



sociologies

# les modèles de l'action

---

sous la direction de

bertrand saint-sernin  
emmanuel picavet  
renaud fillieule  
pierre demeulenaere



puf

# COLLECTION SOCIOLOGIES

Hans ALBERT

*La sociologie critique en question*

Jacques ANTOINE

*Valeurs de société et stratégies des entreprises*

Raymond ARON

*Études sociologiques*

Jean BAECHLER

*La solution indienne*

Daniel BELL

*La fin de l'idéologie*

Quentin BELL

*Mode et société - Essai sur la sociologie du vêtement*

Joseph BEN-DAVID

*Éléments d'une sociologie historique des sciences*

Philippe BESNARD

*L'anomie. Ses usages et ses fonctions dans la discipline sociologique depuis Durkheim*

Philippe BESNARD (sous la direction de)

*Division du travail et lien social*

Pierre BIRNBAUM

*Sociologie des nationalismes*

Raymond BOUDON, Maurice CLAVELIN

*Le relativisme est-il résistible ?*

Raymond BOUDON, Alban BOUVIER,

François CHAZEL (sous la direction de)

*Cognition et sciences sociales. La dimension cognitive dans l'analyse sociologique*

Étienne BOURGEOIS et Jean NIZET

*Pression et légitimation*

François BOURRICAUD

*L'individualisme institutionnel - Essai sur la sociologie de Talcott Parsons*

Alban BOUVIER

*L'argumentation philosophique - Étude de sociologie cognitive*

Giovanni BUSINO

*Critiques du savoir sociologique*

François CHAZEL (sous la direction de)

*Action collective et mouvements sociaux*

Mohamed CHERKAoui

*Les changements du système éducatif en France (1950-1980)*

Augustin COCHIN

*L'esprit du jacobinisme*

Laurent CORDONNIER

*Coopération et réciprocité*

Lewis A. COSER

*Les fonctions du conflit social*

Charles-Henry CUIIN

*Les sociologues et la mobilité sociale*

Charles-Henry CUIIN (sous la direction de)

*Durkheim d'un siècle à l'autre. Lectures actuelles des « règles de la méthode sociologique »*

Maurice CUSSON

*Croissance et décroissance du crime*

*Criminologie actuelle*

Pierre DEMEULÉNAERE

*Homo oeconomicus. Enquête sur la constitution d'un paradigme*

Liliane DEROCHE GURCEL

*Simmel et la modernité*

Annette DISSELKAMP

*L'éthique protestante de Max Weber*

Mattei DOGAN et Robert PAHRE

*L'innovation dans les sciences sociales*

Jean DUBOST

*L'intervention psychosociologique*

Jacques DUPAQUIER

*L'invention de la table de mortalité. De Graunt à War-gentin, 1662-1766*

*L'espérance de vie sans incapacités. Faits et tendances, premières tentatives d'explications*

Shmuel EISENSTADT

*Approche comparative de la civilisation européenne*

Jon ELSTER

*Karl Marx. Une interprétation analytique*

Ernest GELLNER

*La ruse de la déraison*

Anthony GIDDENS

*La constitution de la société*

Roger GIROD

*Le savoir réel de l'homme moderne. Essais introductifs*

Jean-Jacques GISLAIN et Philippe STEINER

## *Les modèles de l'action*

*SOCIOLOGIES*

*Collection dirigée par Raymond Boudon*

*Membre de l'Institut*

*Professeur à l'Université de Paris IV - Sorbonne*

Sous la direction de  
Bertrand Saint-Sernin, Emmanuel Picavet,  
Renaud Fillieule et Pierre Demeulenaere

---

LES MODÈLES  
DE L'ACTION



Presses Universitaires de France

ISBN 2 13 048996 6

Dépôt légal — 1<sup>re</sup> édition : 1998, février

© Presses Universitaires de France, 1998  
108, boulevard Saint-Germain, 75006 Paris

## *Liste des auteurs*

---

**Raymond Boudon**, membre de l'Institut (Académie des sciences morales et politiques), est professeur de sociologie à l'Université de Paris-Sorbonne (Paris IV). Ses nombreuses contributions à la sociologie de l'éducation, à la sociologie de la connaissance et à la méthodologie des sciences sociales ont fait de lui l'un des principaux représentants de l'individualisme méthodologique d'inspiration wébérienne. Parmi ses ouvrages, traduits dans plusieurs langues, citons : *L'analyse mathématique des faits sociaux* (Plon, 1967, 1971), *L'inégalité des chances* (A. Colin, 1973, 1977), *Effets pervers et ordre social* (PUF, 1977, 1979), *La logique du social* (PUF, 1979, 1983), *La place du désordre* (PUF, 1984) et *Le juste et le vrai* (Fayard, 1995).

**Alain Boyer** est professeur de philosophie à l'Université de Caen. Ses recherches concernent l'épistémologie et la philosophie politique et sociale, notamment la pensée de Karl Popper, dont il est l'un des meilleurs spécialistes. Il a également consacré des études à la pensée de Friedrich von Hayek et à la méthodologie historique. On lui doit en particulier : *K. R. Popper: une épistémologie laïque?* (Presses de l'École normale supérieure, Paris) et *L'explication en histoire* (Presses Universitaires de Lille, 1992).

**Pierre Demeulenaere** est maître de conférences en sociologie à l'Université de Paris-Sorbonne (Paris IV). Il a consacré ses premières recherches aux modalités et aux incertitudes de la représentation de l'action dans la tradition de la science économique (*Homo oeconomicus. Enquête sur la constitution d'un paradigme*, Paris, PUF, 1996). Il a élargi cette enquête au domaine de la sociologie (*Histoire de la théorie sociologique*, Paris,

Hachette, 1997). Il se consacre désormais à l'étude de la dimension normative de l'action sociale, notamment dans ses aspects esthétiques.

**Renaud Fillieule** est maître de conférences en sociologie à l'Université de Lille I. Dans sa thèse de doctorat, il a appliqué et développé les méthodes de la sociologie cognitive afin d'éclairer la thématique de l'information et de la décision dans les sciences sociales (*La question de la rationalité dans les sciences sociales: histoire et bilan*, Université Paris IV, 1994). Ses recherches concernent aussi les relations entre psychologie, sociologie et économie. Il conduit par ailleurs des études de sociologie appliquée.

**Emmanuel Picavet** est maître de conférences en philosophie politique à l'Université Panthéon-Sorbonne (Paris I). Ses recherches concernent la méthodologie des sciences sociales et la théorie politique. Dans *Choix rationnel et vie publique* (PUF, 1996), il a étudié la constitution de la théorie moderne de la décision rationnelle (individuelle et collective) et sa contribution à la pensée politique. Il a également publié : *Approches du concret. Une introduction à l'épistémologie* (Ellipses, 1995). Ses travaux actuels concernent la théorie des normes et des droits.

**Thomas Piketty** est chargé de recherches au CNRS, dans l'équipe du CEPREMAP, après avoir été assistant à l'Université de Londres (London School of Economics and Political Science), puis professeur assistant au Massachusetts Institute of Technology (États-Unis). Ses travaux concernent la théorie économique de l'information et des incitations, la fiscalité et la théorie positive de la vie politique. Sa thèse de doctorat, dont une version remaniée a été publiée (*Introduction à la théorie de la redistribution des richesses*, Economica, 1994), lui a valu le prix de l'Association française de sciences économiques en 1993. Il est également l'auteur de *L'économie des inégalités* (La Découverte, 1997).

**Bertrand Saint-Sernin** est professeur de philosophie des sciences à l'Université de Paris-Sorbonne (Paris IV). Ancien recteur d'Académie, il a consacré plusieurs ouvrages à l'analyse de la décision et à la philosophie de l'action, notamment : *Les mathématiques de la décision* (PUF, 1973), *Le décideur* (Gallimard, 1979) et *Genèse et unité de l'action* (Vrin, 1979). Il a également publié plusieurs études sur Antoine-Augustin Cournot, Joseph Conrad et Simone Weil. Dans trois livres récents, il évoque les destins croisés de l'action et de la connaissance dans les temps modernes : *Parcours de l'ombre. Les trois indécidables* (Éditions des Archives contemporaines, 1994) ; *La raison au XX<sup>e</sup> siècle* (Seuil, 1995) ; *Entretiens nocturnes sur la théorie des jeux, la poésie et le « nihilisme » chrétien* (Bruxelles, Le Cri, 1997).

**Bernard Walliser** est professeur d'économie à l'École nationale des Ponts et Chaussées. Ses travaux sont consacrés, d'une part, à la théorie de la décision et à la théorie des jeux, d'autre part, à la méthodologie économique. Il s'intéresse plus particulièrement à la structure et à la révision des croyances des agents économiques dans un cadre logique ou probabiliste. On lui doit plusieurs ouvrages, en particulier: *Systèmes et modèles* (Seuil, 1977), *Anticipations, équilibres et rationalité économique* (Calmann-Lévy, 1985) et *L'intelligence de l'économie* (Odile Jacob, 1994).



# Sommaire

---

<i>Liste des auteurs</i>	v
Présentation, <i>par Emmanuel Picavet</i>	1
Au-delà du « modèle du choix rationnel » ?, <i>par Raymond Boudon</i>	21
Les ambiguïtés constitutives du modèle du choix rationnel, <i>par Pierre Demeulenaere</i>	51
Aspects praxéologiques des principes interpersonnels, <i>par Emmanuel Picavet</i>	67
Modèles de l'action et théorie des prix, <i>par Renaud Fillieule</i>	103
Individualisme méthodologique et économie de l'information, <i>par Thomas Piketty</i>	117
Rationalité cognitive et théorie des jeux, <i>par Bernard Walliser</i>	127
Science ou poésie de l'action? L'espace poppérien du raisonnement historique, <i>par Alain Boyer</i>	143
Complétude et incomplétude de l'action, <i>par Bertrand Saint-Sernin</i>	165



## *Présentation*

Les contributions ici réunies furent initialement présentées et discutées au cours d'une Journée d'étude organisée à la Sorbonne le 28 mars 1996, à l'initiative de M. le P<sup>r</sup> Bertrand Saint-Sernin, dans le cadre de l'École doctorale « Philosophie et sciences sociales » de l'Université de Paris-Sorbonne. Les débats furent présidés par M. le P<sup>r</sup> Maurice Clavelin, à qui les participants adressent leurs sincères remerciements.

Notre souci commun était d'explorer les voies actuelles de l'analyse de la décision et de l'action, afin de mieux comprendre comment s'articulent, dans ce domaine, description, modélisation et compréhension. Cette réflexion était de nature à favoriser un rapprochement des sciences sociales et de la philosophie. De fait, le dialogue s'est établi autour de la double thématique des raisons de l'action et des représentations théoriques que l'on peut en donner. Si la philosophie pratique s'intéresse aux raisons qui doivent conduire à préférer ou à entreprendre une chose plutôt qu'une autre, les sciences sociales ne font pas autre chose lorsqu'elles doivent tenir compte des motifs qui peuvent expliquer que les agents étudiés se comportent de telle ou telle manière. Ce dont on a des raisons de penser qu'il faut le faire, autrui peut également découvrir qu'il y a lieu de le faire. Il existe donc une très forte complémentarité entre les aspects « positifs », traditionnellement dévolus aux sciences sociales et à la psychologie, et la dimension « normative » – plutôt réservée à la philosophie – de l'étude de la décision et de l'action.

Les sciences sociales recherchent des modes d'explication des conduites susceptibles de rendre celles-ci moins surprenantes, moins rebelles

à la prévision. Cette entreprise conduit à se poser la question du type d'explication que l'on peut légitimement considérer comme « ultime » dans les disciplines qui concernent la conduite humaine. Une chose est sûre : on ne peut, dans la pratique des sciences sociales, renvoyer indéfiniment à d'éventuelles explications neurophysiologiques, toujours à venir, des formes singulières de l'action. A moins de faire sienne la croyance en un déterminisme universel, ou tout au moins la croyance en la possibilité de prévoir l'action sans faire référence à l'échelon spécifique des raisons qu'ont les personnes de se résoudre à tel choix ou à telle action, il est d'ailleurs malaisé de se représenter la forme que pourrait prendre l'explication biologique du déclenchement d'une guerre ou du déroulement d'une campagne électorale. Pour qu'une telle explication puisse être considérée comme strictement biologique (ou physico-chimique), ne faudrait-il pas, en effet, la supposer capable de présenter ce qu'elle explique comme le produit prévisible (ou du moins probable) de la conjonction de certaines lois biologiques ou physico-chimiques et de certaines suites d'événements eux aussi de nature physico-chimique ? Or, si même cela était possible, et sans postuler d'emblée que l'explication véritable suppose toujours une forme de « compréhension » singulière de la personnalité de l'agent, on peut se demander si l'on disposerait alors d'une explication pleine et entière.

En effet, dans le domaine de la conduite humaine, la question « Pourquoi ? » renvoie le plus souvent aux motifs des agents, donc à la manière dont ils déterminent et se représentent ce que leur action devrait être. Nous nous intéressons précisément aux étapes qu'ils ont pu parcourir pour arriver à cette représentation : la prise de conscience d'un souhait, le traitement des éléments d'information, l'évaluation des probabilités des événements pertinents, un raisonnement, etc. Or, ces représentations intermédiaires ne nous intéressent pas en tant que déterminantes causales de la résolution d'agir : rien ne dit qu'elles aient cette qualité. Elles nous intéressent parce que nous savons que nous considérons, les uns et les autres, qu'elles sont liées d'une manière définie à certaines intentions d'action. Et aussi, à un échelon plus fondamental encore, parce que nous n'ignorons pas que les agents considèrent en général que leur résolution d'agir de telle ou telle manière pourrait être altérée ou annulée s'ils voyaient des raisons à cela.

Par exemple, nous savons que nous nous représentons habituellement comme « étant en notre pouvoir » le fait d'emporter ou non un parapluie. Nous savons aussi que le fait de prendre un parapluie, souvent, fait l'objet d'une délibération au moins en partie consciente : nous

pesons le pour et le contre, sur la base d'une estimation des chances qu'il pleuve et de la gêne que peut causer le transport de l'objet. Ainsi, à propos d'un homme dont nous savons qu'il n'aime pas s'encombrer d'objets inutiles, la question « Pourquoi a-t-il emporté son parapluie ? » a un sens, et il n'est pas absurde d'y répondre par : « Il croyait qu'il pleuvrait » ou même, plus précisément : « Il a ajouté foi aux prévisions météorologiques de la matinée, qui annonçaient qu'il pleuvrait. » Dans cet exemple, il y a, à l'origine de la réponse, quelque chose qui se laisse appréhender assez naturellement comme la solution d'un problème en tout point analogue à celui que résout plus ou moins consciemment l'agent au moment de la délibération. Cette solution procède d'un certain nombre de « bonnes raisons » que nous pouvons reconstituer d'après les données factuelles disponibles.

Dans le domaine de l'analyse générale de la décision, le recours fréquent à l'analyse mathématique semble souvent avoir pour effet de rendre moins immédiatement compréhensible la place accordée aux raisons des acteurs. Les théories postulent que les agents ont des buts ou des préférences, mais sans dire lesquelles précisément, tout le travail revenant à analyser les conséquences ou l'expression, dans l'existence individuelle ou collective, de cette finalité indéfinie. La forme l'emporte alors sur la matière, et les résultats d'existence ou d'impossibilité prennent le pas sur les lois empiriques et les prévisions testables. Mais on s'expliquerait mal le recours fréquent des sciences économiques et sociales à la théorie de la décision, si cette théorie elle-même n'acceptait de se soumettre, comme à une force de rappel sur ses développements mathématiques, à l'exigence d'une mise en forme convaincante de l'observation des actes de personnes que l'on ne peut présumer plus irrationnelles que l'observateur, ou de l'expérience possible du choix rationnel (appuyé sur de bonnes raisons). En ce sens, l'analyse abstraite de la décision est inséparable d'une démarche d'élucidation de la nature de la rationalité pratique, que l'on veuille concevoir cette ambition sur le mode du dévoilement progressif de ce qui est « en soi » rationnel, ou bien à la manière de l'affinement progressif d'une intuition du choix cohérent.

C'est ainsi que les modèles abstraits de la décision paraissent convaincants ou spécialement intéressants lorsqu'ils permettent de mettre en forme, avec économie et élégance, les réseaux de finalités, croyances et arguments capables de donner sens à certaines situations de choix individuel ou collectif qui présentent une importance particulière. Dans la mesure où les développements mathématiques sont finalement intégrés à une analyse de la conduite personnelle ou de la vie sociale, ils ces-

sent d'être une pure forme, pour épouser les contours de l'action et des institutions, et renouer ainsi avec les raisons de croire et d'agir des acteurs réels.

L'examen entrepris ne pouvait ignorer les doutes fréquemment exprimés au sujet de la pertinence, tant pour les sciences sociales empiriques que pour la philosophie morale et politique, du recours à une modélisation en termes de choix rationnel procédant de prémisses restrictives. Mais les critiques les plus intéressantes sont celles qui aident à délimiter, de manière négative, le domaine de validité potentielle d'une classe de descriptions ou de modèles. Dès lors que l'on s'assigne comme tâche de comprendre ou d'expliquer l'action, en voulant aller au-delà de l'enregistrement et du traitement des régularités empiriques, il faut en effet proposer une reconstitution des motifs, soit de l'action proprement dite, soit de certains aspects particuliers des conduites observées. La recherche des motifs rationnels est également indispensable lorsqu'il s'agit de justifier ou de recommander certaines règles, conventions ou conduites. Comment, donc, envisager une démarche constructive permettant de prendre du recul par rapport aux modèles de rationalité trop restrictifs, et qui suscitent pour cette raison des critiques légitimes, sans pour autant abandonner l'ambition d'expliquer et de comprendre ce qui rend l'action justifiable ou raisonnable aux yeux de l'acteur, du théoricien ou de l'observateur ? Telle est la problématique générale des contributions ici réunies.

Chacune de ces enquêtes concerne un champ disciplinaire défini. Mais le rapprochement des résultats permet de voir émerger un petit nombre de conclusions transversales. En premier lieu, il est à la fois possible (comme le montre l'évolution récente de la théorie des jeux et des sciences économiques) et nécessaire (pour répondre aux accusations parfois justifiées d'irréalisme) d'intégrer dans les modèles de la décision et de l'action la dimension « cognitive » de l'acquisition et du traitement de l'information. Cette dimension de la rationalité pratique l'amène en coïncidence avec le champ d'investigation traditionnel de l'épistémologie générale. Il s'agit en effet de comprendre comment les personnes, au moment d'agir, tâchent de comprendre l'environnement général et les conditions particulières de leur action ou de leur interaction avec autrui, voire le déroulement prévisible de cette interaction elle-même. De fait, la plupart des problèmes de décision supposent, de la part de l'agent, certaines connaissances d'arrière-plan et la résolution de certains problèmes spécifiques. Que les solutions retenues par l'agent soient correctes ou erronées, il demeure qu'en règle générale il est difficile de postuler qu'elles sont totalement arbitraires et ne donnent lieu à aucun raisonnement.

Par ailleurs, notre compréhension de l'action est souvent incomplète parce que nous ne nous interrogeons pas suffisamment sur les raisons compréhensibles que l'on peut avoir de retenir certaines valeurs, d'accepter telles normes, ou d'agir conformément à un principe. Tout dans ce domaine n'est pas affaire de goûts ou de préférences non analysées : on délibère aussi sur les fins. C'est le domaine de la rationalité axiologique, familier aux philosophes, et dans lequel il faut certainement chercher des éléments pour la résolution de certains des grands problèmes en suspens des sciences sociales contemporaines, notamment l'apprentissage des normes de coopération dans les interactions répétées, l'accord sur les règles de partage des avantages et des désavantages, ainsi que l'accord rationnel sur les normes juridiques. Dans ce domaine également, il y a lieu de s'interroger sur les problèmes qui se posent aux agents et sur les éléments dont ils disposent pour parvenir à des solutions justifiables à leurs propres yeux. La rationalité des valeurs, certes, ne peut concerner sans distinction l'ensemble de tous les principes auxquels les personnes manifestent leur attachement. En particulier, n'est pas « rationnel en valeur » tout ce qui conduit à compter pour rien les conséquences de l'action ou de la mise en pratique de certaines règles d'action. Mais certains principes capables de guider l'action ou le choix des normes se prêtent effectivement à une argumentation rationnelle indépendante des conséquences de leur adoption, et qui rend compréhensible leur adoption par des personnes qui pourraient en choisir d'autres, et auraient pour cela certaines raisons tirées des conséquences prévisibles.

Ce double enrichissement, en direction de la cognition et de l'axiologie, donne son visage à un individualisme méthodologique renouvelé, et cependant plus proche que jamais de ses sources wébériennes et parétien-nes. Que l'on veuille partir des désirs, des interrogations ou des choix axiologiques des personnes, l'attitude scientifique reste la même : il s'agit d'essayer de comprendre ce qui peut expliquer les choix effectués, en faisant le pari, aussi longtemps que rien ne force à y renoncer, qu'il y a, précisément, quelque chose à comprendre. A mesure que s'élargit le champ des conduites et des attitudes dont on pense qu'il est possible de les expliquer en termes de choix rationnel, il devient aussi plus clair qu'il est vain de séparer radicalement la rationalité théorique de la raison pratique. Tout devient affaire de coordination entre différents ordres de problèmes, dont la plupart comportent à la fois un versant pratique et un versant théorique. Ainsi, la détermination des moyens convenables pour parvenir à certaines fins relève en général de la connaissance pure, mais la résolution de ce problème exige souvent que l'on y consacre des moyens

qui pourraient servir à autre chose, ce qui pose la question pratique de la valeur comparée de cette activité de recherche et d'autres types d'activités. De même, l'accord sur des valeurs ou des normes communes engage évidemment la pratique, dans la mesure où il s'agit précisément de l'orienter, mais aussi la connaissance, puisqu'il s'agit de résoudre un problème : trouver une solution qui, par ses propriétés, puisse être acceptée par des personnes ayant des convictions diverses.

Enfin, il paraît indispensable – et c'est la troisième leçon générale qui se dégage – de faire évoluer conjointement la description de l'action et le type d'explication ou de justification que l'on envisage. Il faut en effet prêter attention au cadre conceptuel que nous adoptons pour saisir les motifs de l'action, la délibération, l'instant de la décision et le déroulement temporel de l'action. Ainsi, est-il vraiment pertinent d'opposer radicalement les « intérêts » des individus et leurs « valeurs », lorsque les motifs de ces deux catégories se traduisent par le même type de comportement ? L'information est-elle une condition préalable qui guide la décision, ou bien est-elle aussi quelque chose que l'on choisit d'acquérir ou de ne pas acquérir, autrement dit, l'un des objets de la décision ? Quant à la délibération, faut-il la penser comme contemporaine de l'action (comme pourrait le donner à penser l'importance des révisions de plans ou d'intentions au cours de l'action), ou bien la situer par principe avant le début de l'action ? Les multiples perspectives par lesquelles se révèle l'action font hésiter sur la manière correcte de la décrire, et il serait certainement vain de prétendre résoudre complètement ce problème par l'introspection ou l'observation psychologique. Les contributions qui suivent suggèrent qu'il faut surtout veiller à ce que le type de description retenu s'accorde, sans multiplier les distinctions superflues, au type d'enquête ou de délibération que l'on a en vue. Il est toujours possible que nos catégories linguistiques et notre approche ordinaire des phénomènes nous conduisent à figer indûment certaines distinctions. Or, il y a lieu de souhaiter qu'au contraire les distinctions trouvent leur raison d'être dans la théorie.

La philosophie du choix rationnel qui a eu le plus d'influence dans le domaine des sciences de l'homme (et plus spécialement en économie) est certainement la conception dite « instrumentale », d'ascendance humienne, selon laquelle l'action rationnelle est celle qui sert adéquatément des motifs qui, en eux-mêmes, ne sont pas soumis à une norme de rationalité : intérêts, passions ou désirs. Cette approche générale a été vivement contestée, notamment dans les domaines où son influence a été la plus forte (l'économie et la psychologie des choix), comme le montre la

discussion, par Raymond Boudon, de quelques-uns des « paradoxes de la décision » les plus notables. Mais la critique, par ailleurs, ne peut ignorer les aspects qui rendent les explications ou recommandations « instrumentales » convaincantes dans certains cas. En effet, la « mise en cohérence » des actions, des préférences personnelles et des croyances, telle qu'elle est reflétée ou construite par les théories du choix rationnel les plus classiques, n'est pas sans rapport avec l'intelligibilité de l'action, son sens même – ce que l'on peut appeler, d'une manière générale, la raison ou les raisons de l'action. Elle permet, pour certaines classes de situations, de rendre compte des aspects grâce auxquels la décision et l'action peuvent paraître acceptables ou recommandables aux yeux de celui qui agit et, indissociablement, devenir intelligibles pour autrui.

Il faut donc se mettre en quête d'une théorie de l'action, ou à tout le moins d'une forme de coordination entre des théories plurielles, qui permette la saisie rationnelle de l'objet sans réduire arbitrairement sa complexité propre. Prenant implicitement leurs distances par rapport à toute méthodologie « irréaliste » qui compterait pour rien la pertinence descriptive des éléments de base des théories, les auteurs ont tenté de montrer, à partir de questions relevant de plusieurs champs disciplinaires, comment les difficultés et les paradoxes rencontrés par les doctrines ou les modèles classiques de l'action invitent à un renouvellement des postulats à partir desquels on aborde le processus même par lequel les personnes se représentent et comprennent les circonstances de leur action, élaborent ou acceptent des valeurs ou des normes de référence, prennent leurs décisions, et enfin agissent.

L'émergence d'une « théorie de la décision » unifiée au cours du XX<sup>e</sup> siècle, capable d'embrasser du même regard les circonstances du choix individuel et les enjeux de la vie publique, a fait naître l'espoir d'une discipline rigoureuse de la pensée de l'action, digne de s'imposer simultanément à l'économie, à la sociologie et à la politique. Les sciences économiques ont embrassé très tôt la cause des sciences de la décision, en privilégiant la modélisation en termes de choix optimisateur. Mais la relation qui se noue ainsi entre la recherche empirique et la théorie de la rationalité pratique pose de nombreux problèmes. Elle doit certainement, tout d'abord, être envisagée dans la perspective dynamique de l'élaboration et du développement d'un programme de recherche scientifique. Enracinée dans la tradition de l'individualisme méthodologique, la théorie moderne du choix rationnel a progressivement affirmé son pouvoir explicatif et son aptitude à éclairer d'un jour nouveau un grand nombre de problèmes sociaux.

De fait, les branches abstraites que sont la théorie du choix rationnel dans l'incertain, la théorie des jeux (coopératifs et non coopératifs), la théorie des choix collectifs, et plus récemment la théorie de la concrétisation par des mécanismes (ou « implémentation ») constituent ensemble l'épine dorsale d'un vaste programme de recherches en philosophie et dans les sciences sociales. Frank Ramsey, John von Neumann, John Nash, Leonard Savage et Kenneth Arrow, en notre siècle, ont particulièrement contribué à donner une assise conceptuelle à ces développements. Or, ceux-ci ont toujours privilégié une conception de la rationalité que l'on ne peut certes réduire complètement au stéréotype des relations entre moyens et fins, mais qui oppose nettement les préférences et les croyances, et suppose, ou plutôt construit activement, une forme élaborée de cohérence entre les préférences, les croyances et les choix. Il s'agit donc d'une conception assez étroitement définie de la rationalité.

Ce programme est plus vivant que jamais, si l'on en juge par l'étendue de ses conquêtes récentes. Il n'est point de domaine de la vie sociale ou politique dont on ne tente d'analyser les ressorts dans les termes de la théorie classique du choix rationnel. Ainsi, à mesure que les méthodes des sciences économiques sont utilisées avec plus de maîtrise et de profondeur en sociologie et en théorie politique, c'est l'ensemble du domaine traditionnel des sciences de la société qui se trouve amené en coïncidence, à la faveur d'une sorte de postulat méthodologique, avec le domaine d'application concevable de l'analyse abstraite de la décision d'abord élaborée, souvent dans une perspective de simple élucidation de l'idée de rationalité, par les philosophes de l'action et les économistes mathématiciens. Il serait cependant souhaitable que cette emprise croissante des modèles inspirés par la théorie du choix rationnel s'accompagne de progrès réels dans l'élaboration des critères de délimitation théorique de leur champ d'application légitime.

Le premier problème qui se pose est celui de la pertinence descriptive et explicative des modèles par lesquels on se propose de représenter les phénomènes sociaux en termes de choix rationnel appuyé sur l'intérêt, sur l'utilité ou sur des préférences simplement considérées comme des données (ne faisant pas l'objet d'une délibération). Il s'agit d'un enjeu majeur dans les sciences sociales contemporaines parce qu'il est indispensable, dans ce domaine, de chercher à préciser le lien à créer entre deux ambitions essentielles : d'une part, comprendre et expliquer les phénomènes sociaux en décrivant leur enracinement dans les croyances, les désirs, et finalement les conduites des personnes ; d'autre part, donner de ces phénomènes, sinon des lois exactes, du moins une description syn-

thétique ou un modèle fiable des tendances, des séquences régulières et des liens de causalité qu'ils recouvrent. Les progrès de la modélisation dans les sciences de la décision s'accompagnent-ils nécessairement d'une plus grande pertinence des modèles développés dans les sciences sociales ? N'y a-t-il pas lieu de craindre qu'à rebours le souci d'élucidation complète de toutes les conséquences, si lointaines fussent-elles, de telle ou telle conception de la rationalité pratique, ne vienne stériliser l'étude de la vie sociale concrète, en substituant aux processus repérables des interactions idéales, mais imaginaires ? Est-on sûr, au demeurant, de privilégier la bonne conception de la rationalité ? Tel est l'objet, envisagé dans une perspective à la fois critique et constructive, des deux premiers essais.

Raymond Boudon, prenant acte de l'extension croissante du domaine d'application de la théorie « orthodoxe » du choix rationnel, propose cependant un dépassement de celle-ci, selon une démarche de sociologie cognitive et axiologique. La théorie de la décision traditionnelle modélise avant tout la cohérence et l'efficacité du recours à certains actes dans la perspective de certaines fins. Rien ne différencie l'action, dans ce contexte, d'une série d'actes ayant seulement une valeur instrumentale au regard des buts poursuivis par l'agent (ces finalités pouvant se réduire aux simples préférences personnelles subjectives). Or, l'explication de l'action suppose que l'on propose une reconstitution des raisons de l'action, et cela précisément parce que les « agents » qu'étudient les sciences sociales sont des êtres humains, capables, comme chacun de nous, de former des jugements et de déterminer leur conduite d'après des systèmes de raisons. C'est là un levier d'explication propre aux sciences de l'homme et de la société, et il est essentiel d'observer que les raisons de l'action ne concernent pas seulement les résultats de l'action.

Il est difficile, dans ces conditions, d'assimiler la conduite des agents à un simple instrument dont la finalité serait la production optimale des résultats, qu'il faudrait alors considérer comme les seuls supports de l'évaluation et de la motivation. L'explication de l'action par ses raisons doit faire intervenir, au-delà de la dimension instrumentale, d'autres caractéristiques de la motivation ou de la détermination à agir qui procèdent des raisons que se représentent les agents. Il faudra donc s'efforcer de pénétrer la logique propre à l'élaboration des systèmes de croyances, en se référant à la situation des agents. De même, il faudra s'efforcer de faire justice aux considérations déontologiques, aux règles éthiques en particulier, qui entrent dans le répertoire des « bonnes raisons » qu'ont les agents de se conduire de telle ou telle manière.

Les choix méthodologiques de R. Boudon favorisent, d'une part, un rapprochement de la modélisation des conduites et de l'analyse de la rationalité scientifique et, d'autre part, l'intégration, dans le spectre des formes acceptables d'explication des conduites, du recours à des normes sociales et morales fondamentales, à titre de principes guidant l'action sans référence obligée à ses conséquences. Certes, il serait certainement possible d'intégrer dans les propriétés des « conséquences » de la théorie de la décision classique la conformité à certaines normes ou valeurs. Ce sont là des propriétés qui, à l'égal d'autres caractéristiques, peuvent les rendre plus ou moins désirables. Rien ne dit que les « préférences » ne puissent être en partie façonnées par des valeurs ou des normes, qui seraient pour ainsi dire incorporées en elles. C'est là une simple affaire d'interprétation des notions de base de la théorie. Il serait également possible de tenir compte de la conformité de l'action même, et non pas seulement de ses résultats, à des valeurs ou à des normes. L'action ne pourrait plus alors être assimilée à un simple instrument devant permettre la réalisation d'états de fait désirables. Elle serait en effet le support même de la valeur. Dans ce contexte même, il resterait cependant possible de défendre une conception instrumentale de la rationalité, en considérant par exemple que le choix de telle action (plutôt que telle autre) est pour l'agent un moyen de réaliser ce qui a pour lui de la valeur, à savoir la conformité, dans l'action, aux principes (moraux, religieux ou autres) qu'il admet.

Mais il n'en va plus ainsi lorsqu'on considère que les principes moraux se prêtent eux-mêmes à une justification rationnelle. Puisqu'ils ne s'intègrent plus alors aux préférences non analysées (que celles-ci portent sur les conséquences de l'action ou sur l'action même), ils doivent pouvoir être compris comme la réponse adéquate à un problème ou à une situation. En ce sens, la réinterprétation de la rationalité axiologique à laquelle nous sommes conviés peut avoir des implications profondes pour la théorie du choix rationnel. La rationalité axiologique, si l'on suit cette voie jusqu'à son terme, ne serait pas à interpréter simplement comme un type de préférence parmi d'autres (à savoir, le type des préférences qui procèdent de la conformité des résultats à des valeurs plutôt qu'à des intérêts ; ou bien encore, le type des préférences qui privilégient la conformité de l'action à des normes, plutôt que leur efficacité dans la promotion de certaines propriétés des résultats). Il faudrait plutôt dire que la rationalité axiologique n'est pas affaire de préférence, mais de réponse à des problèmes, bien qu'elle intervienne dans les choix par l'intermédiaire de la résolution de ces problèmes.

L'approche défendue par R. Boudon n'interdit pas ce type d'analyse, mais se prête aussi à une lecture moins rigide, dans laquelle le travail explicatif consisterait à faire intervenir tour à tour les préférences indépendantes des valeurs morales, les valeurs proprement dites, et aussi les raisons que les personnes peuvent avoir d'accepter et de mettre en pratique ces valeurs. Appliquée dans plusieurs directions de la recherche empirique, cette approche permet de renouer avec la notion wébérienne de rationalité axiologique (*Wertrationalität*) en lui donnant une pertinence nouvelle, et oblige à renoncer à l'ambition d'une explication universelle des phénomènes sociaux selon les canons de la seule rationalité instrumentale. La distinction opérée entre plusieurs types de rationalité laisse subsister des zones d'interférence importantes, et la possibilité de leur intégration conjointe dans des explications complexes à plusieurs strates. Au demeurant, le nouveau programme de recherche de la sociologie cognitive et axiologique est destiné à tenir compte de l'apport le plus décisif de la théorie usuelle du choix rationnel : la recherche d'explications fondées sur les seules raisons des acteurs, sans recours à des fictions de théoricien dont les agents devraient tenir compte quoi qu'ils en aient, comme à leur insu, par exemple à la faveur d'un ténébreux déterminisme socio-économique. L'explication par les raisons de l'action, rappelle R. Boudon, est complète en son genre, plus complète même pour les faits concernés que ne pourrait jamais l'être une description des processus neurophysiologiques associés aux intentions et à l'action. Mais elle doit être suffisamment complexe pour permettre d'aborder utilement de larges classes de phénomènes.

