère évoqué précédemment sont reconduites (irréversibilité et croissance de l'information). La modification majeure provient du fait que désormais la première décision de l'entrepreneur concerne l'évaluation de la valeur des services de l'expert afin d'accepter ou de rejeter l'achat d'un supplément informationnel.

Une propriété tout à fait intéressante peut être dérivée de ce cadre décisionnel. Il s'agit du lien entre la valeur de l'information accordée par le décideur et le degré d'incertitude dans lequel il se trouve. Les résultats obtenus, par ailleurs, dans ce domaine, montrent la présence d'une forte complémentarité entre le niveau d'incertitude et la valeur de l'information. Cette relation n'est plus valable dans un contexte *pre posteriori* de la décision d'investissement où le prix que le décideur est prêt à payer pour obtenir l'information est indépendant du degré d'incertitude dans lequel il se trouve au départ. La proposition suivante peut être évoquée (voir G. Lambert [1989]) :

Proposition : *Lors d'une procédure de décision pre posteriori d'investissement, la valeur de l'information supplémentaire est indépendante du degré d'incertitude dans lequel se trouve le décideur ; elle dépend exclusivement du caractère discriminant de cette information, c'est-à-dire de sa capacité à faire changer d'avis le décideur.*

Le modèle décrit ici permet aux entreprises d'avoir un critère de décision intégrant l'étape préalable à la décision d'investissement qui correspond à la recherche d'information. L'acquisition du bien informationnel est alors analysée comme un investissement immatériel réalisé par la firme dans le but de choisir la meilleure décision d'investissement. Le couplage de l'effet irréversibilité et de l'évaluation du paramètre informationnel renforce l'incidence de la

procédure d'investissement relativement à l'aspect strictement financier des critères de gestion classiques (valeur actuelle nette) puisque le choix du décideur ne porte plus exclusivement sur les résultats d'un projet d'investissement, mais également sur les éléments constitutifs de ce choix.

Enfin, le glissement du cadre a posteriori à un cadre pre posteriori de la décision, caractérisé ici par le recours à l'expert, débouche sur un statut différent du bien informationnel. Lorsque l'information est naturellement révélée par la nature, la valeur que le décideur accorde à celle-ci est fortement dépendante du paramètre d'incertitude. En revanche, cette dépendance n'est plus valable si l'information relative aux états du monde est révélée par un expert. C'est alors le caractère dissuasif de cette information vis-à-vis des différentes options qui va lui procurer de la valeur. Ainsi, l'apport informationnel du service proposé par l'expert sera d'autant plus important pour le décideur industriel que ses croyances initiales (caractérisées par les probabilités subjectives sur chaque état) sont proches du paramètre d'incertitude qui représente le point d'indifférenciation entre les projets d'investissement (notion de taux pivot d'I. Fischer [1930]).

2. La dimension organisationnelle de l'investissement

L'évocation de phénomènes d'apprentissage dans le domaine de la production industrielle n'est pas un fait nouveau. L'observation des courbes d'apprentissage dans différents secteurs de l'industrie (W.B. Hirschmann [1964]) permettait de mettre en évidence un lien entre le volume d'investissement et les coûts de production. Dans ces travaux, qui reposent sur des mesures empiriques dans différents secteurs de l'industrie américaine, l'auteur montre l'existence d'économies d'échelle se traduisant par des coûts de production moindres et provenant d'un simple accroissement du volume de la production. Si l'effet mécanique de ces phénomènes d'apprentissage rencontre une limite au-delà de laquelle une augmentation du volume de la production va accroître le coût unitaire de production, W.J. Abernathy - K. Wayne [1974] font ressortir d'autres limites comme, par exemple, la perte de flexibilité qui peut entraîner la recherche d'économies d'échelle pour l'entreprise. L'arbitrage entre la flexibilité de la firme et la recherche d'économies d'échelle (provenant de phénomènes d'apprentissage générés dans une structure industrielle fixe) est présent dans l'outil d'aide à la décision, que nous proposons dans notre première partie. Dans un tel critère, l'observation ex post dont faisait l'objet ce type de phénomène était alors directement intégrée dans les plans du manager, c'est-à-dire au niveau du calcul ex ante de la rentabilité économique d'un projet.

2.1. Choix technologiques et apprentissage organisationnel

La nature de l'apprentissage décrit au paragraphe précédent, correspond à ce que R.M. Cyert - J.G. March [1963] appellent l'apprentissage adaptatif, qui consiste en une meilleure adaptabilité de la firme dans une structure industrielle stable. Plus récemment, certains travaux effectués en sociologie des organisations (J. Child [1972]) vont dans le même sens, en mettant en exergue la capacité d'adaptation des organisations face à des changements de l'environnement. Analytiquement, le comportement sous-jacent à ceci repose sur la description d'une interaction entre un décideur et un environnement purement exogène caractérisé par un

ensemble d'états possibles. Dans cette optique, le décideur gère un système en supposant qu'il n'exerce aucune influence sur les variables d'état. La nature de l'information relative à ces états est donc purement exogène. En introduisant la possibilité d'une action possible de l'agent sur certaines variables d'état du système, nous pouvons déboucher sur des règles de comportement rationnel radicalement opposé à celui décrit préalablement, dans la mesure où l'investissement n'est plus vu comme une perte d'option, mais il peut être lui-même créateur d'options technologiques nouvelles.

De ce point de vue, si notre première approche des phénomènes d'adaptation revient à intégrer au moment du calcul d'investissement (dans la mesure du possible), l'effet d'apprentissage relatif au degré de flexibilité véhiculé par l'investissement, nous voudrions, ici, aller plus loin encore dans cette voie. Nous nous attacherons, en effet, dans cette partie, à élargir la prise en considération de tels phénomènes en ne les limitant plus à une structure productive donnée, mais concernant cette fois-ci des choix pouvant entraîner des changements de structure majeurs du système productif. Le problème que pose alors le calcul d'investissement est lié à l'extrapolation ex ante de tels changements pouvant intervenir dans les structures de production. A l'aide d'un exemple, nous montrerons qu'il est possible sous certaines conditions d'incorporer ces changements de structure dans le calcul de rentabilité ex ante, et ainsi se rapprocher d'une analyse de l'impact réel de l'investissement sur la firme.

Ce dernier aspect soulève le délicat problème des limites du champ que doit couvrir le calcul d'investissement dans la mesure où la technologie crée des solutions pouvant se diffuser dans l'entreprise, ou encore provoque des mutations irréversibles. En effet, l'impact d'un projet d'investissement peut être déterminé à différents niveaux. M.F. Cantley- D. Sahal [1980] distinguent cinq niveaux d'impact : celui de l'équipement individuel, du process de production, de la compagnie, de l'industrie et de la société en général. Il est vrai que l'impact de certaines décisions d'investissement dépasse le cadre de l'équipement et touche au fonctionnement du process lui-même. Il semble évident que la rentabilité d'un projet qui résulte de ce découpage peut s'avérer très différente. Le calcul économique se doit de tenir compte, dans la limite de ce qui est quantifiable, de ces changements de structure pouvant survenir à un niveau plus général, suite à un investissement unitaire. Nous illustrerons ceci dans l'exemple qui suit où nous montrerons qu'un investissement unitaire, situé en amont du process de production, peut déboucher sur un processus d'intégration au niveau du process.

2.2. Critère d'investissement et apprentissage technologique : application à l'investissement énergétique

Lorsque l'agent peut, par son action, influencer certaines variables de l'environnement économique, un comportement possible du point de vue de l'objectif de rentabilité peut se ramener à une spécialisation ponctuelle sur des paramètres de l'investissement susceptibles d'être améliorés. Ce comportement s'oppose alors à la recherche de flexibilité évoquée précédemment dans la mesure où, pour un investissement énergétique, il privilégierait l'amélioration du rendement énergétique plutôt que le choix d'un équipement flexible fonctionnant avec l'énergie la

plus avantageuse du moment. Le décideur effectue alors une sélection parmi ses différentes options technologiques (qui correspondent à des couples énergie-technologie) et l'objectif qu'il poursuit alors ne dépend plus que de l'état des variables qu'il contrôle pour ses actions. Pour revenir à l'investissement énergétique, le décideur n'ayant aucune action possible sur le prix des différentes énergies, va placer au centre de ces objectifs l'amélioration des rendements énergétiques de l'équipement. De ce fait, il endogénéise l'incertitude qui dépend alors de sa capacité à rationaliser le fonctionnement de sa technologie. Il s'affranchit par là même de l'aléa principal lié aux variations du prix des énergies. La capacité du décideur à engendrer du progrès technique revient à placer au centre de la décision d'investissement l'apprentissage technologique et relève d'une forme de rationalité procédurale plutôt que substantielle.

L'exemple industriel qui fonde notre raisonnement est tiré d'une étude réalisée par le BETA portant sur l'investissement énergétique dans la Chimie [BETA, 1984]. Observant que le choix de l'équipement énergétique relevait plutôt de la sélection de l'énergie (électricité, gaz, fuel, charbon) la moins onéreuse sur la base d'un critère de sélection tourné vers le strict examen du bilan énergétique (notion de délai de récupération de l'investissement), la rationalité procédurale nous suggérait d'autres types de stratégies industrielles (G. Lambert [1987] et [1988, Partie III]). En effet, de ce point de vue, les énergies ne sont pas équivalentes car leurs potentialités d'évolution technologique respectives qu'elles renferment diffèrent. Les technologies basées sur l'électricité offrent notamment des possibilités de décentraliser la production de thermies nécessaires au fonctionnement de la production, par emploi de systèmes de chauffage direct du produit. La réduction des réseaux de vapeur, vecteur classique de transfert de la chaleur dans l'entreprise, permet l'élimination de pertes énergétiques et ouvre des possibilités nouvelles d'automatisation des process. Ce progrès technique peut s'expliquer en terme de **processus d'intégration** de la source d'énergie au process de production. Si cette intégration est parfaite, le système technique peut évoluer idéalement vers un optimum connu, proche de l'efficacité de Carnot, définie selon le second principe de la thermodynamique. Le choix ne relève plus alors d'une sélection de l'énergie dont le prix est le plus intéressant, mais de celle susceptible de conduire le système technique au plus près de l'optimum énergétique défini par les principes de thermodynamique. Le paramètre d'incertitude devient donc purement **endogène** puisqu'il va dépendre des efforts déployés par l'industriel en vue de réaliser cette intégration. Dans ce processus dynamique, le choix de l'énergie n'est pas neutre dans la mesure où l'intégration de l'énergie au process est une opération délicate, voire impossible, avec les différentes énergies fossiles (gaz, fuel et charbon).

Une procédure d'ingénierie dans le but de réaliser cette intégration couplée au choix de l'énergie qui pourrait engendrer les phénomènes d'apprentissage les plus prometteurs de ce point de vue (énergie électrique), nous permet d'aboutir à un critère de sélection. Celui-ci (G. Lambert [1987]) fait appel aux mêmes bases théoriques que précédemment fondées sur un critère bayésien avec irréversibilité des décisions, mais il intègre de surcroît un processus d'apprentissage technologique probabilisable par l'agent (structure croissante d'information) dépendant à la fois du choix de l'énergie et des efforts d'ingénierie déployés par l'industriel (efforts mesurables ex ante par le montant de R&D consenti).

Formalisation du critère: Soit un horizon de production à deux périodes ; une période présente dont les paramètres sont parfaitement connus et une période future comportant quelques paramètres incertains. L'horizon est supposé fini, excluant ainsi toutes possibilités d'investissement au delà de la seconde période. En début d'horizon, l'entrepreneur, supposé neutre vis-à-vis du risque, se trouve face à la nécessité d'effectuer un investissement énergétique de remplacement ou de modernisation. L'équipement a atteint un seuil d'obsolescence critique qui se traduit par une chute significative des rendements énergétiques. Parallèlement à ce type d'équipement fonctionnant avec une énergie fossile (noté f), sont disponibles sur le marché des équipements à énergie électrique (notés e).

Le prix futur de l'énergie fossile est incertain, tandis que, le prix de l'énergie nouvelle est supposé constant sur toute la période. Nous considérons qu'à la seconde période un changement du prix de l'énergie fossile peut avoir lieu. Deux états du monde sont alors possibles :

- l'état i dans lequel le prix de l'énergie fossile (p_f) reste inchangé,
- l'état j dans lequel le prix de cette énergie (p'_f) augmente de façon importante.

Ces deux états du monde sont affectés par le décideur de probabilités subjectives telles que :

$$\begin{aligned} \text{prob}(i) &= \Pi \text{ (maintien du statu-quo énergétique)} \\ \text{prob}(j) &= 1 - \Pi \text{ (renchérissement de l'énergie fossile).} \end{aligned}$$

La structure d'information est supposée parfaite, ce qui se traduit par le fait qu'à la fin de la première période, l'entrepreneur connaît parfaitement l'état du monde qui se réalisera à la seconde période. Le prix de l'énergie électrique est tel que :

$$p_f < p_e < p'_f$$

Plusieurs possibilités d'action s'offrent à l'entrepreneur.

- La première solution (d) consiste à acheter en début d'horizon un équipement neuf à énergie fossile, et rend impossible toute possibilité de passage ultérieur à l'énergie nouvelle durant la seconde période. Cette irréversibilité qui fige les choix sur l'ensemble des deux périodes provient de l'existence d'une contrainte de financement qui limite le nombre d'investissements pouvant être réalisés. Elle est renforcée par une absence de marchés de l'occasion pour les équipements productifs.

- La seconde solution (d) consiste à différer l'investissement en connaissance de l'information sur l'état du monde. Toutefois, pour maintenir son activité à un niveau de rentabilité suffisant avec son ancienne technique, l'industriel réalise un investissement de modernisation dont le montant est bien inférieur à celui nécessaire pour l'achat d'équipement neuf.

Nous raisonnerons en termes de bénéfices espérés actualisés dégagés à chaque période. L'incidence de la manière dont est financé l'investissement, ainsi que les effets de l'amortissement sur l'assiette de l'impôt, sont supposés résolus dans la valeur du bénéfice espéré actualisé qui représente un bénéfice net après impôt.

L'achat en début d'horizon d'un nouvel équipement à énergie fossile engendre un bénéfice espéré actualisé (noté r_f^N) à la première période. L'équipement neuf ayant en toute logique un meilleur rendement que l'ancien, l'opération de réfection de l'équipement existant procure à l'industriel un bénéfice noté r_f tel que, sur tout l'horizon, l'on ait :

$$r_f^N > r_f$$

L'investissement dans l'énergie électrique procure un bénéfice noté r_e tel que :

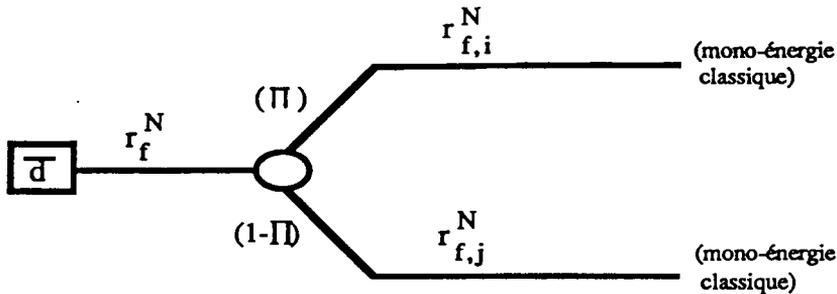
$$r_f^N > r_f > r_e$$

Etant donnés les prix relatifs entre les différentes énergies, le passage à l'énergie électrique ne sera pas envisagé avant la seconde période (en l'absence d'un délai d'installation important).

À la seconde période, les états du monde influencent la valeur de r_f^N de telle sorte que :

$$r_{f,i}^N = r_f^N > r_f > r_e > r_{f,j}^N$$

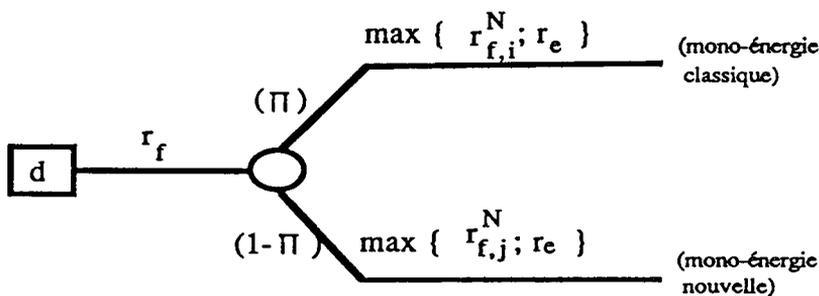
La détermination du gain total espéré (noté B) est obtenue en faisant la somme des bénéfices dégagés sur chacune des périodes. Pour la décision irréversible d , nous pouvons tracer l'arbre de décision suivant :



Le gain total espéré s'écrit :

$$\overline{B(d)} = r_f^N + (\Pi) r_{f,i}^N + (1-\Pi) r_{f,j}^N$$

Pour la décision flexible notée d nous avons :



Le gain total espéré pour la solution flexible se note :

$$B(d) = r_f + (\Pi) r_{f,i}^N + (1-\Pi) r_e$$

Pour être en mesure de trancher entre les deux solutions qui se présentent à l'entrepreneur, nous pouvons utiliser le concept de valeur d'option. Elle est alors obtenue si le décideur choisit en début d'horizon une décision irréversible d'investissement. La valeur d'option est égale à la valeur maximale du regret que pourrait avoir l'entrepreneur dans la seconde période s'il a pris une décision irréversible en début d'horizon. Elle s'écrit :

$$V = (1-\Pi) [r_e - r_{f,j}^N]$$

Cette valeur, toujours positive, correspond à la différence entre les deux branches inférieures (état j) des arbres de décision. Pour que la solution d soit choisie, il faut que cette valeur soit supérieure aux gains certains que l'entrepreneur ferait s'il avait opté pour la solution irréversible en $t = 0$.