Aussi les révisions du paradigme du choix rationnel instrumental destinées à prendre en compte la rationalité limitée apparaissent-elles insuffisantes. C'est ce dont témoignent un certain nombre de « paradoxes » persistants, parmi lesquels le problème du vote revêt une importance méthodologique particulière : pourquoi vote-t-on, en l'absence de toute incidence causale, et en l'absence d'enjeux personnels considérables ? Le problème est de savoir si les paradoxes font partie des petites énigmes à résoudre dont pourrait se nourrir la théorie du choix rationnel dans son processus dynamique d'approfondissement et de consolidation, ou bien, à l'inverse, d'une catégorie de problèmes profonds que cette théorie ne parvient pas à surmonter de manière convaincante, et qui doivent provoquer sa chute, à terme, au profit d'un nouveau paradigme ou d'un nouveau programme de recherche. La contribution de R. Boudon propose de privilégier cette seconde hypothèse, et donne des pistes pour bâtir un nouveau paradigme, sur des fondements qui per-

mettent déjà de rendre compte d'une classe étendue de phénomènes sociaux. Il faudra examiner si le programme de recherche qui se dessine est capable de concurrencer l'ancien en s'avérant plus général et plus « progressif » (selon la terminologie de Lakatos), grâce à la résolution des paradoxes, et peut-être par la découverte et l'explication de faits nouveaux.

L'essai de Pierre Demeulenaere permet de prendre conscience avec lucidité des difficultés sémantiques importantes qui affectent les vocables d'intérêt et d'utilité, associés à des notions techniques qui jouent un rôle central dans la théorie traditionnelle du choix rationnel, notamment en économie. La contribution de l'auteur est critique au sens philosophique du terme. Il s'agit d'apprécier les limites qu'impose à notre compréhension du réel le recours à des notions dont le champ d'application et les relations mutuelles ne sont pas définis avec l'exactitude désirable. Les conséquences méthodologiques de cette analyse sont claires : non, les sciences sociales dans leur ensemble ne progresseront pas de manière certaine en tâchant d'appliquer à l'ensemble des phénomènes sociaux la théorie classique du choix rationnel, qui a fait ses preuves en économie dans les cas particuliers où, pour des raisons notamment institutionnelles, le jeu que jouent les acteurs les conduit en effet à rechercher la promotion d'intérêts particuliers et repérables (ainsi que le rappelle R. Boudon à propos de Tocqueville). Ce que l'on nomme parfois plaisamment l'« impérialisme » des sciences économiques contemporaines n'est donc pas justifié ; il faut même y résister, à moins que l'on songe à réviser le paradigme dominant dans les sciences économiques elles-mêmes.

Le versant constructif des propositions de P. Demeulenaere s'articule étroitement au paradigme de la sociologie cognitive au sens de R. Boudon, auquel il apporte la contribution d'une analyse fine de la structure de l'action, notamment des relations de moyens à fins, rejoignant ainsi la tradition praxéologique jadis représentée par l'école autrichienne d'économie, notamment dans le sillage de Ludwig von Mises. La théorie du choix rationnel comme optimisation est sous-déterminée ou opaque dans sa description des fins de l'action. Comme le suggère R. Boudon, elle aborde les préférences ou leur structure (par exemple l'aversion pour le risque) comme de simples faits pouvant servir à expliquer d'autres faits, mais n'ayant pas besoin d'être eux-mêmes expliqués. Or, les motifs qui orientent l'action sont inévitablement complexes, de sorte que, pour les décrire, le recours à un « intérêt » s'attachant plus particulièrement aux résultats de l'action ne pourra être seulement générique.

C'est pourtant ce que l'on observe dans la théorie du choix rationnel lorsqu'elle postule seulement l'existence de fonctions d'utilité ou d'ordres de préférences dont l'optimisation serait une norme de l'action quelle que soit la nature de celle-ci.

En réalité, l'action peut être guidée par bien d'autres choses que les résultats associés à ce qui est causalement produit par elle. Certes, il est parfois possible d'invoquer des intérêts précis, visés par les agents, pour expliquer telle classe de comportements, mais il faut alors préciser comment on acquiert la certitude que les agents sont motivés de la sorte. Le statut du théoricien doit faire partie de la théorie. P. Demeulenaere invite ainsi le lecteur à aller au-delà d'une conception purement formelle de la rationalité pratique, pour enraciner plus profondément notre compréhension de celle-ci dans la structure même de l'action, ce qui suppose tout d'abord un effort de description réaliste de celle-ci.

Ma contribution reprend, en insistant davantage sur les problèmes de modélisation des actions conjointes des personnes, le thème de la rationalité axiologique déjà abordé par R. Boudon, et tente de préciser en quel sens on peut évoquer, à partir de raisons tirées des conditions de la coexistence des personnes, l'idée de la rationalité des systèmes de normes (ou de l'acceptation rationnelle de tels systèmes). Il s'agit, tout d'abord, d'analyser dans quelle mesure on peut formuler certaines normes sociales fondamentales et caractériser leur éventuelle rationalité (au sens de la rationalité des valeurs) dans le cadre d'une description synthétique des actions simultanées qui ne privilégie pas systématiquement les « conséquences » au détriment des processus. J'espère ainsi avoir montré qu'aucun obstacle de principe n'est à opposer à la formulation conjointe, dans un même cadre descriptif simple, des principes dits « déontologiques » et des principes fondés sur les résultats ou conséquences de l'action ou de l'interaction. Il n'y aurait donc pas lieu d'opposer à ce propos deux grandes classes de raisonnements pratiques.

Dans un second temps, je m'attache à préciser quelques critères, appliqués à un certain nombre d'exemples, permettant de caractériser la rationalité axiologique des normes interpersonnelles (qu'elles soient morales, économiques, politiques, juridiques ou autres) sur la base de la résolution de conflits typiques entre grandes classes d'intentions des agents, sans frustration arbitraire des intentions d'actions. Le principe essentiel est le suivant : nombre de principes interpersonnels importants (normes ou valeurs guidant la conduite) peuvent être compris comme des solutions à des problèmes et, plus particulièrement, comme des manières de régler certaines classes de conflits. Or, ces solutions ont la propriété

singulière de ne faire porter les restrictions ou interdictions que sur des intentions capables par elles-mêmes (via leur adoption par plusieurs) de provoquer les conflits en question, ou sur des intentions qui ne sont réalisables que si ces conflits ont lieu. Ne sont pas frustrés, par contre, les projets qui, tout en étant impliqués dans les conflits à régler, sont réalisables une fois ces conflits réglés : ce sont les projets, en somme, dont on pourrait caractériser la frustration comme arbitraire.

Cette méthodologie a d'abord été développée à propos de la justification des droits de l'homme. Appliquée à divers contextes d'interaction et d'argumentation, elle permet de rendre compte du statut privilégié de normes souvent tenues pour rationnelles, sinon universellement, du moins dans des classes importantes de situations : ne pas voler, ne pas mentir, contribuer à proportion de ce que l'on reçoit, etc. C'est donc un modèle, probablement bien imparfait, de la rationalité axiologique. Il montre en tout cas qu'il est possible de relier entre eux de manière systématique, en les justifiant conjointement, un certain nombre d'exemples d'une rationalité axiologique ne se laissant pas réduire aisément à la rationalité instrumentale.

Les trois contributions suivantes développent les perspectives d'une approche en termes de rationalité cognitive dans le domaine de l'explication des phénomènes économiques et stratégiques. L'un des problèmes posés dans ces contributions est celui des relations entre la modélisation des interactions sociales et la structure des actions intentionnelles telles que les prépare ou les vise chaque agent. Les contributions de Renaud Fillieule, Thomas Piketty et Bernard Walliser nous éclairent sur le caractère extrêmement délicat, à cet égard, de la modélisation des structures d'information, mais aussi sur le rôle décisif que cette entreprise peut jouer dans le dévoilement de la logique interne de certains processus sociaux.

La contribution de Renaud Fillieule s'inscrit dans la perspective d'un dépassement du paradigme traditionnel du choix rationnel dans le contexte de l'explication des phénomènes économiques. L'économie est un domaine où l'approche classique en termes de choix rationnel optimisateur, encore très largement dominante, doit répondre aux défis venus d'autres courants de pensée, en particulier le courant « évolutionniste » issu notamment des travaux de Richard Nelson et Sidney Winter. Le modèle élaboré et étudié par R. Fillieule s'inscrit dans cette lignée, et explore les possibilités offertes à la théorie économique (ou, mieux, à la théorie sociale des phénomènes économiques) lorsqu'elle ne s'appuie plus sur les deux hypothèses cruciales de la science néo-classique : l'apu-

rement des marchés et la rationalité instrumentale. L'auteur, à travers son analyse de l'idée d'action recadrée, a su montrer la parenté qui existe entre ce type de travail et le nouveau paradigme de la sociologie cognitive.

Selon l'approche « évolutionniste », il faut s'efforcer de comprendre la vie sociale dans son épaisseur temporelle, en la saisissant comme un processus que l'on n'a aucune bonne raison d'assimiler *a priori* à une suite continue de situations d'équilibre où, dans une harmonie étonnante, le comportement de chacun serait automatiquement la réponse la plus rationnelle (au sens instrumental) à celui de tous les autres. La vie sociale – et l'intuition courante n'en disconvient pas – est plutôt un processus complexe perpétuellement perturbé par de nouvelles forces et par des événements inattendus (par exemple de nouvelles découvertes ou de nouvelles formes de sociabilité).

On ne peut négliger pour autant les gages de compréhension associés, dans certains cas, aux approches en termes d'équilibre. En effet, « comprendre » une situation sociale, n'est-ce pas donner les raisons de l'ajustement réciproque des conduites ? Théorie des jeux et science économique y excellent, et la théorie de l'équilibre économique général, de Léon Walras à Kenneth Arrow, Frank Hahn et Gérard Debreu, joue à cet égard le rôle d'un modèle particulièrement achevé. Ce courant de pensée a réussi à développer une théorie systématique et convaincante des possibilités de l'échange mutuellement avantageux. Cependant, ces approches posent un problème de pertinence descriptive. Si même il faut penser que la vie sociale passe par une suite de moments privilégiés d'« équilibre », comment éliminer l'hypothèse plausible d'un déséquilibre temporaire dans les intervalles ? De plus, le recours à des fictions théoriques telles que le commissaire-priseur de Walras menace la cohérence du programme de l'individualisme méthodologique dans les sciences sociales. Les doutes se renforcent si l'on veut bien prêter attention au fait que l'évolution sociale et économique des sociétés modifie continuellement le cadre de référence, peut-être aussi les motifs, des acteurs.

Le modèle de l'action recadrée développé par R. Fillieule repose sur cette idée simple : les agents découvrent progressivement les nouvelles conditions qui serviront de cadre à leurs calculs, à leurs prévisions, et donc à leurs décisions. Cela s'oppose au modèle complexe, lourd de paradoxes, qu'il faudrait développer pour appliquer systématiquement à la vie sociale concrète la théorie du choix rationnel la plus traditionnelle. Mais il ne s'agit pas de verser dans des explications fondées sur un postulat implicite d'irrationalité des acteurs. Au contraire, Renaud Fillieule a mon-

tré comment envisager de manière fructueuse une « rationalité du cadre », de type fondamentalement épistémologique, capable de guider les changements dans la théorisation des situations.

L'essai de Thomas Piketty constitue une synthèse épistémologique sur l'apport de l'économie de l'information, telle qu'elle s'est développée en particulier sous l'impulsion de Friedrich von Hayek, Jacob Marschak, Edmund Phelps et Roy Radner, avant d'orienter, comme elle le fait aujourd'hui, une part notable de la recherche en sciences économiques. Les décisions économiques dépendent de l'information des acteurs. Au-delà de cette évidence, l'économie de l'information est parvenue à élaborer une théorie novatrice de la dépendance de sens contraire : l'information des acteurs est en partie façonnée par les conditions de leur interaction sociale, en particulier par les possibilités de contrat et d'échange, et par les coûts associés à l'acquisition d'information. En dernière analyse, les formes particulières que prennent certaines institutions économiques et politiques importantes se laissent expliquer à partir des impératifs qu'impose la structure du partage de l'information à la poursuite individuelle de l'intérêt propre.

Ainsi, telle classe de modèles en économie financière se propose de rendre compte de l'apparition, et surtout de la permanence, d'un système bancaire, sur la base de l'analyse des conséquences de l'information dissymétrique des parties dans la négociation des contrats de prêt. D'une manière générale, les économistes théoriciens parviennent à rendre compte de l'anatomie de larges classes de contrats en synthétisant leurs propriétés les plus saillantes à partir de la structure de l'information, et surtout des modalités d'acquisition d'information, des contractants potentiels. Mais des analyses voisines permettent également de renouer avec certaines intuitions perdues et retrouvées de la théorie sociale, notamment le « théorème du jury » de Condorcet.

Bernard Walliser propose un examen d'ensemble des possibilités qu'offre la théorie des jeux contemporaine en ce qui concerne la modélisation des structures d'information. Loin des archétypes souvent critiqués de l'information complète et des prévisions parfaites, la théorie des jeux, dans ses développements récents, est l'exemple d'une prise en charge consciente et assumée de la différenciation des strates de l'information et de son partage. Elle permet aussi, notamment grâce aux liens qui se sont créés avec la théorie des automates, une modélisation approfondie des processus d'apprentissage. Enfin, le recours de plus en plus fréquent à la logique épistémique rend possible, dans ce domaine, la découverte de contraintes fortes qui résultent, pour les possibilités d'in-

teraction des agents, de la rationalité théorique de ceux-ci (leur aptitude aux inférences correctes).

La théorie des jeux fait désormais intervenir l'itération des interactions et la sélection graduelle des conduites, comme dans les travaux célèbres d'Axelrod (discutés par R. Boudon) sur la dynamique de l'adoption des normes de comportement coopératif, par lesquelles chacun se met en position de profiter de possibilités d'interaction mutuellement avantageuses, pourvu qu'autrui fasse de même. Cela pose le difficile problème de la sélection des postulats de rationalité pertinents dans de tels contextes. Plus généralement encore, la contribution de B. Walliser a pour enjeu le passage problématique de la modélisation de la rationalité individuelle à la modélisation des comportements en contexte stratégique, les aspects essentiels de l'incertitude portant notamment sur la conduite des autres agents.

Au-delà des applications à l'explication des phénomènes sociaux, cette transition présente également un intérêt philosophique considérable pour l'analyse des normes éthiques et politiques. Ainsi, le rôle des promesses et des menaces est central dans nombre d'analyses modernes du contrat social. Dois-je obéir aux clauses du contrat si je n'ai pas l'assurance qu'il sera également respecté par autrui ? Et puis-je me fier aux promesses d'y obéir que me font les autres, étant donné le type des menaces de sanctions effectivement en vigueur en cas de désobéissance ? On peut aussi songer à des situations plus concrètes. Ainsi, dans le cas de la dissuasion stratégique entre les nations, le rôle des menaces est essentiel pour assurer la stabilité des conduites. Comment, alors, modéliser la structure des croyances croisées des agents sur les intentions des uns et des autres, afin de rendre compte de manière adéquate des virtualités réelles de stabilité et des facteurs d'instabilité ? Tel est le type des questions abordées par la théorie des jeux, dont B. Walliser indique les grands enjeux actuels.

Les deux dernières contributions affrontent directement la nature de l'action et de la compréhension que nous pouvons en avoir, en l'envisageant telle qu'elle s'inscrit dans la nature et dans l'histoire. Au-delà de la concurrence entre les stratégies particulières d'explication ou de justification de l'action, il y a l'action même, telle qu'elle se déploie dans l'expérience des personnes et des peuples, telle aussi qu'elle se dévoile dans l'échange ordinaire des personnes sur leur vie privée ou leur vie commune. Ainsi, la philosophie de l'action ne saurait se réduire à la réflexion sur la méthodologie des sciences de la décision, ni même à une étude de la nature de l'action, saisie dans ses aspects événementiels et psychologi-

ques. Elle est toujours aussi, comme nous l'indiquait déjà la problématique de la découverte des raisons de l'action, une philosophie de l'intersubjectivité. L'histoire occupe à cet égard une position privilégiée, étant à la fois discipline d'explication, expérience partagée et référence commune : trois occasions d'être un lieu d'objectivité.

A partir de la comparaison aristotélécienne de la poésie et de l'histoire, Alain Boyer pose le problème de l'explication de l'action en histoire. L'histoire a paru être un genre moins élevé que la poésie, en particulier parce que cette dernière parvient au typique et s'étend au possible. Est-ce à dire que l'histoire soit condamnée à la pure description, sans ordre nécessaire ou intelligible, de ce qui s'est passé et de ce que les hommes ont fait ? Probablement non, car l'histoire peut, en dépit de la difficulté bien connue de la reconstitution des intentions des acteurs, tenter d'insérer les événements dans une problématique, en tâchant de comprendre la situation des acteurs et les réponses qu'elle appelait pour chacun d'eux. D'une certaine manière, on peut même penser que la comparaison aristotélécienne avec la poésie donne des clefs pour comprendre en quoi une forme d'imitation de celle-ci peut être profitable à l'histoire.

L'historien peut faire appel à des modèles illustrant des cas privilégiés d'intelligibilité de l'action, en faisant appel à une sorte de logique propre des situations. Il peut rechercher les causes en usant de raisonnements. Il peut, également, grâce à la connaissance des conséquences, préciser le sens de l'action mieux que ne le peut l'acteur lui-même, et composer une intrigue. L'historien, en somme, peut s'élever à la compréhension à l'égal du poète. Il semble même possible d'admettre qu'il ait en vue, sans rompre avec sa vocation scientifique, l'idéal d'unité organique du récit. Il faut seulement se garder de présenter comme une hypothèse empirique confirmée l'existence d'une contrepartie réelle pour les interprétations et les notions générales qui permettent de donner au récit historique cette unité.

Sur la base d'une telle approche méthodologique, dans laquelle la narrativité se mêle aux exigences de l'objectivité, la reconstitution des séquences d'actes et d'événements se prête à une analyse de style poppérien. Répondant à un certain nombre d'objections concernant notamment la testabilité des hypothèses (en l'absence d'expérimentation) et la prévision, Alain Boyer réaffirme la possibilité de conjuguer interprétation et science, raisonnement contrefactuel (« que se serait-il passé si... ? ») et rejet des prétendues « lois » historicistes, et enfin, singularité de l'objet et légitimité des hypothèses testables. Selon cette approche ouverte et pluraliste, l'action se prête à une reconstitution qui, si elle n'est évidemment

pas une pure construction, satisfait les exigences de la compréhension ou de l'explication comme ne pourrait jamais le faire la simple accumulation des données. Or, une telle reconstitution doit accorder une place centrale à la détermination des objectifs, des croyances ou des intentions des agents, voire de leurs « plans de vie ». L'individualisme méthodologique est donc ici très pertinent, et l'approche ici défendue, comme les contributions précédentes, incite à en retenir une version « large », privilégiant la résolution par les agents eux-mêmes (telle qu'on peut se la représenter) des problèmes de toute nature qu'ils sont amenés à se poser.

L'essai de Bertrand Saint-Sernin est une méditation sur l'opacité du vouloir, sur l'espace qui sépare la décision volontaire de l'action complète ou achevée, cette dernière s'inscrivant inévitablement dans une histoire qui échappe pour partie à celui ou celle qui en demeure cependant responsable. C'est dans cet espace, montre B. Saint-Sernin, que se noue le lien étroit entre la responsabilité et l'édification de l'homme intérieur. Comment, puisqu'il le faut, se porter volontaire pour ce qui échappe tôt ou tard à la volonté ? La chose est impossible, selon l'auteur, si l'on ne s'efforce de construire grâce à l'action une force capable de porter toujours celle-ci en quelque manière au-devant d'elle-même. Telle est en effet la vraie manière de soutenir l'action et de lui donner consistance, alors qu'elle est toujours, dans sa forme élémentaire, « prise dans un réseau d'interactions qui en modifient la portée et le sens ». L'opacité du vouloir et la difficulté de son affermissement ne laissent pas intacte la raison pratique, puisque, si la raison doit être opérante, elle doit habiter la volonté et la guider d'une manière ou d'une autre.

Ce que la raison ne peut accomplir qu'avec difficulté peut-il être réussi avec plus d'éclat par la nature ? B. Saint-Sernin pose le problème de la complétude organique de l'action, en des termes qui invitent à réfléchir sur le statut des interactions sociales au sein du monde vivant en général. N'avons-nous pas vu, dans la lignée des travaux de John Maynard Smith, la théorie des jeux s'appliquer au monde animal ? Et ne pourrait-on penser que la théorie des interactions rationnelles des hommes indique, davantage que les voies d'un perfectionnement de la raison pratique, la réalisation d'un plan de coexistence moins imparfait des êtres selon les lois de la nature ? Si, comme le veut Cournot, la nature ne réalise que partiellement ce qui est possible d'après ses propres lois, n'y a-t-il pas lieu de croire que l'agencement rationnel des actions ressortit au parachèvement de la nature, ou, pour reprendre la formule de Jean-Pierre Sérís, à la réalisation technique d'une nature « à plus forte raison » ? La constitution d'un élément stable à la source de l'action impose cependant une distance

vis-à-vis des flots d'événements naturels, ce qui amène l'auteur à conclure au dualisme au moins opératoire de la volonté et de la nature.

Au décideur dont B. Saint-Sernin avait brossé le portrait, les forces peuvent manquer. L'action elle-même s'entend à déjouer les plans de son auteur. Plus que tout, l'appel pressant à la responsabilité épuise l'acteur : exigence qui a pour corrélat la rareté, mais aussi la stabilité de l'action véritable. Il est ainsi possible de retrouver une sorte d'unité organique de l'action, lorsque les actes et les pensées, pleinement assumés, s'offrent mutuellement l'appui qui résulte du sentiment de n'avoir rien à modifier.

## *Au-delà du « modèle du choix rationnel »<sup>1</sup> ?*

### DEUX FAÇONS DE VOIR LE COMPORTEMENT DE L'ACTEUR SOCIAL

Il existe dans les sciences sociales deux perspectives contrastées sur la manière d'analyser le comportement de l'acteur social. Celle des économistes, selon laquelle le comportement obéirait à des principes précis ; celle des sociologues, selon laquelle il peut être inspiré par les principes, les motifs, les forces, les mécanismes les plus divers. Peu satisfaits par ce côté « attrape-tout » de la théorie implicite du comportement mise en œuvre par leur discipline, certains sociologues, comme J. Coleman<sup>2</sup>, ont proposé à leurs collègues, non sans succès, de rallier la perspective des économistes.

L'économie propose en effet une théorie bien articulée de la rationalité, celle héritée de la tradition économique néo-classique, qui veut que l'action vise la maximisation de la différence entre avantages et coûts qu'elle entraîne pour le sujet. Les auteurs américains la désignent par l'expression de *rational choice model*, expression qu'on peut traduire littéralement par « modèle du choix rationnel » et abrégé « MCR » ou par « modèle de l'utilité espérée ». Cette théorie a été assouplie dans les dernières décennies. C'est parce qu'il l'a jugée trop rigide pour être applicable

1. Ce texte s'appuie fortement sur ma communication (« Au-delà de la rationalité limitée ») au séminaire *Transaction et théorie sociologique*, FUL d'Arlon, 28-30 mars 1996.
2. J. Coleman, *Foundations of Social Theory*, Cambridge/London, The Belknap Press of Harvard University Press, 1990.

au monde réel que H. Simon a proposé de remplacer la notion de « maximisation » par celle de « satisfaction », et de substituer à l'image d'un acteur omniscient celle d'un acteur ne disposant que d'une information imparfaite et lacunaire, doté d'une « rationalité limitée ».

En fait, le modèle « néo-classique » et le modèle de la rationalité limitée doivent être conçus comme des cas particuliers d'une théorie générale, s'appliquant à certaines plaques de phénomènes, non à d'autres. Mais il est des phénomènes qui requièrent encore d'autres conceptions de la rationalité.

## LA SÉDUCTION DU MCR

Le MCR tient une partie de sa force de séduction à ce qu'il promet des explications dépourvues de boîtes noires : le piéton qui traverse la rue regarde à droite et à gauche, de manière à éviter de se faire écraser. Coût : une certaine perte de temps ; avantage : rester en vie. Une telle explication est entièrement dépourvue de boîtes noires. Elle est complète. Même si un biologiste pouvait se promener dans le cerveau du piéton comme entre les roues du moulin de Leibniz et nous dire ce qui s'y passe exactement lorsque le piéton hésite sur le bord du trottoir, cela n'ajouterait strictement rien à l'explication que le MCR propose de son comportement. En revanche, lorsque le sociologue nous indique que l'acteur se comporte de telle manière par un effet de socialisation, son explication débouche sur une boîte noire. Quels mécanismes la notion vague de « socialisation » recouvre-t-elle au juste ?

Cette promesse du MCR, fournir des explications sans boîtes noires, est si intéressante, elle le rend si séduisant qu'on a tendance à en abuser et à l'appliquer à des situations où il apparaît comme manquant de réalisme et par conséquent comme dépourvu de véritable intérêt scientifique. Qu'est-ce en effet qu'une explication scientifique qui ne viserait pas le réalisme ?

Ce diagnostic aboutit à une question cruciale. Le fait que le MCR promette des explications sans boîtes noires n'est pas dû au *contenu* de ses principes (maximiser, alternativement optimiser la différence entre avantages et coûts), mais au fait qu'il explique le comportement par des raisons. C'est parce qu'on peut expliquer le comportement du piéton par des raisons que l'explication est définitive, que le biologiste ne peut la

compléter. Par contraste, dès que le comportement est expliqué par des causes (n'ayant pas le statut de raisons), celles-ci créent inévitablement des boîtes noires. D'où l'intérêt de rechercher une théorie de la rationalité plus générale que celle du MCR, qui permettrait de conserver les avantages de l'explication par les raisons, tout en évitant les limitations du MCR.

Le caractère insatisfaisant pour le sociologue de la théorie de la rationalité contenue dans le MCR ou dans les versions assouplies qu'en proposent un Simon ou un Spence (prise en compte des coûts de l'information) provient, je crois, d'une raison simple, à savoir que les situations que le sociologue analyse se distinguent les unes des autres par un ensemble de traits structurels : certaines actions se développent dans un contexte où l'acteur dispose de l'information dont il a besoin ; dans d'autres cas, il sait au contraire qu'il aurait avantage à s'informer, mais aussi que cette information est coûteuse. Dans certains cas, la situation est immédiatement lisible pour l'acteur ; dans d'autres cas, il doit l'appréhender à l'aide d'une « théorie ». Dans certains cas, l'action s'inscrit dans l'instant ; dans d'autres cas, elle s'inscrit dans le temps. Dans certains cas, l'action est évaluée à partir de ses conséquences ; dans d'autres cas, à partir des principes qui l'inspirent. Pourquoi ? Dans quelles circonstances ? Ces questions soulèvent des problèmes considérables que je ne ferai qu'effleurer. Mais il suffit de les poser pour voir qu'elles dessinent un territoire qui échappe à la juridiction du MCR. Le MCR est en effet une théorie de type conséquentialiste (une action est jugée appropriée ou non à partir de ses conséquences, telles du moins que l'acteur les perçoit). Les raisons qui incitent, dans certaines circonstances, à considérer les principes plutôt que les conséquences de l'action ne peuvent donc être reconstruites dans le cadre du MCR.

Ainsi, la diversité structurelle des situations envisagées par le sociologue fait que, dans tel cas, telle théorie de la rationalité est plus pertinente et dans tel cas telle autre. Proposition complémentaire : la conception néo-classique de la rationalité, la théorie de la rationalité limitée, la théorie « incrémentaliste » (encore qualifiée de « gradualiste »), la notion wébérienne de « rationalité axiologique », s'efforcent de capter des situations typiques.

J'essaierai d'abord d'étayer à l'aide d'exemples ma thèse qu'à la pluralité des théories correspond une pluralité des situations de décision et d'action. Ensuite, je poserai la question de savoir si ces théories peuvent être fédérées en une théorie unique, qui en constituerait la matrice. Conscient de ce que cette typologie peut avoir de schématique, je distinguerai quatre théories de la rationalité et chercherai à montrer qu'elles ont chacune une valeur explicative s'agissant de grands ensembles de phénomènes.

## LA RATIONALITÉ DE TYPE R1 (LE MCR)

Certains ont proposé d'ériger le MCR en une axiomatique générale. On peut démontrer que cela est impossible<sup>1</sup>. Mais il faut d'autre part reconnaître qu'il existe de multiples situations qui s'analysent de façon convaincante à l'aide de ce type de rationalité, et qui sont loin d'être marginales. Elles apparaissent lorsque les acteurs sociaux sont effectivement invités par le contexte à agir en fonction de leur intérêt et lorsqu'ils disposent des informations essentielles pour agir en connaissance de cause.

J'évoquerai ici un exemple que j'ai évoqué ailleurs<sup>2</sup> : Tocqueville s'interroge dans *L'Ancien Régime* sur une différence macroscopique entre la France et l'Angleterre à la fin du XVIII<sup>e</sup> siècle. Bien que les physiocrates soient alors fort influents, l'agriculture française stagne. Dans le même temps, l'agriculture se modernise rapidement en Angleterre. Pourquoi ? En raison de l'absentéisme des propriétaires fonciers français. Pourquoi cet absentéisme ? Parce que la centralisation administrative fait qu'en France occuper une charge royale est plus facile qu'en Angleterre, tout simplement parce que ces charges sont plus nombreuses et par suite plus accessibles, et qu'elles donnent à ceux qui les occupent un supplément de pouvoir, d'influence et de prestige. La centralisation administrative fait que se mettre au service de l'État central représente un moyen sûr d'exercer une influence politique. En Angleterre, les charges officielles sont moins nombreuses ; de surcroît, le pouvoir local y étant beaucoup plus indépendant du pouvoir central qu'en France, la vie locale offre toutes sortes d'opportunités aux ambitieux. Le propriétaire foncier anglais est donc beaucoup moins incité à quitter ses terres pour aller à la capitale servir le roi que ne l'est son homologue français.

Dans cette analyse, les propriétaires fonciers sont dans une situation qui les invite à décider en toute autonomie de ce qu'ils jugent être de leur

1. P. Demeulenaere, « Intérêts, rationalité et explication de l'action », communication à la Journée d'études *Philosophie et modèles de l'action (Sciences de l'homme, Théorie de la décision, Philosophie)*, Paris, 28 mars 1996 (version remaniée dans ce volume : « Les ambiguïtés constitutives du modèle du choix rationnel ») ; R. Fillieule, « Frames, inferences, and rationality: some light on the controversies about rationality », *Rationality and Society*, 8 (2), 1996 ; L. Lévy-Garboua, C. Montmarquette, « Cognition in seemingly riskless choices and judgments », *ibid.* ; R. Boudon, « The "cognitivist model" : a generalized "rational choice model" », *ibid.*, 123-150.
2. R. Boudon, *La logique du social* (1979), Paris, Hachette, 1983.

intérêt, à raisonner en termes de coûts et d'avantages, et où de surcroît les informations nécessaires à la prise de décision leur sont facilement accessibles. Ils connaissent le prix des charges et les avantages qui s'y attachent. Ce n'est donc pas un hasard si Tocqueville, qui ne doit évidemment rien à la théorie économique, prête instinctivement à ses acteurs une rationalité de type R1.

On pourrait relever dans la littérature sociologique contemporaine maints exemples de même type. Ainsi, Root se demande pourquoi, au XVIII<sup>e</sup> siècle, la politique économique anglaise favorise de façon chronique les producteurs de grains, tandis qu'en France elle favorise les consommateurs<sup>1</sup>. De nouveau, il s'agit d'expliquer une importante différence macroscopique. L'explication de Root fait de cette asymétrie un autre effet de la centralisation administrative. Les « consommateurs » que sont les Parisiens voient bien qu'en manifestant dans les rues de la capitale, ils peuvent exercer une pression efficace sur le pouvoir et obtenir des prix favorables à leurs intérêts. En revanche, une manifestation à Westminster n'a guère de chances de succès, les députés des Communes étant des propriétaires fonciers soucieux surtout de ne pas décevoir leurs électeurs de province, lesquels sont, vu le caractère censitaire du suffrage, pour la plupart des propriétaires fonciers comme eux-mêmes. C'est pourquoi les « journées » d'action sont fréquentes à Paris, mais rares à Londres (raison pour laquelle il est difficile de traduire le mot en anglais). C'est pourquoi la politique économique anglaise favorise les producteurs, la française les consommateurs. Root discerne ainsi une cause de la stagnation de l'agriculture française qui vient compléter celle de Tocqueville. Mais le point qu'il importe de souligner, c'est que la théorie de la rationalité R1 apparaît ici comme naturelle parce qu'elle épouse les conditions dans lesquelles se trouvent les acteurs responsables de l'effet macroscopique que Root se propose d'expliquer : comme dans le cas de l'exemple de Tocqueville, leurs décisions s'inscrivent dans un cadre d'information complète, d'autonomie décisionnelle, où rien ne s'oppose à ce que l'acteur social se donne un objectif de maximisation de son « utilité ».

Mais, si bien des phénomènes peuvent s'analyser dans le cadre d'une théorie de la rationalité de type R1, il en existe aussi beaucoup d'autres qui ne peuvent être expliqués à partir de ce type de rationalité. Il peut être utile de mentionner quelques exemples de cas échappant de façon flagrante à la théorie de la rationalité de type R1.

1. H. L. Root, *La construction de l'État moderne en Europe : la France et l'Angleterre*, Paris, PUF, 1994.

## LES « PARADOXES » QUI DÉFIENT LA RATIONALITÉ DE TYPE R1

Les premiers à avoir montré que la rationalité de type R1 ne peut légitimement émettre de prétention à la généralité sont les économistes eux-mêmes. Allais publie en effet dès 1953 un article opposant un certain nombre de défis à la théorie économique de la rationalité<sup>1</sup> : certains comportements économiques élémentaires paraissent ne pas relever de la rationalité que la théorie économique prête à l'acteur économique. L'un de ces paradoxes est le suivant : des sujets sont invités à choisir entre un gain sûr  $G$  et une loterie d'espérance mathématique  $G + b$  ( $b$  étant une quantité positive). S'ils jouent à la loterie suffisamment longtemps, comme il leur est proposé de le faire, les sujets sont pratiquement assurés d'un gain  $G + b$ . Or la plupart d'entre eux optent pour le gain sûr. *Ils préfèrent donc moins à plus*, un résultat contradictoire avec le principe de maximisation. Ainsi, leur comportement n'est pas guidé par la seule considération de l'utilité espérée, mais peut-être par une « aversion à l'égard du risque ». A moins qu'une insuffisance de l'intuition en matière de calcul des probabilités n'entraîne l'apparition de certains « biais cognitifs » et qu'ils tendent à sous-estimer l'espérance mathématique d'une loterie.

Ces paradoxes suggèrent que, même s'agissant des comportements économiques les plus simples, le MCR se révèle parfois impuissant. La notion même d'une « aversion à l'égard du risque » implique que la psychologie utilitariste du MCR ne permet pas de comprendre des comportements économiques élémentaires ; elle suggère même qu'il faut introduire des facteurs « irrationnels », mystérieux de surcroît : que désigne en effet au juste la notion de l'« aversion à l'égard du risque » ? Ne s'agit-il pas là d'une étiquette purement descriptive, recouvrant des mécanismes sur lesquels les psychologues eux-mêmes ne paraissent pas avoir grand-chose à nous apprendre ? Les mêmes remarques peuvent être faites à propos de l'hypothèse des « biais cognitifs » : une étiquette, en aucun cas une explication. Dans les deux cas, l'explication crée une boîte noire.

Un autre paradoxe classique est celui du vote. Si l'on prend le MCR au sérieux, on ne comprend pas pourquoi les gens votent : puisque mon vote n'a qu'une chance pratiquement nulle d'influencer le résultat d'une consultation populaire, pourquoi voterais-je plutôt que de me consacrer à

1. M. Allais, « Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine », *Econometrica*, 1953, 21, 4, 503-546.

des activités plus intéressantes ou plus efficaces ? Pourtant, les gens votent. Le paradoxe du vote tient un peu, dans l'histoire des sciences sociales récentes, la place que le théorème de Fermat a tenu pendant des siècles dans l'histoire des mathématiques : régulièrement, paraissent des articles promettant de réconcilier les principes de la rationalité économique avec le fait têtue du vote.

En dehors du vote et des paradoxes d'Allais, il existe bien sûr d'autres immenses ensembles de faits qui paraissent difficilement pouvoir être expliqués par des théories de type R1. J'y reviendrai.

#### LA RATIONALITÉ DE TYPE R2 (LA « RATIONALITÉ LIMITÉE »)

Ces « paradoxes » ne sont pas directement à l'origine de la révision que H. Simon fait subir à la théorie de la rationalité. Celui-ci a été motivé dans sa réflexion plutôt par le fait que la rationalité R1 rend mal compte des décisions prises dans les organisations. C'est pourquoi il propose de substituer l'idée de « satisfaction » (*satisficing*) à celle de « maximisation » : l'acteur social cherche à obtenir, non pas les résultats les meilleurs, mais des résultats satisfaisants. En outre, Simon introduit l'idée que l'acteur ne dispose souvent que d'une information lacunaire. C'est le chercheur en heuristique qui inspire ici le théoricien des organisations.

Considérons la décision banale qui consiste à recruter une personne pour occuper un poste. Le responsable de ce recrutement aura le souci de choisir un candidat susceptible de s'adapter aussi facilement que possible à la vie de l'entreprise et aux tâches auxquelles il est destiné, mais il ne cherchera pas à débusquer la décision qui serait *la meilleure possible* de ce point de vue. Il recherchera une solution « satisfaisante », non la meilleure solution possible. D'un autre côté, le décideur est ici dans une situation où l'information est, de par la nature des choses, lacunaire. Un entretien donnera au décideur des impressions sur la vivacité de l'impétrant, ses succès scolaires et professionnels lui permettront de préciser son pronostic, des tests lui apporteront des informations supplémentaires ; mais il ne s'agira que d'indicateurs. La seule information qui intéresse vraiment le décideur est en elle-même inaccessible, puisqu'elle ne peut se révéler que dans l'avenir. Le décideur est donc confronté à une question essentielle : celle du niveau des coûts de collecte de l'information qu'il est raisonnable pour lui d'accepter. En multipliant entretiens, tests, observations, mises à

l'épreuve diverses, l'incertitude du décideur tend à décroître, mais elle ne peut s'effacer ; la recherche d'informations supplémentaires obéira à la loi des rendements décroissants, chaque unité supplémentaire de temps passé à la recherche d'informations nouvelles en rapportera moins. La recherche d'information cessera, nous dit la théorie économique, à partir du moment où le coût de la dernière information recueillie correspondra au bénéfice qu'elle engendre.

Les théoriciens du « signal » ont en effet proposé de repenser la notion de rationalité limitée dans le cadre de la rationalité R1<sup>1</sup> : ne suffit-il pas de compliquer le modèle R1 en y introduisant l'hypothèse que l'acquisition d'informations pertinentes rentre dans les coûts de la décision ? En fait, cette conciliation avec la théorie R1 n'est possible que dans des cas particuliers, car l'action est fréquemment tributaire, non seulement d'« informations », mais d'une théorisation par l'acteur de la situation de décision dans laquelle il se trouve.

La notion de « rationalité limitée » est souvent illustrée par le cas du joueur d'échecs. Toute partie d'échecs est associée à un arbre logique. Si les joueurs pouvaient le consulter, l'incertitude sur l'issue de la partie disparaîtrait. Mais l'arbre est si ramifié que sa réalisation et sa consultation excèdent les capacités humaines. Faute de pouvoir effectivement s'y référer, le joueur va donc utiliser des détours plus ou moins subtils. On constate sur cet exemple qu'en parlant de « rationalité limitée », Simon ne vise pas seulement les situations où l'acteur souffre d'un déficit d'information. Le joueur d'échecs ne peut maîtriser la situation de décision dans laquelle il se trouve en recherchant des informations supplémentaires ; il doit plutôt échafauder une véritable théorie (composée dans ce cas de *rules of thumb* telles que « consentir à perdre un fou plutôt qu'une tour », etc.). Mais peut-on alors conserver le cadre du MCR ?

### LA RATIONALITÉ DE TYPE R3 (LA « RATIONALITÉ COGNITIVE »)

On peut en effet et l'on doit sans doute aller plus loin que Simon. Car l'acteur social peut ne pas être confronté seulement à un simple déficit d'information, mais être placé en face d'un problème qu'il doit

1. M. Spence, « Competition in salaries and signalling prerequisites for jobs », *Quarterly Journal of Economics*, vol. XC, n°1, 1976, 51-74.

résoudre à l'aide d'une *théorie*, mobilisant pour cela les ressources cognitives dont il dispose, bref, agissant, si l'on consent à mettre à part les différences de degré, comme le fait le scientifique. Dans de nombreuses situations, les acteurs ne peuvent en d'autres termes déterminer la solution la meilleure possible à leurs yeux au problème auquel ils sont confrontés qu'en esquissant, fût-ce sur le mode intuitif, une théorisation du problème. Plus techniquement, ils tentent de trouver un système de raisons leur apparaissant comme satisfaisant eu égard au problème qui leur est posé. *La rationalité de l'acteur banal rappelle alors davantage celle qu'évoquent les philosophes des sciences que celle de l'économie néo-classique.* Ce qu'il s'agit ici de maximiser ou d'optimiser, ce n'est plus une différence entre coûts et avantages, mais la force d'un système d'arguments. C'est pourquoi il est utile de s'arrêter un instant sur le sujet de la rationalité des croyances scientifiques.

S'agissant des croyances scientifiques, on part généralement du principe que ceux qui les endossent ont des raisons fortes de le faire ; la décision d'opter pour une théorie scientifique plutôt que pour une autre est déterminée par des raisons et ces raisons sont perçues par l'acteur, en l'occurrence le scientifique, comme fortes. Ainsi, pour reprendre un exemple rendu célèbre par Kuhn, lorsque Priestley croit en l'existence du phlogistique, il a le sentiment que sa croyance est fondée sur des raisons fortes. *Raisons fortes ne veut pas dire raisons vraies* (plus personne ne croit aujourd'hui à la théorie du phlogistique). Mais « raisons fortes non vraies » ne veut pas dire non plus « raisons liées à des idiosyncrasies personnelles ». A l'évidence, Priestley n'était pas le seul à croire au phlogistique. C'était au contraire une croyance répandue. Et elle était répandue parce que, étant donné l'état du savoir, les arguments de Priestley représentaient des arguments forts. Ils le restèrent jusqu'à ce que l'accumulation d'autres considérations, encore inconnues à l'époque, affaiblisse irréversiblement les arguments de Priestley au profit de ceux de Lavoisier. Les décisions de Priestley (endosser la théorie du phlogistique) et de Lavoisier (ne pas l'endosser) sont contradictoires entre elles ; l'avenir devait démontrer que l'une des croyances était fautive et l'autre vraie ; mais dans le contexte dans lequel la « décision » de Priestley et de Lavoisier est prise, l'une est aussi rationnelle que l'autre. Surtout, il s'agit d'un type de rationalité qui ne se ramène ni à la rationalité de type R1 ni à la rationalité de type R2.

On retrouve ici une des intuitions les plus étonnantes, les plus justes et les plus fécondes sans doute de Durkheim : les théorisations du « primitif », avance-t-il, obéissent aux mêmes ressorts que celles du

savant<sup>1</sup>. Il faut surmonter notre étonnement devant l'« irrationalité » apparente du primitif ; telle croyance à laquelle le « primitif » tient et qui nous paraît dépourvue de sens ne l'est pas *pour lui*. Retraduisant son idée dans les termes de mon exemple : aussi bizarres soient-elles, les croyances du primitif ne s'expliquent pas autrement que celles de Priestley (qui ne sont pas bizarres mais fausses) ou que celles de Lavoisier (qui ne sont ni bizarres ni fausses). Dans tous les cas, la croyance s'explique parce qu'elle « fait sens » pour celui qui l'endosse, parce que le croyant a des raisons fortes d'accepter sa croyance, étant entendu que des raisons qui sont fortes aujourd'hui peuvent cesser de l'être demain (si le contexte cognitif se modifie), et que ce qui est vrai ici ne l'est pas forcément là (si le contexte cognitif n'est pas le même ici et là).

L'histoire de la science fait donc clairement apparaître comme banale la situation où s'installe dans les milieux scientifiques la croyance que « X est vrai » parce qu'il existe des raisons fortes, bien que fausses, fragiles ou douteuses d'y croire, étant donné l'état de la science *hic et nunc*. Ici, des raisons fortes et partagées sont bien les causes de la croyance collective.

Ce qui est vrai des croyances scientifiques peut l'être aussi des croyances ordinaires : on les analyse souvent de façon satisfaisante comme le produit de raisons qui apparaissent au sujet comme fortes, étant donné le contexte cognitif qui est le sien. Cette rationalité de type R3 (« rationalité cognitive ») prend au sérieux l'idée que l'on ne peut se contenter, sauf cas particulier, de concevoir l'action comme située dans un contexte d'information lacunaire. Souvent, l'acteur doit construire, sur un mode plus ou moins intuitif, une « théorie » lui permettant de faire face à une situation de décision. Cette théorie, il l'endossera s'il a l'impression qu'elle est fondée sur des raisons fortes.

J'ai analysé ailleurs de multiples résultats surprenants de la psychologie cognitive à l'aide de ce schéma<sup>2</sup>. Les expériences de la psychologie cognitive présentent un grand avantage : les croyances fausses qu'elles produisent ne peuvent s'expliquer ni par des causes affectives, ni par des effets de socialisation, d'imitation ou de contagion. Souvent, les psychologues cognitifs eux-mêmes rendent compte des résultats de leurs expériences en introduisant des hypothèses « de type Lévy-Bruhl », en postulant que la pensée ordinaire obéit à des règles « prélogiques » ou qu'elle

1. Comme l'a fort bien vu R. Horton, « Lévy-Bruhl, Durkheim and the scientific revolution », in R. Horton, R. Finnegan, *Modes of Thought*, London, Faber, 1973. Voir aussi R. Boudon, *L'art de se persuader* (1990), Paris, Fayard./Le Seuil, coll. « Points », 1992.

2. R. Boudon, *L'art de se persuader*, op. cit. ; R. Boudon, *Le juste et le vrai*, Paris, Fayard, 1995.

est sujette à l'effet de « biais cognitifs » (dont l'origine reste bien sûr très mystérieuse). Or, dans la totalité des cas, on peut expliquer ces résultats comme on analyse les croyances des scientifiques : vraies ou fausses, elles sont fondées sur des raisons fortes. Et c'est parce que ces raisons sont fortes étant donné le contexte cognitif, qu'une majorité de répondants donnent la même réponse. Ces expériences de la *psychologie* cognitive intéressent donc directement le *sociologue*, puisqu'elles créent *in vitro* des croyances collectives.

Je me contente de reprendre un de ces exemples. Imaginons une situation où l'on interroge des psychiatres, nous dit Tversky. S'ils apparaissent comme persuadés que la dépression est cause de tentatives de suicide, lorsqu'on leur demande d'où ils tirent cette certitude, ils répondront que, selon leur expérience, il n'est pas rare d'observer les deux traits chez les patients qu'ils reçoivent dans leur cabinet<sup>1</sup>. (Tversky synthétise ici à partir d'une expérience fictive les résultats d'expériences réelles.)

Bien entendu, la réponse révèle que les répondants utilisent une seule information dans le tableau de contingence ci-après. Leur argument est en effet : « Si  $a$  est élevé, il existe une relation de causalité entre dépression et tentative de suicide. » Or tout apprenti en statistique sait bien que l'argument est faux, en principe du moins. Pour conclure à une relation de causalité entre dépression et tentative de suicide, il faut considérer, non une, mais quatre informations élémentaires, non seulement  $a$  mais la différence entre deux rapports :  $a/e - c/f$ . Selon Tversky, la réponse des psychiatres révélerait que le cerveau est câblé selon des principes erronés et que l'intuition statistique contredit les règles de la statistique.

L'hypothèse explicative est à la fois lourde et fragile. Pourquoi notre intuition statistique serait-elle par principe défaillante, alors que notre intuition géométrique ou arithmétique ne l'est pas ? Si l'intuition statistique est le résultat d'une évolution phylogénétique<sup>2</sup>, pourquoi celle-ci aurait-elle (en contradiction avec les théories de l'évolution) sélectionné des principes non valides ? En fait, on peut interpréter la réponse des psychiatres d'une tout autre façon et montrer qu'ils ont des raisons fortes de croire ce qu'ils croient.

Supposons que  $e$  dans le tableau ait une valeur réaliste : qu'il soit par exemple égal à 20 %, en d'autres termes que 20 % des patients des psychiatres manifestent des symptômes de dépression. Cette valeur est pro-

1. D. Kahneman, A. Tversky, « Availability : a Heuristic for Judging Frequency and Probability », *Cognitive Psychology*, 5, 1973, p. 207-232.
2. Une hypothèse fréquemment avancée par la psychologie cognitive.

bablement un plafond. Supposons par ailleurs que  $g$  soit aussi égal à 20 % (20 % des patients ont commis une tentative de suicide) : également une valeur sans doute maximum. Avec ces hypothèses, si le pourcentage  $a$  de personnes présentant les deux caractères était supérieur à 4, les deux variables seraient corrélées entre elles et l'on pourrait interpréter la corrélation comme une présomption de causalité. Ainsi, un psychiatre qui aurait observé que, sur 100 patients, 10 présentent les deux caractères aurait des raisons fortes de présumer que la dépression est cause de tentative de suicide.

	<i>Tentative de suicide</i>	<i>Pas de tentative de suicide</i>	<i>Total</i>
Symptômes de dépression	$a$	$b$	$e = a + b$
Pas de symptômes de dépression	$c$	$d$	$f = c + d$
Total	$g = a + c$	$h = b + d$	$i = a + b + c + d$

Cette expérience est particulièrement instructive. Si l'on suit mon analyse, les psychiatres s'appuient, de façon intuitive, sur une analyse subtile, tenant implicitement compte des valeurs marginales *plausibles* du tableau. Si l'on a des présomptions sur les limites des valeurs marginales, une seule information, comme la valeur de  $a$ , peut effectivement fonder une présomption de causalité. Les expérimentateurs, eux, perçoivent les résultats de leur expérience avec les ressources cognitives qui sont les leurs : ils appliquent mécaniquement leur savoir statistique. Les raisons des psychiatres, bien que non élucidées, sont donc sans doute plus solides que celles de leurs juges : nouvelle version de l'arroseur arrosé. En tout cas, les cobayes comme leurs observateurs fondent leur impression sur des raisons fortes. Mais l'on note aussi que les deux groupes ont des appréciations différentes parce qu'ils sont dans des situations différentes d'un point de vue cognitif. On retrouve bien, dans le contexte de la « connaissance ordinaire » une opposition comparable à l'opposition Lavoisier-Priestley.

Ce type d'analyse peut être facilement illustré, non seulement par des exemples tirés de la sociologie cognitive, mais tout aussi facilement par d'innombrables exemples sociologiques<sup>1</sup>.

1. R. Boudon, *L'idéologie ou l'origine des idées reçues* (1986), Paris, Fayard/Le Seuil, coll. « Points », 1991, et *L'art de se persuader*, *op cit.*

Après la Seconde Guerre mondiale, le keynésianisme fait figure de dogme<sup>1</sup>. L'inflation est chronique ; le keynésianisme fournit des recettes techniques permettant de lutter contre l'inflation. En augmentant la fiscalité, nous dit Keynes, on ponctionne le pouvoir d'achat, on réduit la demande et on comprime les prix. Donc, un alourdissement de la fiscalité a des effets *déflationnistes*. On demande alors à des responsables économiques ce qu'ils en pensent dans une enquête par questionnaire. A une forte majorité, ils répondent, en contradiction avec le dogme régnant, que l'augmentation de la fiscalité est génératrice d'inflation. Cette croyance est fondée sur des raisons fortes : elle est même *vraie* au niveau local. Pour répondre à la question posée, les entrepreneurs évoquent naturellement leur expérience : « Si la fiscalité augmente, mes coûts de production vont augmenter, se disent-ils ; j'essaierai donc, comme mon concurrent, de transférer cette augmentation sur les prix. *Ergo*, l'augmentation de la fiscalité est cause d'inflation. » On peut ajouter que cette « vérité locale » peut devenir générale, s'agissant des produits caractérisés par une élasticité nulle de la demande. A supposer qu'il n'y ait que des produits de ce genre, la vérité serait même vraie dans l'absolu.

On a là une situation comparable à celle, que j'évoquais tout à l'heure, des discussions sur le phlogistique à la fin du XVIII<sup>e</sup> siècle, au sens où l'on observe des croyances incompatibles de la part de diverses catégories d'acteurs, toutes étant fondées cependant sur des raisons fortes. La différence entre les deux cas est que la question de l'existence du phlogistique est dépourvue d'ambiguïté, alors que la question de la relation de causalité entre fiscalité et inflation est ambiguë : elle suggère l'existence d'une loi, alors qu'il ne s'agit que d'une relation conditionnelle. Ou le phlogistique existe ou il n'existe pas ; en revanche, un alourdissement de la pression fiscale peut, *selon les cas*, être inflationniste ou au contraire déflationniste.

En tout cas, cet exemple démontre bien que le caractère contextuel de certaines croyances, que l'apparition de croyances incompatibles entre elles sur un sujet n'interdisent pas d'y voir l'effet de raisons fortes. Les explications en termes de rationalité de type R3 sont préférables aux explications faciles et creuses qui font des croyances des entrepreneurs les effets d'un câblage cérébral propre aux responsables économiques, lequel serait lui-même l'effet de mécanismes biologiques, sociologiques ou psychologiques conjecturaux.

1. R. Boudon, *L'idéologie*, *op. cit.*

Comme les croyances scientifiques, toutes ces croyances plus ou moins solides qui sont constamment mobilisées dans l'action quotidienne s'analysent avantageusement comme le produit de raisons fortes, en d'autres termes comme l'effet de ce que j'appelle ici la « rationalité cognitive » (type R3). Il n'y a bien souvent entre les systèmes de raisons qui fondent ces croyances « ordinaires » et celles qui fondent les croyances scientifiques qu'une différence de degré, non de nature.

#### RAISONS DE CHOISIR ENTRE DES SYSTÈMES DE RAISONS CONCURRENTES

Les travaux d'Axelrod<sup>1</sup> font apparaître une autre plaque de phénomènes pour lesquels les théories de la rationalité R1 et R2 sont inopérantes. Dans son livre sur les dynamiques de la coopération, Axelrod suppose que deux joueurs jouent à un jeu du dilemme du prisonnier (DP) répété. Dans un jeu de DP simple, un individu rationnel (au sens R1 qui est celui de la théorie des jeux) est incité à jouer sa stratégie de « défection », laquelle aboutit à un résultat sous-optimal. Dans un jeu de DP répété, comme le montre Axelrod, les individus jouent plutôt la stratégie du donnant-donnant : je coopère tant que l'autre coopère aussi, sinon je fais immédiatement défection.

Au fond, Axelrod suppose que le sujet se tient le raisonnement suivant : « Si je joue ma stratégie de “défection” j'évite le pire, mais si l'autre joue aussi sa stratégie de “défection” parce qu'il veut éviter le pire, nous obtenons tous deux un résultat peu satisfaisant, moins bon en tout cas que celui que nous obtiendrions en coopérant. Donc, je fais l'hypothèse que l'autre percevra cette difficulté, et qu'il jouera plutôt sa stratégie de “coopération”, car il espérera que j'aie vu moi-même la difficulté, et que par suite je fasse de même. » Bref, chacun des deux joueurs parie sur l'intelligence de l'autre : voyant qu'un choix rationnel à courte vue engendre un résultat peu satisfaisant, chacun fait l'hypothèse que, par une coopération tacite, l'autre fera également le choix de la coopération. Cela revient à dire que les individus sont ici, non pas « rationnels » au sens de l'axiomatique de la théorie des jeux (type R1), mais intelligents ! En termes plus

1. R. Axelrod, *The Evolution of Cooperation*, Harmondsworth, Penguin, 1990 ; trad. franç. M. Garène, *Donnant donnant : une théorie du comportement coopératif*, Paris, O. Jacob, 1992.

précis, il existe un type de rationalité qui, dans certaines circonstances, contredit les prédictions qu'on peut effectuer sur le comportement des acteurs sociaux à partir de R1. Cette rationalité n'est autre que la rationalité « cognitive » de type R3 qu'on vient de définir. Elle repose sur une théorisation de la situation. Le choix découle d'une argumentation forte au sens où chacun des arguments entraînerait un assentiment immédiat, par exemple : « il est absurde d'obtenir un résultat médiocre quand on peut en obtenir un meilleur », « sans doute l'autre est-il de mon avis sur ce point », « comme moi, il va voir que la coopération produirait un résultat qui nous est favorable à tous deux », etc.

En interprétant le comportement des joueurs comme fondé sur des choix raisonnés (*i.e.* reposant sur des raisons fortes pour l'acteur), on rend compte de la variabilité des solutions qu'on observe dans la réalité sociale lorsque des acteurs sont dans la situation de jouer à un jeu de DP répété.

Dans les cas qu'envisage Axelrod, la stratégie du « donnant-donnant » émerge facilement. Toutefois, si on en fait le résultat des raisons fortes que je viens de décrire, on voit aussi que ces raisons ne seront perçues par l'acteur comme fortes que dans le cas où les risques qu'il encourrait au cas où l'autre choisirait la « défection » là où lui-même choisirait la « coopération » ne dépasseraient pas un certain seuil. C'est pourquoi la stratégie du « donnant-donnant » n'a pas fonctionné du temps de la guerre froide. Ici, les deux acteurs jouèrent de façon répétée la stratégie de la « défection » (accepter la course aux armements), car il était trop risqué, eu égard à l'importance des enjeux, d'abaisser sa garde tant qu'on n'avait pas la certitude que l'autre l'abaisserait aussi. Axelrod n'évoque guère ces cas de jeux de DP répété où la stratégie du « donnant-donnant » n'est *pas* choisie dans la réalité et qui, le cas de la guerre froide le démontre, sont loin d'être d'une importance marginale.

Ainsi, lorsque les risques de la coopération sont trop grands, les acteurs cherchent surtout à « éviter la catastrophe ». Lorsque les risques sont plus modérés, ils tendent à utiliser une rationalité plus subtile. Dans le premier cas, la magnitude du risque fait que l'acteur a des raisons fortes de le minimiser. Dans le second, le risque étant modéré, l'acteur a des raisons fortes de minimiser plutôt les regrets qu'il encourrait en évitant tout risque. Dans le premier cas, le critère dit « de Wald » lui apparaîtra comme pertinent ; dans le second, ce sera le critère dit « de Savage ». En d'autres termes, l'acteur théorise la situation : faut-il être « waldien » ou « savagien » ? Faut-il minimiser les risques ou les regrets éventuels ? La réponse dépend bien entendu de la grandeur relative du risque et des regrets. Pour

revenir à la course aux armements, elle s'est effectivement terminée, non parce que les duellistes, États-Unis et URSS, auraient appliqué la stratégie du « donnant-donnant » avec persévérance, mais par la menace de la « guerre des étoiles ». Car, avant que n'apparaisse cette fin heureuse, les duellistes avaient des raisons fortes de chercher avant toutes choses, comme tout gouvernement *doit* le faire, à minimiser les risques encourus par les citoyens, quitte à chercher simultanément les moyens de sortir de la structure de DP. L'importance des risques était telle que la solution ne pouvait venir dans ce cas que d'une innovation ayant pour effet de briser la structure du système d'interaction : ce fut le résultat du gigantesque coup de bluff de Reagan.

Le choix stratégique suivi par les grandes puissances du temps de la guerre froide pourrait, on le voit, être présenté comme la conclusion d'un raisonnement composé d'une suite de propositions fortes, bien articulées entre elles : « Un gouvernement ne peut exposer la population à un risque sérieux s'il a les moyens de l'éviter ; une stratégie "donnant-donnant" n'est pas dépourvue de risque ; on peut l'éviter en jouant la stratégie alternative de "défection" ; mais le résultat est "sous-optimal" ; il faut donc chercher les occasions, s'il en est, ou imaginer des innovations, si l'on peut, conduisant à la seule solution satisfaisante : briser à terme la structure de DP. »

Cette discussion d'Axelrod aboutit à une conclusion essentielle : confrontés à une situation de type « DP répété », les acteurs doivent se demander s'il faut minimiser les risques ou les regrets éventuels. C'est seulement lorsqu'ils auront répondu à cette question qu'ils pourront déterminer le meilleur moyen pour y parvenir. Or, à cette question, une théorie de type R1 n'apporte pas de réponse (sauf à introduire l'hypothèse très théorique d'une commensurabilité entre risques et regrets). Elle peut seulement déterminer les meilleurs moyens d'atteindre l'objectif (minimiser X ou Y), une fois que celui-ci est fixé. Il ne s'agit donc pas ici seulement pour les acteurs de déterminer les moyens les meilleurs pour satisfaire des préférences supposées données, mais de déterminer l'objectif le plus souhaitable et de le faire à partir de raisons fortes. Ici encore, les acteurs « théorisent » la situation avant de se déterminer selon les principes de la théorie de la rationalité de type R1.

Les travaux des « incrémentalistes » amèneraient aux mêmes conclusions que ceux d'Axelrod. Face à certaines situations, l'acteur, nous disent Braybrooke et Lindblom<sup>1</sup>, tente de gagner du temps, il tâtonne, essaie des

1. D. Braybrooke, C. E. Lindblom, *A strategy of decision, policy evaluation as a social process*, New York, The Free Press/Londres, Collier-Macmillan, 1963.

conjectures alternatives sur les motivations d'autrui, etc. Dans ces situations complexes, l'acteur procède par une suite de théorisations conjecturales de la situation dans laquelle il se trouve, essayant d'aboutir à un état où un système de raisons fortes à ses yeux justifiera son choix.

#### LA RATIONALITÉ DE TYPE R4 (LA « RATIONALITÉ AXIOLOGIQUE »)

Il est des cas où toute théorie s'appuyant sur une rationalité de type R1 conduit à des prédictions qui se trouvent brutalement démenties par le réel. J'ai déjà évoqué le « paradoxe du vote ». La théorie R1 nous dit que l'électeur ne devrait pas voter. *Et pourtant il vote*, comme aurait dit Galilée.

Un mot au sujet de ce paradoxe, qui a fait couler beaucoup d'encre. Il peut paraître un peu artificiel, mais il pose un problème sérieux, qui concerne la méthodologie des sciences sociales dans son ensemble.

Encore une fois, le paradoxe résulte de ce que, si on se donne une axiomatique de type R1 supposant le sujet égoïste, soucieux de maximiser l'utilité espérée de ses décisions, on ne comprend pas pourquoi il vote. A supposer même qu'il éprouve une passion pour le candidat A et un sentiment de répulsion pour son adversaire B, il devrait voir que la probabilité que son bulletin en faveur de A fasse basculer le résultat de la consultation en faveur de A est nulle. Diverses « solutions » tentent de venir à bout de ce paradoxe.

Les adversaires de la théorie du choix rationnel y ont vu tout simplement la preuve de l'irrationalité de l'acteur social. Les gens votent parce qu'ils *sont* irrationnels. C'est la manière la plus facile de résoudre le paradoxe du vote. Quand on ne sait pas expliquer tel comportement, il suffit de déclarer que celui qui s'y livre obéit à des motivations irrationnelles. Les électeurs votent parce qu'ils aiment voter, nous dit-on, et le paradoxe du vote n'existe que par l'entêtement des théoriciens du choix rationnel. Comme souvent, ces explications de caractère irrationnel sont à la fois faciles et inconsistantes ; elles renvoient à des conjectures difficilement testables et imaginent des mécanismes plus ou moins improvisés pour rendre compte de ce qu'elles cherchent à expliquer : au lieu d'éclairer l'*explanandum*, elles multiplient les boîtes noires.

C'est parce qu'ils sont conscients des faiblesses de ces explications irrationnelles que les théoriciens du choix rationnel se sont acharnés à trouver des solutions plus satisfaisantes. Ainsi, dans l'article le plus subtil

sans doute qui ait été produit sur la question, Ferejohn et Fiorina<sup>1</sup> ont proposé une solution qui évoque le pari de Pascal : même s'il y avait peu de chances que Dieu existe, mieux vaut parier qu'il existe ; même si mon vote a peu de chances d'être décisif, j'aurais des regrets si grands s'il s'avérait l'être, que je vote par précaution, d'autant plus facilement que les « coûts » du vote sont faibles. Le vote serait ainsi une assurance à très bas prix permettant de couvrir des risques très improbables, mais aux enjeux considérables. L'assurance contre l'incendie n'est pas obligatoire. Elle n'a pas lieu de l'être, car beaucoup y souscrivent. Ils le font parce que, le risque étant faible, le coût individuel est faible, alors que l'enjeu individuel est considérable. Le risque a bien ici une structure de pari de Pascal : coût faible de l'assurance, regrets intenses si l'assurance n'a pas été contractée. La facilité avec laquelle on souscrit à l'assurance-incendie démontre donc que les circonstances incitant l'acteur social à faire le « pari de Pascal » n'ont rien d'exceptionnel. On peut observer à titre de comparaison qu'il est d'autres situations de risque qui échappent à cette figure : ainsi, le coût des assurances-automobile est si élevé qu'il faut ici contraindre le public à les contracter. La difficulté de l'explication proposée par Ferejohn et Fiorina est toutefois que, dans le cas du vote, le risque d'être exposé à des regrets est inexistant, puisque la probabilité qu'une voix quelconque soit décisive est pratiquement nulle. Or un nombre considérable de gens votent, même dans le cas où ils ont l'impression que les candidats en présence soit ne se distinguent guère entre eux, soit sont peu attrayants<sup>2</sup>. L'explication de Ferejohn et Fiorina est brillante mais peu convaincante.

D'autres ont voulu résoudre le « paradoxe du vote » en introduisant l'idée que l'abstention entraînerait des coûts du point de vue de la réputation sociale, dus à ce que l'abstention est mal vue<sup>3</sup>. L'abstentionniste risquerait sa réputation. Il s'exposerait ainsi à des coûts sociaux qu'il préférerait éviter. L'astuce consiste ici à conférer un coût à l'abstention. Ce qui permet d'expliquer que les gens votent tout en conservant la théorie éco-

1. J. A. Ferejohn, M. P. Fiorina, « The paradox of not voting : A decision theoretic analysis », *The American Political Science Rev.*, 68, 2 juin 1974, 525-536.
2. On pourrait facilement imaginer, en prolongeant les hypothèses de Ferejohn et Fiorina des modèles mathématiques (ou des modèles de simulation) complexes : si je pense que les autres, étant rationnels, s'abstiendront de voter, mon vote devient décisif. Mais chacun peut raisonner de même. Si tel est le cas, aucun vote ne sera décisif, mais tout le monde voterait, etc. On peut alors imaginer des variantes avec effets de miroir, etc. Il n'est pas sûr que de tels exercices nous fassent beaucoup avancer, puisqu'on obtient une solution beaucoup plus simple et congruente avec les données de la conscience en sortant du cadre de la rationalité de type R1.
3. E. Overbye, « Making a case for the rational, self-regarding, "ethical" voter... and solving the "Paradox of not voting" in the process », *European Journal of Political Research*, 27, 1995, 369-396.

nomique du vote. Outre que cette interprétation est *ad hoc*, elle est défectueuse dans sa construction même : pourquoi faut-il que le public considère l'abstention d'un mauvais œil s'il est composé d'individus rationnels qui devraient voir qu'il est inutile de voter ?

D'autres encore ont essayé de réconcilier la théorie R1 avec le fait du vote en distinguant la « causalité-diagnostic » de la « causalité instrumentale ». Après Pascal, Calvin... Le calviniste est convaincu qu'il est soit élu soit damné, Dieu en ayant décidé de toute éternité. S'il est élu, le calviniste sera vertueux *et* il ira au paradis. La vertu n'a aucun effet réel sur son destin dans l'au-delà, mais, s'il se comporte de façon vertueuse, il peut avoir l'impression d'être élu. Vertu et élection sont liées par une relation de « causalité-diagnostic ». Mais il y a plus : on peut faire l'hypothèse que le dogme calviniste incite davantage à la vertu que le dogme catholique. Le catholique admet que la vertu a un effet sur le destin de l'individu dans l'au-delà : Dieu tient compte de son comportement ici-bas. Il en résulte qu'une faute qu'il commet aujourd'hui peut être rachetée demain par une bonne action. En revanche, le calviniste peut redouter qu'une seule faiblesse ne soit le signe de sa damnation éternelle, Dieu n'étant guère porté à envoyer de faux signaux. L'électeur traiterait en fin de compte son propre vote comme ayant valeur de diagnostic. S'il vote, c'est que les électeurs du même bord que lui vont voter. S'il ne vote pas, c'est qu'ils risquent de ne pas voter. Même s'il sait que son destin dans l'au-delà ne dépend que de Dieu, le calviniste est incité à la vertu parce que celle-ci a pour lui une valeur de « diagnostic ». Même s'il sait que son vote n'affectera pas l'élection, le partisan de A votera parce qu'il aura l'impression que son vote a valeur de diagnostic quant à la probabilité pour que votent les partisans de A. La psychologie que nos auteurs prêtent aux calvinistes est certes plausible, mais son application au cas du vote paraît un peu artificielle. Enfin, à supposer qu'elle ait une validité quelconque, l'explication ne sauverait pas la théorie du choix rationnel pour la simple raison qu'elle échappe entièrement à ce cadre<sup>1</sup>.