La valeur de ces gains s'écrit :

$$r_f^N - r_f$$

Ils représentent le coût de la flexibilité, c'est-à-dire la perte certaine que le décideur est prêt à subir pour conserver la possibilité de choisir en connaissance des états du monde. L'entrepreneur choisira de reporter l'investissement de remplacement si :

$$V = (1-\Pi) [r_e - r_{f,j}^N] > r_f^N - r_f$$

Nous introduisons ici une hypothèse supplémentaire de progrès technique possible pour la technologie à énergie électrique. Ce progrès technique consiste en une intégration de l'énergie au processus. Il est supposé être le

résultat d'une activité de recherche et développement interne à l'entreprise et donne lieu à une incertitude endogène probabilisée de la manière suivante :

- Etat g : $\text{prob}(g) = \rho_\alpha(\alpha, I)$
 où α : durée d'utilisation annuelle de la technologie à énergie électrique ($0 \leq \alpha \leq 1$) ;
 I : dépense en recherche et développement .

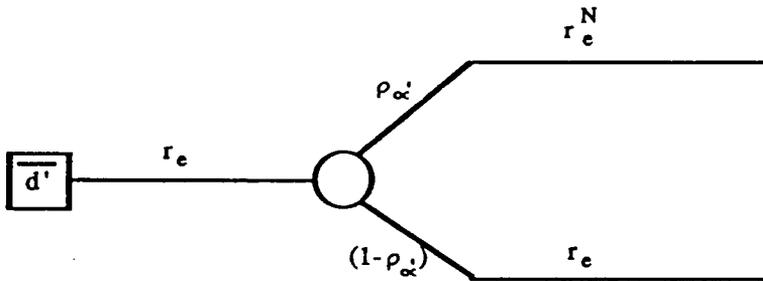
ρ_α représente la probabilité de réussite du processus d'intégration, elle est une fonction croissante de α et de I . Remarquons que l'activité de recherche développement pourrait être réalisée par une société d'ingénierie externe à l'entreprise. Dans ce cas alors la probabilité de réussite ρ correspondrait à une incertitude exogène à l'entreprise.

$$\begin{aligned} d\rho_\alpha / d\alpha &> 0 \\ d\rho_\alpha / dI &> 0 \end{aligned}$$

- Etat h : $\text{prob}(h) = 1 - \rho_\alpha$

C'est la probabilité complémentaire de ρ_α qui caractérise l'échec du processus de recherche - développement pour l'intégration de l'énergie au process.

Relativement aux stratégies évoquées précédemment, le décideur peut réaliser un investissement irréversible en début d'horizon en investissant dans l'énergie électrique qu'il utilise sur toute l'année ($\alpha' = 1$). Dans une telle stratégie, la probabilité de réussite de l'intégration sera supérieure à une solution de type bi-énergie ou le processus de recherche et développement ne peut fonctionner que pendant la période tarifaire d'été. L'arbre de décision sera le suivant :



Le bénéfice d'une telle décision s'écrit :

$$B(d') = r_e - I + \rho_{\alpha'} \cdot r_e^N + (1 - \rho_{\alpha'}) r_e$$

Supposons, conformément aux hypothèses faites sur les résultats de l'intégration et le niveau des prix relatifs, que :

$$r_e^N > r_e^E > r_{f,i}^N > r_e > r_{f,j}^N$$

La condition de rentabilité exige que cette valeur dépasse la dépense initiale engagée par l'entreprise dans l'activité de recherche-développement.

Si la décision d est optimale sans progrès technique, la décision d' peut ici dominer si : $B(d') > B(d)$

ou encore :

$$\rho_{\alpha'} r_e^N + (\Pi - \rho_{\alpha'}) r_e - \Pi r_{f,i}^N > r_f - (r_e - I)$$

Le membre gauche de cette inégalité correspond au concept de valeur d'option qui mesure la perte anticipée d'avoir choisi l'équipement flexible et non l'équipement mono-énergie électrique pour réaliser le processus d'intégration. Cette valeur d'option est d'un intérêt non négligeable sur le plan théorique car contrairement au modèle précédent, elle vient cette fois créditer une décision irréversible d'investissement.

Il nous semble que les résultats auxquels nous parvenons sont intéressants dans la

mesure où ils montrent qu'une réponse optimale à l'incertitude exogène sur les prix des énergies peut ne plus l'être si l'on considère la capacité d'évolution de la technologie. En effet, une pure stratégie d'adaptation aux aléas de l'environnement de la firme met en exergue des stratégies flexibles qui conduisent à des comportements de report d'investissement (flexibilité financière) ou de bi-énergie (flexibilité des procédés). Si l'on ajoute le paramètre informationnel endogène du progrès technique, un investissement mono-énergie, rigide relativement à l'incertitude sur les prix de l'énergie, peut s'avérer être le plus intéressant. Ce résultat s'explique par le fait que la solution technique adaptée renferme une plus grande flexibilité entre les différents états que peut prendre la technologie (dont l'étape ultime est connue ici et correspond au second principe de la thermodynamique), en vue d'une solution parfaitement intégrée. Avec de telles considérations, la procédure d'investissement devient primordiale puisqu'elle permet de positionner le système technique sur une trajectoire optimale tendant à minimiser les coûts de production à long terme conformément à une rationalité de type procédurale.

2.3. La dimension organisationnelle de l'investissement : exemples d'investissements en rationalité procédurale

Depuis la dernière décennie le rôle des actifs incorporels de la firme, et plus particulièrement ceux relatifs à la constitution et l'organisation d'une équipe de recherche et développement, s'est considérablement accru. Dans bon nombre de secteurs d'activité la part des dépenses en investissements immatériels est devenue comparable à celle vouée aux investissements matériels. Le problème alors posé sur le plan du calcul économique consiste en sa mesure, et plus spécifiquement dans l'évaluation de son impact sur les résultats de l'entreprise. En effet, la difficulté provient du fait que la capacité technologique et organisationnelle nécessaire au maintien sur les périodes futures des compétences technologiques de la firme dans sa spécialité, correspond à une variable de fonds non identifiable à une simple variable de stock. Elle s'en distingue, en effet, par le temps qui est nécessaire à sa constitution, son caractère indivisible (notion de taille critique), et la nécessité de l'entretenir par un flux de maintenance afin de conserver son efficacité.

Plusieurs approches financières ont tenté de mieux cerner cette notion d'investissement immatériel. D. Pene[1979], dans la lignée des travaux de R. Coase[1937], l'assimile à la notion de "Goodwill". Celle-ci correspond à la différence des rendements de l'entreprise et la valeur des capitaux nécessaires s'il fallait reconstituer la firme ex nihilo. La valeur de l'investissement immatériel ainsi défini reflète de façon indifférenciée l'image de marque de la firme sur laquelle se fonde l'attachement de la clientèle, le savoir technologique, et la valeur organisationnelle qui caractérise la firme. Les coûts supplémentaires appelés "coûts de transaction" qui proviennent de la mise en relation d'entités différentes justifient la création d'entreprises au milieu des mécanismes purs de transactions marchandes (R. Coase [1960], O.E. Williamson [1979], D. Teece [1980, 1982]). Avec cette analyse, l'organisation, qui caractérise la firme et qui incorpore différentes fonctions présentent sur le marché, représente une source d'économies (voir l'analyse de P. Cohendet & P. Llerena ci-dessus). L'organisation est alors un moyen d'internaliser et d'intégrer les transactions ayant normalement lieu sur le marché entre les différentes fonctions de

l'entreprise (marketing, département R&D,...).