Bref, aucune des théories proposant de résoudre le paradoxe du vote n'est bien crédible. Est-ce à dire qu'il faille en donner une interprétation irrationnelle ? Il existe sans aucun doute des gens qui aiment voter, des gens que l'accomplissement de leur « devoir électoral » tire de leur ennui. Mais il serait hardi de supposer que tous les électeurs votent simplement parce qu'ils aiment voter. De surcroît, une telle théorie n'expliquerait pas

1. G. A. Quattrone, A. Tversky, « Self-deception and the voter's illusion », in J. Elster (ed.), *The Multiple Self*, Cambridge, Cambridge University Press, 1987, 35-58.

les faits de conscience qui accompagnent le vote, lesquels sont aussi des *faits*. Les gens déclarent en effet qu'ils votent parce qu'il *faut* voter, non parce qu'ils *aiment* voter. Une théorie affective du vote (« les gens votent parce qu'ils aiment voter ») non seulement serait facile ; elle ne serait complète qu'en acceptant une hypothèse très lourde, celle de la « fausse conscience » : les gens pensent qu'ils *doivent* voter alors qu'en fait ils *aiment* voter, mais ils ont une conscience fautive de leurs motivations. Ce qui amènerait à la question difficile de savoir pourquoi les gens ont une conscience fautive de leurs motivations. Les théories du type « les gens votent parce qu'on leur a appris à voter » serait exposée aux mêmes objections : théorie « facile », *ad hoc* et, en même temps, contradictoire avec les données de la conscience. Pour utiliser les catégories de Weber, il existe bien des actions « affectives » et des actions « traditionnelles » ; mais en s'engageant sur ces pistes dans le cas du paradoxe du vote, on produit des théories qui, au vu des critères habituellement utilisés pour juger de la validité d'une théorie scientifique, apparaissent immédiatement comme défectueuses. Une règle essentielle de la déontologie scientifique impliquant qu'on ne se débarrasse pas d'un ensemble de faits sous prétexte qu'ils refusent de rentrer dans une théorie, on ne voit pas comment on pourrait ignorer que la plupart des gens déclarent voter parce qu'il *faut* voter.

Bref, la rationalité instrumentale échoue à expliquer le vote, mais on ne peut y voir non plus l'effet d'actions « traditionnelles » ou d'actions « affectives ». Je laisse de côté les hypothèses de type « biais cognitif », qui ne sont pas prises en compte dans la typologie wébérienne. Ici, elles prendraient la forme : en raison de certains biais cognitifs résultant d'un câblage cérébral défectueux, les électeurs tendent à surestimer la probabilité pour que leur vote soit « pivotal ». Reste la quatrième catégorie wébérienne, celle de la « rationalité axiologique » (*Wertrationalität*) : les gens votent parce qu'ils estiment qu'ils doivent voter. Et ils croient qu'ils doivent voter parce qu'ils ont des raisons fortes de le croire, bien que ces raisons ne soient pas celles qu'envisage la rationalité instrumentale.

Le paradoxe du vote peut finalement s'expliquer de la manière suivante. Les gens ont des raisons de croire que la démocratie est un bon régime. Les élections ont pour but de faire que les gouvernants tiennent compte de la volonté du public. Ce résultat n'est pas assuré, mais il est plus probable dans ce type de régime. Les élections sont donc une bonne chose. Si personne ne votait, cette bonne chose cesserait d'exister. Il faut donc que chacun vote. Mon abstention n'a aucun effet sur le nombre de votants ; elle est sans conséquence. Mais elle contredit le principe « il faut voter ». Si j'ai une préférence claire, si j'ai le sentiment

que telle équipe gouvernante serait meilleure, je dois l'exprimer, comme doivent le faire les autres. Sans doute n'ai-je qu'un contrôle pratiquement nul sur le résultat de la consultation, mais en adoptant le principe « il faut voter » et en l'appliquant (sauf si j'ai des raisons fortes de ne pas l'appliquer, au cas où j'aurais par exemple l'impression qu'aucun candidat n'est valable et où je voudrais le signifier par mon abstention), je mobiliserai ma part de responsabilité et ferai ce qui est en mon pouvoir pour que la consultation ait effectivement lieu. La rationalité axiologique nous dit que les électeurs votent parce qu'ils perçoivent le vote comme important. Comme dans le cas d'Axelrod, la rationalité R1 est corrigée par une rationalité supérieure. Des raisons fortes conduisent en d'autres termes à rejeter les principes qui guident la rationalité de type R1. C'est parce que le principe « il faut voter » est fondé sur des raisons solides que j'ai le sentiment que *je dois* m'y conformer. J'ai en d'autres termes des raisons solides de voter, *bien que*, je le sais, ma voix n'ait guère de chances de faire de différence. Plus besoin ici de supposer que le sujet s'illusionne sur l'importance de son vote. Car, paradoxalement, en voulant peindre l'électeur comme rationnel au sens étroit, les explications du vote de type R1 le dépeignent en fait comme sujet à des illusions variées (purement conjecturales). L'électeur de Ferejohn s'illusionne sur la magnitude du risque en faisant de l'électeur un parieur pascalien (que la France soit présidée par tel ou tel homme est tout de même moins grave pour moi que de savoir si, ayant fait le mauvais choix, je me trouve promis aux flammes éternelles). L'électeur d'Overbye s'illusionne sur les autres électeurs, qu'il voit comme irrationnels, alors que chacun, pris isolément, est rationnel. L'électeur de Tversky-Quattrone est sujet au piège des « corrélations fallacieuses ».

Remarque incidente : à partir du moment où l'on prend la notion de « rationalité axiologique » au sérieux, où l'on admet en d'autres termes que l'acteur puisse être mû par des raisons fortes ne relevant pas de la catégorie de la « rationalité instrumentale », on accepte une théorie de la rationalité qui s'écarte des axiomes de base du pragmatisme. D'un autre côté, cette théorie n'endosse en aucune façon la doctrine kantienne. Elle admet en effet par exemple qu'on peut mentir avec de bonnes raisons : il suffit que le mensonge soit commis dans une situation telle qu'il existe des raisons fortes de mentir. Une telle proposition aurait sans doute fait dresser les cheveux sur la tête de Kant.

La distinction entre « rationalité instrumentale » et « rationalité axiologique » est fondée sur le fait que, dans certains cas, l'action est guidée par des principes et non par les conséquences qu'elle risque d'entraîner.

Lorsqu'on considère qu'une chose est bonne, les élections par exemple, on ne peut sans raisons agir en contradiction avec ce principe.

Sans cette distinction, d'immenses plaques de phénomènes sociaux deviennent inintelligibles. Je prendrai deux exemples pour illustrer mon propos.

Celui du jeu dit « de l'ultimatum » d'abord : l'expérimentateur a la capacité de distribuer une somme de 100 F. Le sujet A est appelé à faire une proposition sur la manière dont les 100 F devraient être partagés entre lui-même, A et B. B de son côté a seulement la capacité d'approuver ou de rejeter la proposition de A. S'il l'approuve, le partage se fait selon la proposition de A. S'il la rejette, les 100 F restent dans la poche de l'expérimentateur. Si les individus étaient seulement calculateurs et égoïstes, on devrait observer de la part de A des propositions de partage telles que « 70 F pour moi (A), 30 F pour B ». En effet, B aurait dans ce cas un intérêt à accepter la proposition de A, même si elle le désavantage beaucoup. En fait, la plupart des sujets choisissent le partage égal. Ils refusent en d'autres termes de profiter de la supériorité en matière de pouvoir de décision que l'expérimentation leur confère par construction. Pourquoi ? Parce que les différences de rétribution doivent se justifier par des différences de contribution. On ne peut profiter des circonstances pour s'enrichir aux dépens d'autrui. Ce système de raisons fortes explique non seulement les réactions des sujets au jeu de l'ultimatum, mais bien des résultats obtenus par la psychologie sociale dans les innombrables expériences qu'elle a conduites sur les « sentiments de justice »<sup>1</sup>.

Mon second exemple est un exemple d'école, celui du sentiment de rejet que provoque normalement le vol. Le philosophe anglais Mandeville a proposé une rigoureuse démonstration par l'absurde du fait que le vol ne peut être condamné au vu des conséquences qu'il entraîne. Sans doute le vol nuit-il à certains individus, à savoir les victimes des vols, nous dit Mandeville avec humour. Mais le vol fournit aussi du travail aux serruriers et aux avocats. On pourrait y adjoindre de nos jours les assureurs. Aujourd'hui, le théorème de Mandeville apparaît, si l'on peut dire, comme plus vrai que jamais. Le vol donne en effet naissance à un marché dual aux *conséquences* heureuses. Le marché alimenté par le coulage et le vol permet en effet aux plus démunis – qui ne savent pas toujours qu'ils acquièrent de la marchandise volée au titre de « la bonne occase » – de se procurer à bas prix les produits convoités de la « société de consumma-

1. A. Tversky, P. Slovic, D. Kahneman (eds), *Judgment under Uncertainty*, Cambridge, Cambridge University Press, 1982.

tion», produits électroniques par exemple, que les couches favorisées acquièrent au prix du marché. Ici, en contradiction avec le célèbre théorème sociologique de Caplovitz<sup>1</sup>, ce n'est pas le pauvre, mais bien le riche qui paie davantage, et la redistribution se fait dans le « bon » sens. Mais le vol n'est pas bon seulement pour les avocats, les assureurs et les couches sociales défavorisées. Il augmente aussi la demande, et il stimule l'offre. Sans le vol, le taux de chômage serait donc encore plus important qu'il ne l'est. Il est donc possible que le vol comporte d'heureux effets macro-économiques et qu'il provoque des effets de redistribution que les politiques fiscales recherchent sans toujours parvenir à les atteindre.

La démonstration par l'absurde esquissée par Mandeville indique bien que l'on ne saurait tirer d'une argumentation conséquentialiste la certitude morale que le vol est mauvais. Le vol comportant des effets mitigés, ce n'est pas parce que ses conséquences seraient mauvaises qu'on le condamne.

En fait, sous son apparente rigueur, le théorème de Mandeville oublie qu'il existe d'autres raisons de rejeter ou d'accepter quelque chose que celles qui ont trait aux conséquences que ce quelque chose entraîne. Le fait qu'on ne puisse démontrer que le vol est mauvais – c'est-à-dire rendre compte du sentiment collectif qui s'exprime par le jugement de valeur « le vol est mauvais » – à partir de raisons conséquentialistes n'implique pas en d'autres termes qu'il n'y ait pas de raisons fortes pour condamner le vol. Le vol est mauvais parce qu'il contredit des *principes* essentiels sur lesquels la société est fondée. En effet, l'ordre social est fondé sur le fait que toute rétribution sociale consentie à un individu doit en principe correspondre à une certaine contribution de sa part. Sinon, c'est le principe même du lien social qui se trouve remis en cause. Or le vol est une rétribution positive que le voleur s'attribue aux dépens de la victime, en ayant recours à une contrainte illégitime. Ce faisant, il contredit une règle essentielle qui est à la base de tout système social. Par là, il viole la notion même d'échange social. Le vol est donc bien condamné parce qu'il existe des raisons fortes pour le condamner. Mais ces raisons ne sont pas de type conséquentialiste.

Cette théorie offre une explication rationnelle des sentiments négatifs engendrés par le vol. Elle permet d'expliquer non seulement le caractère universel de la condamnation du vol, mais aussi des phénomènes que nous pouvons facilement observer tous les jours. Elle rend par exemple compte du fait apparemment paradoxal qu'un vol même sans grande

1. D. Caplovitz, *The Poor Pay More*, London, Macmillan/New York, Free Press, 1967.

conséquence sur le bien-être de la victime est normalement ressenti par elle comme insupportable, ou du fait que le spectateur d'un vol éprouvera un sentiment d'indignation pour le fait même du vol, indépendamment du préjudice subi par la victime. En fin de compte, c'est bien parce le vol atteint au cœur même du système social que ses avantages ne sauraient compenser ses inconvénients : les raisons conséquentialistes qu'on pourrait avoir de préférer le vol sont subordonnées aux raisons de principe qui convainquent de les rejeter. Il y a en d'autres termes des raisons fortes de préférer les raisons de principe aux raisons instrumentales.

Ceux qui veulent à tout prix ramener la rationalité au type R1 proposent parfois une interprétation utilitariste de l'évaluation négative du vol : le vol serait condamné parce que chacun aurait peur d'être volé : l'indignation que j'éprouve devant le fait qu'un inconnu de moi soit victime d'un vol ne serait qu'une manifestation de la peur que j'ai d'être moi-même volé. L'hypothèse n'est pas déraisonnable, mais elle n'est pas démontrée. De surcroît, j'éprouve un sentiment *d'indignation* et non *de peur* : comment expliquer cette « fausse conscience » ? D'autre part, mon indignation est d'autant plus grande que la victime est plus faible ; comment l'expliquer ? Je devrais au contraire être rassuré par le fait que les aigrefins s'attaquent plutôt aux faibles. Ainsi, la théorie « je condamne le vol parce que j'ai peur d'être volé » conduit à des conséquences difficilement compatibles avec l'observation.

La condamnation d'autres comportements s'explique encore plus difficilement par une théorie de type R1. Je pense aux infractions qui, comme le plagiat, profitent à ceux qui en sont victimes. Le plagiaire provoque un sentiment de dégoût, pourtant, il ne se nuit souvent qu'à lui-même. Il peut même contribuer à attirer l'attention sur l'auteur qu'il plagie et, par là, lui rendre service. Pourquoi le plagiaire est-il l'objet d'une condamnation morale bien que son comportement ne nuise à personne ? Parce qu'il viole les principes qui fondent le jeu de l'invention et de la création, exactement comme le voleur viole les principes de base de l'échange social. En même temps, le plagiaire trompe le lecteur.

Les manquements à la politesse s'analysent de la même façon. Ils sont généralement sans grande conséquence. Le fait que quelqu'un me bouscule dans le métro sans s'excuser n'a guère de conséquence appréciable sur mon bien-être. Il ne provoque le plus souvent aucune douleur physique et, même s'il en provoque, ce n'est pas en raison de cette douleur que je porterai un jugement négatif sur le comportement du « malappris ». Ici encore, la réaction du sujet qui est victime d'un manquement à la politesse s'explique parce que celui-ci représente une violation d'un

principe essentiel de la vie sociale, celui de la dignité de tout individu. C'est cette violation qui provoquera chez la plupart des gens un sentiment d'irritation plus ou moins facilement contrôlable.

L'imposteur, qui réussit avec la complicité de médiateurs ou de collègues complaisants, à faire passer pour scientifiques des travaux qui contredisent les principes élémentaires de la méthode scientifique provoque en général une réaction d'exclusion très forte à partir du moment où l'imposture est perçue. L'imposteur démasqué n'a en général d'autre issue que de se retirer de la vie scientifique et d'exercer, s'il le peut, une autre activité. Mais ce n'est pas par les dégâts qu'il cause que l'imposteur provoque un sentiment de rejet. Car les conséquences de ses méfaits sont souvent provisoires et limitées et il finit généralement par être découvert. La répulsion qu'il inspire provient plutôt de ce qu'il contredit les règles fondamentales du jeu scientifique et prétend se voir attribuer une reconnaissance qu'il ne mérite pas.

Ce sont aussi des raisons analytiques, des raisons de principe, qui fondent par exemple les réactions contre le resquilleur. Il a des chances d'être rappelé à l'ordre, même s'il n'impose à ceux qui font la queue qu'un désagrément mineur. Le rejet auquel s'expose le resquilleur provient de ce que l'avantage qu'il s'attribue est injustifié. Une théorie de type R1 ne rend pas compte de la réaction elle-même : pourquoi une telle réaction à un désagrément aussi mineur ? De surcroît, en contradiction avec R1, l'irritation à l'égard du resquilleur tend à être d'autant plus forte que la file d'attente est plus importante, alors que le désagrément marginal qu'il inflige à ses compagnons est d'autant plus faible. Cette liaison fonctionnelle est en revanche correctement expliquée par une théorie de type R4. Du point de vue des critères d'appréciation de la validité d'une théorie scientifique mis en évidence par Popper, une théorie de type R4 explique beaucoup mieux les réactions de l'*homo sociologicus* au vol, au plagiat, à l'imposture ou à la resquille qu'une théorie de type R1.

Et comment expliquer que le compliment non mérité ne nous fasse guère plaisir si l'on entend s'accrocher à une rationalité de type R1 ? Comment rendre compte du vague dégoût que suscite le carriériste, même s'il parvient à ses fins, et l'insatisfaction persistante qui le taraude ? Il cherche à accumuler honneurs sur honneurs, grandeurs d'établissement sur grandeurs d'établissement. Mais, sachant fort bien lui-même que, loin d'attendre l'apparition de ces signes de reconnaissance, il les a quelque peu provoqués, il lui faut à tout prix obtenir un signe supplémentaire de reconnaissance. Vu le sentiment d'urgence qui l'habite, il ne pourra s'abstenir de chercher à nouveau à aider le destin. Le comporte-

ment de fuite en avant de ces prétendus « éternels insatisfaits » s'analyse fort bien dans le cadre d'une rationalité de type R4. Il constitue un paradoxe de plus pour la rationalité de type R1.

Les phénomènes banals que je viens d'évoquer sont trop peu pris en considération par les sciences sociales. Leur banalité fait qu'ils tendent à être perçus comme des données de fait, comme des phénomènes naturels qu'il suffirait d'enregistrer. Il faut reconnaître aux tenants du MCR qu'ils ont le mérite de les « dénaturiser » et d'en tenter l'explication. Mais, dès qu'on s'arrête aux explications de type R1 que le MCR en propose et qu'on tente d'en apprécier la validité à partir des critères poppériens et non poppériens à partir desquels on apprécie habituellement une théorie scientifique, leurs faiblesses apparaissent. Or, tous ces phénomènes familiers s'expliquent facilement par des théories de type R4 qui satisfont beaucoup plus facilement lesdits critères. Seules, elles sont capables de rendre compte notamment de la palette des sentiments bien déterminés qui apparaissent dans les exemples que j'ai évoqués. Pourquoi les sentiments suscités par telle ou telle situation ne représenteraient-ils pas des *faits* dont il faille tenir compte ? Faits subjectifs certes, mais faits tout de même, dont le sociologue ne peut se désintéresser, puisque les sentiments en question ne sont en aucune façon le résultat d'idiosyncrasies personnelles. Ces sentiments, telle l'indignation qu'on éprouve à l'égard du malfrat qui s'attaque à un être faible, tendent à être partagés par tous. On ne voit donc pas pourquoi le sociologue pourrait se dispenser d'en tenir compte et de chercher à en rendre compte.

#### VERS UNE THÉORIE GÉNÉRALE DE LA RATIONALITÉ ?

Les théories de la rationalité présentes sur le marché s'appliquent à des régions particulières du réel. La théorie R1 est importante et applicable dès que certaines conditions sont réunies (contexte social autorisant le raisonnement en termes utilitaristes, informations accessibles et complètes, etc.). C'est le cas de l'exemple de Tocqueville que j'évoquais au début. Dans d'autres cas, l'information dont dispose l'acteur est lacunaire. Dans d'autres cas encore, il doit théoriser le problème qui se présente à lui. Dans d'autres cas, un conflit apparaît entre des systèmes de raisons concurrentes et il doit trouver des raisons fortes de choisir entre ces systèmes. Quand ce conflit se combine avec une information lacunaire, les

variantes « gradualistes » de la théorie de la rationalité ont des chances d'être mises en jeu par l'acteur. Dans d'autres cas, l'action découle de la mise en œuvre de principes auxquels l'acteur social adhère sur la base de raisons fortes. Il faut alors comprendre les raisons qui le conduisent à considérer les principes plutôt que les conséquences.

On peut donc partir du postulat que, sauf à faire la preuve du contraire, l'acteur a des raisons fortes de faire ce qu'il fait, de croire ce qu'il croit, etc. C'est pourquoi ses actions ou croyances ont un *sens pour lui*. Mais, selon les situations qu'il rencontre, ces raisons sont de nature variable. Ainsi, il peut avoir des raisons fortes de considérer (ou de ne pas considérer) les différences entre les avantages et les coûts des alternatives qui s'offrent à lui comme décisives. Bref, le modèle général qui émerge de cette discussion se résume : l'acteur adhère à une croyance, prend une décision donnée, accomplit une action donnée, lorsqu'il a l'impression que ladite croyance, que ladite action ou que ladite décision reposent sur un système de raisons fortes, sur un système de raisons que l'Autre généralisé<sup>1</sup> endosserait.

C'est en partie par l'effet d'un fétichisme technique qu'on a cru que la rationalité de type R1 pouvait être tenue comme expliquant indistinctement tout comportement, alors qu'elle n'explique ni que les gens votent, ni qu'ils évitent le risque, ni qu'ils soient indignés par le vol ou le plagiat, ni d'innombrables comportements familiers. Au-delà de ce probable fétichisme, les modèles d'inspiration utilitariste ont un avantage sur les modèles qui proposent des explications du comportement à partir de causes irrationnelles et voient par exemple dans toute action un effet de la socialisation. Cet avantage, c'est qu'ils fournissent des explications dépourvues de boîtes noires. Lorsque j'affirme : « Les propriétaires fonciers quittent leurs terres parce qu'ils ont des raisons fortes de le faire », l'explication est complète. Elle ne laisse subsister aucune zone d'ombre. Ces raisons étant irrécusables, l'explication est de surcroît solide. En revanche, si je dis « un tel s'est comporté de cette manière parce qu'il a l'habitude d'agir ainsi », l'explication se perd dans une boîte noire, sauf si je puis décliner les raisons sous-jacentes à ladite habitude, ce qu'il est parfois possible de faire : j'ai l'habitude de prendre mon parapluie quand il pleut ; loin d'être dépourvue de sens pour l'acteur, cette habitude s'explique par les raisons solides qui la fondent. Ce sont en d'autres termes des

1. G. H. Mead, *Mind, Self and Society. From the Standpoint of a Social Behaviorist*, Chicago, The University of Chicago Press, 1934 ; en français, *L'esprit, le soi et la société*, Paris, PUF, 1963.

raisons qui sont ici la cause de l'habitude, même si elles sont enfouies dans la conscience.

Mais il faut voir que cette capacité à fournir des explications dépourvues de boîtes noires, qui explique pour partie la séduction du modèle utilitariste de type R1, n'est pas propre à ce type de rationalité. Toute explication sociologique faisant de l'*explanandum* le résultat d'actions, d'attitudes, de croyances fondées sur des raisons est « finale ».

Bien entendu, l'acteur peut ne pas avoir de raisons de faire ce qu'il fait ou de croire ce qu'il croit. Les actions *affectives* existent aussi. Elles peuvent ne pas créer de boîtes noires. C'est le cas lorsqu'on évoque des causes affectives de caractère familial et dont on peut s'assurer empiriquement qu'elles entrent bien en jeu dans les comportements dont on cherche à rendre compte. Les actions *traditionnelles*, celles qu'on exécute *parce que* la tradition le veut, existent aussi bien sûr. Toutefois, avant de recourir à ces catégories explicatives, il faut être sûr que toute explication en termes de rationalité est exclue. Surtout, il faut se méfier de toutes ces causes irrationnelles souvent verbeuses (« frames », « mentalité », « habitus », « fausse conscience », « pensée magique », « âme nationale », « personnalité de base », « biais cognitifs », « représentations sociales », « aversion pour le risque », « amour du vote », etc.) qu'on place trop facilement à l'origine du comportement des acteurs dès que le sens de celui-ci nous échappe.

Il faut conserver de la théorie économique classique l'idée que l'acteur social est maximisateur. Cette notion de maximisation ou d'optimisation donne un contenu à la notion de rationalité. Mais il faut aussi admettre que ce que l'acteur maximise ou optimise c'est un système de raisons. C'est seulement dans des cas particuliers qu'il optimisera la balance coûts-avantages : lorsqu'il a des raisons fortes de considérer cette dimension de la situation.

Enfin, je voudrais souligner à nouveau quelques points majeurs. Tout d'abord, la rationalité ne se réduit pas à la rationalité instrumentale. En second lieu, l'activité cognitive du sujet peut ne pas se limiter à une recherche d'informations, mais viser une théorisation du problème qui lui est posé. La meilleure théorie à ses yeux sera alors celle qui sera fondée sur les raisons qui lui paraîtront suffisamment fortes. La connaissance ordinaire ne suit pas ici des procédures différentes dans leur nature de celles de la connaissance scientifique. Cette proposition n'est en aucune façon contradictoire avec les résultats de la psychologie cognitive. Elle permet au contraire d'en proposer une interprétation satisfaisante. En troisième lieu, il faut préciser que la théorie que j'ai proposée ici n'est en

aucune façon intellectualiste. Non seulement l'acteur social n'est pas dans mon esprit dépourvu de passions, mais je suis tout prêt à souscrire à l'adage de Hume selon lequel la raison est la servante des passions. A quoi il faut immédiatement ajouter qu'on ne connaît pas de société où le sujet social puisse toujours exprimer librement ses passions : dans certains cas, il le peut ; dans d'autres cas, non. C'est pourquoi il doit toujours théoriser la situation dans laquelle il se trouve et pourquoi ses décisions ne sont pas compréhensibles sans cette théorisation. En quatrième lieu, la théorie proposée est par construction *individualiste* (au sens méthodologique), *mais non atomique* : les raisons des individus n'ont de valeur pour eux que s'ils ont le sentiment qu'elles ont potentiellement une valeur pour l'Autre généralisé. En cinquième lieu, cette théorie est contextuelle et historique : des raisons fortes ici et maintenant ne le seront pas forcément ailleurs et demain ; si des faits ou des arguments nouveaux apparaissent, un système de raisons fort peut devenir faible (et réciproquement). Cela n'implique pas qu'il ne puisse y avoir de vérités positives ou normatives. Le fait que la science ait une histoire n'implique pas qu'il n'y ait pas de vérités scientifiques. Il en va de même de la morale.



*Les ambiguïtés constitutives du modèle  
du choix rationnel*

Les relations complexes entre la tradition de la sociologie et la tradition de la science économique se sont organisées autour d'un certain nombre de concepts déterminants, en particulier celui de rationalité. Ce mot difficile, polysémique du point de vue des multiples aspects de ces deux traditions, a été largement associé au domaine de l'économie, tandis que fréquemment des sociologues comme Marcel Mauss mettaient au contraire en avant une « irrationalité » économique<sup>1</sup> pour caractériser les phénomènes mêmes traités par la science économique.

Contre cette opposition des perspectives, un courant important de la pensée sociologique contemporaine a cherché à réunifier les deux disciplines en important le modèle d'analyse économique en sociologie. James Coleman peut être considéré comme l'adepte le plus systématique de cette démarche<sup>2</sup>. Ce programme de recherche peut être associé au terme générique du « choix rationnel ».

Ce modèle a été fortement critiqué d'un point de vue externe, par la mise en avant de phénomènes sociaux qui ne pouvaient pas être expliqués par ce type de modèle. Notre propos ici sera différent : il consistera à montrer que le modèle souffre d'une forte indétermination sémantique due à la complexité des éléments qu'il convoque sous sa bannière. Il s'agit donc pour nous de décomposer, ou de déconstruire, ces différents éléments, afin de les isoler et de les distinguer, non pour les discréditer, mais pour en permettre un usage plus pertinent parce que mieux maî-

1 M. Mauss, *Essai sur le don*, in *Sociologie et anthropologie*, Paris, PUF, coll. « Quadrige », 1995, p. 271.

2. J. Coleman, *Foundations of social theory*, Cambridge, The Belknap Press of Harvard University Press, 1990.

trisé, loin des confusions qui caractérisent actuellement cette entreprise théorique.

Il y a essentiellement trois aspects qui apparaissent dans la théorie du choix rationnel, telle qu'elle est appliquée à la sociologie, trois aspects qui ne sont généralement pas distingués, mais au contraire associés, et davantage, confondus : tous trois jettent les bases d'une définition de la rationalité de l'action, qui apparaît alors comme un phénomène univoque, alors qu'en réalité des aspects tout à fait différents sont mis en avant en fonction des situations.

Le premier aspect est celui de l'action que l'on pourrait qualifier de stratégique : il s'agit de la relation entre les moyens et les fins, l'efficacité de l'action portant sur la pertinence des moyens pour atteindre ces fins.

Le deuxième aspect concerne cette fois-ci les fins elles-mêmes : les acteurs ayant des intérêts, ils cherchent à les satisfaire de la meilleure manière possible. Certes, ici les moyens ont encore de l'importance, mais l'accent est mis sur la nature des fins, à savoir les *intérêts*, qui commandent ensuite le choix de certains moyens. Autrement dit ce seront les intérêts qui permettront de rendre compte de l'existence d'un choix rationnel.

Enfin, le troisième aspect, moins évident du point de vue du sens commun, mais décisif pour la formalisation des théories du choix, insiste sur une mise en ordre des préférences qui permet la constitution d'une fonction d'utilité. Nous préciserons ultérieurement les liens de cette troisième dimension avec les deux premières.

Notre objectif est d'indiquer que ces trois dimensions sont distinctes à de nombreux égards, et que si, fréquemment, il est possible de les réunir en un modèle commun, il y a aussi possibilité de les dissocier, en sorte que la question de la rationalité soit interprétée en des termes différents.

## 1 | L'ACTION STRATÉGIQUE

Commençons donc par la relation stratégique qui unit la fin et les moyens. La mise en évidence de cette relation est évidemment très ancienne, puisqu'elle remonte à Aristote. Cette relation est promue par Vilfredo Pareto et Max Weber au rang de modèle d'action logique ou rationnelle suivant les terminologies de chacun de ces deux auteurs. Elle

apparaît de surcroît, chez Weber, comme un modèle d'action intelligible par l'analyste. Cette référence à l'action instrumentale se retrouve évidemment chez un auteur comme Coleman, qui cherche à en systématiser l'application.

Deux remarques très importantes peuvent être faites au sujet de l'emploi de ce modèle d'action, du point de vue de sa stricte définition : premièrement, *n'importe quel type de fin* peut donner naissance à une action de ce type. Les fins pouvant donner lieu à la recherche de moyens efficaces peuvent en particulier avoir un caractère religieux, éthique ou esthétique. Max Weber précise bien cela, et le statut de fin, par rapport aux moyens, est un statut de position, qui n'implique aucune spécification particulière du contenu des fins. Les fins peuvent donc très bien relever de ce qu'intuitivement on appelle les « valeurs ».

Deuxièmement, cette relation entre les fins et les moyens représente un procédé très fécond pour l'explication des activités sociales. L'explication consiste alors, précisément, à dévoiler un lien, non perceptible à première vue, entre des situations qui ont un statut de fin et des situations qui ont un statut de moyen. La banalité du principe masque le caractère non évident, pour le chercheur, de telles relations, qui sont à découvrir dans l'expérience, sans qu'elles soient données immédiatement. Bien avant la thématization de ce procédé explicatif par Weber, Montesquieu en avait fait un usage abondant : pourquoi se laisse-t-on pousser les ongles dans certains pays ? demande Montesquieu. C'est un moyen de montrer que l'on ne travaille pas<sup>1</sup>. Les exemples de ce type d'explication sont innombrables chez lui.

L'explication peut alors être de deux sortes dans ce contexte : soit on montre, ce qui n'apparaissait pas spontanément, qu'un certain type d'action, à première vue surprenant, s'explique par le fait qu'il s'agit là d'un moyen pour atteindre une fin. Soit, compte tenu d'une fin poursuivie, l'analyste montre que, dans certaines situations, c'est tel moyen, et non pas tel autre, qui permet de l'atteindre. Par exemple Tocqueville explique pourquoi l'argent est tellement important dans les régimes démocratiques. Son explication a deux aspects : d'une part, il s'agit de montrer que l'argent est en fait un moyen pour obtenir du prestige. C'est donc la recherche du prestige qui est la finalité dominante, l'argent n'étant qu'un moyen. Mais, d'autre part, il s'agit de montrer que, dans une société démocratique où les individus recherchent le prestige, c'est l'argent, et non pas un autre moyen, qui permet de l'obtenir le plus communément,

1. Montesquieu, *L'Esprit des Loix*, Paris, Seuil, 1964, p. 642.

alors que, dans une société inégalitaire, ce ne saurait être l'argent. Là, il ne s'agit plus de révéler la fin, mais de montrer que celle-ci implique, dans certaines situations, tels moyens plutôt que tels autres<sup>1</sup>.

La rationalité, thématifiée par Hume, puis Weber et Pareto, concerne ici alors, dans une acception étroite, l'adéquation des moyens aux fins, non les fins elles-mêmes, qui sont laissées dans l'indétermination. Pourquoi parler de rationalité ? à cause de la compréhensibilité intersubjective de telles situations d'adéquation qui établissent un lien repérable objectivement entre deux catégories d'événements. Aucune référence n'est donc faite à la notion d'*intérêts*. Comme nous l'avons dit, les fins peuvent être de toutes sortes, et la rationalité relève simplement de l'intelligibilité d'une relation entre deux séquences du réel, sans qu'il soit, en aucune manière, question d'intérêt.

## 2 | INTÉRÊTS

Pourquoi donc faire intervenir la notion d'intérêt ? parce que les auteurs classiques comme Pareto et Weber et les auteurs contemporains comme Coleman établissent une liaison spéciale entre le thème de la rationalité et celui de l'intérêt. Pourtant, à cet égard, la notion d'intérêt reste vague et difficile à préciser. Il y a de surcroît une différence entre le français, qui ne connaît que le seul mot d'intérêt, tandis que l'anglais distingue le *self-interest* de l'*interest*. Une tentative de précision de cette notion vague d'intérêt peut être confrontée à quatre types différents de définition. Deux de celles-ci sont négatives, c'est-à-dire qu'elles s'appuient sur une opposition à une autre catégorie que celle de l'intérêt, tandis que les deux autres sont positives et essaient de rendre compte directement du contenu de la notion.

Une première manière de distinguer la catégorie des intérêts consiste à les opposer aux sentiments. C'est ce que fait Pareto, qui partage en gros les fins en deux catégories permettant par suite la division des sciences sociales en deux domaines, la sociologie et l'économie, correspondant, plus ou moins, à ces deux catégories.

Une deuxième distinction s'appuierait sur une lecture hâtive de Weber, qui opposerait les « valeurs » aux intérêts. Il y a à cet égard une

1. A. de Tocqueville, *De la démocratie en Amérique*, II, Paris, Robert Laffont, 1986, p. 584.

ambiguïté chez Weber, dans la mesure où, en effet, il oppose quelquefois valeur et intérêt : ainsi, dans le cas des procédures de garantie d'un ordre tenu pour légitime, la valeur se distingue de l'intérêt, la valeur assurant une garantie purement intime, ou intérieure, tandis que l'intérêt relève de l'expectation de certaines conséquences spécifiques d'ordre externe<sup>1</sup>. Pourtant, un peu plus loin<sup>2</sup>, Weber précise d'une tout autre manière la notion d'intérêts, en indiquant que ceux-ci peuvent relever soit d'une rationalité en valeur, ou axiologique, soit d'une rationalité en finalité. Dans ce passage l'intérêt relève alors d'une catégorie plus générale que l'opposition entre la valeur de la *Wertrationalität* et le but de la *Zweckrationalität*.

Toujours est-il qu'on oppose généralement, à tort semble-t-il, chez Weber, les intérêts aux valeurs, et on qualifie alors la rationalité instrumentale de rationalité utilitaire, ce que se garde bien de faire R. Boudon dans *L'origine des idées reçues*<sup>3</sup>. Cette opposition entre intérêts et valeurs se retrouve d'ailleurs chez Pareto<sup>4</sup>, qui lui-même distingue un *homo oeconomicus* mû par l'intérêt, d'un *homo religiosus* et d'un *homo ethicus* mû par les valeurs.

On a là un lieu commun de la tradition des sciences sociales, lieu commun qui est l'occasion de multiples développements présupposant l'évidence de ces distinctions. On peut toutefois essayer d'indiquer que cette opposition est peu claire et même assez difficile. Son examen permet de la considérer comme peu valide à certains égards, même si elle peut être retenue d'un autre point de vue.

Si on prend l'exemple de la recherche du prestige, dans quelle catégorie faudra-t-il le mettre ? La sociologie contemporaine n'hésite pas à classer le prestige dans la catégorie des *intérêts*. Mais l'on voit bien alors, du point de vue des oppositions précédentes, que le prestige peut relever aussi bien d'un sentiment (vanité, jalousie, sentiment de sa propre excellence) que d'une valeur : il y a une grande variété de prestiges correspondant à une pluralité de valeurs, le prestige du chercheur en sciences physiques ne se substituant pas à celui de la star de football : ces prestiges-là ne sont pas interchangeables. Le prestige n'est alors éventuellement recherché qu'en tant qu'il dérive de la valeur visée, et non l'inverse.

On est alors tenté de laisser tomber la catégorie des intérêts, en tant qu'elle serait opposée à celle de la valeur ou des sentiments, pour ne rete-

1. M. Weber, *Économie et société*, Plon, 1995, t. 1, p. 68.

2. M. Weber, *ibid.*, p. 78.

3. R. Boudon, *L'art de se persuader*, Paris, Fayard, 1990, p. 406-407.

4. V. Pareto, *Manuel d'économie politique*, t. VII des *Œuvres complètes*, Genève, Librairie Droz, p. 18.

nir qu'un seul ensemble, celui des *motivations* de l'acteur, toujours associées à certaines valeurs ou à certains sentiments. Il faudrait alors indiquer que ces motivations des différents acteurs sont, en fonction de la variété des individus et des situations, soit semblables, soit dissemblables ; et, dans les deux cas, soit compatibles, soit opposées. On insistera alors sur la convergence ou l'opposition des intérêts, quels que soient les contenus de ceux-ci en termes de valeur ou de sentiment. Ainsi les intérêts du fumeur, dans un café, sont opposés aux intérêts du non-fumeur gêné par la fumée, et reposent sur la valorisation ou la dévalorisation de l'acte de fumer. De même, deux membres de deux différentes religions prosélytes, essayant chacun de convertir ceux qui n'appartiennent pas à l'une de ces deux religions, ont des intérêts opposés. Il y a là opposition d'intérêts, non pas à l'écart du domaine religieux, mais d'un point de vue spécifiquement religieux, lequel est à l'origine des intérêts, ce qui ne rentre pas du tout dans le cadre de l'opposition de Pareto entre intérêts et valeurs.

Faut-il alors abandonner celle-ci, qui a pourtant un contenu intuitif important ? A vrai dire, les choses sont plus complexes. D'un certain point de vue, comme il vient d'être indiqué, les motivations des différents acteurs, qui ont les contenus les plus variables, relèvent d'*intérêts visés*, indépendamment de toute question de valeur ou de sentiment, qui sont pourtant toujours plus ou moins présents à l'origine de ces visées. Mais, par ailleurs, et en envisageant les choses sous un autre angle, il se trouve qu'à propos d'une décision concrète de l'acteur sur un sujet quelconque, des motivations d'ordre opposé, d'un point de vue interne à l'acteur, peuvent advenir, et être confrontées à sa délibération : or, dans ces cas, certains avantages visés, confrontés à des valeurs ou à des sentiments qui les condamnent explicitement, pourront alors prendre le nom d'intérêt : par exemple, le sentiment de fidélité à son pays et l'achat d'une voiture étrangère moins chère ; par exemple, la fidélité à une norme religieuse et le désir du divorce. A cet égard, valeurs ou sentiments, d'une part, et intérêts, d'autre part, s'opposent, dans la mesure où certaines normes ont pour effet d'empêcher certains avantages qui pourraient être poursuivis par l'acteur.

On peut voir là sans doute une origine aux ambiguïtés de Weber à ce sujet, qui voit, sans les thématiser explicitement, les deux aspects : d'une part les motivations les plus diverses représentent des intérêts, complémentaires ou opposés pour les différents acteurs, ces intérêts pouvant avoir des contenus tout à fait variables en termes de valeur ou de sentiment. Mais d'autre part, à propos de certaines actions, il y a précisément opposition (ou d'ailleurs convergence) entre intérêts et valeurs, les secon-

des prohibant (ou au contraire recommandant) ce que favoriseraient les premiers.

Ainsi donc, le même choix effectif, la même finalité retenue par un acteur, pourra avoir tantôt le statut d'intérêt, tantôt le statut de valeur. Le nationalisme en est un bon exemple qui peut représenter certains *intérêts* pour l'acteur, éventuellement opposés à d'autres intérêts nationalistes, mais qui peut tout aussi bien s'opposer, comme valeur cette fois-ci, à certains intérêts. De même pour la poursuite de la richesse : elle représente pour Tocqueville une valeur des sociétés démocratiques opposée à cet égard à certains intérêts (la tranquillité) ; mais elle représente aussi un intérêt qui peut être opposé à certaines valeurs (le dévouement à de grandes causes).

On comprend mieux alors certaines difficultés de la tradition des sciences sociales : Pareto lie la science économique à la thématization d'un *homo oeconomicus* opposé aux *homo religiosus* et *homo ethicus*. Mais, en même temps, il insiste beaucoup, et de manière décisive, sur le fait qu'en économie, le critère de l'utilité individuelle, c'est le choix<sup>1</sup>. A cet égard, les choix, constitutifs des intérêts, peuvent avoir n'importe quel contenu en termes de valeur ou de sentiment.

On retrouve alors la troisième manière de décrire les intérêts, qui consiste à les relier à de l'*utilité*, au sens ordinaire du terme, qui n'est pas du tout clair. Ce sens ordinaire (c'est-à-dire celui qui ne tient pas compte des fonctions d'utilité) a longtemps signifié, comme l'indique Hayek<sup>2</sup>, le statut du moyen par rapport à une fin. L'utilité n'existe pas en tant que telle, il y a de l'utilité en vue de quelque chose. Le fait de considérer l'utile comme une fin en soi représente un déplacement du sens du mot. Or l'on voit bien que l'idée d'une utilité finale universelle, en tant que telle, ne peut exister, puisque les satisfactions sont variables à ce sujet.

La quatrième définition des intérêts dérive de cette difficulté. Dans la mesure où les finalités sont variables et où ces finalités correspondent à des intérêts, on peut référer ceux-ci à un choix effectué par les acteurs. Un intérêt, dans cette quatrième définition, n'est donc plus opposé à des sentiments ou à des valeurs, non plus qu'il est référé à une utilité substantielle. L'utilité dérive simplement du fait que quelque chose est choisi par un acteur, pour sa satisfaction, qui peut être de n'importe quelle sorte. C'est la théorie de l'ophélimité de Pareto. On retrouve ici l'idée de

1. V. Pareto, *Traité de sociologie générale*, t. XII des *Œuvres complètes*, Genève, Librairie Droz, 1968, p. 1330.

2. F. Hayek, *Droit, législation et liberté*, Paris, PUF, 1982, t. 2, p. 20.

motivation générale de l'action. L'idée suivant laquelle l'utilité ou les intérêts peuvent être si différents que nous ne pouvons les connaître qu'à travers le choix des acteurs fut d'abord mise en avant par J. S. Mill : celui-ci insistait, d'un point de vue moral opposé au paternalisme, sur le fait que nul mieux que l'individu n'était à même de définir ses intérêts<sup>1</sup>. Si cela est vrai d'un point de vue moral, il apparaîtra que cela est vrai aussi d'un point de vue méthodologique (bien que la réciproque de cette proposition ne soit pas nécessairement vraie : d'un point de vue moral je peux quand même considérer que les libres choix d'autrui sont mauvais). Je connaîtrai, comme théoricien du choix, les préférences d'un acteur si elles me sont données, ou si je les suppose données. C'est ainsi d'ailleurs que Pareto délimite le champ de l'économie, comme science des choix.

Mais alors, il n'y a plus lieu, comme nous l'avons vu, dans cette acception, d'opposer valeurs ou sentiments aux intérêts, puisque ceux-ci, référés aux choix, peuvent les inclure. Il n'y a alors pas de sens de parler d'une démarche utilitariste de l'économie, ou des théoriciens du choix rationnel, ou de la sociologie lorsqu'elle se réfère à ces méthodes. Comment expliquer alors une telle réputation ? C'est qu'à certains égards il existe, en sociologie, une telle démarche utilitariste, ce mot étant pris en un certain sens parmi d'autres<sup>2</sup> : en effet, référer l'utilité à des choix impose nécessairement de considérer indifférent le contenu des choix. Mais, en même temps, l'effort de sociologues contemporains ayant été, à la suite de Hobbes ou de Hume, de montrer comment certaines normes pouvaient dériver de certains intérêts spécifiés, certaines valeurs dérivait alors de ces intérêts plutôt que d'une reconnaissance propre. Il y avait alors une démarche utilitariste, non pas au sens de cette utilité toute formelle des préférences, mais au sens plus réduit de certaines utilités particulières définies qui, pour être mieux satisfaites, exigent la mise en place de normes en apparence contraires à ces intérêts.

1. J. S. Mill, *De la liberté*, Paris, Gallimard, 1990, p. 178.

2. Nous avons essayé ailleurs de définir six sens généralement associés, et confondus, à la notion d'utilitarisme. L'usage de ce mot requiert la plus grande prudence, dans la mesure où sa nature polysémique prête à confusion. Cf. Pierre Demeulenaere, « Justice et maximisation de richesse : une analyse de sociologie cognitive de la théorie de R. A. Posner », in R. Boudon, A. Bouvier, F. Chazel, *Cognition et sciences sociales*, Paris, PUF, coll. « Sociologies », 1997.

### 3 | INTÉRÊTS ET RATIONALITÉ

Quel est alors le lien entre le concept de rationalité et ces différentes notions d'intérêt ?

Pareto a établi un lien spécial entre les intérêts, par opposition aux sentiments et valeurs, et les actions logiques, c'est-à-dire les actions stratégiques dans le sens précédemment donné à ce mot. La rationalité était alors associée à une situation où le choix des moyens permet d'atteindre les fins visées, quelles que soient ces fins. Ici donc Pareto indique que c'est une certaine catégorie de fins seulement, à savoir les intérêts, qui permet de donner lieu à ces actions logiques.

Ceci peut d'ailleurs être contesté de deux points de vue : d'une part, des sentiments et des valeurs peuvent très bien donner naissance à des actions stratégiques. Le criminel par vengeance prépare son coup froidement et la volonté de vengeance dérive d'un sentiment ou d'une valeur d'honneur. On retrouve d'ailleurs ici le problème de la définition des intérêts : en tant qu'il s'agit de motivations de l'action, quel que soit le contenu de ces motivations, celles-ci peuvent très bien donner lieu au choix de moyens efficaces.

Deuxièmement, si on considère un intérêt tenu pour tel par Pareto, dans un sens étroit opposé aux sentiments ou aux valeurs, la poursuite de la richesse, il n'y a certainement pas de moyens univoques qui permettent de l'atteindre. Pareto décrit d'ailleurs fort bien l'incertitude à laquelle est confronté le spéculateur, thème que reprendront les analystes autrichiens de l'action économique, von Mises et Hayek au premier chef. Il n'y alors pas d'action logique au sens étroit défini par Pareto.

Par ailleurs, nombre de « sentiments » que Pareto laisse inexplicés comme motifs ultimes de l'action humaine peuvent assez aisément être interprétés dans les termes de l'action stratégique rationnelle : par exemple le sens de la hiérarchie. Celle-ci peut en effet être quelquefois aimée pour elle-même, comme le suppose Pareto, mais il est bien plus vraisemblable que, la plupart du temps, les gens admettent et respectent une hiérarchie bien plutôt parce qu'elle est un moyen fonctionnel pour coordonner l'action lorsque cela est possible et nécessaire.

Mais tout ceci ne change rien au fait fondamental : Pareto a voulu associer la rationalité à la poursuite d'intérêts. Il le fit parce qu'il considé-

rait que certaines motivations, nommées intérêts, étaient plus à même que d'autres (les sentiments ou les valeurs) de donner lieu à des actions logiques ou rationnelles. Mais il ne dit jamais qu'il était logique, ou rationnel, de poursuivre ses intérêts. La rationalité concerne toujours chez lui les moyens, non la fin, et la rationalité n'est jamais transférée vers les fins : la rationalité, ce n'est pas un conseil donné recommandant de suivre ses intérêts, c'est seulement, compte tenu de ceux-ci, l'efficacité de leur satisfaction. Les finalités ultimes de l'action, chez Pareto, dans une perspective humienne, échappent au critère de rationalité, c'est-à-dire qu'elles ne lui sont pas conformes, mais ne lui sont pas opposées non plus : on est dans une sphère qui ne relève ni du logique, ni de l'illogique en tant que celui-ci désignerait le contraire du logique. Il y a du logique, il y a de l'illogique, contraire à lui, mais il y a aussi du non-logique qui concerne des finalités pour lesquelles le critère logique/illogique n'est pas pertinent.

#### 4 | UN GLISSEMENT DE SENS

Pourtant, un glissement de sens était possible, et il advint : lorsque l'on dit que « la poursuite des intérêts est rationnelle », cette proposition peut être comprise de deux manières tout à fait différentes. Elle peut vouloir dire : 1 / la poursuite des intérêts donne naissance à des choix rationnels, c'est-à-dire efficaces – c'est la thèse de Pareto ; elle peut vouloir dire aussi : 2 / il est rationnel de poursuivre ses intérêts, plutôt que des sentiments ou des valeurs. Il y a donc là un déplacement de la rationalité du choix des moyens vers le choix des fins, comprises en un sens restreint et spécifique comme intérêts (non pas les motivations de l'action en général, mais certaines motivations).

On peut relier ce glissement de sens à Weber, bien que ce dernier soit, sur ce sujet, infiniment nuancé et complexe. Nous avons dit que, chez lui, la notion d'intérêt se trouvait tantôt en amont de la différenciation entre *Zweckerationalität* et *Wertrationalität*, tantôt en aval, et située alors du côté de la *Zweckerationalität*. Max Weber n'indique donc pas que la rationalité soit du côté de l'intérêt comme finalité, mais il considère toutefois, en une formule célèbre, que la *Wertrationalität* reste toujours affectée d'une irrationalité par rapport à la *Zweckerationalität*, « et cela d'autant plus que l'on donne une signification plus absolue à la valeur d'après laquelle

on oriente l'activité »<sup>1</sup>. On peut directement indiquer qu'un tel déplacement de la rationalité vers la catégorie de l'intérêt conçu dans un sens étroit n'est pas admissible, même si l'on peut comprendre les raisons du glissement de sens ; celles-ci sont essentiellement au nombre de cinq :

- il y a tout d'abord l'ambiguïté linguistique indiquée précédemment. Dire que « la poursuite des intérêts » est rationnelle peut signifier deux choses différentes ;
- de surcroît, si une action est rationnelle dans le choix des moyens, l'action entreprise, complète, contient aussi nécessairement la fin visée, qui est alors incluse dans un ensemble, fins et moyens. La rationalité apparaît alors comme une rationalité de l'ensemble, incluant fins et moyens, donc bien sûr les fins ;
- certains intérêts, comme la recherche de l'argent, peuvent apparaître comme plus universels, ou plus généraux, que certaines valeurs, comme le fait de respecter la mémoire de ses ancêtres. Il y aurait à cet égard une plus grande compréhensibilité due à une plus grande fréquence. Mais on ne saurait en déduire nécessairement une plus grande rationalité, par opposition à ces valeurs de fréquence moindre ;
- il est vrai que de très nombreuses valeurs s'appuient sur des propositions positives qui les fondent. Or, fréquemment, ces propositions positives ayant un caractère erroné, l'adhésion aux valeurs qui en dérivent peut être considérée comme irrationnelle. C'est le cas, par exemple, de l'astrologie. De très nombreuses valeurs dérivent de propositions positives fausses, et elles sont à cet égard irrationnelles. Toutefois, on ne saurait en déduire que *toutes les* valeurs dérivent de propositions positives fausses. Deux autres solutions sont possibles : elles dérivent de propositions positives vraies, elles ne dérivent pas de propositions positives ;
- enfin, la catégorie de l'intérêt est souvent associée à l'idée de calculabilité, ce qui d'ailleurs ne va pas de soi : en effet, face à deux intérêts de nature hétérogène et opposée, l'acteur n'a pas les moyens de calculer son plus grand intérêt, il doit *choisir ce* qu'il préfère. Or, la calculabilité est généralement placée du côté de la rationalité, et, conséquemment, l'intérêt.

Pourtant, si l'on retient cette hypothèse d'une rationalité des fins de l'intérêt conçues en un sens étroit, au détriment d'autres finalités, on est confronté à un dilemme : si l'on admet que les intérêts sont les mêmes

1. M. Weber, *Économie et société*, op. cit., t. 1, p. 57.

pour tout le monde, et qu'ils sont opposés aux valeurs, pourquoi le sociologue devrait-il considérer que le fait de suivre son intérêt est plus rationnel que l'adhésion à une valeur, dans le cas où l'on ne pourrait pas établir que celle-ci dérive d'une proposition positive fausse ?

Au contraire, si l'on considère que les intérêts ne sont pas les mêmes pour tout le monde, et que le sociologue ne les connaît qu'à travers le choix des individus, il sera tautologique, le plus souvent, de dire que les acteurs poursuivent leur intérêt, puisque celui-ci ne sera connu qu'à travers le choix qui en est fait. Le sociologue du choix rationnel n'est pas ici le théoricien abstrait du choix ou l'économiste qui s'intéresse simplement à la cohérence interne des choix, car ce sont les contenus de ces choix qui lui importent. Dans certains cas, en effet, sur la base de certains intérêts connus par le choix, on pourra déduire certains comportements cohérents et conformes à ces intérêts. Mais dans ce cas, qui représente une possibilité réelle mais limitée de l'analyse, la rationalité ne pourra pas être interprétée comme relevant du choix des fins, puisqu'il y aura seulement établissement d'un lien déductif entre différents aspects d'une même finalité, sans que ceux-ci bénéficient d'un privilège quelconque en termes de rationalité.

Des remarques similaires peuvent être faites à propos du choix interprété en termes de coût et d'avantage. On peut tout d'abord indiquer que tout choix correspond, d'une certaine manière, à la sélection de l'option la meilleure pour l'acteur, sinon on ne comprendrait pas pourquoi le choix serait fait. Cette idée, du moins, est défendue par L. von Mises, et constitue un aspect central de sa praxéologie<sup>1</sup>. La définition générale du choix n'implique pas une telle sélection du meilleur : mais ici on prend la mesure de la dimension intentionnelle du choix. On voit mal pourquoi l'acteur choisirait l'option qui n'est pas celle qui lui paraît la meilleure (quels que soient les critères de ce meilleur, qui peuvent être les plus divers).

A partir de là, on peut interpréter tout choix en termes de coût et d'avantage, mais finalement, dans la plupart des cas, de manière simplement métaphorique : cette dimension métaphorique est dépassée (indépendamment de la constitution d'une fonction d'utilité sur laquelle nous reviendrons dans un moment), lorsque l'on délimite certains avantages et certains coûts qui ont un aspect quantitatif : c'est le cas par exemple dans les paris et les jeux, qui représentent un modèle central du choix chez Coleman ; c'est aussi le cas de certaines actions effectives qui sont direc-

1. L. von Mises, *L'action humaine*, Paris, PUF, 1985.

tement liées à des problèmes quantitatifs de coût et d'avantage : le montant des amendes, la durée d'une peine pour un prisonnier, etc.

Mais la plupart des choix possibles pour un acteur sont irréductibles à ce type de situations, qui représentent des schémas particuliers et limités des contextes d'action, même s'ils sont fréquents.

Dans le cadre de ces situations où on arrive, du point de vue des données de l'action, à repérer des coûts et des avantages à minimiser ou maximiser, on peut parler de maximisation de la satisfaction. Mais ceci ne correspond pas du tout à l'idée que la satisfaction est liée ici au choix d'une finalité retenue *contre* des sentiments ou des valeurs, et que cela exprimerait justement la rationalité de l'acteur.

Il y a en effet deux situations possibles. Soit l'action est reliée à une seule dimension, par exemple éviter la prison dans le dilemme du prisonnier. La rationalité ne correspond pas du tout alors au choix de ne pas aller en prison, alors qu'on est coupable, mais au fait, alors que l'on est coupable et que l'on a décidé de ne pas aller en prison, de choisir ensuite les meilleurs moyens de ne pas y aller.

En revanche, fréquemment, certains coûts et avantages objectifs de l'action sont liés à d'autres aspects qui n'ont pas directement cet aspect quantitatif, par exemple la dimension psychologique pour la décision d'avoir ou de ne pas avoir des enfants, qui s'ajoute aux coûts ou bénéfices financiers envisagés. Ceux-ci peuvent être comparés, en sorte que certaines situations soient, de ce point de vue, plus avantageuses que d'autres, comme l'a bien montré Gary Becker. Toutes choses égales par ailleurs, ces coûts et ces avantages peuvent être comparés par les parents qui en tiendront compte éventuellement. Mais la rationalité ne saurait signifier que ces avantages sont sélectionnés prioritairement par rapport aux autres possibles.

On en revient encore à la nécessité de référer ces coûts et avantages aux choix des acteurs : mais si ce sont les choix (variables) qui déterminent des coûts et des avantages, on ne peut pas dire que telle sélection des coûts et des avantages soit, en tant que telle, rationnelle par rapport à une autre, pas plus d'ailleurs que l'on ne puisse associer tel ou tel aspect de l'action nécessairement à tel ou tel coût. Coleman se demande à un moment pourquoi les membres d'une équipe sportive jouent le jeu plutôt que d'être des *free-riders*<sup>1</sup>. Il présuppose par là que jouer le jeu représente un coût, ce qui est une absurdité, il suffit qu'ils aiment jouer. On retrouve le même problème dans la théorie du vote. Certes, une voix de plus ou de

1. James Coleman, *Foundations...*, p. 274.

moins ne fait pas de différence, et du coup je suis confronté à l'inutilité de mon vote. Mais cela implique-t-il que mon vote soit coûteux pour moi ? il peut très bien se faire qu'il soit, à la fois, inutile (du point de vue du résultat) et plaisant.

Il faut donc, si l'on renvoie à des coûts et des avantages substantiels, ne pas considérer que la rationalité dérive ici d'une mise à l'écart, par les acteurs, des valeurs. En revanche, si l'on renvoie à des choix effectués par les acteurs, il faut prendre garde à ne pas par ailleurs juger *a priori* des coûts et avantages, et supposer par exemple que le footballeur devrait avoir envie de laisser tomber le jeu.

Des considérations similaires peuvent être formulées à propos de la notion d'égoïsme. Il n'est pas du tout nécessaire d'associer les notions de rationalité et d'égoïsme. Il est certain que l'étude des conséquences de la poursuite de certains intérêts égoïstes est éclairante : à partir du caractère décevant de certaines stratégies rationnelles à finalités égoïstes, comme dans le célèbre dilemme du prisonnier, on assistera à l'émergence de normes en apparence altruistes sur la base de motivations égoïstes. Coleman a généralisé cette procédure d'analyse. Mais ceci ne permet pas de conclure :

- 1 / qu'une rationalité instrumentale a nécessairement un lien avec de l'égoïsme. Des finalités altruistes peuvent fort bien donner lieu à des stratégies rationnelles ;
- 2 / que la rationalité implique l'égoïsme. Il est vrai qu'en maintes occasions les acteurs ont des intérêts égoïstes, et qu'alors, agissant rationnellement pour les satisfaire, cette rationalité converge alors avec l'égoïsme. Mais si la rationalité est souvent associée à l'égoïsme, parce qu'elle se fait la servante de finalités égoïstes, cette liaison n'est que de rencontre. Croire qu'il y a là une unité absolue de dispositions relèverait à nouveau d'un glissement de sens ;
- 3 / on ne peut pas non plus en conclure qu'une norme, pour pouvoir être expliquée en liaison avec la rationalité des acteurs, doit alors nécessairement dériver de motifs égoïstes.

## 5 | LA MISE EN ORDRE DES PRÉFÉRENCES

Le troisième élément auquel se réfère la théorie du choix rationnel en sociologie, sur la base d'un développement plus général de celui-ci, correspond au principe de la constitution de fonctions d'utilité dérivées de la

mise en ordre des préférences. Il est quelquefois possible que cette mise en forme des choix corresponde à l'effectivité (du moins dans sa possibilité) d'actions individuelles : dans le fameux exemple du dilemme du prisonnier, les acteurs mettent en ordre leurs préférences, qui correspondent par ailleurs à une utilité substantielle quantitative, passer le moins de temps en prison. Ils calculent alors la meilleure solution pour eux, de manière égoïste.

Cet exemple remarquable peut renforcer l'idée d'une unité intrinsèque entre les trois aspects de la théorie du choix rationnel tels que nous les avons distingués. Il permet de confirmer par ailleurs le sentiment d'un lien fort entre rationalité et égoïsme.

Pourtant, comme l'on sait, ce procédé des fonctions d'utilité fut promu en partie pour éviter les difficultés d'une comparaison interpersonnelle d'avantage, et pour renvoyer donc cet avantage au choix des acteurs. Dès lors, la notion d'égoïsme ne peut plus avoir qu'un sens formel : un choix personnel peut s'opposer à un autre choix personnel. Mais ce choix formellement égoïste ne l'est pas nécessairement d'un point de vue substantiel, puisque la mise en ordre des préférences peut très bien concerner des motifs altruistes d'action. Par exemple, une organisation charitable doit toujours sélectionner, parmi le lot des malheureux, ceux qui seront aidés en priorité. Comme le principe de la mise en ordre des préférences est nécessairement lié au principe général du choix qui permet de les connaître, il n'y a alors aucune spécification *a priori* du contenu de ces choix.

La rationalité correspond simplement au fait que, compte tenu d'une mise en ordre des préférences, et de contraintes auxquelles le choix est soumis, l'acteur maximise son utilité. Mais ce n'est pas parce qu'il maximise son utilité, en tant que celle-ci représenterait une fin en soi, opposée à des valeurs ou des sentiments, qu'il est rationnel. L'utilité ici ne représente pas un motif distinct d'autres motifs, mais *présuppose* ces motifs.

Il y a à nouveau un glissement de sens possible, qui va associer l'utilité comme telle au choix rationnel : on peut trouver un tel glissement de sens chez Harsanyi<sup>1</sup>.

L'acteur apparaît chez lui rationnel parce qu'il se réfère à une finalité, l'utilité. L'utilité, pourtant, n'est pas dans ce type de représentations une réalité naturelle, mais une réalité construite à partir des préférences individuelles. La rationalité ne vient donc pas de ce qu'il existerait quelque

1. J. Harsanyi, *Rational behavior and bargaining equilibrium in games and social situations*, Cambridge, Cambridge University Press, 1977, p. 8.

chose comme de l'utilité à maximiser pour être rationnel, cette utilité fournissant, par nature, le plus de satisfaction possible : la satisfaction est en fait définie antérieurement, au niveau des préférences, ce qui permet la procédure ultérieure de maximisation sous contrainte qui présuppose nécessairement une définition préalable des termes de la satisfaction.

Dès lors, partant de tels principes, la maximisation d'avantage ou minimisation de coût ne correspond en rien à des avantages ou coûts substantiels, tels que l'argent, le pouvoir, le prestige ou le temps passé en prison. Le principe d'un mise en ordre des préférences est bien plus large, et indépendant, formellement, de telles réalités concrètes, pouvant même, éventuellement, s'opposer à elles. Il peut certes y avoir convergence de fait entre cette procédure formelle et ce type de réalités substantielles qui seraient retenues par le choix, mais évidemment ces réalités substantielles particulières n'épuisent pas le registre des choix possibles.

Dès lors, dans la mesure où ces procédures de mise en ordre des préférences soulèvent des difficultés d'ordre épistémologiques considérables, il est peu utile de tenter de soumettre l'analyse sociologique à ce type de démarche. Vouloir le faire à tout prix, comme dans le cas de Coleman, conduit à des ambiguïtés importantes. Devant la difficulté qu'il y a à retrouver les courbes d'utilité des acteurs décrits (par exemple la jeune fille qui va s'aventurer dans les bois en compagnie d'un jeune homme qu'elle croit galant, mais qui sera dangereux et la violera), Coleman se rabat toujours sur un modèle beaucoup plus étroit que celui du choix rationnel en général, et associe les actions à des intérêts spécifiés qu'il cherche à mettre en évidence. Lorsque cela n'est pas possible, il y a alors référence à la structure générale du choix qui permet d'interpréter toute décision en termes de coût et d'avantage. Le problème est alors que ces métaphores ne sont d'aucune utilité scientifique.

La fécondité, si elle existe, des distinctions ici faites doit être de permettre d'éviter certaines confusions. Elle doit permettre aussi de ne pas enfermer la rationalité dans un modèle unique et limité, ne serait-ce que parce que celui-ci représente lui-même des tendances diverses, soit légitimes quand elles ne reposent pas sur des confusions, soit illégitimes lorsque celles-ci interviennent. La rationalité apparaîtra alors comme un concept plus général que la dimension stratégique de l'action, plus général que la référence à des intérêts et plus général que les procédures formelles de choix qui ne permettent pas de rendre compte du choix des finalités. Il convient de renvoyer ici au programme de R. Boudon qui s'inscrit lui-même à la suite d'une tradition sociologique vénérable d'explication des actions individuelles.

## *Aspects praxéologiques des principes interpersonnels<sup>1</sup>*

### INTRODUCTION

L'existence collective peut être régie, sur le mode normatif, par des principes moraux, politiques, religieux ou juridiques qui, s'ils devaient être respectés, impliqueraient certaines restrictions sur ce que les uns ou les autres peuvent faire sans en être empêchés, ou encore, beaucoup plus généralement, des restrictions sur le type de souhaits que les uns ou les autres peuvent voir réalisés sans que cela contredise les principes considérés.

Ces « principes interpersonnels » peuvent relever, par exemple, d'une éthique ou d'une politique fondée sur le respect d'autrui, ce qui implique normalement l'absence d'interférence avec les décisions d'autrui en certaines matières<sup>2</sup>. Ce sont alors des principes qui sélectionnent, parmi les configurations possibles de l'état de la société, ou parmi les modes possibles d'action conjointe des agents formant la société, ceux qui se conforment à certaines normes exprimant le type d'autonomie ou d'indépen-

1. Je remercie Alban Bouvier, Pierre Demeulenaere, Serge Diébolt et Philippe Mongin pour leurs commentaires très utiles d'une première version de ce texte. Je garde seul la responsabilité des insuffisances qui demeurent.
2. Probablement pas, cependant, l'absence totale d'interférence. Même lorsque le respect de l'autonomie ou de la liberté d'autrui trouve sa source dans la reconnaissance de sa nature d'être rationnel, la contrainte normative extérieure n'est évidemment pas toujours à penser comme une restriction de la liberté ou de l'autonomie personnelle. J'ai abordé ce problème dans *Choix rationnel et vie publique*, Paris, PUF, 1996, chap. X, sect. 2. Le même problème est analysé par Allan Gibbard dans *Wise Choices, Apt Feelings* (Cambridge, Mass., Harvard University Press, 1990), trad. franç. par S. Laugier, *Sagesse des choix, justesse des sentiments* (Paris, PUF, 1996), p. 311-313.

dance que l'on reconnaît aux personnes. Telle est la démarche la plus habituelle que l'on trouve à la source de la spécification ou de la revendication des droits individuels. Une telle approche est « déontologique », au sens où elle privilégie, parmi les propriétés de l'interaction des personnes, la conformité à certaines normes, dont le respect aurait une importance intrinsèque<sup>1</sup>.

D'autres principes relèvent plutôt d'une évaluation des états ou des conduites sur la base d'un but à promouvoir. Ce sont les principes dits « téléologiques ». Il peut s'agir, par exemple, de principes relevant d'une méthodologie de l'« information-utilité » (le *welfarism* des auteurs de langue anglaise), selon laquelle le bien-être des personnes (à l'exclusion de tout autre élément d'information) est le fondement de l'évaluation des états de la société ou des actions conjointes modifiant (ou susceptibles de modifier) l'état de la société. Plus généralement, il peut s'agir de principes relevant du « principe de la conséquence » en vertu duquel l'évaluation des conduites conjointes possibles des membres de la société et la non-exclusion de ces conduites conjointes par le système de normes que l'on retient doivent se fonder sur les conséquences de ces actions conjointes (et non pas sur leur concordance ou discordance avec certains principes éthiques portant sur l'action même, abstraction faite de ses conséquences)<sup>2</sup>.

Cette distinction de deux grandes catégories de principes interpersonnels n'est pas universellement acceptée. Ainsi, Kelsen a défendu de manière convaincante la thèse de la priorité du déontologique dans tous les cas, à partir des considérations suivantes. Tout d'abord, il est faux de prétendre identifier la nécessité normative du devoir-être qu'instaure une norme à la nécessité téléologique. En effet, une fin est un effet voulu, et n'a de sens qu'en relation avec quelque chose d'autre qui est un « moyen » ; or, « [l]a norme, qui pose comme obligatoire un certain comportement, ne se réfère en aucune manière à un moyen vis-à-vis duquel le comportement obligatoire est une fin, ou à une fin vis-à-vis de laquelle le comportement obligatoire est un moyen ». Certes, l'éthique téléologique suppose qu'un comportement puisse être dit moralement bon s'il correspond à une fin présumée (par exemple le plaisir, le bonheur ou l'excel-

1. Pour un examen de la spécificité de ces systèmes, et de l'opposition entre morale déontologique et morale téléologique, voir notamment : Philippe Mongin et Claude d'Aspremont, « Utility Theory and Ethics », à paraître dans le *Handbook of Utility Theory*, sous la direction de S. Barbera, P. Hammond et C. Seidl (Dordrecht et New York, Kluwer Academic Press).
2. Sur l'information-utilité et le principe de la conséquence, et sur leur rôle dans la doctrine utilitariste, voir notamment les analyses classiques d'Amartya Sen dans *On Ethics and Economics* (Oxford, B. Blackwell, 1987), trad. franç., *Éthique et économie* (Paris, PUF).

lence humaine). Mais dans ce cas même, la norme est à distinguer de la fin que pose celui qui l'énonce, et c'est bien la conformité du comportement à la norme (non pas immédiatement le statut instrumental relativement à la fin posée) qui permet de dire qu'un acte est « bon » ou possède une valeur morale. Convenablement analysée, la formule générale de l'éthique téléologique (est « bon » ce qui est un moyen pour réaliser une certaine fin) se ramène donc à l'exigence de conformité à une norme posant qu'il faut faire des choses ayant une valeur instrumentale au regard de telle ou telle fin. On retrouve un principe déontologique : il faut faire ce qui est considéré comme un devoir. Ainsi, « il n'y a qu'une éthique, à savoir l'éthique normative ou déontologique »<sup>1</sup>.