E. Sage [1977] estime dans la même optique la valeur des brevets et du potentiel de recherche de la firme par la quantité de salaires qui serait nécessaire à la reconstitution de ce patrimoine technologique. Plus récemment à propos de l'investissement immatériel, A. Marion [1987] propose la notion de "rentabilité fonctionnelle" qui est atteinte lorsque l'entreprise a la capacité de maintenir, voire de renforcer, les fonds de savoir-faire nécessaires à la pérennité de son activité. La dimension organisationnelle joue un rôle essentiel dans cette notion dynamique de rentabilité. Toutefois, le problème de la mesure d'un tel investissement est loin d'être résolu.

L'ingénierie simultanée nous fournit un exemple intéressant en ce qui concerne la dimension organisationnelle de l'investissement. Cette stratégie industrielle correspond à un contrat de collaboration passé entre un producteur et un fournisseur d'équipement en vue d'élaborer une solution commune. Ce concept a été mis en place par COMAU (concepteur de système de production italien) pour l'assemblage automatique des chaînes de montage automobile. Cette démarche repose sur la décision d'association avec un concepteur de procédés de production le plus tôt possible, c'est-à-dire avant que l'objet technique à réaliser soit entièrement défini par le producteur.

Le but recherché est d'adapter l'objet à fabriquer, dès sa phase de conception, aux contraintes de la production automatisée seules réellement connues de l'équipementier. Relativement à une procédure classique d'appel d'offre de la part du constructeur en vue de sélectionner l'équipementier qui proposerait la solution la meilleure, l'ingénierie simultanée consiste à choisir une collaboration technologique précoce, avant même que la solution technique ne soit élaborée. L'investisseur accepte donc une perte de flexibilité vis-à-vis du choix de l'équipementier et parie sur une activité d'ingénierie commune permettant d'aboutir à la solution technologique la plus satisfaisante. Le développement simultané d'une pièce et d'un outil doit permettre de mieux adapter le produit fabriqué à la production automatisée, et ainsi, d'économiser du temps de développement pour l'équipementier.

L'application de cette démarche à la construction du moteur du modèle Fiat Uno aurait permis, selon le constructeur automobile, de réaliser un gain de temps d'environ 10 mois, alors qu'un déroulement classique aboutit à la réalisation de l'unité de production au bout d'une période de gestation estimée à 46 mois. Nul doute que, dans ce secteur industriel, le gain de temps évoqué provenant d'un processus interactif d'apprentissage constitue un avantage comparatif majeur. Cet exemple nous montre que la flexibilité n'est pas uniquement un attribut du bien d'investissement, mais peut résulter de la dimension organisationnelle qui fonde la procédure d'investissement : *"L'organisation est analysée systématiquement en termes d'apprentissage, individuel ou collectif, pour des (groupes d') individus-types, procéduralement rationnels"* (O. Favereau, 1989, p.340).

Si ce choix ne relève pas d'un calcul économique explicite de la part du décideur, il est intéressant de noter que le motif qui guide un tel comportement correspond à la prise en

considération d'un processus d'apprentissage généré par un arrangement contractuel entre producteur et équipementier. La procédure décisionnelle ne consiste plus à investir pour du progrès technique exogène mais, à mettre en place dans l'entreprise des mécanismes qui le génère. Cette endogénéisation du paramètre d'incertitude place au tout premier plan la façon dont la décision d'investissement se prend (partenariat technologique) en pariant sur des processus d'apprentissage, conformément aux principes de la rationalité procédurale. Paradoxalement, l'irréversibilité dans un tel contexte perd la connotation négative acquise face à un environnement marqué par l'instabilité, pour devenir une source de restructuration de la firme permettant de répondre aux perturbations technico-économiques. L'adaptation ne se fait plus "passivement" par la gestion de la variété, mais de manière active par la mise en place de conditions propices à l'apprentissage.

3. Conclusion

L'élargissement du champ de la rationalité économique des décideurs permet d'intégrer deux caractéristiques majeures des investissements pourtant absentes des critères de choix en matière d'investissement. Il s'agit, d'une part, de rendre sensibles les critères de sélection aux **effets d'irréversibilité** qu'entraînent certaines décisions d'investissement et, d'autre part, de ne pas occulter la **dimension organisationnelle** (ou organisatrice) de l'investissement qui peut générer des apprentissages technologiques pouvant déserrer l'emprise des variables environnementales sur la firme. Ces deux axes de développement, nous les traitons à l'aide du concept simonien de rationalité procédurale.

D'autres formes d'apprentissage organisationnel ont été répertoriées. Dans sa typologie des systèmes d'apprentissage organisationnel, P. Shrivastava [1983] en dénombre quatre formes possibles. Deux d'entre elles au moins recourent les processus d'adaptation décrits ici, à savoir l'apprentissage organisationnel en tant qu'**adaptation au changement** et qu'**institutionnalisation des effets d'expérience**. Les deux autres formes d'apprentissage organisationnel correspondent à la **modification des présupposés communs** et au **développement des connaissances sur la base des relations entre actions et résultats**. Si ces formes n'ont pas été directement étudiées ici, il nous semble qu'un critère bayésien avec une structure d'information croissante les intègre déjà au niveau du calcul économique. En effet, ces deux types d'apprentissage se situent à la base de la constitution d'une structure d'information initiale (expérience passée) et de son reparamétrage dans le temps en fonction des résultats obtenus (probabilités conditionnelles). Ils ont fait l'objet de développement plus poussés en ce qui concerne le fonctionnement de la mémoire humaine à travers la notion d'auto-organisation (voir B. Ancori ci-dessus). L'apprentissage organisationnel aboutissant à la révision des présupposés communs qui fondent les principes de toute organisation est présent dans la théorie de l'action (C. Argyris - D.A. Schon [1978]), comme dans le modèle de promulgation - sélection - maintien (K. Weick [1969]), dans lesquels l'adaptation nécessite parfois de l'individu qu'il bouleverse les règles, invente de nouvelles stratégies dans le but de pérenniser l'organisation. Du point de vue des problèmes de gestion, ces changements majeurs ayant lieu de façon conjoncturelle dans les orientations de la firme

sortent du cadre du calcul de rentabilité des investissements proprement dit, et donc de notre champ d'investigation. Parallèlement, l'organisation peut être vue comme un lieu de transformation (des inputs en outputs) qui génère un processus continu d'apprentissage organisationnel (R.L.M. Dunbar et al. [1982], J. Dutton - R. B. Duncan [1981]). Chaque décision dans la firme est simultanément analysée et profite à la connaissance globale de l'organisation. A travers cette vision, l'organisation industrielle est idéalement perçue comme un champ d'expériences et, est capable, sur la base d'actions/résultats, de faire face à toutes les situations. Or, nous nous apercevons qu'il faut souvent du temps pour que la firme puisse caractériser les facteurs de changement survenus dans son environnement, et que l'inertie des structures en place renforce l'incapacité de celle-ci à s'adapter. Ce problème s'observe de façon notoire en gestion lorsqu'il s'agit de mettre en place une nouvelle comptabilité analytique reposant sur un découpage différent de la firme (voir V. Giard [1988]). Il est présent en ce qui concerne la mise en place d'une gestion à flux tendus, ou encore dans la prise en compte dans les calculs de rentabilité de l'investissement immatériel (voir A. Marion [1987]).

Les développements autour des notions d'apprentissages adaptatif et technologique dans le cadre du calcul économique devraient permettre de jeter une passerelle entre praticiens et théoriciens des investissements dans la mesure où les dimensions auxquelles sont sensibles les premiers (irréversibilité, apprentissage technologique) entrent dans le processus d'évaluation au même titre que les arguments financiers qui fondent les critères de décision. Les réflexions autour de la rationalité procédurale pour l'évaluation des investissements devraient venir combler le vide théorique existant entre les critères d'extrême prudence, souvent employés par les décideurs et les critères actuariels centrés autour de la notion de valeur actuelle d'un projet. Ces procédures conditionnent déjà fortement les prises de décision, il nous incombe désormais de les traduire au niveau du calcul économique pour leur donner du sens et une rationalité.

BIBLIOGRAPHIE

ABERNATHY W.J. - WAYNE K. (1974), "Limits of the learning curve", *Harvard Business Review*, September-October, n°5.

ALLAIS M. (1953), "Le comportement de l'homme rationnel devant le risque : critique des postulats de l'Ecole américaine", *Econometrica*, 21, pp. 503-546

ALLAIS M. (1984), "The foundations of the theory of utility and risk", dans O. HAGEN - F. WENSTOP eds., *Progress in utility and risk theory*, D. Reidel, Dordrecht/Boston.