Il reste néanmoins couramment admis qu'il y a lieu de continuer à évoquer un pôle « déontologique » et un pôle « téléologique » dans la théorie morale ou politique, même si l'on pense que les systèmes téléologiques peuvent incorporer une conception des devoirs, et les systèmes déontologiques, une conception du bien<sup>2</sup>. On reconnaît que la différence entre l'éthique téléologique et l'éthique déontologique consiste essentiellement en ceci, que la première s'exprime par des normes qui font intervenir une conception du bien que l'on peut définir indépendamment de la conformité de l'action à ces normes, et qui recommandent ou commandent exclusivement d'agir en vue de ce bien<sup>3</sup>. A l'opposé, la détermination de « ce qu'il faut faire » dans le contexte d'une éthique déontologique repose sur des critères – le devoir, l'obligation, la convenance, l'honnêteté, la beauté morale, le droit ou le correct, le respect d'autrui, etc. – qui caractérisent l'action en propre (non pas sa relation à autre chose).

Il est sans doute possible d'admettre la priorité du déontologique à l'échelon le plus fondamental de la normativité, tout en reconnaissant qu'il existe entre les différents principes interpersonnels des différences de formulation ou de construction qui justifient l'identification d'un pôle « téléologique ». Le vrai problème est de déterminer si ce pôle téléologique appelle des méthodes de description et d'analyse différentes de celles qui conviennent à la politique et à l'éthique déontologiques. En d'autres termes, le vocabulaire des devoirs, des droits et de l'obligation et le

1. Hans Kelsen, *Allgemeine Theorie der Normen*, Vienne, Manz Verlag, 1979 ; trad. franç., *Théorie générale des normes*, Paris, PUF, 1996, chap. 2 (p. 13-14) et n. 16 (p. 381-382).
2. C'était déjà la thèse de W. David Ross dans *Foundations of Ethics*, Oxford, 1939, objet des remarques critiques de Kelsen (ouvr. cité, p. 381).
3. John Rawls, *A Theory of Justice*, Oxford, Oxford University Press, 1972, p. 25 ; trad. franç., *Théorie de la justice*, Paris, PUF.

vocabulaire du bonheur, de l'excellence et de l'utile sont-ils à ce point hétérogènes que l'on ne puisse en aucun cas décrire dans un cadre théorique commun les situations et les problèmes auxquels ils s'appliquent ? En outre, y a-t-il des modalités de justification rationnelle propres à la dimension « déontologique », et d'autres modalités de justification rationnelle qui conviendraient exclusivement à la dimension « téléologique » ? Faudrait-il donc évoquer deux types séparés d'adoption rationnelle des normes ?

Cet essai est destiné à montrer que l'on peut avoir de bonnes raisons d'accepter, à propos de cas particuliers importants, les deux thèses suivantes :

- 1 / il est parfois possible d'aborder simultanément l'étude des deux grandes catégories de principes dans un cadre unifié, en privilégiant l'analyse de la compatibilité ou de l'incompatibilité entre les intentions individuelles (projets ou souhaits) que peuvent envisager les personnes ;
- 2 / ce type d'approche permet d'identifier une propriété générique de l'adoption des systèmes de normes, dont on peut estimer qu'elle donne une certaine force de conviction (au sens de l' « aptitude à être justifié ») à plusieurs principes fondamentaux qui se rattachent aux deux grandes catégories mentionnées : du côté « déontologique », certains principes souvent présentés comme faisant l'objet d'un devoir (au moins en l'absence de très bonnes raisons allant en sens contraire : ne pas voler, ne pas agir en faussaire, etc.) ; du côté « téléologique », l'optimum de Pareto ; enfin, l'attribution des libertés individuelles fondamentales, qui fait l'objet de tentatives de justification, comme on le sait, soit d'après des principes déontologiques (le devoir de respecter l'autonomie d'autrui, le devoir de consacrer la dignité d'êtres libres que possèdent les citoyens, etc.), soit de manière téléologique (en invoquant le bonheur de l'humanité par exemple).

Il sera utile d'esquisser un cadre d'analyse simple destiné à faciliter l'étude des principes interpersonnels considérés dans leur rapport aux modalités de coexistence des personnes. Nous étudierons ensuite les possibilités de traduction de certains principes interpersonnels classiques dans le vocabulaire et le cadre conceptuel des intentions individuelles et des restrictions qui peuvent peser sur elles. Un procédé de délimitation théorique d'un « domaine réservé » des personnes sera proposé, puis rapproché d'éléments de justification d'autres types de principes, ce qui donnera lieu à une formulation plus générale.

## 1 | ACTIONS ET INTENTIONS DANS L'INTERACTION SOCIALE

a) *L'action est-elle un bon guide pour la philosophie des normes ?*

Les normes qui ont de l'importance pour l'existence collective se réfèrent aux conduites des uns et des autres, soit directement parce qu'elles prévoient des restrictions sur les conduites qu'il sera possible de tenir, soit indirectement parce qu'elles prévoient des restrictions sur les états du monde et de la société qui résultent des conduites. Or, l'action semble une réalité quelque peu insaisissable, de sorte qu'il est légitime de se demander si elle peut être un bon guide pour l'étude des principes interpersonnels. Bergson, notamment, a proposé des analyses qui expriment une profonde défiance à l'égard des concepts et de la langue dans laquelle on tente ordinairement d'exprimer la nature ou les finalités de l'action : rapportant nos concepts à la dimension du « spatial », Bergson donne à penser que l'action vécue échappe toujours, par nature, aux moules fixes et grossiers dans lesquels on aimerait la couler.

Mais il y a probablement une part de résignation inutile dans la thèse selon laquelle toute théorie de l'action trahit nécessairement son objet. Il importe surtout de se demander *dans quelle mesure* la théorie a réellement besoin d'une description fidèle de l'action. Considérons comme un fait qu'une description d'action ou de conduite est susceptible de correspondre ou non à la situation qui prévaut finalement, aux yeux d'une certaine personne. On doit alors considérer que les difficultés affectant la description de l'action n'empêchent pas de se demander si une personne a réalisé ou non le projet qui, à ses yeux, pouvait être résumé par une certaine description d'action ou de conduite (ou plus généralement encore, de modalité d'existence au sein de la société). Or, la réalisation ou non-réalisation des projets ou intentions est clairement un élément important pour l'étude des normes interpersonnelles. Les normes prévoient en effet que certains projets imaginables ne pourront être réalisés, ou ne pourront être réalisés conjointement avec certains projets d'autrui, ou encore, que certaines personnes *devront* mettre en œuvre des projets auxquels elles n'auraient peut-être pas songé spontanément.

La notion même d'« intention » enveloppe la possibilité d'une comparaison entre ce qui est visé en intention et ce qui est finalement accompli, obtenu ou simplement vécu. On ne parlerait pas des « intentions » de l'acteur si l'on ne jugeait possible, pour celui-ci, de constater l'accord ou

la discordance entre ce qu'il projetait *avant l'action* et ce qu'il admet avoir fait, obtenu ou vécu *une fois l'action (ou l'inaction) accomplie*. Puisque l'accord ou le défaut d'accord peut être constaté et exprimé, il s'agit bien d'un point de départ envisageable pour les théories ayant trait à la conduite humaine. Il faut simplement prendre acte du fait que les termes mis en correspondance par les relations « accord » et « défaut d'accord » ne sont pas les actions telles qu'en elles-mêmes, mais des descriptions d'actions<sup>1</sup>.

Par ailleurs, l'étude des *conflits d'intentions*<sup>2</sup> donne prise sur l'action, selon des modalités dont on ne peut manquer d'apercevoir qu'elles présentent un réel intérêt pour l'étude d'un grand nombre de questions éthiques ou politiques. Tout d'abord parce que les normes interpersonnelles (qu'elles soient éthiques ou politiques), loin d'être seulement des commandements, assurent des solutions définies à ces conflits. Qu'il soit légitime ou non de prétendre que telle est leur « fonction », il est constant que c'est là ce qu'elles font dans le contexte de l'existence collective. Si incertaine que soit la reconstitution linguistique des traits les plus importants de l'action concrète vécue, un fait demeure : il nous est parfois possible de dire que deux individus ont (ou n'ont pas) la faculté de réaliser simultanément leurs projets<sup>3</sup>.

Ainsi, analysant ce que peuvent comporter d'arbitraire, malgré leur contenu intuitif, les doctrines qui opposent les « intérêts » aux « sentiments » (Pareto) ou aux « valeurs » (Weber), Pierre Demeulenaere a envisagé, à un certain niveau de son analyse, de regrouper ces différentes enti-

1. Ainsi, l'accord peut se transformer en défaut d'accord si la description gagne en précision, l'intention demeurant inchangée : si mon intention est d'aider mes amis de manière honnête, il est possible que je juge finalement ma conduite contraire à mon intention, alors même que ma conduite s'accorde avec une description plus grossière de mon intention (« aider mes amis »).
2. Le mot « intention » est pris ici (et dans la suite) en un sens très large : ce qu'un agent peut viser, soit en termes d'état personnel, soit en termes de conduite à tenir, au cours d'une interaction avec d'autres personnes. Le mot « conduite » permet de renvoyer indistinctement à l'action et à l'absence d'action (l'inaction étant de fait, bien souvent, une possibilité offerte aux acteurs). L'action elle-même peut se composer de plusieurs actes. Le mot « conduite » permet aussi d'incorporer dans les intentions associées (les « intentions de conduite ») certaines conditions relatives aux circonstances, sur le modèle : « Si telles circonstances prévalent, agis (ou n'agis pas) de telle manière, et si telles autres circonstances prévalent, agis (ou n'agis pas) de telle autre manière, etc. » On peut considérer que les descriptions d'actions (à la différence des descriptions d'actes ponctuels) incorporent des conditions relatives au temps (sur le modèle : « A tel moment, fais ceci ; ensuite, fais cela, etc. »). Je m'en tiens à ce point de vue. Il est alors possible de considérer que les intentions d'actes interviennent dans les intentions d'action, les intentions d'action dans les intentions de conduite, et les intentions de conduite dans les intentions au sens large que l'on retient ici.
3. C'est ce que nous appellerons plus loin la possibilité de réalisation de la « configuration d'intentions conjointes » (en abrégé, CIC) qui est la leur.

tés sous l'appellation de « motivations » (toujours associées à certaines valeurs ou à certains sentiments), et a montré qu'il reste alors possible de distinguer d'une manière significative les situations dans lesquelles les motivations des uns et des autres sont compatibles, et les situations dans lesquelles elles s'opposent entre elles. Or, le critère permettant d'opérer cette distinction n'est autre que le conflit interpersonnel, autrement dit, l'impossibilité pour les uns et les autres de réaliser simultanément ce qui est visé en intention par chacun :

« Ainsi, les intérêts du fumeur, dans un café, sont opposés aux intérêts du non-fumeur gêné par la fumée, et reposent sur la valorisation ou la dévalorisation de l'acte de fumer. De même, deux membres de deux différentes religions prosélytes, essayant chacun de convertir ceux qui n'appartiennent pas à l'une de ces deux religions, ont des intérêts opposés. »<sup>1</sup>

Le conflit entre les intentions des uns et celles des autres (par exemple « fumer » – « ne pas respirer de l'air enfumé », ou bien « convertir » – « ne pas se convertir »), au sens précis de l'impossibilité d'une réalisation complète simultanée, introduit une polarité au sein des intentions d'action : les intentions d'un certain groupe s'opposent à celles d'un autre groupe d'une manière inévitable. Cette polarité semble échapper, dans une large mesure, aux incertitudes qui affectent la compréhension ou la représentation de l'action en tant que telle. Elle permet de circonscrire de manière intelligible une certaine catégorie de motifs d'action, non pas à partir de la nature psychologique ou morale intrinsèque de ces motifs, mais sur la base d'une pure relation de contradiction ou d'incompatibilité révélée dans l'interaction possible.

Dans ces exemples, les intentions considérées sont appréhendées *sous une certaine description*, suffisante pour apercevoir la contradiction qui est en jeu. Il n'est pas question de leur nature ultime en tant que faits naturels ou psychologiques. Certes, il n'est pas facile de communiquer à autrui ce qu'est exactement pour moi l'expérience vécue de l'action de fumer ou du fait de respirer de l'air pur. Et pourtant, chacun comprendra que, si deux actions (ou modalités d'insertion dans l'environnement naturel et social) peuvent respectivement être décrites respectivement, fût-ce de manière partielle, par « fumer du tabac » et « respirer de l'air exempt de fumée de tabac » et si deux personnes sont dans une même pièce, la première ayant l'intention de réaliser la première action, et la seconde, la seconde action, alors ces deux personnes ne peuvent simultanément réaliser leur intention.

1. P. Demeulenaere, « Les ambiguïtés constitutives du modèle du choix rationnel » (dans ce volume).

L'éthique sociale et la philosophie politique ont très souvent à examiner des conditions de possibilité ou impossibilité conjointe de ce type, puisqu'elles traitent notamment des conditions de règlement de certains types de conflits possibles entre les conduites ou les aspirations des personnes. Il y a donc lieu de réaffirmer que la prétention de ces disciplines à une argumentation de type rationnel ne peut être réputée nulle au motif de l'inaptitude supposée du langage et de nos concepts à transmettre une idée juste de l'action concrète et vécue. Il est concevable qu'une théorie prenant pour objet les conduites et leurs aspects interpersonnels reste possible, quand bien même la nature intime de l'action serait, de soi, ineffable. A cause des deux raisons qui ont été mentionnées (intelligibilité des comparaisons entre intentions et actions accomplies, et intelligibilité des conflits d'intentions), il y a lieu de penser qu'il ne faut pas négliger les possibilités d'analyse procédant des conditions de compatibilité entre les intentions. Il faut seulement se garder de confondre les intentions et les descriptions que l'on en donne. En d'autres termes, il faut éviter de croire que nos descriptions nous donnent prise sur l'action même, sur la totalité vécue d'une fraction de l'existence humaine. L'action vécue, idéalement préservée de toute déformation, de toute réduction symbolique ou linguistique, ne se livre intégralement qu'à celui qui la vit. Elle n'intéresse pas directement la théorie politique (ni même probablement l'éthique sociale), puisqu'elle ne peut intervenir telle qu'en elle-même dans la discussion ou l'argumentation. Par contre, les descriptions d'intentions sont à parité de niveau avec les conditions usuelles de la communication humaine, la formulation des règles de la vie commune et la justification de celles-ci.

*b) Exemples de principes interpersonnels*

Les principes interpersonnels ne légifèrent pas tous relativement aux mêmes objets, mais cela ne rend pas nécessairement impossible de les appliquer simultanément à certaines situations ou certaines modalités d'interaction. Le principe de Pareto, par exemple, concerne les niveaux d'utilité ou de bien-être des personnes concernées, et il en va de même des divers principes inspirés par l'approche utilitariste du problème de l'évaluation éthique des états d'une société. Par contre, dans la formulation d'un principe énonçant un droit individuel – tel que :

(M) « Toute personne célibataire peut se marier avec une autre personne célibataire consentante, et nul ne peut l'en empêcher »,

les niveaux de bien-être n'interviennent pas, bien que l'on puisse évidemment s'intéresser *par ailleurs* à la vérification ou non-vérification d'un principe tel que le principe de Pareto dans des situations reflétant la mise en œuvre de (M). Si important qu'il puisse être par ailleurs du point de vue du bonheur ou de la satisfaction des aspirations des individus, le principe (M) prend appui sur une réalité d'un genre bien différent : le domaine d'exercice ouvert à la faculté de choix des personnes. En effet, c'est parmi les actions qu'une personne peut désirer entreprendre que le principe (M) opère une double sélection (évidemment non exhaustive) : d'un côté, celles qui pourront à coup sûr être menées à bien (« épouser quelqu'un qui le veut bien ») ; de l'autre, celles qui ne pourront en aucun cas être menées à bien (« empêcher quelqu'un de se marier avec une personne célibataire consentante »).

Le principe de Pareto, lui, opère un tri parmi les états sociaux possibles, envisagés comme configurations des niveaux de bien-être des individus, et donne un critère partiel pour apprécier la légitimité de certaines actions conjointes ayant des conséquences pour la situation d'une société. Ainsi, relativement à un état de référence, le principe, dans l'une de ses versions (le « principe des améliorations parétiennes »), pourra être formulé de la manière suivante :

(PP) « Relativement à l'état de référence, tout état réalisable dans lequel le niveau de bien-être d'au moins un individu est supérieur sans que soit inférieur celui d'aucun individu est dit "meilleur au sens de Pareto". La substitution d'un état de la société à un autre ne doit pas se produire lorsqu'elle est évitable et lorsque l'état initial peut être dit "meilleur au sens de Pareto" que le nouvel état. Doit se produire, par contre, la substitution à l'état initial d'un état meilleur que lui au sens de Pareto, lorsque cette substitution est possible. »<sup>1</sup>

La théorie formelle des choix collectifs a fréquemment recours à des principes auxquels on rattache également le nom de Pareto. Par exemple, le « principe faible de Pareto » invoqué dans cette théorie signifie que, si tous les membres d'une société préfèrent unanimement un état *a* de la société à un autre état *b*, l'état *a* doit également être « socialement » pré-

1. On dit d'un état qu'il constitue un « optimum de Pareto » si aucun autre état accessible à la société n'est meilleur que lui au sens de Pareto. C'est un état à partir duquel l'accroissement du bien-être de quiconque implique la dégradation de celui d'au moins une personne. Voir, pour un exposé original : Vilfredo Pareto, *Traité de sociologie générale*, trad. franç. de P. Boven, t. XII des *Œuvres complètes*, Genève, Droz, 1968, p. 1339.

féfé à l'état *b*. Autrement dit (pour une interprétation en termes de choix plutôt qu'en termes d'évaluation), si les états *a* et *b* étaient accessibles à la société (il peut s'agir, par exemple, de deux candidats à une élection), et si la société devait évoluer vers l'un ou l'autre de ces deux états, le « principe faible de Pareto » conduirait à choisir l'état *a*. Ce principe peut donc être formulé ainsi :

(P) « Si la société doit évoluer de son état présent vers l'un ou l'autre de deux états *a* et *b*, dont le premier est unanimement préféré au second, la société doit évoluer vers l'état *a*. »<sup>1</sup>

Cette différence des objets contribue à rendre difficile la comparaison ou l'arbitrage entre des principes fondés sur les niveaux de bien-être et des principes fondés sur l'attribution de droits, dans les cas où l'on peut mettre en évidence un conflit net entre des critères relevant de l'une et l'autre catégorie, par exemple dans le « paradoxe du parétien libéral » exposé par A. Sen en 1970<sup>2</sup>.

## 2 | ÉLÉMENTS POUR L'ANALYSE DES RÉALISATIONS CONJOINTES D'INTENTIONS

### a) Définitions

Dans ce qui suit, on considère une société *N* formée de *n* individus.

**Définition 1.** On appellera *circonstance* la description de certains aspects du monde extérieur qui peuvent affecter, à un instant donné, la

1. Pour la commodité de l'exposé, je formule ce principe sur la base de deux options concurrentes. La formulation en termes d'action conjointe ne trahit pas, semble-t-il, l'esprit du principe évoqué. Celui-ci est présenté, lorsqu'on privilégie les choix collectifs plutôt que l'évaluation des situations collectives, comme une propriété de la manière dont le processus gouvernant l'évolution de la société (la « fonction de choix social » au sens de K. Arrow) met en relation les ordres de préférence « individuels » et l'ordre de préférence « social » : si les premiers privilégient *a*, le dernier doit aussi privilégier *a*. Mais dans cette interprétation, le fait que l'ordre « social » privilégie *a* ne signifie pas autre chose que ceci : lorsque la société peut évoluer soit vers *a*, soit vers *b*, elle évolue vers *a*. La « société » n'est pas un être susceptible de posséder des préférences au sens plein du mot, et les principes de classement « sociaux » envisagés par la théorie des choix collectifs ne font que résumer les directions que prendrait (ou devrait prendre) une société dans l'hypothèse où son évolution serait déterminée – au moins partiellement – par certains choix de ses membres (hypothèse qui ne contredit qu'un déterminisme intégral).
2. A. K. Sen, « The Impossibility of a Paretian Liberal », *Journal of Political Economy*, 1970. Cet article fournit la matière du chap. 6\* de l'ouvrage du même auteur : *Collective Choice and Social Welfare*, Amsterdam, North Holland, et Edimbourg, Oliver & Boyd, 1970.

situation d'au moins un membre de la société (aspects dont l'émergence n'est pas due à la conduite des membres de cette société).

**Définition 2.** On appellera *intention* (ou *description d'une intention sociale*, ou *intention sociale*) la donnée de diverses propriétés d'une façon d'être d'un membre de la société au cours d'un certain intervalle de temps (ces propriétés pouvant éventuellement être spécifiées conditionnellement à la réalisation de certaines circonstances ou conditionnellement à la vérification de certaines propriétés de l'interaction des membres de la société).

**Définition 3.** On appellera *configuration d'intentions conjointes* (CIC) la donnée de  $n$  descriptions d'intentions sociales (une pour chaque individu).

Par une généralisation immédiate, on peut aussi considérer des CIC associées à des sous-ensembles d'individus.

**Définition 4.** On appellera *réalisation d'une intention sociale* le fait, pour la personne à laquelle on attribue cette intention, de connaître effectivement des modalités d'existence au sein de la société qu'elle juge conformes à ce que décrit l'intention. On appellera *réalisation d'une CIC* la réalisation de toutes les intentions individuelles composant cette CIC.

**Définition 5.** Pour une CIC donnée, on appellera *interaction associée à cette CIC* une description s'appliquant au mode d'être des membres de la société au cours d'une période de temps donnée, suffisamment précise pour permettre de dire, pour chacun des membres de la société, s'il a pu (ou non) connaître au cours de cette période des modalités d'existence conformes à ce que décrit l'intention sociale qu'on lui prête dans la CIC considérée.

**Définition 6.** Pour une CIC donnée, et une interaction donnée associée à cette CIC, on appellera *concrétisation de la CIC au cours de l'interaction* l'ensemble des membres de la société qui, dans cette interaction, réalisent leur intention (celle qui est prévue pour chacun d'entre eux dans la CIC).

**Définition 7.** Pour une CIC donnée, on appellera *contexte d'interaction associé à cette CIC* (ou simplement *contexte*) la donnée des composantes suivantes d'une interaction quelconque associée à la CIC :

- a) une *période de temps* ;
- b) éventuellement, des éléments descriptifs relatifs à une *situation de référence* des membres de la société et qui, conjointement avec d'autres éléments de l'interaction, doivent être pris en compte pour préciser si les intentions des uns ou des autres sont ou non réalisées. Par exemple : une certaine description de la situation initiale (si la CIC com-

prend des intentions du type : « connaître une situation meilleure que la situation initiale »).

- c) éventuellement, les *circonstances* qu'il peut être nécessaire de prendre en compte pour préciser la possibilité ou l'impossibilité pour les membres de la société de réaliser leur intention respective.

On peut aussi considérer des contextes d'interaction sans préciser qu'ils sont associés à une CIC. Un *contexte d'interaction* est alors simplement la donnée d'une période de temps, d'éléments descriptifs relatifs à la situation initiale des membres de la société, et d'éléments descriptifs relatifs à l'interaction de ces personnes.

**Définition 8.** On appellera *correspondance d'interaction* la correspondance  $F$  qui, au couple formé par une CIC  $X$  et un contexte d'interaction  $C$ , associe l'ensemble des concrétisations possibles de  $X$  étant donné  $C$  (*i.e.* l'ensemble des concrétisations de  $X$  qui sont possibles dans des interactions diverses respectant les limites qu'impose le contexte d'interaction considéré).

b) *Commentaire des définitions*

Une « circonstance » peut être déterminée soit par des événements naturels, soit par les actes de personnes extérieures à la société, soit par un mixte de ces divers actes ou événements. La description que nous appelons « circonstance » est composée d'éléments relatifs à des événements, actes ou états qui ne sont pas purement mentaux (qui appartiennent par certains traits au moins à ce qui est habituellement pensé comme le « monde extérieur »). Ces éléments, bien qu'ils ne soient pas eux-mêmes des événements mentaux, peuvent déterminer, pour certains agents, l'accessibilité des différents états mentaux qui peuvent être les leurs (par exemple, des degrés de satisfaction ou un sentiment de liberté). Une circonstance est donc une description de certains aspects du « monde extérieur » dans lequel évoluent, coexistent et interagissent les membres de la société.

Les définitions proposées conduisent à donner un statut méthodologique privilégié à la notion d'« intention sociale », et à étudier les conditions de la réalisation conjointe des intentions sociales. Les conflits auxquels on s'intéresse peuvent être identifiés à l'impossibilité de la réalisation, dans certains contextes d'interaction, de certaines configurations d'intentions conjointes des membres d'une société. Ces conflits concernent les conduites, soit parce qu'ils caractérisent directement la

coexistence des conduites (ou actions), soit parce qu'il s'agit de conflits entre des souhaits ou aspirations dont la concrétisation dépend au moins partiellement des conduites des uns et des autres. Les conflits sont de deux sortes, selon qu'ils sont dus à l'incompatibilité entre les intentions d'un individu donné (contradiction *interne*), ou bien à l'impossibilité de la réalisation simultanée des intentions de divers individus, en l'absence de contradiction interne dans les systèmes d'intentions des protagonistes (contradiction *externe*).

Les « circonstances » nous intéressent, d'une part, parce qu'elles peuvent intervenir dans les intentions sociales – ces dernières pouvant comprendre des embranchements du type : « si la circonstance C prévaut, faire (ou ne pas faire) ceci ; sinon faire (ou ne pas faire) cela » – et, d'autre part, parce qu'elles peuvent déterminer au moins partiellement les possibilités de concrétisation à un instant donné (ou d'accomplissement sur une période de temps) des intentions des membres de la société. Les « circonstances » que nous devons prendre en compte sont donc celles qui interviennent explicitement dans les intentions considérées et celles qui, lorsqu'elles surviennent, peuvent compromettre (dans le cas général, à cause de leur conjonction avec d'autres propriétés de l'interaction) la possibilité de réalisation de l'intention d'au moins un agent.

Les propriétés mentionnées dans les descriptions d'intentions ou d'interactions peuvent comprendre des éléments tels que des actes, l'obtention de tel résultat dans l'action, certains aspects de la situation personnelle des membres de la société, l'évolution de cette situation personnelle sous tel ou tel rapport, l'état des rapports entre l'agent et les autres membres de la société (ou l'évolution de ces rapports), etc. Ces aspects peuvent être purement psychologiques (par exemple, un état de contentement ou un sentiment de liberté). Ce sont alors les agents eux-mêmes, plutôt que le théoricien, qui peuvent juger qu'ils sont effectivement présents dans l'interaction vécue. Mais le théoricien peut accorder de l'importance au fait que les agents jugent ou non que ces aspects de leurs intentions sont réalisés. Il ne peut donc s'en désintéresser dans tous les cas.

Si une description d'intention sociale peut être décomposée en plusieurs éléments, on peut appeler « intentions » ces éléments, et considérer la description globale comme un ensemble regroupant ces diverses intentions. En ce sens, on peut parler de l'« ensemble des intentions » de chaque agent. L'« intention sociale » est la projection dans l'avenir d'un certain mode d'être d'une personne au sein d'un groupe de personnes. Une intention sociale peut également être envisagée comme la description d'un mode d'être effectif (et non d'un mode d'être visé ou projeté). Dans

l'un et l'autre cas, on considère la même entité linguistique. Seule l'interprétation diffère, et cette différence peut être ramenée à l'opposition entre point de vue *ex ante* (avant l'interaction sociale concernée) et point de vue *ex post* (supposant l'interaction achevée).

Conçue d'après ce modèle, la description d'une intention sociale comprend *a priori* une référence aux circonstances qui peuvent prévaloir (soit que l'agent « pâtisse » ou « bénéficie » de ces circonstances, soit que celles-ci conditionnent la direction de la conduite de l'agent). Le mode d'être prévu ne sera pas nécessairement la même suivant la « circonstance » qui prévaut à un instant donné ou suivant la physionomie du *continuum* de circonstances affectant la situation des membres de la société au cours d'une période de temps. En l'absence de mention explicite des circonstances, on suppose que le mode d'être décrit est envisagé quelles que soient les circonstances qui prévalent. La présence ou l'absence de conditions relatives aux états du monde dans la description des intentions sociales peut évidemment affecter les conclusions relatives à la compatibilité ou à l'incompatibilité de ces intentions.

Un mode d'être social peut souvent être décrit de plusieurs manières différentes, au moyen d'énoncés plus ou moins précis. On ne considère jamais ici des intentions « ultimes » dont la connaissance permettrait de se représenter jusque dans le plus infime détail la conduite telle qu'elle est effectivement tenue et vécue (avec toutes les nuances des états de conscience de l'acteur). Il n'est d'ailleurs pas certain que de telles entités se prêtent à la communication. Est-on certain de pouvoir se représenter adéquatement la situation et la conduite d'autrui telles qu'elles sont vécues par autrui, avec tous les attributs qui leur sont associés du fait de l'intime union du sujet et de la vie qui est la sienne ? Et si cela était possible, pourrait-on traduire cette représentation dans les termes d'un langage ? Pour éviter d'affronter directement ces problèmes, on ne considère que des descriptions d'intentions, en les supposant toutes exactes (même si elles sont très imprécises). C'est ce qui explique que l'on puisse identifier, en fait, intentions et descriptions d'intentions. Cela étant, on n'impose aucune limite *a priori* au niveau de détail des descriptions. Tout ce qui est exprimable peut être pris en compte. La description des intentions sociales ne se voit imposer aucun « moule » préconçu.

Les intentions sociales envisagées doivent permettre une comparaison, par l'agent lui-même (et secondairement par le théoricien), entre ce qui est projeté et ce qui s'accomplit ou, en d'autres termes, entre les aspirations (projets ou souhaits) et les réalisations. Il n'y aurait donc guère de sens, ici, à considérer des intentions « inconscientes », même si l'on

estime par ailleurs qu'il s'agit de ressorts importants de l'action. Il peut être question d'intentions que l'on considère pour les besoins de la théorie, et que l'on n'attribue à aucun agent en particulier. Même lorsqu'elles sont envisagées sur le mode hypothétique, toutefois, ces intentions sont supposées présentes à l'esprit des agents, au sens suivant : ils peuvent constater, une fois dénouées les incertitudes propres à l'interaction, si ce qui était projeté en intention s'est réalisé ou non.

Plus la description d'une intention sociale se précise, plus rigoureuse est la sélection qu'opère cette description au sein de l'ensemble de tous les modes d'être imaginables d'un individu en société. Le type de mode d'être social sélectionné coïncide alors avec un ensemble plus restreint de modes d'être possibles. Cette « sélection » opère sur les modes d'être ; elle ne se confond pas avec le choix individuel des actions au cours d'une interaction sociale. Les modes d'être sélectionnés par une description de l'intention ne sont pas nécessairement uniques, car il est en général possible de rendre la description de l'intention plus détaillée. On devra donc se borner à dire, si l'intention se réalise, que le mode d'être qui prévaut « fait partie » de l'ensemble sélectionné par le critère que constitue l'intention sociale telle qu'on la décrit.

Enfin, l'intention sociale peut comprendre à la fois des éléments descriptifs pertinents pour apprécier la conformité de la conduite visée à certaines valeurs et, d'autre part, des éléments qui concernent la réalisation de certains états de chose (fins, finalités, résultats, buts, conséquences escomptées). Le cadre conceptuel retenu ne sépare pas *a priori*, pour une intention sociale donnée, les qualités morales, les aspects psychologiques (succession des états mentaux de l'agent au cours de la période considérée) et les « conséquences » (ou résultats). On s'intéresse à un certain mode d'être visé en intention, qui peut incorporer le cas échéant, à côté d'autres aspects, une description de tel ou tel trait de la conduite habituellement représenté comme « conséquence » ou « résultat » d'une action (qu'il s'agisse d'un résultat observable – hauteur d'un bâtiment ou profit d'une firme, par exemple – ou d'un état psychologique résultant d'une action menée à bien ou de tel type d'interaction).

La donnée des intentions qui composent une CIC n'exclut pas les délais dans l'exécution d'actions ou de séries d'actions. Dans ce cas cependant, on considère que c'est « dès à présent » que l'on envisage d'agir plus tard, suivant un calendrier prévu dans la description de chaque intention sociale. L'accomplissement des actions, s'il est ainsi différé, doit faire l'objet d'un calendrier explicite dans la description de l'intention sociale. A défaut de calendrier explicite, on suppose que l'exécution (ou,

du moins, la tentative d'exécution) des actions mentionnées est immédiate. Ainsi, la « conduite conjointe » que l'on considère peut faire intervenir des actions des uns et des autres qui ne se superposent pas dans le temps. Quant à l'accomplissement nécessaire pour pouvoir parler de « réalisation », il n'est évidemment pas nécessaire qu'il ait les effets habituellement escomptés dans les actions que prévoit (si elle en prévoit) l'intention sociale, si du moins ces effets ne sont pas explicitement mentionnés dans cette intention.

La donnée de l'« interaction » permet – aux agents ou au théoricien, suivant les cas – de dire si tel agent a réalisé (ou n'a pas réalisé) son intention sociale au cours d'une période donnée de coexistence avec les autres membres de la société. Puisque la connaissance de l'interaction doit permettre de dire qui réalise (et qui ne réalise pas) son intention, l'interaction doit comprendre au moins la mention du fait, pour un agent, d'évoluer dans une société où se vérifie à un certain moment telle ou telle circonstance (extérieure), ou telle ou telle caractéristique de l'interaction, capable de rendre impossible la réalisation de son intention sociale (si une circonstance de ce genre se vérifie en effet). Et cela, pour tout agent.

Le rapport entre CIC et interaction est très extérieur, de sorte que l'idée de « concrétisation » émerge d'une comparaison entre deux réalités bien distinctes. Il se peut fort bien, en effet, que l'interaction considérée ait lieu entre des individus dont les maximales d'action n'ont que très peu de rapport – voire aucun rapport – avec les intentions formulées dans la CIC qu'on leur impute. On considère les interactions *associées* à une CIC donnée afin de garantir que la « concrétisation » ait un sens. On impose en effet que l'interaction soit une description poussée jusqu'à un degré de détail suffisant pour permettre d'identifier les membres de la société qui réalisent l'intention que leur prête la CIC.