ALLAIS M. (1986), "The general theory of random choices in relation to the invariant cardinal utility function and the specific probability function", *Centre d'Analyse Economique*, Paris.

ANDERSON R.J.Jr. (1981), "A note on option value and the expected value of consumer's surplus", *Journal of Environmental Economics and Management*, Vol.8, pp.187-191.

ARGYRIS C. - SCHÖN D.A. (1978), "*Organizational Learning : A Theory of Action Perspective*", Reading, Mass. : Addison- Wesley.

ARROW K.J. (1970), "*Exposition of the theory of choice under uncertainty* ", in ARROW K. J., Chap 2, pp. 44-89.

BELL D. (1982), "Regret in decision making under uncertainty", *Operations Research*, 30, pp.961-981.

BELL D. (1985), " Disappointment in decision making under uncertainty ", *Operations Research*, 33, pp. 1-27.

BENARD J. (1989), "Calcul économique public : conflit des paradigmes et résistance des organisations et des mentalités", *Revue d'Economie Politique*, n°2, pp.307-321.

BERNANKE B.S. (1983), "Irreversibility, uncertainty and cyclical investment", *Quarterly Journal of Economics*, Vol.98, n°1.

B.E.T.A. (1974), "*Chimie et Electricité*", rapport réalisé pour le Ministère de l'Industrie, EdF et cinq groupes de l'Industrie Chimique Française.

BISHOP R.C. (1982), "Option value : an exposition and extension", *Land Economics*, Vol.58,n°1, pp.3-15.

BOHM P. (1975), "Option demand and consumer's surplus : comment", *The American Economic Review*, Vol.65,n°4,pp.733-736.

CANTLEY M.F. - SAHAL D. (1980), "*Who learns what ? A conceptual description of capability and learning in technological systems*", R.R.-80-42 ; International Institute for Applied Systems Analysis, Austria.

CHARREAUX G. (1980), "Critères de choix des investissements : une clarification des critères de la valeur nette actualisée et du taux interne de rentabilité", *Revue Française de Gestion*, 1980.

CHILD J. (1972), "Organizational structure, environment and performance. The role of strategic choice", *Sociology*, n°6.

CICCHETTI C.J. - FREEMAN A.M. III. (1971), "Option demand and consumer surplus : further comment", *Quarterly Journal of Economics*, n°85.

COASE R. (1937), "*The nature of the firm* ", Edt. Economica.

COASE R. (1960), "*The problem of social cost*", *Journal of Law and Economics*, octobre.

COHENDET P. (1983), "*Le concept de temps dans l'oeuvre de K. ARROW* ", miméo BETA, Programme STS.

CYERT R.M. - MARCH J.G. (1963), "*A Behavioral Theory of the Firm*", Englewood Cliffs, N.J. : Prentice-Hall.

DASGUPTA A.K. - PEARCE D.W. (1972), "*Cost Benefit Analysis : Theory and Practice*", McMillan.

DUNBAR R.L.M. - DUTTON J.M. - TORBERT W. R. (1982), "Crossing mother; ideological constraints on organizational improvement", *Journal of Management Studies*, Vol.19,pp.91-108.

DUTTON J. - DUNCAN R. B. (1981), "*The process and threats to sensemaking and their relationship to organizational learning*", Working paper, Kellogg Graduate School of Management, Northwestern University.

- FAVEREAU O. (1989), "Vers un calcul économique organisationnel ?", *Revue d'Economie Politique*.
- FELLNER W. (1961), "Distortion of subjective probabilities as a reaction to uncertainty", *Quarterly Journal of Economics*, pp.670-689.
- FISHER I. (1930), "*Theory of interest*", MacMillan.
- FREEMAN A.M.III. (1979), "*The benefits of environmental improvement*", John Hopkins University Press, Baltimore.
- GIARD V. (1988), "Gestion de production : évaluation économique et prise de décision", *Revue Française de Gestion*.
- HENRY C. (1974 a), "Investment decisions under uncertainty : the irreversibility effect", *The American Economic Review*.
- HENRY C. (1974 b), "Option value in the economics of irreplaceable assets", *Review of Economic Studies*.
- HICKS J.R. (1940), "*The valuation of social income*", *Economica*.
- HIRSCHMANN W.B. (1964), "Profit from the learning curve", *Harvard Business Review*.
- KALDOR N. (1969), "Welfare comparisons of economics and interpersonal comparisons of utility", in K. ARROW and T. SCITOVSKY (eds), *Readings in Welfare Economics*, London, pp.387-389.
- LAMBERT G. (1987), "Modèle de décision avec information endogène : application aux investissements énergétiques", BETA/CNRS, "*L'évolution économique et théories de l'information*".
- LAMBERT G. (1988), "Choix d'investissements : un nouvel outil de décision", *Revue Française de Gestion*, n°68.
- LAMBERT G. (1989), "*Une analyse bayésienne 'pre posteriori' de la décision d'investissement*", working papers BETA, n°8906.
- LLERENA P. (1985), "*Décision avec incertitude et irréversibilité: Fondements de la théorie de la valeur d'option et application aux investissements productifs*", Thèse de Doctorat d'Etat, ULP Strasbourg.
- MACHINA M.J. (1983), "Generalised expected utility analysis and the nature of observed violations of the independence axiom", in STIMUM et WENSTOP (eds), *Foundations of Utility and Risk Theory with Applications*, pp. 263-293.
- MARION A. (1987), "Problématique financière de l'investissement immatériel", *Analyse Financière*.
- MASSE P. (1968), "*Le choix des investissements*", Dunod.
- MISHAN E. J. (1988), "*Cost benefit analysis : an informal introduction*", Unwin Hyman Ltd, London
- MODIGLIANI F. - MILLER R. (1958), "The cost of capital, Corporation Finance and the Theory of Investment", *The American Economic Review*.
- PENE E. (1979), "*Valeur et regroupements des entreprises*", Edt Dalloz.
- QUINTART A.- ZISSWILLER R. (1982), "*Investissements et désinvestissements de l'entreprise. Pratiques et méthodes*", Dalloz.
- RICHARD A. (1982), "Eléments de synthèse entre valeur nette actualisée et délai de récupération : l' "Effet Irréversibilité" ", *Revue d'Economie Politique*, n°1.

RUNGE C.F. (1984), "Energy exploration on wilderness : 'privatization' and public lands management", *Land Economics*, Vol.60, n°1, pp.56-68.

SAGE E. (1977), "*Comment évaluer une entreprise*", Edt Sirey.

S AVAGE L.J. (1954), "*The foundations of statistics* ", New York, John Willey and Sons.

SCHMALENSEE R. (1975), "Option demand and consumer's surplus : reply", *The American Economic Revue*, Vol.65., n°4, pp.736-737.

SHACKLE G.L.S. (1949) " A non additive measure of uncertainty ", *Review of Economic Studies*, 17, pp.70-74.

SHAFFER G. (1976), "*A mathematical theory of evidence* ", Princeton University Press.

SHRIVASTAVA P. (1983), "A typology of organizational learning systems", *Journal of Management Studies*, n°20.

SIMON H. (1976), "From Substantive to Procedural Rationality", in Latsis, S., ed., *Method and Appraisal in Economics*, Cambridge University Press.

SIMON H. (1978), "Rationality as process and as product of thought", *American Economic Review*, Vol.68.

SMITH V. K. (1983), "Option value : a conceptual overview", *Southern Economic Journal*.

SMITH V.K. - DESVOUGES W.H. - FISHER A. (1983), "Estimates of the option values of water quality improvements", *Economic Letters*, Vol.13,pp.81-86.

TEECE D. J. (1980), "Economics of scope and the scope of the enterprise", *Journal of Economic Behaviour and Organization* 1, pp. 223-247.

TEECE D. J. (1982), "Towards an economic theory of the multiproduct firm", *Journal of Economic Behaviour and Organization* 3, pp. 39-63.

THOMPSON R.A. - THUESEN J. (1985), "Dynamic investment criteria for capital budgeting decisions", *Engineering Economist*, Vol.31.

VEDRINE J.P. (1985), "*Techniques quantitatives de gestion* ", Editions Vuibert Gestion.

Von NEUMANN J. - MORGENSTERN O. (1947), "*Theory of games and economic behavior* ", Princeton University Press.

WEICK K. (1969), "*The Social Psychology of Organizing*", Reading, Mass.: Addison-Wesley.

WEISBROD B. A. (1964), "Collective Consumption Services of Individual Consumption Goods", *Quaterly Journal of Economics*, Vol.78,pp.471-477.

WILLIAMSON O.E. (1979), "*Transactions-cost economies : The governance of contractual relations*", *Journal of Law and Economics*, octobre.

WILLINGER M. (1988), "*Information et fondements de la rationalité dans l'incertain : Trois essais sur le statut de l'information dans la théorie de la décision en incertain* ", Thèse de Doctorat de Sciences Economiques, ULP Strasbourg.

WILLINGER M. (1990), " La rénovation des fondements de l'utilité et du risque ", *Revue Economique*, Vol. 41, n°1.

YAARI M. E. (1987), " The dual theory of choice under risk ", *Econometrica*, 55, pp.95-116.

CONCLUSION GENERALE

Dans *Le Cahier bleu*, Ludwig Wittgenstein soulève l'un des principaux paradoxes tenant à la réflexivité du langage, à savoir que chaque sujet parlant constitue un point de perspective privilégié sur le monde - une limite de ce dernier, plutôt que l'un de ses contenus. Selon ce philosophe en effet, on peut distinguer deux modes d'usage du mot "je" (ou "moi") : un usage objectif, désignant une personne particulière et laissant une porte ouverte à l'erreur (comme dans "j'ai grandi de douze centimètres"), et un usage subjectif, ne désignant aucune personne particulière et où il est impossible de se tromper (comme dans "j'ai mal aux dents"). Et il poursuit :

"Cependant le mot 'je' se réfère bien à la personne qui parle, qui se désigne, et bien souvent celui qui dit 'je' se montre lui-même du doigt. Mais cette désignation est parfaitement superflue, il pouvait aussi bien lever la main que pointer le doigt vers lui-même (...) Le mot 'je' n'a pas la même signification que l'expression : 'la personne qui parle en ce moment'. Mais cela ne veut pas dire que L.W. et 'je' désignent des choses différentes. Cela signifie simplement que ces deux termes sont des instruments différents de notre langage." [L. Wittgenstein, 1965, p. 127].

Ainsi L.W. constitue l'un des contenus du monde, mais non 'je' : en tant que sujet, celui-ci est *atopos*, sans place assignée dans le monde dont il parle. Et pourtant, quand "je" parle, "je" fait quelque chose, comme quand "je" grandis de douze centimètres. Commentant ce paradoxe, P. Ricoeur en montre le seul dénouement possible, c'est-à-dire l'*ancrage* du sujet parlant dans le monde par son inscription en celui-ci :

"On ne peut en sortir en effet qu'en admettant une certaine corrélation entre ce *je* qui, d'une certaine façon, n'appartient pas au monde dont il parle, et un certain événement du monde dont il est parlé (...) L'énonciation est dans ce sens elle-même un événement qui arrive. Cette énonciation à la fois participe du statut du *je*, situé en quelque sorte hors du monde, et consiste en une des singularités désignées par voie ostensive dans ce monde (...) Ne faut-il pas mettre en corrélation le *je* vécu et ancré, et une tranche de l'histoire du monde? (...) la corrélation est assurée par le corps propre, organe propre de l'ancrage d'un *je*, qui, de vacant (*shifter*), devient *moi*, *Un tel*. Quand je dis : 'moi, P.R., né à...le...', je désigne à la fois mon existence insubstituable et ma place dans l'état-civil. Cette double désignation désigne l'ancrage lui-même." [P. Ricoeur, 1987, pp. 63-65] (1).

Avec ces quelques remarques d'ordre philosophique, nous ne nous sommes qu'apparemment éloignés du sujet traité dans les pages qui précèdent. En réalité, le paradoxe soulevé par L. Wittgenstein, et son dénouement par P. Ricoeur, nous installent au cœur même de la problématique donnant son unité à l'ouvrage qui s'achève ici. En effet, au-delà de la distinction entre systèmes artificiels (dont nous comprenons la structure et le fonctionnement, parce que c'est nous qui les avons construits) et systèmes naturels (que nous observons, mais dont nous n'avons qu'une compréhension imparfaite, surtout s'ils sont organisés hiérarchiquement), il existe une autre distinction, tout aussi fondamentale que la précédente. Écoutons une dernière fois H. Atlan, qui nous invite à découvrir cette distinction séparant les systèmes naturels que nous observons sans en faire partie de ceux où tel n'est pas le cas.

Les systèmes naturels que nous observons tout en en faisant partie ne sont évidemment rien d'autre que les

“systèmes naturels humains où l'observateur est en même temps partie ou totalité du système. Nous avons vu que notre approche suppose que nous ne connaissons pas l'information que le système a sur lui-même, avec ses différentes significations possibles. C'est dire que, transposée aux systèmes humains, sociaux en particulier, elle implique un point de vue particulier où nous faisons comme si nous ne connaissons pas le sens pour nous de ce que nous vivons nous-mêmes, soit comme individus organisés, soit comme éléments du système social. Ce point de vue n'est pas autre chose que le postulat ou parti pris d'objectivité, conséquence de l'extension de la méthode scientifique aux phénomènes de notre vie. On voit en quoi ce parti pris néglige une part importante, peut-être essentielle, de l'information dont nous pouvons disposer. On ne fait ici que toucher du doigt les limites de cette méthode transposée à l'analyse des phénomènes humains; là, volontairement, et alors même que nous avons le choix et pouvons faire autrement, nous négligeons le subjectif pour ne considérer ces phénomènes que du point de vue d'un observateur extérieur, qui n'aurait aucune information du type de celle que le système a sur lui-même. Comme si l'observateur ne se confondait pas avec la totalité du système, s'il s'agit de l'individu, ou avec un de ses composants, s'il s'agit du système social.” [1979, p. 94].

Le parti pris d'objectivité scientifique dont H. Atlan soulignait ainsi les limites n'est plus aujourd'hui qu'une forme de scientisme héritée du siècle dernier: au moment même où le concept d'auto-organisation faisait une percée décisive dans notre épistémè, I. Prigogine et I. Stengers (dont nous avons vu qu'ils sont à l'origine d'un concept d'auto-organisation légèrement différent de celui de H. Atlan) montraient avec éclat le caractère obsolète d'un tel scientisme, en proposant les voies d'une “nouvelle alliance” situant l'homme dans le monde qu'il décrit [I. Prigogine et I. Stengers, 1979]. C'est très exactement une telle démarche, déjà bien engagée dans les sciences traitant de systèmes “naturels”, que nous avons voulu suivre dans notre ouvrage analysant, à la lumière du concept de rationalité contextuelle et du paradigme de l'auto-organisation, certains aspects de l'apprentissage et de l'évolution dans ce système éminemment social qu'est le système économique.

En effet, c'est bien le souci d'insérer l'*homo oeconomicus* au coeur même du système économique (de gommer son caractère *atopos*) qui a inspiré l'ensemble des chapitre précédents. Car *atopoi*, les micro-sujets le sont encore le plus souvent dans la littérature économique - ainsi lorsque l'on parle d'analyser leurs comportements dans un environnement incertain, ce pourquoi nous préférons parler de *monde* incertain. A nos yeux, si la rationalité des agents a ceci de contextuel qu'aucune de leurs décision n'est prise en dehors de contextes précis qui pèsent sur la nature même de leurs choix, c'est bien parce que l'évolution de chacun, rendue incertaine par celle des autres, contribue à son tour à rendre incertaine celle de ces derniers. Nous avons voulu étudier le sujet économique en tant qu'il est *partie prenante* du monde dans lequel il est ancré - au sens linguistique de ce terme, rappelé par P. Ricoeur; et puisqu'il en va de même pour

chacun, c'est le monde dans son ensemble, c'est-à-dire le composé sujet-et-environnement, que nous voyons baigner dans l'incertitude, et non simplement l'environnement de chacun de ces sujets considérés tour à tour comme parfaitement déterminés et *atopoi*:: plutôt qu'une juxtaposition d'incertitudes partielles, nous avons essayé de penser une incertitude *générale*.

Bien qu'à l'état implicite, une telle préoccupation était présente dès l'ouverture de ce volume par le chapitre de M. Willinger. En rapportant de manière critique les expériences montrant la non invariance des systèmes de préférences individuelles des agents, ce chapitre débouchait logiquement sur la recommandation d'ouvrir la "boîte noire" du décideur afin de mieux appréhender la "technologie de la décision". Le système ici envisagé était l'agent individuel confronté à un choix, et révélant à cette occasion la variabilité de préférences dont l'ordre changeait avec le contexte de présentation des options possibles voire d'expression de telles préférences. Tenant lieu d'environnement au système individuel considéré, ce contexte imprimait donc sa marque aux réponses de celui-ci, qui acquérait par là-même une dimension évolutive en synchronie avec un contexte lui-même mouvant. A condition de ne pas considérer le décideur comme *atopos*, mais immergé au contraire dans le monde économique, il est alors facile d'admettre que l'évolution de son système de préférences ne peut pas rester sans impact sur les systèmes de préférences d'autres agents, et qu'il engendre nécessairement en retour une bonne part de l'incertitude à laquelle se trouvent désormais confrontés ces derniers.