L'incidence d'un événement exogène sur la possibilité de réalisation de l'intention sociale d'un membre de la société peut emprunter deux canaux, qui justifient chacun l'inclusion d'un tel événement dans l'ensemble des circonstances mentionnées dans le contexte d'interaction. Selon une voie directe, l'événement peut constituer en lui-même un obstacle, ou au contraire lever un obstacle. Par exemple, si je veux cueillir de beaux fruits dans mon jardin, des conditions climatiques désastreuses (contre lesquelles je ne puis rien) peuvent évidemment m'en empêcher. L'autre voie est indirecte : ma faculté de réaliser une intention peut être inhibée (ou, au contraire, rétablie) à cause de l'accomplissement par autrui d'un acte selon une intention qui posait, comme condition de cet acte, la réali-

sation d'un événement exogène. Par exemple, si Paul a dans ses intentions de ne pas venir à ma réception à moins qu'il fasse trop mauvais pour faire de la voile (ce dernier projet ayant sa préférence dans l'autre cas), le mauvais temps permet, de manière indirecte, la réalisation de mon intention d'organiser une réception dans laquelle Paul figure parmi les invités présents.

Lorsque C est un contexte « associé à X », la détermination de la concrétisation qui prévaut effectivement dépend seulement du schéma d'interaction. En effet, les différents éléments de F (X, C) sont obtenus pour des interactions ne différant entre elles que par le schéma d'interaction (puisque le contexte C incorpore tous les éléments descriptifs relatifs à la situation initiale et aux circonstances extérieures de l'interaction qui peuvent être utiles pour déterminer la concrétisation qui prévaut). A un contexte d'interaction associé peuvent correspondre diverses interactions, et chaque interaction reflète un certain schéma d'interaction possible entre les membres de la société. Parmi les facteurs déterminant la réalisation d'une interaction particulière, on peut songer aux suivants : le moment précis auquel tel agent prend telle décision prévue (mais non datée) dans son intention sociale ; l'intensité de la résistance physique opposée à un membre de la société par un autre membre de la société (en cas de desseins opposés) ; l'ordre dans lequel s'accomplissent les actes des uns et des autres ; le mode d'organisation par lequel les agents coordonnent leurs actions, etc.

Si par exemple j'ai l'intention de faire du piano pendant une heure, et si les circonstances sont « deux heures avec mon frère dans l'appartement familial », la réalisation ou non-réalisation de mon intention peut dépendre du moment que choisit mon frère pour réaliser une intention semblable, et éventuellement (s'il y a conflit) de l'identité de celui qui cédera le premier. Lorsqu'il y a ainsi plusieurs schémas d'interaction envisageables, la « correspondance d'interaction » précise, pour chaque CIC et pour un contexte donné, l'ensemble des concrétisations (c'est-à-dire des listes de membres réalisant leur intention) associées à ces divers schémas.

### *c) La révision des configurations conjointes*

Certaines CIC, comme on l'a vu, ne sont pas réalisables. C'est le cas lorsque les intentions des uns et des autres sont conflictuelles pour toutes les circonstances envisageables, et le restent quel que soit le schéma d'in-

teraction auquel on puisse penser. Or, l'interaction doit avoir lieu, et avoir lieu d'une certaine façon. Puisque toutes les intentions composant la CIC « conflictuelle » ne peuvent être réalisées ensemble, il faut qu'un tri s'effectue entre celles qui seront réalisées et celles qui ne le seront pas.

Ce tri s'effectue en partie par l'opération des événements exogènes affectant la société considérée. Il peut s'effectuer également en vertu du fait que certains agents l'emportent sur les autres par la force ou la ruse. Par ailleurs, si certaines règles éthiques, juridiques ou politiques sont appliquées, cela ne reste pas sans conséquence, en général, sur les possibilités de réalisation conjointe des intentions. S'il se donne pour mission d'étudier les propriétés et les mérites comparés des systèmes de normes, le théoricien peut supposer que certaines normes s'appliquent en effet, de manière à observer la manière dont ces normes exercent une contrainte sur la conduite conjointe des personnes<sup>1</sup>.

Pour chaque contexte d'interaction, les normes en vigueur effectuent un tri parmi les interactions *a priori* envisageables : certaines respectent ces normes, d'autres non. Cela induit un *processus de révision des CIC*, que l'on peut décrire, en termes très généraux, de la manière suivante. Dans chaque contexte d'interaction envisageable, et pour chaque CIC que l'on considère, de deux choses l'une :

- soit la CIC se réalise dans l'une des interactions sélectionnées par les normes pour ce contexte, et alors la CIC considérée reste la même ;
- soit non : on sait alors que la CIC ne peut pas être conforme aux normes dans le contexte considéré. Dans ce cas, les membres de la société ne peuvent réaliser leur CIC qu'en substituant à la CIC initiale une CIC *révisée* qui, pour le contexte d'interaction considéré, se trouve réalisée dans l'une au moins des interactions respectant les normes.

C'est de l'étude systématique de cette transition pour différents systèmes de normes que l'on peut attendre la reconstitution « sans fiction » de

1. On peut évidemment défendre la thèse selon laquelle la conformité des conduites aux normes est nécessairement problématique, et doit le rester. Kelsen, ainsi, a proposé une déduction théorique de l'existence d'un seuil minimal de tension en dessous duquel on ne peut descendre : « Il doit être possible d'agir contre l'ordre, l'ordre doit pouvoir être *enfreint*, faute de quoi il n'a nullement le caractère d'un ordre normatif » (H. Kelsen, *Das Wesen des Staates*, 1926 ; trad. franç. P.-H. Tavoillot, *L'essence de l'État*, in *Cahiers de philosophie politique et juridique*, n° 17, 1990, « La pensée politique de Hans Kelsen », p. 23). Mais cela n'empêche pas de s'interroger sur les propriétés des dispositifs normatifs en tant que procédures idéales de règlement de certains conflits, ce qui oblige à considérer la « légalité » comme un critère de sélection d'une classe de conduites conjointes.

certaines des analyses habituellement formulées au moyen de la fiction théorique du « contrat social », lesquelles reposent en effet aussi, en dernière analyse, sur la comparaison d'une situation exempte de norme juridique positive (l'état de nature) et d'une situation contrainte par des normes juridiques positives (l'état civil).

#### *d) Typologie élémentaire des principes interpersonnels*

Les principes interpersonnels sont des principes qui, s'ils étaient appliqués, auraient des conséquences pour la délimitation des modes d'existence conjointe envisageables des membres d'une société. Ils excluent certains modes d'être conjoints qui seraient *a priori* envisageables. Il peut s'agir de principes moraux (par ex., « tu ne voleras point », « tu pardonneras à ton ennemi »), de règles déontologiques (« tu collecteras du sang sans offrir de rétribution », « tu respecteras le secret bancaire »), de principes de philosophie politique (« tu légifèreras de manière à aider d'abord les plus pauvres », « tu diviseras les pouvoirs ») ou, encore, de normes juridiques. Il est certainement utile de distinguer les principes selon qu'ils concernent directement ou indirectement les interactions sociales.

#### *Premier type*

Le type le plus simple est celui des principes portant directement sur les interactions, donc sur les conduites. La loi et le règlement peuvent ici servir de modèle. Selon l'opinion que Xénophon prête au sophiste Hippias, les lois de la cité sont les prescriptions ou les défenses établies par les citoyens<sup>1</sup>. Dans la mesure toutefois où l'on considère les « modes d'être » en général (et non pas simplement les actions), on peut assimiler l'obligation d'agir d'une certaine manière à l'interdiction de mettre en œuvre une intention sociale complète quelconque (qui peut comprendre une forme d'action ou l'absence d'action) ne prévoyant pas d'agir de cette manière ; il est loisible de ne réserver aucun sort particulier aux prescriptions, et de s'en tenir aux interdictions (la convention contraire étant d'ailleurs également légitime)<sup>2</sup>. Un assez grand nombre de princi-

1. Xénophon, *Mémoires*, IV, 4, 13.

2. Cela ne veut pas dire, évidemment, que les devoirs et les interdictions aient le même sens. Seule nous intéresse ici l'équivalence du choix des devoirs et du choix des interdictions du point de vue de la description du système normatif.

pes moraux et de principes de gouvernement peuvent ainsi prendre la forme de la conjonction de deux groupes d'énoncés :

- a) *principes de législation*, du type : « Telle norme s'applique à tels individus et à telle catégorie de conduites humaines (individuelles ou conjointes) – ou plus généralement à telle catégorie de modes d'être – dans la société considérée » ;
- b) *principes d'interdiction*, du type : « Telle intention sociale, lorsqu'elle est le fait de telle personne dans telle ou telle circonstances, est exclue par telle norme, au sens suivant : cette norme, qui s'applique effectivement dans le domaine concerné, d'après un principe du type a), exclut que le mode d'être que décrit l'intention soit effectivement vécu par l'agent. »

En vertu des énoncés du type b), les CIC comprenant des intentions visées dans ces énoncés voient leur réalisation tomber dans le « hors-norme » (l'illégalité, s'il s'agit de règles juridiques, et l'immoralité, s'il s'agit de normes éthiques). On ne peut réaliser ces intentions sans contrevenir aux normes qui s'appliquent.

### *Deuxième type*

Un second type que l'on peut distinguer ne s'applique pas directement aux conduites, mais aux normes qui peuvent s'appliquer aux conduites. Nous pouvons aussi formuler des principes d'ordre constitutionnel (par exemple : règles générales concernant l'organigramme des pouvoirs, règles garantissant la démocratie, « droits de l'homme »...) qui permettent, tout comme les « principes moraux » dépourvus d'expression proprement juridique, de porter un jugement sur les lois et le règlement. Les principes relevant de ce second type semblent devoir se fondre dans le moule suivant :

- a') *domaine de législation* : un ensemble de normes (non précisé) devra s'appliquer à tel ou tel aspect (précisé) des modes d'être dans la société considérée ;
- b') *exclusion* : certaines CIC (par exemple, celles qui figurent dans une certaine liste, ou celles qui ont telle ou telle propriété) devront être exclues par les normes qui s'appliqueront ;
- c') *exclusion d'exclusion* : il est exclu que telle ou telle propriété d'une CIC (par exemple, la présence dans cette CIC d'une certaine intention) soit, d'après les normes qui s'appliqueront, un motif suffisant d'exclusion de cette CIC.

Il semble clair que certains grands principes de morale ou de politique trouvent une expression distincte dans chacun de ces deux grands types. Il en va ainsi des droits de l'homme, qui sont à la fois, d'une part, des règles opposables à la conduite des personnes (ce qui rend possible – ce qui ne rend pas tout à fait impossible – leur rattachement à une hypothétique « loi naturelle » supposée contraignante, au même titre que le droit positif, pour les conduites) et, d'autre part, des règles d'échelon constitutionnel contraignant les normes juridiques, lesquelles contraignent à leur tour, plus en aval, les conduites.

*e) Nouvelle approche des exemples de principes interpersonnels*

Les définitions précédentes permettent d'envisager, pour les principes interpersonnels qui ont servi d'exemples, de nouvelles formulations :

(M') « n'est soumise à aucune restriction l'intention, pour un individu non marié, de se marier avec une personne célibataire consentante ; est exclue, par contre, la réalisation de l'intention d'un individu qui voudrait empêcher une personne célibataire de se marier avec une personne célibataire consentante » ;

(PP') « si les membres de la société se trouvent dans un état de référence *a*, et s'ils se conduisent conjointement (soit en agissant, soit en n'agissant pas) d'une manière qui peut influencer sur l'état de la société :

- est exclue la réalisation de toute configuration d'intentions de ces personnes, si cette réalisation conjointe est de nature à rendre moins bonne la situation d'au moins un membre de la société sans améliorer celle de personne ;
- est également exclue la réalisation de toute configuration d'intentions des membres de la société, si cette réalisation conjointe n'est pas de nature à améliorer la situation d'un membre de la société sans détériorer celle d'aucun autre, alors qu'une modification correspondant à cette description serait possible grâce à la réalisation d'une autre configuration d'intentions des membres de la société »<sup>1</sup> ;

1. Cette formulation « négative » (*exclusion* de certaines intentions de conduite) revient en fait à *prévoir* un certain type d'action (ou inaction) ou de conduite conjointe des acteurs. Dans la formulation donnée, en effet, on suppose d'emblée que les membres de la société vont adopter conjointement certaines conduites. Exclure « négativement » certaines configurations de conduite conjointe revient donc à prévoir « positivement » que les membres de la société vont adopter l'une ou l'autre des configurations de conduite conjointe qui ne sont pas exclues.

(P') « si un état *a* est unanimement préféré (par tous les membres de la société) à un état *b*, et si la société se trouve dans un état à partir duquel elle peut accéder (grâce à la conduite de ses membres) soit à l'état *a*, soit à l'état *b*, est exclue la réalisation conjointe de toute configuration d'intentions des membres de N, si cette réalisation conjointe est de nature à provoquer le passage de la société à l'état *b* plutôt qu'à l'état *a* ».

Pour les principes évoqués, la transposition du principe initial dans le vocabulaire des restrictions sur les intentions suppose que l'on puisse préciser en quoi consiste le fait, pour l'intention d'un individu (ou pour une configuration d'intentions conjointes de plusieurs individus), d'entrer en conflit avec les intentions d'autres individus, ou avec la réalisation d'un état « meilleur » (que l'on oppose à un état « moins bon »). Par ailleurs, dans de nombreux problèmes d'éthique sociale, l'évaluation porte entièrement sur les conduites individuelles ou conjointes des uns et des autres. C'est le cas lorsque la « réalisation » (ou actualisation) d'un état possible de la société (que l'on privilégie par rapport à d'autres pour une raison quelconque) dépend seulement des conduites des uns et des autres, sans qu'il soit nécessaire de prendre en compte la réalisation d'événements naturels encore inconnus au moment où se pose le problème de l'évaluation. On peut alors traduire tout ce qui est dit au sujet des modifications ou des virtualités de l'état de la société en termes d'actions ou de modalités d'existence sociale des agents dont la « conduite » importe.

Les exemples précédents suggèrent que certains principes d'éthique sociale significatifs peuvent s'exprimer comme exclusion de certaines intentions *au motif de leur conflit possible avec certaines autres intentions*. On peut exprimer de manière unifiée, d'après cette idée, des principes ayant trait aux droits et aux utilités. Les uns et les autres excluent toute possibilité de réalisation pour certaines configurations collectives d'intentions. Mais allons plus loin : afin d'obtenir des éléments de comparaison de ces différents principes, il semble opportun d'essayer de sonder la profondeur des motifs d'exclusion dans chaque cas.

Pour aborder l'étude conjointe de différents principes interpersonnels, on peut tenter d'identifier des différences dans les types de contradictions pragmatiques qu'ils recouvrent. Plus spécifiquement, on peut s'intéresser au type de causalité (ou responsabilité) impliqué dans l'émergence de ces contradictions ou conflits. Ce projet conduit à proposer une nouvelle formulation des principes (M') (P') et (PP'), faisant cette fois intervenir de manière explicite les incompatibilités entre intentions à titre de motif d'exclusion de certaines configurations d'intentions.

Le principe (M') signifie au fond la chose suivante : n'est soumise à aucune restriction l'intention, pour un individu non marié, de se marier avec une personne célibataire consentante ; est exclue, par contre, la réalisation de l'intention d'un membre de la société, si cette réalisation est de nature à causer l'impossibilité, pour l'une ou l'autre des personnes célibataires de cette société, de se marier avec une personne célibataire consentante. On peut donc le reformuler ainsi :

(M'') Est exclue la réalisation de toute configuration d'intentions sociales conjointes comprenant, pour un individu au moins, une intention ou un ensemble d'intentions dont la réalisation causerait l'impossibilité, pour un autre individu, de réaliser une intention du type « se marier avec une personne consentante ».

En ce qui concerne le principe (PP'), distinguons deux cas :

- *cas 1* : il existe une configuration de conduites conjointes ayant pour résultat l'amélioration de la situation d'au moins une personne sans détérioration de celle d'aucune autre personne<sup>1</sup> ;
- *cas 2* : une telle configuration n'existe pas.

On peut alors reformuler (PP') de la façon suivante :

(PP'') Est exclue la réalisation de toute configuration réalisable X d'intentions de conduite qui vérifie :

- a) il existe au moins un membre de la société pour lequel la réalisation par les autres individus de la conduite prévue en X rend impossible la réalisation de l'intention : « Faire ce que prévoit X et obtenir une situation au moins aussi bonne pour soi que la situation initiale » ; et : la réalisation de X n'améliore la situation de personne ;

ou qui ne vérifie pas a), mais vérifie, si l'on est dans le cas 1 :

- b) pour chaque membre de la société, la réalisation par les autres individus de la conduite prévue en X *rend impossible* la réalisation de l'intention : « Faire ce que prévoit X et obtenir une situation meilleure pour soi que la situation initiale. »

1. Selon la terminologie de certains auteurs : il existe un état qui « domine faiblement au sens de Pareto » l'état initial. Le nouvel état dominerait l'état initial au sens « fort » s'il faisait l'objet d'une préférence stricte unanime. Voir, par exemple : Annund Hylland, « Purpose and Significance of Social Choice Theory », in *Foundations of Social Choice Theory*, sous la direction de J. Elster et A. Hylland, Cambridge University Press, 1986, p. 59-60.

Les conditions *a)* et *b)* peuvent être lues de la façon suivante. Est exclue toute intention conjointe qui conduit inévitablement à la détérioration de la situation de quelqu'un sans amélioration de celle de personne. Par ailleurs, si la chose est possible dans une CIC qui ne conduit à une détérioration pour personne, est exclue toute CIC qui ne conduit pas à l'amélioration de la situation d'au moins un membre de la société. Autrement dit, l'intention « faire ce que prévoit X et obtenir une situation meilleure pour soi que la situation initiale » ne pourra être frustrée pour tous les individus, lorsqu'elle est réalisable pour certains X, que dans le cas où ces X seraient tous de nature à rendre inévitable une détérioration de la situation de quelqu'un. Ce que l'on peut exprimer de manière moins exacte, mais plus ramassée, en disant que *le désir d'améliorer son propre sort ne peut être universellement frustré que dans le cas où cela causerait une détérioration du sort d'autrui.*

En ce qui concerne (P'), selon les mêmes principes, on peut proposer la reformulation suivante :

(P'') « Si la société N se trouve dans un état à partir duquel elle peut évoluer vers deux états, dont l'un est unanimement préféré à l'autre, est exclue la réalisation de toute configuration d'intentions de conduite X des membres de N, lorsque cette configuration est telle que, pour au moins un membre de la société, la réalisation par les autres de la conduite prévue en X rende impossible la réalisation de l'intention : « Faire ce que prévoit X et vivre dans celui des deux états que je préfère. »

### 3 | LA MISE EN COMPATIBILITÉ DES INTENTIONS PAR LES SYSTÈMES NORMATIFS

#### *a) Aspects formels de l'implication des intentions dans leur conflit avec d'autres intentions*

Les reformulations proposées plus haut pour certains principes interpersonnels reposent sur l'exclusion de certaines configurations d'intentions en raison du conflit de certaines intentions avec d'autres. Par ailleurs, ces principes font intervenir des éléments donnés des intentions des membres de la société, éléments qui servent de référence (par exemple, se marier ou améliorer son sort). A ce titre, les principes considérés peuvent être dits *matériels*. Or, il y a lieu de se demander, lorsque deux intentions sont en conflit, si l'une des deux joue un rôle plus accusé que l'autre dans l'émergence du conflit. Peut-on appréhender d'une manière

*formelle* (indépendante du domaine d'action et de l'objet précis des intentions individuelles) le type d'implication de certaines intentions dans l'émergence des conflits qui les opposent à d'autres intentions ?

Certains arguments classiques de Kant peuvent servir à illustrer un projet de ce genre. Les intentions d'actions et l'attention portée aux incompatibilités éventuelles entre ces intentions (ainsi qu'aux moyens de les dépasser ou de les éviter) jouaient un rôle central dans la formulation, par Kant, d'un certain nombre de principes éthiques fondamentaux en termes de conditions sur les « maximes » subjectives de l'action : impératif catégorique, principe universel du droit, critère du contrat social, principe de publicité (formule transcendantale du droit public). Par exemple, le principe de publicité représente comme injustes les actions mettant en cause le droit des autres lorsque les maximes de ces actions ne souffrent pas d'être rendues publiques :

« Toutes les actions relatives au droit d'autrui dont la maxime est incompatible avec la publicité sont injustes. »<sup>1</sup>

Ce principe, dont la portée est à la fois éthique et juridique (ainsi que le précise Kant), a pour objet les actions, mais se formule en termes d'intentions d'actions (celles qui sont compatibles avec la maxime subjective, c'est-à-dire avec le critère de l'action). Il fait intervenir la possibilité ou l'impossibilité de la concrétisation de ces desseins. Le critère fait intervenir, précisément, la propriété suivante : dans certains cas, la divulgation de la maxime fait échouer les intentions d'actions (les « desseins ») s'accordant avec cette maxime. Soit, dans les termes de Kant :

« Une maxime, en effet, que je ne puis *divulguer*, sans faire échouer du même coup mon propre dessein qu'il faut absolument *dissimuler* pour réussir, et que je ne puis *avouer publiquement*, sans provoquer par là inévitablement la résistance de tous à mon dessein, une semblable maxime, dis-je, ne peut faire naître cette opposition nécessaire, générale, par suite saisissable *a priori*, de tous contre moi, que du fait de l'injustice dont elle menace un chacun. »<sup>2</sup>

Ce principe a la forme d'une restriction que l'on place sur les intentions d'actions. Sont écartées comme injustes toutes celles qui ne vérifient pas le critère proposé. Telle est la raison pour laquelle on peut dire, à la suite de Kant, qu'il s'agit d'un principe « purement négatif ».

1. Immanuel Kant, *Projet de paix perpétuelle*, Appendice, II. Trad. franç. J. Gibelin, Paris, Vrin, 1988, 6<sup>e</sup> éd., p. 76.

2. I. Kant, *ibid.* Ce principe est, précise Kant, « comme les axiomes, évident sans démonstration et, de plus, d'une application facile » (p. 77).

Le principe qui permet d'opérer la discrimination d'une manière précise repose sur le fait d'une contradiction pragmatique au moins virtuelle : si j'ai l'intention d'accomplir une action injuste A *et* d'en divulguer la maxime, l'opposition rencontrée (concrétisée par les actions des autres) est de nature à m'empêcher de mener à bien (si l'on ose dire) l'action A. La simple analyse de l'incompatibilité mutuelle de certains desseins permet ainsi de « faire un tri » parmi les intentions d'action envisageables, et ce tri n'est pas arbitraire. En effet, la perspective d'un conflit au moins virtuel entre les desseins des uns et des autres induit un partage naturel entre les intentions qui sont plus spécialement « responsables » d'un conflit et celles qui n'en suscitent pas. Les intentions exclues sont celles qui, d'une certaine manière, sont spécialement impliquées dans le conflit.

Il n'est certes pas aisé de donner la formule exacte de la dissymétrie qui est ici en jeu. On peut toutefois proposer l'interprétation suivante. Si chacun a des intentions de conduite conformes au droit, l'exercice par chacun de son libre arbitre s'accorde avec celui de tous selon une loi universelle. C'est l'exigence de juridicité qu'exprime le « principe universel du droit » formulé dans la *Rechtslehre*. En d'autres termes, il n'y a pas de conflit entre mon intention de conduite et l'intention d'autrui de jouir de son droit. Il n'est pas sûr qu'autrui *veuille* effectivement jouir de son droit, mais il s'agit d'une intention de référence, possédant une importance singulière pour l'analyse juridique et politique, à titre d'intention possible. Si maintenant mon intention est injuste, et si elle est rendue publique, elle échoue à cause de l'opposition des autres. Mais en un certain sens je suis plus responsable que les autres de la frustration de ma propre intention. Mon intention ne pouvait en effet se concrétiser qu'en frustrant l'intention que l'on prête aux autres de vouloir jouir de leur droit. Mon action, de cette manière, enveloppe dans sa conception même un conflit qui n'existait pas sans elle (puisque l'intention de chacun de jouir de son droit n'est source d'aucun conflit)<sup>1</sup>. La responsabilité des conflits est du côté des intentions qui ne sont pas conformes au droit.

Cette interprétation concerne le rapport de l'action à la juridicité en général. Mais la spécification même des droits particuliers peut elle aussi donner lieu à une recherche du degré d'implication des intentions dans

1. Il est évident qu'il peut y avoir conflit d'interprétation, dans le cas même où chacun est loyal. Mais, précisément, le règlement du conflit d'interprétation (par un juge, une Cour ou un souverain...) permettra de préciser de quelle manière le droit *correctement interprété* permet la réalisation simultanée des intentions des uns et des autres de jouir de leur droit.

les conflits qui les opposent les unes aux autres. Par exemple, (M'') exclut les configurations comprenant des intentions telles que : « imposer à autrui un conjoint différent de celui de son choix » ou « empêcher quelqu'un de se marier avec une personne consentante qu'il veut épouser ». Or, intuitivement, ces intentions peuvent apparaître « davantage responsables » du conflit avec d'autres intentions que ne peuvent l'être ces autres intentions. Essentiellement pour les deux raisons suivantes.

Tout d'abord, l'intention « imposer à autrui un conjoint » peut suffire à créer *par elle-même* une contradiction interpersonnelle lorsqu'elle est présente chez deux individus. C'est le cas si deux individus forment des projets matrimoniaux incompatibles au sujet d'un troisième. D'autre part, une intention du type « empêcher quelqu'un de se marier avec une personne consentante qu'il veut épouser » suppose un conflit avec les intentions d'autrui. Une intention de ce genre ne peut être réalisée que s'il y a un conflit dont on sort victorieux. En ce sens, le conflit n'est pas symétrique.

*b) La délimitation théorique d'un « domaine réservé » des personnes*

Si l'on entend par « intentions » ou « projets » non seulement des « actions », mais plus généralement des actions, des attitudes et des modes d'existence au sein de la société (et c'est le parti que nous avons pris), alors tous les « droits fondamentaux » peuvent être considérés comme garantissant, si le système juridique les respecte, une solution juridique à certaines classes de conflits ou incompatibilités des intentions. Tous ont en commun de prévoir que certaines configurations conjointes de projets seront nécessairement illégales.

Une approche traditionnelle du problème des droits conduit à donner un statut spécial à une certaine sphère d'activité ou de conduite, au motif qu'elle concerne des choses privées ou personnelles, ou encore, des choses n'ayant rien à voir avec les intérêts bien compris d'autrui. Or, on peut tenter de déterminer au moins partiellement, par le biais de la structure des intentions conjointes, quels sont les types de projets que l'on peut faire figurer, pour des raisons compréhensibles, dans un tel « domaine réservé ». A cette fin, il est utile de procéder de manière inductive, en réfléchissant sur certaines des libertés fondamentales garanties par la loi dans les pays s'engageant au respect des droits de l'homme, avant de se demander si ces principes ont en commun de protéger la réalisation d'intentions d'un type repérable.

Les libertés publiques traditionnelles concernent des domaines d'action très variés : le culte, l'association, le voyage, la réunion, l'expression

littéraire ou artistique, la jouissance de la propriété, etc. Pourtant, il y a un certain degré de ressemblance formelle dans la manière dont les libertés publiques conduisent à envisager le régime des droits et des interdictions dans ces différents domaines d'activité. En particulier, le législateur devra prévoir des restrictions sur la réalisation d'intentions qui semblent éminemment problématiques au point de vue interpersonnel : par exemple, imposer à autrui un déplacement, un culte ou un discours à prononcer, ou encore, empêcher autrui de pratiquer tel culte, de s'associer à telle personne, ou de jouir de sa propriété de telle ou telle manière.

Plus précisément, les libertés publiques traditionnelles peuvent être interprétées comme des propriétés des systèmes de normes juridiques en vertu desquelles, d'une part, ces systèmes de normes offrent une solution juridique à certaines classes de conflits d'intentions individuelles, et, d'autre part, cette solution est offerte grâce à des restrictions qui portent de manière sélective sur les intentions « problématiques » plutôt que sur les intentions « non problématiques » ; au contraire, est expressément garantie la possibilité de réaliser les intentions non problématiques sans tomber dans l'illégalité. En d'autres termes, il est possible de repérer, grâce à des principes uniformes (indépendants du domaine d'action concerné), une classe d'intentions dont les libertés publiques protègent la possibilité de réalisation. J'ai proposé d'appeler « classe A » ce type d'intentions<sup>1</sup>.

Concentrons notre attention sur les situations à deux personnes<sup>2</sup>. Considérons donc des individus *a* et *b*. Lorsque l'individu *a* possède un ensemble d'intentions  $X_a$  et l'individu *b* un ensemble d'intentions  $X_b$ , on considère la configuration d'intentions conjointes (CIC) :  $X = \{ X_a, X_b \}$ .

$X_a$  sera dit *contradictoire* (au sens interne) relativement à un certain contexte d'interaction *C* si, quel que soit  $X_b$ , et quelle que soit l'interaction *I* associée à  $X$  et compatible avec les données de *C*, la concrétisation de  $X$  dans l'interaction *I* ne comprend pas *a*. En d'autres termes : quel que soit  $X_b$ , l'image de  $(X, C)$  par la correspondance d'interaction *F* ne comprend aucune concrétisation comprenant *a*.

1. J'ai proposé cette délimitation dans mon article : « Sur la justification des droits de l'homme », *Archives de philosophie*, 59 (1996), n° 2, p. 249-271.
2. C'est parmi ces contradictions que l'on trouve celles qui sont les plus irréductibles. En effet, lorsque les contradictions pragmatiques à *n* personnes ne se laissent pas ramener à la juxtaposition de contradictions à deux personnes, cela semble dû en général à des circonstances sociales, économiques ou institutionnelles qui pourraient être autres qu'elles ne sont. Ainsi, il semble que les contradictions profondément enracinées dans la logique de l'existence collective soient à chercher plutôt dans les contradictions impliquant deux personnes. J'ai développé cet argument dans l'article cité.

Si cela est vrai *quel que soit le contexte d'interaction considéré*, on pourra dire que  $Xa$  est *intrinsèquement contradictoire*. Dans ce cas, les intentions de l'individu  $a$  enveloppent quelque problème d'ordre logique, ne tenant pas à la nature des circonstances.

$Xa$  et  $Xb$  peuvent être dits *en contradiction externe relativement à un certain contexte C* si  $F(X, C)$  ne comprend pas la paire  $\{a, b\}$ .

Si  $Xa$  et  $Xb$  sont en contradiction externe pour tout contexte associé à  $X$ , on peut dire de  $Xa$  et  $Xb$  qu'ils sont en *contradiction externe* (au sens absolu).

Le caractère « problématique » des intentions tient à leur rôle dans certains conflits avec d'autres intentions. Il faut donc partir des situations de conflit et étudier le rôle qu'y jouent les intentions. Les critères suivants aident à circonscrire le domaine des intentions « protégées » par les libertés publiques. De plus, les libertés publiques obligent à rendre illégaux des intentions qui violent l'un ou l'autre de ces critères, et seulement des intentions de ce genre.

On considère un ensemble de contextes d'interaction  $C$ . Soit donc  $Z$  une intention de conduite.  $Z$  est « de classe A » si et seulement si elle vérifie *a)* et *b)*<sup>1</sup> :

*a)* Supposons que  $Xa$  et  $Xb$  ne soient affectés ni de contradiction interne, ni de contradiction externe, et que  $Z$  n'appartienne ni à  $Xa$ , ni à  $Xb$ .

Considérons des sous-ensembles d'intentions  $A$  et  $B$  quelconques vérifiant :

$(Xa \setminus A) \cup \{Z\}$  non contradictoire (au sens interne) pour  $C$ , si  $C$  est dans  $C$ .

$(Xb \setminus B) \cup \{Z\}$  non contradictoire (au sens interne) pour  $C$ , si  $C$  est dans  $C$ .

Considérons aussi des sous-ensembles d'intentions  $C$  et  $D$  quelconques vérifiant :

$(Xa \setminus A \cup C)$  et  $(Xb \setminus B) \cup \{Z\}$  non contradictoires au sens externe pour  $C$ , si  $C$  est dans  $C$ .

$(Xa \setminus A) \cup \{Z\}$  et  $(Xb \setminus B \cup D)$  non contradictoires au sens externe pour  $C$ , si  $C$  est dans  $C$ .

1. Dans l'énoncé de ces critères, si  $A$  et  $B$  sont deux ensembles,  $A \cup B$  désigne l'ensemble formé par la réunion des éléments de  $A$  et des éléments de  $B$ , tandis que  $A \setminus B$  est l'ensemble des éléments de  $A$  qui ne sont pas en même temps éléments de  $B$ .

Alors on est certain que  $(Xa \setminus AUC) \cup \{Z\}$  et  $(Xb \setminus BUD) \cup \{Z\}$  sont non contradictoires au sens externe pour  $C$  dans  $C$ .

Autrement dit : l'intention  $Z$  n'est pas telle que son adoption simultanée par deux individus suffise à rendre incompatibles leurs intentions lorsque, initialement, elles ne sont pas incompatibles, et ne sont pas susceptibles de le devenir du fait de l'adoption de  $Z$  par un seul des individus (c'est-à-dire du fait de la rencontre de  $Z$  avec d'autres types d'intentions). Exemple d'intention ne passant pas le test : dicter à quelqu'un son itinéraire ou son lieu de culte.

*b)*  $Z$  n'est pas telle qu'il existe une intention  $\zeta$  vérifiant :

- si  $Xa = \{Z\}$  et si  $Xb$  (non contradictoire au sens interne) ne comprend pas  $\zeta$ , alors, quel que soit  $C$  dans  $C$  :  
 $a$  n'appartient à aucun des éléments de  $F(X, C)$  ;
- si  $Xa = \{Z\}$  et si  $Xb$  (non contradictoire au sens interne) comprend  $\zeta$ , alors, quel que soit  $C$  dans  $C$  :  
 $\{a, b\}$  n'appartient pas à  $F(X, C)$ .

En d'autres termes : l'intention  $Z$  n'est pas telle que sa réalisation suppose à la fois la présence chez autrui d'une certaine intention et la frustration de cette intention d'autrui. Exemples d'intentions ne passant pas le test : empêcher autrui de pratiquer tel ou tel culte, empêcher autrui de s'exprimer de telle ou telle manière ou de s'associer à certaines personnes, etc. Ce sont des intentions dont la réalisation *suppose* un conflit et une victoire.

Pour que ces critères soient significatifs, il faut considérer une classe de contextes d'interaction  $C$  vraisemblable et adaptée aux questions traitées, mais aussi large que possible. En considérant plusieurs spécifications de  $C$ , on obtient différentes délimitations d'une sphère privée des personnes. Si  $C$  est simplement la classe de tous les contextes possibles pour une interaction entre des êtres humains dans le monde tel que nous le connaissons, les critères *a)* et *b)* délimitent une classe d'intentions qui, dans plusieurs domaines de l'activité humaine, coïncide avec l'ensemble des intentions dont les libertés publiques classiques garantissent la possibilité de réalisation. Au contraire, les intentions qui violent l'un ou l'autre de ces critères sont de celles qui subissent des restrictions légales si les normes juridiques concrétisent les libertés publiques classiques.