Notre deuxième chapitre est venu expliciter davantage cette double dimension, puisque la conception rosenfeldienne de la mémoire, entièrement centrée sur l'importance du contexte pour une remémoration relevant de la même fonction que la perception, était montrée par B. Ancori être étonnamment proche du paradigme auto-organisationnel naguère proposé par H. Atlan. Fonctionnant selon un processus d'apprentissage non programmée, dont l'évolution avait pour cause efficiente son interaction avec son environnement et pour produit la catégorisation de stimuli, le système ici considéré était celui de la mémoire individuelle. Ouvert par construction même aux signaux provenant de son environnement, un tel système s'inscrit tout naturellement dans une société formée de sous-systèmes analogues à lui, et ainsi liés entre eux sur un mode auto-organisateur produisant ses effets à un niveau hiérarchique englobant.

Désormais central dans la conception de notre ouvrage, le paradigme de l'auto-organisation voyait les multiples facettes de ses applications à l'économie recensées par la contribution de P. Garrouste et J.-A. Héraud. Si le trend de la complexité des systèmes économiques concrets est vraisemblablement ascendant depuis l'apparition de la pensée abstraite en Occident, cette complexité nous est aujourd'hui rendue davantage sensible parce qu'elle est toujours plus largement incorporée dans des objets matériels produits avec une sophistication croissante par la technologie. Telle est probablement la raison essentielle pour laquelle les concepts d'auto-organisation mentionnés dans ce troisième chapitre ont surtout été utilisés dans l'analyse du changement technique.

Dynamique de l'industrie, compétition technologique, structuration du système de production, technologie en tant que telle ou évolution globale de la société: à tous les niveaux d'agrégation, la sphère de la production est saisie comme un organisme évolutif, lié à un environnement changeant selon une multitude de flux auto-organisationnels. Lors de sa co-évolution avec son environnement, chaque système auto-organisateur de ce type n'est pas nécessairement voué à la mort inéluctable que connaît l'individu biologique, mais peut très bien prolonger indéfiniment une trajectoire imprimant son empreinte à la dynamique de son environnement autant qu'elle est marquée par l'influence de cette dernière.

A l'issue de cette première partie, le concept de rationalité contextuelle dont la pertinence était suggérée par les expériences portant sur l'invariance des systèmes de préférences individuelles, puis confortée par le rapprochement des conceptions rosenfieldienne et atlanienne de la mémoire et de l'apprentissage, avait ainsi trouvé sa véritable dimension avec l'examen des applications de l'auto-organisation à l'économie. Inséparable d'un processus évolutif mû par les interactions établies entre deux systèmes susceptibles de mutuellement se modifier, la rationalité contextuelle devait alors voir son fonctionnement explicité à différents niveaux d'abstraction et dans diverses régions de l'analyse économique.

Tel était l'objet de notre deuxième partie, et il était logique d'ouvrir celle-ci par un chapitre se situant à un degré élevé d'abstraction et de généralité. Dans le quatrième chapitre de notre ouvrage, B. Ancori proposait ainsi un modèle formel d'évolution d'un système complexe avec création d'information. Ce modèle incorporait tous les ingrédients progressivement rassemblés dans la partie précédente, puisqu'il analysait l'évolution d'une société de systèmes cognitifs fonctionnant sur un mode rosenfieldien et accédant à une conscience d'eux-mêmes (formalisée par la présence de l'ensemble vide dans toutes les mémoires individuelles). Pareille évolution témoignait fidèlement de la double dimension du processus auto-organisationnel (auto-information interne du système, et action auto-organisatrice de celui-ci sur son environnement) en montrant tous les éléments d'un groupe quelconque de sous-système susceptibles de s'influencer mutuellement lors de leurs communications. Enfin, la société de systèmes cognitifs ainsi formalisée était le lieu d'un accroissement virtuellement indéfini de complexité comme de redondance: bien qu'ils n'aient pas été explicitement introduits dans le modèle, les effets de renforcements et/ou le renouvellement du lexique initial éloignait le système global de l'éventualité d'une forme de mort entropique; et l'accroissement du poids de certaines conventions finissait par faire accéder celles-ci, telle la monnaie, à un niveau *meta* par rapport aux catégories élémentaires - le niveau où se détermine le mode même sur lequel s'opèrent de pareilles catégorisations.

Privilégiant la première dimension de l'auto-organisation, qui voit le système considéré utiliser à son profit les chocs perturbateurs en provenance de son environnement, le cinquième chapitre venait alors préciser le contenu d'une telle utilisation de la part de la firme moderne

constituant ici le système considéré. Après avoir montré ce qui sépare un régime d'information stable d'un régime d'informations perturbées, P. Cohendet et P. Lerena développaient les implications du récent passage du premier au second dans nos économies, et ceci tant au niveau du comportement contextuellement optimal de l'unité empirique considérée qu'à celui des conséquences épistémologiques d'un tel renouvellement des conditions de la production. Du premier point de vue, il est ainsi devenu clair que l'adaptation dynamique de la firme à un environnement imprévisible passe par la mise en place d'une organisation qui minimise les coûts de transaction entre les sous-systèmes la constituant tout en maximisant les effets d'apprentissage au sein de ceux-ci. Du second point de vue, la principale implication consiste en la disparition du caractère tranché de la séparation traditionnelle entre activité économique tournée vers le marché et gestion interne de l'entreprise.

Dans un sixième et dernier chapitre, G. Lambert venait alors précisément apporter le point de vue du gestionnaire confronté à cette turbulence nouvelle de l'environnement, et soucieux de forger des outils aussi bien adaptés aux exigences inédites que pareille situation suggère au théoricien qu'aux nécessités opérationnelles qu'elle rend impérieuses aux yeux des praticiens. La décision d'investissement étant saisie dans la perspective de son impact sur la capacité d'adaptation de la firme donnait ici lieu, dans la veine de la rationalité procédurale introduite par H. Simon, à la formulation d'un critère allant bien au-delà de l'analyse financière traditionnelle. Etant ensuite envisagée du point de vue de son action structurante sur les variables d'état de l'appareil productif global, la décision d'investissement, appliquée à l'investissement énergétique, montrait la possibilité pour le décideur de s'opposer à la solution flexible traditionnellement retenue en présence de certaines conditions d'irréversibilité et d'incertitude. Capacité d'adaptation de la firme et dimension organisationnelle de l'investissement: à travers ce type de décision toujours très difficile à prendre, c'est bien la double dimension d'un processus auto-organisationnel qui se trouvait ici impliqué.

En définitive, l'évolution empirique récente de nos économies industrialisées comme les conséquences théoriques liées à la prise en compte de l'incertitude et de l'irréversibilité se rejoignent pour inciter l'économiste à bouleverser les frontières traditionnelles de sa discipline et à enrichir son stock de concepts. Construits pour analyser le comportement optimal d'un *homo oeconomicus* atomistique, défini comme une moyenne omnisciente au sein d'un espace homogène et entièrement calculatoire (au sens strict où il se prête uniquement à des opérations arithmétiques portant sur des prix et des quantités), les concepts traditionnels fonctionnent remarquablement bien dans un monde statique, où le contexte de la décision n'a pas véritablement à être pris en compte car il ne se modifie pas, ou se modifie peu. Au contraire, dans le monde évolutif et incertain dans lequel nous le voyons aujourd'hui plongé, l'agent économique doit plutôt être conçu comme une structure cognitive singulière, effectuant *explicitement* bien d'autres types de calculs que des opérations arithmétiques (tri, classement, permutation, combinaison, comparaison, substitution, transcodification, etc.), afin d'évoluer de

manière qu'il juge satisfaisante étant donné le contexte précis de cette évolution. C'est à effectuer un pas en cette direction que s'est attaché le présent ouvrage, en montrant diverses articulations théoriques existant entre les deux concepts liés de rationalité contextuelle et d'auto-organisation.

Note

(1) La proposition selon laquelle l'énonciation est un "événement qui arrive" renvoie ici à la proposition, souvent citée, qui ouvre le "Tractatus logico-philosophicus" de L. Wittgenstein : "Le monde est tout ce qui arrive".

BIBLIOGRAPHIE

ATLAN H. [1979], *"Entre le cristal et la fumée. Essai sur l'organisation du vivant"*, Seuil

PRIGOGINE I. et STENGERS I. [1979], *"La nouvelle alliance. Métamorphose de la science"*, Gallimard

RICOEUR P. [1987], "Individu et identité personnelle", in *"Sur l'individu"*, Colloque de Royaumont, Seuil

WITTGENSTEIN L. [1965], *"Le Cahier bleu et le Cahier brun. Etudes préliminaires aux Investigations Philosophiques"*, suivi de *"Ludwig Wittgenstein"* par Malcom N., Gallimard

TABLE DES MATIERES

	Page
INTRODUCTION GENERALE	1
PARTIE 1 : VERS UNE RATIONALITE CONTEXTUELLE	
CHAPITRE 1 : L'HYPOTHESE D'INVARIANCE DES PREFERENCES: LES ENSEIGNEMENTS DE L'ECONOMIE EXPERIMENTALE (par Marc WILLINGER)	10
Introduction	11
1. L'inversion des préférences (Préférence reversal)	13
1.1. Le résultat expérimental	14
1.2. Des interprétations divergentes	15
1.3. Les implications du phénomène d'inversion	17
2. Le "Framing"	18
2.1. L'expérience de McNeil & al. (1982)	18
2.2. Le "framing" des conséquences	19
2.3. Les implications du "framing"	20
3. Biais liés au traitement de l'information	22
3.1. L'expérience de Ronen (1971)	23
3.2. L'expérience de Tversky & Kahneman (1986)	23
4. Le problème de la représentativité de l'expérience	25
5. Vers une alternative à l'hypothèse d'optimisation ?	27
Notes	31
Bibliographie	32
CHAPITRE 2 : MEMOIRE ET APPRENTISSAGE : DE LA NEUROBIOLOGIE A L'AUTO-ORGANISATION (par Bernard ANCORI)	35
1.l'invention de la mémoire	37
1.1. Mémoire et théorie de la localisation fonctionnelle	38
1.2. Contre la localisation fonctionnelle	41
1.2.1. L'homme qui ne savait pas lire ce qu'il avait écrit	42
1.2.2. Hughlings-Jackson, Freud, Penfield et Gloor : l'importance du contexte pour la remémoration	44
1.2.3. Contexte et signification	47

1.3. Darwinisme neuronal et mémoire inventive	50
1.3.1. Limites de la reconnaissance automatique	50
1.3.2. Darwinisme neuronal et mémoire inventive	52
2 L'auto-organisation d'après H. Atlan	57
2.1. L'auto-organisation selon H. Atlan	59
2.2 Signification de l'information dans un système hiérarchisé	64
2.2.1. Ordre par le bruit versus bruit organisationnel	65
2.2.2. Complexité et complication	66
2.2.3. Signification de l'information dans un système hiérarchisé	69
3. Vers un modèle d'apprentissage non dirigé	71
Notes	78
Bibliographie	80
CHAPITRE 3 : AUTO-ORGANISATION ET CREATION TECHNOLOGIQUE	81
(par Pierre GARROUSTE et Jean-Alain HERAUD)	
1. Introduction à la notion d'auto-organisation	82
1.1. L'émergence de l'auto-organisation en économie	83
1.2. L'approche du changement technologique par le modèle de l'auto-organisation	85
2. Les conceptions originelles de l'auto-organisation	86
2.1. L'auto-organisation chez H. Atlan	87
2.2. L'auto-organisation chez H. Maturana et F. Varela : autopoïèse et autonomie	88
2.3. L'auto-organisation chez I. Prigogine et I. Stengers`	89
3. L'analyse des changements technologiques en termes d'auto-organisation	91
3.1. La dynamique d'une industrie comme processus d'auto-organisation	91
3.2. La compétition technologique comme processus d'auto-organisation	94
3.3. L'industrie comme système autonome	96
3.4. La technologie comme système auto-organisé	98
3.5. L'évolution comme processus global d'auto-organisation	101
4. Auto-organisation et stratégie d'innovation	103
Notes	106
Bibliographie	107

PARTIE 2 : AUTO-ORGANISATION, APPRENTISSAGE ET THEORIE DE LA FIRME

CHAPITRE 4 : APPRENTISSAGE AUTO-ORGANISATIONNEL ET PROBLEMATIQUE DU CHOIX INDIVIDUEL 111

(par Bernard ANCORI)

1. Apprentissage et création d'information	114
1.1 Première approche d'un état du système et de ses transformations	114
1.2 Création et signification de l'information : seconde approche d'un état du système et de ses transformations	119
2. La problématique du choix dans un système complexe	127
2.1 Position du problème	127
2.2 Evolution globale et choix individuels	129
Notes	140
Bibliographie	142

CHAPITRE 5 : NATURE DE L'INFORMATION, COUTS DE TRANSACTION ET ORGANISATION DE L'ENTREPRISE 144

(par Patrick COHENDET et Patrick LLERENA)

1. L'évolution des systèmes de production : d'un régime d'information stable à un régime d'information perturbée	146
1.1. Le modèle de standardisation ou le monde de la certitude	148
1.2. Le modèle de variété ou le monde du divers	149
1.3. Le modèle de réactivité ou le monde de l'incertain	151
2. La recherche de flexibilité: une stratégie structurante	153
2.1. Flexibilité statique	153
2.2. Flexibilité dynamique	154
3. Une interprétation de l'évolution des systèmes de production et du rôle de l'intégration	156
3.1. Le rôle de l'intégration comme facteur-clé de fonctionnement des systèmes de production	157
3.2. Le choix stratégique de l'intégration dans le modèle de réactivité	162
3.2.1. Les modifications de l'organisation en régime d'information perturbée	162
3.2.2. L'importance du flux informationnel	163
4. Les coûts de transaction comme mode d'évaluation de l'intégration	167
4.1. Les principes de l'économie des coûts de transaction	167
4.2. Coûts de transaction et organisation interne de la firme	169
4.3. Vers un nouveau modèle économique de l'organisation de la firme	173
Notes	176

Bibliographie	176
CHAPITRE 6	
APPRENTISSAGE INFORMATIONNEL ET RATIONALITE ECONOMIQUE DES CRITERES D'INVESTISSEMENT	178
(par Gilles LAMBERT)	
1. Flexibilité d'adaptation de la firme et analyse bayesienne de la décision d'investissement	181
1.1. Critères séquentiels a posteriori avec irréversibilité et croissance de l'information : une analyse alternative du risque	182
1.1.1. Expression du critère	183
1.1.2. Une représentation probabiliste de l'aversion au risque	187
1.2. Critère séquentiel avec recherche d'information préalable	190
2. La dimension organisationnelle de l'investissement	194
2.1. Choix technologiques et apprentissage organisationnel	194
2.2. Critère d'investissement et apprentissage technologique : application à l'investissement énergétique	195
2.3. La dimension organisationnelle de l'investissement : exemples d'investissements en rationalité procédurale	200
3. Conclusion	202
Bibliographie	203
CONCLUSION GENERALE	207
TABLE DES MATIERES	214

S O M M A I R E

	Page
INTRODUCTION GENERALE	1
PARTIE I : VERS UNE RATIONALITE CONTEXTUELLE	
CHAPITRE 1 : L'HYPOTHESE D'INVARIANCE DES PREFERENCES: LES ENSEIGNEMENTS DE L'ECONOMIE EXPERIMENTALE	10
(par Marc WILLINGER)	
CHAPITRE 2 : MEMOIRE ET APPRENTISSAGE : DE LA NEUROBIOLOGIE A L'AUTO-ORGANISATION	35
(par Bernard ANCORI)	
CHAPITRE 3 : AUTO-ORGANISATION ET CREATION TECHNOLOGIQUE	81
(par Pierre GARROUSTE et Jean-Alain HERAUD)	
PARTIE II : AUTO-ORGANISATION, APPRENTISSAGE ET THEORIE DE LA FIRME	
CHAPITRE 4 : APPRENTISSAGE AUTO-ORGANISATIONNEL ET PROBLEMATIQUE DU CHOIX INDIVIDUEL	111
(par Bernard ANCORI)	
CHAPITRE 5 : NATURE DE L'INFORMATION, COUTS DE TRANSACTION ET ORGANISATION DE L'ENTREPRISE DE L'ENTREPRISE	144
(par Patrick COHENDET et Patrick LLERENA)	
CHAPITRE 6 APPRENTISSAGE INFORMATIONNEL ET RATIONALITE ECONOMIQUE DES CRITERES D'INVESTISSEMENT	178
(par Gilles LAMBERT)	
CONCLUSION GENERALE	207
TABLE DES MATIERES	214