Ces deux critères délimitent un « domaine réservé » d'une manière non arbitraire. Ils permettent d'identifier des intentions impliquées dans des conflits qu'il faut s'efforcer de résoudre par le droit, mais sur lesquel-

les il importe de s'efforcer de ne faire peser aucune restriction juridique. Étant admis qu'une classe de problèmes doit être résolue, et que l'on doit offrir une garantie de cette résolution, le choix de certains principes contraignant le droit apparaît plus « rationnel » que le choix d'autres principes, parce qu'ils recouvrent un emploi moins arbitraire des moyens de résolution. La « rationalité axiologique », si l'on entend par cette expression la rationalité qui s'attache à l'adoption préférentielle de certains principes, est ici à comprendre comme le choix d'une solution *dépourvue d'arbitraire* sous un certain rapport au moins (car fondée sur une dissymétrie objective).

*c) Un principe plus général*

En généralisant la démarche précédente, il est possible de formuler un principe commun qui trouve une expression partielle à la fois dans l'attribution de grandes classes de droits et dans d'autres exemples d'adoption raisonnée de normes (comme nous le verrons dans la dernière section) :

Pour un ensemble d'individus pouvant avoir des intentions mutuellement incompatibles, énoncer qu'il faudra adopter, s'il existe, un schéma de frustration applicable en cas d'incompatibilité effective (et prévoyant alors la possibilité de réalisation d'un ensemble d'« intentions restreintes » et la non-réalisation de certaines « intentions frustrées ») tel que, simultanément :

- 1 / toutes les configurations possibles des intentions ainsi restreintes soient nécessairement exemptes des incompatibilités considérées ;
- 2 / soient frustrées les intentions dont l'adoption conjointe par plusieurs suffit à créer des incompatibilités du type considéré ;

et, s'il n'existe pas d'intentions telles que celles mentionnées en 2 / :

- 3 / ne soit frustré qu'un groupe d'intentions impliquées dans les incompatibilités considérées tel que, pour tout schéma vérifiant 1 / et laissant ces intentions non frustrées, celles-ci seraient irréalisables dans toutes les configurations restreintes.

On peut aussi reformuler 3 / en disant que ne doivent pas être frustrées (si du moins elles existent) les intentions impliquées dans les incompatibilités considérées, et pour lesquelles il existe un schéma vérifiant 1 / , ne les frustrant pas, et tel que ces intentions soient réalisables dans les configurations restreintes par ce schéma.

Ce principe possède une interprétation « normative », dans la mesure où il caractérise la rationalité axiologique de l'adoption de certaines règles à titre de solution à des problèmes définis. En effet, si certaines intentions suffisent à créer par elles-mêmes (via leur adoption par plusieurs) les conflits considérés, il n'est assurément pas arbitraire d'exclure leur réalisation si l'on veut régler ces problèmes. De même, il n'est pas arbitraire de préférer, si la chose est possible, que les conflits soient réglés en ne frustrant que des intentions qui, si elles constituaient la classe des intentions « non frustrées », seraient de toute manière irréalisables (à cause de la frustration des autres intentions nécessaire pour que les conflits soient réglés). Il s'agit d'une rationalité *relative* au type de problème face auquel il faut être rationnel.

Le même principe est susceptible d'une interprétation « positive », pour l'explication de l'adoption de certaines normes ou de certaines conduites à cause de leur conformité à des normes. Il est alors nécessaire de supposer : 1 / que les acteurs veulent résoudre certains problèmes interpersonnels selon un ordre de priorité déterminé ; 2 / qu'ils veulent le faire en considérant seulement les intentions (non pas l'identité de celui qui les forme) et le type d'implication de ces intentions dans les problèmes à régler, et 3 / qu'ils cherchent à justifier les restrictions (qui ne doivent donc pas être arbitraires) plutôt que l'absence de restriction (comme ce serait le cas si les agents considéraient les restrictions comme un but en soi). Il s'agit cependant d'un principe moins immédiatement « intuitif » que les divers raisonnements partiels qu'il recouvre et réunit (« si tout le monde agissait de même... », « la liberté des uns commence où finit celle des autres... », etc.). Se pose donc un problème de réalisme : veut-on privilégier les raisonnements dont on peut penser que les personnes les effectuent véritablement, ou privilégie-t-on la description unitaire ?

#### d) *Autres exemples de rationalité axiologique*

Ce principe général s'applique à plusieurs cas importants, comme on peut s'en apercevoir en suivant l'évocation, par Raymond Boudon, du champ couvert par la méthodologie contractualiste traditionnelle<sup>1</sup> :

« Bien des jugements axiologiques – *mais non tous* – peuvent être considérés comme des théorèmes tirés de l'axiomatique contractualiste, exactement comme les théorèmes de l'arithmétique sont tirés de ses principes premiers. Selon les cas,

1. Chap. 8 in R. Boudon, *Le juste et le vrai*, Paris, Fayard, 1995 ; en particulier, p. 352 et s.

la méthode contractualiste légitimera une analyse conséquentialiste pure ou au contraire la récusera. La méthode contractualiste légitime les feux rouges urbains sur une base conséquentialiste ; mais elle montre que la perspective conséquentialiste ne peut être acceptée s'agissant du vol.<sup>1</sup>

Dans le cas que l'on pourrait appeler « contractualisme avec conséquentialisme » (exemple des feux rouges urbains), l'argumentation éthique se cristallise dans l'exclusion de certaines configurations d'intentions conjointes de conduite – à savoir, toutes celles qui ne mènent pas, à titre de conséquence, à des états de la société ayant les propriétés présentées comme souhaitables. Dans le cas du « contractualisme sans conséquentialisme » (ou même avec anticonséquentialisme), illustré par l'exemple du vol, l'argumentation éthique conduit à l'exclusion de la réalisation de certaines intentions au motif de leur « contradiction » avec certaines valeurs ou avec le sens de l'existence même de certaines institutions :

« Le vol introduit une rétribution forcée sans contrepartie : il est donc incompatible avec la notion même de l'échange social. » [...]

« [...] un théorème général de morale nous dit que, quand un comportement a des effets destructeurs sur le système dans lequel il apparaît, il est normalement jugé mauvais. » [...]

« Le tricheur détruit l'intérêt même du jeu. Le plagiaire provoque un sentiment de dégoût, non parce qu'il se pare des plumes du paon, mais plutôt parce qu'il détruit le jeu de l'invention et de la création. L'impoteur qui réussit, avec la complicité de médiateurs ou de collègues complaisants, à faire passer pour scientifiques des travaux qui contredisent les principes élémentaires de la méthode scientifique détruit l'intérêt même du jeu scientifique. »<sup>2</sup>

Ces différents cas de rationalité axiologique peuvent être considérés comme des applications, pour certains types de problèmes, du principe général énoncé dans la section précédente. Plus précisément, ils concernent la condition 2 /, à la différence des droits fondamentaux qui concernent la condition 3 /, laquelle synthétise l'analyse donnée plus haut pour le « domaine réservé » des personnes.

En ce qui concerne le vol, on peut observer que l'adoption conjointe par plusieurs individus de l'intention « voler autrui et jouir du butin comme de son bien » suffit à engendrer une classe de conflits interpersonnels. C'est l'une des manières de dire que la maxime « voler » n'est pas universalisable : si A veut voler X à B et le posséder, et si C veut voler X à A et le posséder, il y a incompatibilité. La norme « ne pas voler » règle

1. R. Boudon, *op. cit.*, p. 352.

2. *Ibid.*, p. 353.

ces conflits en excluant l'intention unique qui en est la source. Relativement à ce type de problème, c'est une solution rationnelle.

Si deux joueurs veulent chacun tricher *et* être reconnu par l'autre comme vainqueur dans un jeu honnête (où personne ne triche), leur attente sera frustrée. Les conflits virtuels de ce type sont réglés rationnellement par la norme « Il ne faut pas tricher », en application du même principe.

De même, si deux scientifiques veulent chacun être en mesure de s'informer auprès de l'autre tout en annonçant, pour ce qui les concerne, des résultats falsifiés à l'autre, alors ils ne peuvent réaliser simultanément leur intention. C'est qu'elle était en elle-même source de conflits. Relativement aux conflits de ce type, l'adoption de la norme « Il ne faut pas falsifier ses résultats » est rationnelle.

On peut faire valoir que la norme « Il ne faut pas plagier » exclut la réalisation des intentions du type : « faire un emprunt (non avoué) à l'œuvre d'autrui et en être considéré comme le seul auteur », dont l'adoption conjointe par plusieurs suffit – engendrer une classe de conflits interpersonnels : si A est copié par B qui est copié par C, alors nécessairement au moins l'un des deux individus B et C sera frustré dans son désir d'être reconnu comme le seul auteur de l'œuvre.

De même, en ce qui concerne le mensonge : le principe « Il ne faut pas mentir » règle les conflits liés à l'adoption conjointe de l'intention : « dire le faux et ne pas être trompé par les paroles d'autrui » (par exemple, l'incompatibilité de la réalisation des intentions de deux individus qui ont cette intention et se parlent l'un à l'autre). Il accomplit cette opération en excluant l'intention qui est la source des conflits. Sur une scène de comédie, par contre, on « dit le faux » dans un contexte où nul ne s'attend à entendre le vrai : le problème précédent n'est plus pertinent, et la rationalité de la norme « Il ne faut pas mentir » disparaît.

L'argument est semblable pour la fausse promesse. Deux individus qui ont chacun l'intention de faire une fausse promesse à l'autre, tout en voulant pouvoir compter sur la promesse de l'autre, n'ont pas des projets simultanément réalisables.

On peut caractériser de la même manière la rationalité de la norme selon laquelle il faut « contribuer », dans l'interaction sociale, à hauteur de la « rémunération » qu'on en tire (norme de « réciprocité », ou d'égalité entre contribution et rétribution). En effet, cette norme règle, en excluant l'intention qui en est la source unique, les conflits liés à l'adoption conjointe (par plusieurs) de la maxime : « profiter des bienfaits de la coopération sociale tout en refusant d'y apporter une certaine contribution (défi-

nie, dans le cas général, en fonction de l'avantage ou de la rétribution) telle que, si un certain nombre de personnes refusaient de l'apporter, il n'y aurait plus d'interaction mutuellement profitable ». A titre de corollaire : « Il ne faut pas agir en passager clandestin. »

Par ailleurs, on peut identifier (et justifier) de manière similaire une composante « axiologique » du principe conseillant de ne pas quitter un optimum de Pareto. En effet, ce principe règle des conflits qui font intervenir, pour un certain état social X de référence (initial, en particulier), des intentions telles que :

- (i) être au moins « aussi bien » qu'en X ;
- (ii) être « mieux » qu'en X.

Le principe retenu recommande, s'il y a un conflit inévitable entre la réalisation de (ii) par un individu et la réalisation de (i) par les autres, de rester en X. Or, l'adoption par plusieurs de (ii) suffit à engendrer des conflits de la classe considérée : en effet, s'il y a des conflits dans les cas qui voient coexister (i) et (ii), il y a *a fortiori* conflit lorsque chacun adopte (ii).

Mais on peut préférer s'intéresser à une autre dimension, qui ressortit plutôt à la rationalité instrumentale : tant que l'on ne rencontre pas les problèmes d'incompatibilité entre l'amélioration de la situation des uns et le maintien de celle des autres, il est rationnel de prendre comme but le bien-être d'une personne quelle qu'elle soit, et de se donner les moyens de promouvoir ce but.

## CONCLUSION

L'étude des conditions de compatibilité des intentions d'action permet de donner une expression significative à divers principes classiques de politique ou d'éthique sociale. Mais elle permet aussi d'approfondir l'examen de certains problèmes importants pour l'existence collective. Son mérite propre est de permettre de saisir clairement la signification des normes juridiques ou des principes interpersonnels du point de vue de la frustration ou de l'absence de frustration de certaines aspirations ou de certains projets. Il semble possible d'envisager une théorie générale des intentions conjointes des personnes et des processus de révision par lesquels elles se mettent en accord avec certaines normes. Une telle théorie devrait pouvoir se fonder sur l'analyse abstraite de la compatibilité et

de l'incompatibilité des intentions dans divers domaines de l'expérience. Il n'est pas indispensable, pour développer ce type d'approche, de s'appuyer sur une anthropologie des désirs et des croyances. Il ne semble pas utile non plus de dissocier par principe les intentions individuelles qui concernent la conduite personnelle (les projets ou stratégies) et celles qui concernent l'insertion de la conduite personnelle dans la vie sociale (les désirs, préférences ou souhaits). Dans ce contexte, on peut aborder conjointement l'étude des normes de type déontologique et des normes de type téléologique. Mais, surtout, il s'avère possible de chercher à caractériser la rationalité des systèmes de normes d'une manière relativement précise, et qui permet de valider simultanément (pour des types de problèmes différents) des critères divers et largement adoptés. Cette caractérisation axiologique du rationnel fait ressortir le rôle fondamental des dissymétries entre les intentions dans la production des conflits que les normes résolvent, et présente l'avantage singulier d'être compatible avec une forme radicale de pluralisme.

## *Modèles de l'action et théorie des prix*

La théorie des prix est le cœur de l'analyse économique. Or, depuis la révolution marginaliste des années 1870, cette théorie a essentiellement été élaborée à l'aide d'un modèle de l'action bien particulier qui est le modèle de l'*action optimale*. L'objectif de cet article est de mettre en évidence certaines limites de ce modèle, et de proposer d'appliquer à la théorie des prix un modèle plus général de l'action que j'appellerai le modèle de l'*action recadrée*<sup>1</sup>.

### 1 | DEUX MODÈLES DE L'ACTION

Une action « optimale » est une action qui consiste en la maximisation d'un indice, par exemple un indice d'utilité ou de profit. Le modèle de l'action optimale prend toute sa valeur en économie quand on le combine avec le concept d'*équilibre* : un équilibre économique est une situation dans laquelle les actions de tous les agents sont optimales simultanément. La combinaison du modèle de l'action optimale et de la notion d'équilibre a donné naissance au paradigme dit « néo-classique », dont on

1. Je tiens à remercier Thomas Piketty, Christian Schmidt et Bernard Walliser pour les remarques qu'ils m'ont adressées sur une version antérieure de cet article, présentée lors de la Journée d'étude Philosophie et modèles de l'action du 28 mars 1996 à la Sorbonne. La présente version tient compte de certaines de leurs remarques, étant bien entendu qu'elle n'engage que moi.

peut attribuer la paternité entre autres à Léon Walras<sup>1</sup>, et qui s'est développé de façon considérable depuis la fin du XIX<sup>e</sup> siècle.

Les économistes néo-classiques ont dépensé des trésors d'ingéniosité pour démontrer l'existence d'un équilibre dans des cas toujours plus nombreux et sous des conditions toujours moins restrictives. Emportés par leur élan, ils ont largement perdu de vue que le concept d'équilibre ne pouvait se suffire à lui-même en ce sens qu'une situation d'équilibre est nécessairement le résultat d'un *processus de déséquilibre* – à moins de faire l'hypothèse invraisemblable d'une harmonie préétablie entre les actions des agents. Dès lors que l'on déplace son attention de la situation d'équilibre vers le processus de déséquilibre qui la précède et la conditionne, on ne peut plus se contenter du modèle de l'action optimale et on doit faire appel à des modèles plus larges de l'action.

L'un de ces autres modèles est celui de l'action « recadrée » : *une action peut être dite recadrée si elle résulte de la rectification d'une erreur portant sur le cadre constitutif d'une décision.*

Commençons par préciser la notion de *cadre* constitutif d'une décision. Considérons un cas général d'optimisation de l'utilité sous contraintes :

$$\begin{aligned} & \max u(q_1, q_2, \dots, q_n) \\ & \text{sous les contraintes : } g_i(q_1, q_2, \dots, q_n) \geq 0, \quad i = 1, \dots, m. \end{aligned}$$

Cette optimisation s'effectue bien dans ce que l'on peut appeler un « cadre ». Ce cadre possède une structure : l'opération de maximisation, le nombre de variables ( $n$ ), le nombre et la forme des contraintes. Il possède également des paramètres : les fonctions  $u$  et  $(g_i)$ . Un tel cadre peut comporter un certain nombre d'erreurs : l'agent peut se tromper sur ses propres goûts tels qu'il les anticipe (ses goûts peuvent avoir changé sans qu'il s'en rende compte, ce qui modifie sa fonction d'utilité), il peut découvrir l'existence d'un autre bien (ce qui modifie le nombre de variables et par contrecoup les fonctions  $u$  et  $g_i$ ), il peut s'apercevoir que les prix des biens sont différents de ce qu'il croyait (ce qui modifie sa contrainte de budget), et ainsi de suite. Quand il découvre qu'il a effectivement commis l'une de ces erreurs, l'agent modifie en conséquence le cadre de sa décision, et il opère ce que l'on peut appeler un « recadrage ».

Une action recadrée se décompose en trois phases : une phase de *découverte* de l'erreur de cadre, une phase de *rectification* de l'erreur et une phase de *décision* au sein du nouveau cadre. La phase de décision peut bien

1. L. Walras, *Éléments d'économie politique pure*, 1<sup>re</sup> éd., Lausanne, L. Corbaz & Cie éd., 1874.

sûr prendre la forme d'une optimisation par laquelle l'agent fait ce qui lui paraît le mieux compte tenu de ce qu'il vient d'apprendre. Les phases de découverte et de rectification de l'erreur, en revanche, sont difficiles à concilier avec le modèle de l'action optimale puisqu'elles modifient le cadre dans lequel l'agent va *par la suite* pouvoir effectuer une optimisation.

Cette difficulté à concilier rectification d'une erreur de cadre et optimisation vaut tout autant pour les différentes extensions du modèle de l'action optimale, et en particulier pour l'économie de l'information au sens de Stigler<sup>1</sup>. La recherche optimale d'information, telle qu'elle est formalisée par Stigler, s'effectue dans un cadre donné qui n'est pas censé comporter la moindre erreur. Prenons l'exemple le plus simple possible de recherche optimale. Supposons que l'acquisition d'une quantité  $x$  d'information ait un coût de recherche  $C(x)$  et rapporte un revenu  $R(x)$ . Si l'agent connaît les fonctions  $R$  et  $C$ , il peut calculer la quantité optimale d'information  $x^*$  qu'il doit acquérir. Il lui suffit de maximiser l'expression  $(R(x) - C(x))$ , c'est-à-dire de résoudre l'équation :  $R'(x) - C'(x) = 0$ . Dans ce modèle, l'agent découvre de l'information, mais sans pour autant rectifier une erreur de cadre. Si par contre il s'apercevait que la quantité d'information qu'il acquiert ne lui rapporte pas autant qu'il le croyait, alors il lui faudrait modifier son cadre de décision en rectifiant l'erreur qu'il constate sur la fonction  $R$ . Or cette rectification peut difficilement être assimilée à une optimisation puisqu'elle porte sur le cadre prérequis pour qu'une optimisation puisse être effectuée. On pourrait à la rigueur concevoir la découverte de l'erreur dans la fonction  $R$  comme le résultat d'une recherche optimale d'information de niveau supérieur. Mais d'une part cela compliquerait beaucoup les choses, et d'autre part cela risquerait de nous entraîner dans une régression à l'infini<sup>2</sup>. Une solution beaucoup plus parcimonieuse et réaliste consiste à supposer que l'agent découvre les erreurs de cadre sans les avoir délibérément recherchées<sup>3</sup>. De telles

1. G. J. Stigler, « The economics of information », *The Journal of Political Economy*, 1961, 69, p. 213-225.
2. Voir J. Elster, *Sour grapes : Studies in the subversion of rationality*, Cambridge, Cambridge University Press, 1983, p. 18, et P. Mongin, « Simon, Stigler et les théories de la rationalité limitée », *Information sur les sciences sociales*, 1986, 25, p. 591-596.
3. I. M. Kirzner fait très justement remarquer que : « Dans la masse de connaissance, de croyances, d'opinions, d'attentes et de conjectures que l'on détient à un moment donné et qui inspire et donne forme à l'action, seule une fraction peut être décrite comme le résultat d'une recherche délibérée ou d'une activité d'apprentissage. Il est certain qu'une très grande part de la conscience que nous avons de notre environnement et de nos attentes concernant le futur est le résultat d'expériences d'apprentissage qui ont eu lieu entièrement sans avoir été planifiées » (*Perception, opportunity, and profit. Studies in the theory of entrepreneurship*, Chicago, The University of Chicago Press, 1979, p. 142).

découvertes sont totalement irréductibles à une optimisation au sens économique du terme, vu qu'elles sont imprévues et imprévisibles. On peut donc considérer que le modèle de l'action recadrée englobe celui de l'action optimale avec la phase de décision, et le transcende avec les phases de découverte et de rectification d'erreur.

Le modèle de l'action recadrée est superflu pour étudier les situations d'équilibre économique, puisqu'à l'équilibre les cadres de décision des agents sont tous donnés et compatibles entre eux, et n'ont donc pas besoin d'être rectifiés. Ce modèle d'action se combine en revanche tout naturellement avec le concept de processus, et cette combinaison constitue un paradigme de la pensée économique que l'on peut appeler le *paradigme évolutionniste*. Ce paradigme n'est pas nouveau en économie : il a commencé à être développé par les deux écoles autrichiennes, celle de Schumpeter<sup>1</sup> d'une part, et celle de von Mises<sup>2</sup> d'autre part. Chez Schumpeter, le processus porte sur l'évolution technologique, et les actions de recadrage sont accomplies par les entrepreneurs qui réalisent des innovations techniques ou organisationnelles majeures. Chez von Mises, le processus consiste en une réduction des déséquilibres entre les prix des différents biens sur les différents marchés, et les actions recadrées sont elles aussi accomplies par les entrepreneurs qui créent et s'approprient du profit en corrigeant ces déséquilibres.

## 2 | UN PROBLÈME CONCEPTUEL POUR LE PARADIGME NÉO-CLASSIQUE

Le paradigme néo-classique, qui s'appuie sur les deux piliers de la rationalité optimisatrice et de l'équilibre, me paraît être le paradigme le plus élaboré de toutes les sciences sociales. Il a donné naissance à une multitude de modèles formalisés de façon rigoureuse, et constitue sans doute l'association la plus fructueuse qui se soit établie entre mathématiques et sciences sociales. On peut en outre constater qu'il résiste très bien au temps puisqu'il continue aujourd'hui, plus d'un siècle après sa création, à être activement développé.

1. J. A. Schumpeter, *Theorie der wirtschaftlichen Entwicklung*, 2<sup>e</sup> éd., Leipzig, Düncker & Humblot, 1926 ; trad. franç., *Théorie de l'évolution économique*, Paris, Dalloz, 1935.
2. L. von Mises, *Human action. A treatise on economic*, 3<sup>e</sup> éd., New Haven, Yale University Press, 1963 ; trad. franç., *L'action humaine. Traité d'économie*, Paris, PUF, 1985.

Ce paradigme n'est pourtant pas sans défaut. Il pose, en ce qui concerne la théorie des prix, un problème conceptuel que l'on doit considérer comme important : il n'explique pas, et ne peut pas expliquer, la *formation* des prix. Une théorie des prix devrait en principe être capable d'expliquer pourquoi, dans une situation donnée (à l'équilibre ou non), les prix sont ce qu'ils sont. Or les modèles néo-classiques en sont incapables. Tout ce qu'ils peuvent faire, c'est nous indiquer ce que seraient les prix s'il y avait équilibre, sachant qu'à l'équilibre les prix sont d'ores et déjà formés. Ce défaut constitutif tient à la notion même d'équilibre, et ne peut donc en aucun cas être corrigé dans le cadre du paradigme.

Ce problème conceptuel nous amène à nous interroger sur l'hypothèse d'équilibre. Le principal argument en faveur de cette hypothèse est qu'il existe des forces qui conduisent une économie vers une situation d'équilibre, si elle ne s'y trouve déjà<sup>1</sup>. Si de telles forces existent, alors les économistes néo-classiques ont peut-être raison de considérer les processus de convergence vers l'équilibre comme des phénomènes d'importance secondaire, comme des phénomènes transitoires sous-tendus par une réalité plus profonde qui est celle de l'équilibre. Mais ces forces de convergence, quelles sont-elles ? Elles font l'objet de travaux classiques qui remontent à Walras et à Edgeworth. Je ne vais pas entrer dans un examen détaillé de ces travaux<sup>2</sup>. Je me contenterai d'une critique très générale portant sur la méthode qu'ils emploient : les différents modèles de processus de convergence *{n'obéissent pas aux canons de l'individualisme méthodologique}*.

C'est le cas par exemple du processus de « tâtonnement walrasien ». Dans ce processus, où les échanges n'ont lieu qu'une fois que l'équilibre a été atteint, les règles d'ajustement sont définies globalement pour chaque marché d'un bien donné : si tel marché connaît un excès d'offre, alors le prix du bien tend à diminuer ; s'il connaît un excès de demande, alors le prix tend à augmenter. On démontre mathématiquement qu'un tel processus fait converger le système vers l'équilibre, mais on voit que ce modèle ne remonte pas jusqu'aux choix individuels qui donnent naissance au processus.

Les économistes sont tout à fait conscients de cette difficulté méthodologique et ne sous-estiment pas sa gravité. En exagérant à peine, on

1. K. J. Arrow et F. H. Hahn, *General Competitive Analysis*, Amsterdam, North-Holland, 1971, p. 263.  
 2. Voir par exemple H. R. Varian, *Microeconomic analysis*, 2<sup>e</sup> éd., New York, Norton, 1984, p. 244-249, et A. Takayama, *Mathematical economics*, 2<sup>e</sup> éd., Cambridge, Cambridge University Press, 1985, p. 339-347.

peut identifier la science sociale à l'individualisme méthodologique. Or l'individualisme méthodologique a été fondé par le paradigme néo-classique, en particulier par la théorie de l'échange économique, avant d'être considérablement généralisé par des sociologues comme Max Weber<sup>1</sup>, Georg Simmel<sup>2</sup> et beaucoup d'autres. En modélisant les processus de convergence comme ils le font, les économistes néo-classiques renient donc une partie essentielle de leur héritage. Au lieu de placer les mathématiques au service de la science sociale, ils font en quelque sorte disparaître la science sociale au profit des mathématiques.

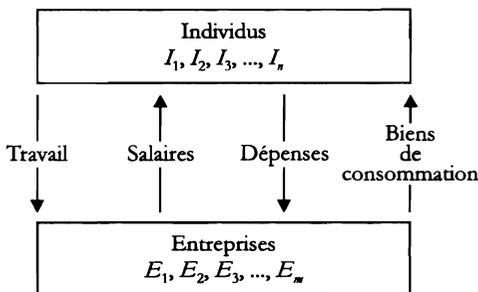
### 3 | UN MODÈLE ÉVOLUTIONNISTE EN THÉORIE DES PRIX

Les difficultés que je viens d'indiquer ouvrent le champ au paradigme évolutionniste en théorie des prix, et nous invitent à faire appel à une conception plus large de l'action, que j'ai appelée le modèle de l'action recadrée. Les modèles évolutionnistes formalisés sont assez peu nombreux, et les plus célèbres d'entre eux ne concernent pas directement la théorie des prix, mais plutôt la théorie du profit<sup>3</sup>, la théorie du duopole<sup>4</sup>, la théorie de l'évolution technologique<sup>5</sup>, et la théorie de la coopération<sup>6</sup>. Je vais donc présenter un modèle évolutionniste formalisé portant sur la théorie des prix. Ce modèle est très élémentaire<sup>7</sup>, mais il permet de se représenter ce que peut être un processus économique fondé sur la conception de l'action recadrée.

1. M. Weber, *Gesammelte Aufsätze zur Wissenschaftslehre*, Tübingen, Mohr, 1951 ; trad. franç., *Essais sur la théorie de la science*, Paris, Plon, 1965.
2. G. Simmel, *Die Probleme der Geschichtsphilosophie. Eine erkenntnistheoretische Studie*, 5<sup>e</sup> éd., Leipzig, Dücker & Humblot, 1923 ; trad. franç., *Les problèmes de la philosophie de l'histoire. Une étude d'épistémologie*, Paris, PUF, 1984.
3. A. A. Alchian, « Uncertainty, evolution and economic theory », *The Journal of Political Economy*, 1950, 58, p. 211-221 ; repris dans A. A. Alchian, *Economic forces at work*, Indianapolis, Liberty Press, 1977, p. 15-35.
4. R. M. Cyert et J. G. March, *A behavioral theory of the firm*, 2<sup>e</sup> éd., Oxford, Blackwell, 1992.
5. R. R. Nelson et S. G. Winter, *An evolutionary theory of economic change*, Cambridge, Mass., Harvard University Press, 1982.
6. R. Axelrod, *The evolution of cooperation*, New York, Basic Books, 1984. Chez les économistes de l'école autrichienne de von Mises, il y a eu un approfondissement conceptuel de ce que pourrait être une théorie des processus de marché, mais très peu de modèles formalisés ; voir F. A. von Hayek, *Individualism and economic order*, Chicago, The University of Chicago Press, 1948, chap. 2, et I. M. Kirzner, *Competition & Entrepreneurship*, Chicago, The University of Chicago Press, 1973.
7. Il demande à être généralisé en introduisant des biens supplémentaires, du capital, des choix intertemporels, un système de prêts et d'emprunts, etc.

## a) Un modèle évolutionniste de base

Je considère un système économique constitué de deux types d'agents, des individus et des entreprises. L'interaction entre ces agents au cours du temps constitue le processus. Pour simplifier, je suppose qu'il n'y a qu'un seul bien de consommation. Les individus travaillent dans les entreprises, en échange de quoi ils reçoivent un salaire qui leur permet d'acheter ce bien. On peut représenter le système de façon très idéalisée sous la forme suivante :



Comme toutes les entreprises produisent le même bien, il n'y a pas de division du travail à proprement parler. Ce système se caractérise en revanche par une très grande *division de la connaissance* au sens de Hayek<sup>1</sup>. Les informations utiles ne sont pas données aux agents. Elles doivent être découvertes par eux au cours du processus. Les individus veulent maximiser leur utilité, mais ils ignorent *a priori* à quelle entreprise ils doivent acheter et dans quelle entreprise ils doivent travailler pour atteindre cet objectif. Ils sont donc à la recherche de conditions économiques plus favorables pour eux. Les entreprises ont également un comportement d'adaptation en fonction de la demande à laquelle elles font face et des fonds dont elles disposent pour payer leurs employés. Ces décisions des différents agents donnent naissance à un processus évolutionniste, dans lequel le temps joue évidemment un rôle essentiel.

Je vais passer en revue les règles de décision des deux types d'agents, les individus et les entreprises, puis je présenterai les différentes phases du processus d'interaction.

*Les actions des individus.* — Chaque individu cherche à augmenter son utilité d'une période à l'autre. Pour cela, il contacte deux entreprises, l'une dont il pourrait devenir client, l'autre dont il pourrait devenir employé. Quatre possibilités lui sont alors ouvertes : soit il reste dans les mêmes entreprises qu'auparavant, soit il devient client dans la nouvelle entreprise mais reste employé dans l'ancienne, soit il devient employé dans la nouvelle entreprise mais reste client dans l'ancienne, soit il devient respectivement client et employé dans les deux nouvelles entreprises. Pour pouvoir choisir, il lui faut estimer l'utilité potentielle qui correspond à chacune de ces quatre possibilités.

Comme il n'y a qu'un seul bien, l'utilité peut être définie comme la quantité consommée ( $c$ ) de ce bien. La quantité de loisir n'intervient pas car je suppose, pour simplifier, que la durée du travail est la même dans toutes les entreprises. Les individus dépensent en principe la totalité de leurs revenus au cours de chaque période (sauf en cas de rationnement de l'offre : voir plus bas).

Pour estimer leur consommation potentielle, les individus doivent se livrer à des calculs qui prennent en compte les prix et salaires affichés par les entreprises. Supposons qu'un individu  $i$  envisage d'acheter le bien de consommation à l'entreprise  $E_j$ , et de devenir employé de l'entreprise  $E_i$ . L'entreprise  $E_j$  propose pour la période  $t$  à venir un prix  $p'_j$ , et l'entreprise  $E_i$  propose un salaire  $s'_i$ . La contrainte de budget de l'individu s'écrit :  $p'_j c = s'_i$ . Sa consommation potentielle (pour la période  $t$ ) est donc :

$$\bar{c}'_i(j, t) = s'_i / p'_j.$$

Ceci est la consommation maximale à laquelle il pourrait prétendre s'il gagnait le salaire  $s'_i$  et achetait le bien au prix  $p'_j$ .

Il faut tenir compte d'une petite complication : la possibilité d'un rationnement de l'offre de bien. Si une entreprise ne produit pas suffisamment pour faire face à la demande lors d'une période quelconque, alors ses consommateurs vont être rationnés pendant cette période. Ils peuvent estimer de façon assez grossière ce possible rationnement  $\bar{c}'_j$  en calculant la quantité moyenne par acheteur produite par l'entreprise à la période précédente :

$$\bar{c}'_j = f_j(n_j^{t-1}) / a_j^{t-1}$$

où  $f_j$  est la fonction de production de l'entreprise  $j$ ,  $n_j^{t-1}$  est le nombre d'employés de la période précédente, et  $a_j^{t-1}$  le nombre d'acheteurs de la période précédente.

Si une entreprise rationne l'offre, ses consommateurs ne peuvent pas dépenser la totalité de leur revenu courant. Ce revenu non dépensé est transmis à la période suivante sous la forme d'un *reste* qui pourra alors éventuellement être dépensé. La consommation anticipée s'écrit finalement :

$$\tilde{c}'_j(j, l) = \min((s'_j + \text{reste}'_{j-1})/p'_j, \bar{c}'_j).$$

Cette consommation n'est que potentielle, car elle dépend de la valeur  $\bar{c}'_j$  qui n'est pas encore connue mais seulement anticipée. Il se peut également que, face à un afflux de nouveaux employés, l'entreprise soit obligée de diminuer le salaire prévu  $s'_j$  pour pouvoir les payer.

En faisant ce calcul, l'individu peut estimer l'utilité potentielle que lui procurera chacune des possibilités qui s'offrent à lui. Il ne lui reste plus qu'à comparer les différents indices et à choisir l'option qui correspond à l'indice le plus élevé.

*Les actions des entreprises.* — Chaque entreprise  $E_i$  prend deux décisions à la fin de chaque période : elle fixe le prix de vente  $p'_i$  et le salaire  $s'_i$  pour la période suivante ( $t$ ). Je suppose qu'elle obéit à des règles très simples : si, lors de la période qui vient de s'achever, elle a fait face à une demande excédentaire par rapport à sa production ( $D_i^{t-1} > f_i(n_i^{t-1})$ ), alors elle augmente le prix ( $p'_i = p_i^{t-1} + \Delta p$ ) ; si elle a fait face à une demande insuffisante ( $D_i^{t-1} \leq f_i(n_i^{t-1})$ ), alors elle diminue le prix ( $p'_i = p_i^{t-1} - \Delta p$ ). Elle fixe le salaire en fonction des fonds dont elle dispose :

$$s'_i = F_i^{t-1} / n_i^{t-1}.$$

Les fonds  $F_i^{t-1}$  représentent ce qu'a rapporté la vente du produit pendant la période précédente, soit :

$$F_i^{t-1} = p_i^{t-1} \min(f_i(n_i^{t-1}), D_i^{t-1}) + \text{reste éventuel de la période précédente.}$$

Le salaire  $s'_i$  fixé par l'entreprise ne sera pas nécessairement le salaire effectivement versé aux employés. S'il y a un afflux d'employés, il se peut que l'entreprise ne dispose pas de suffisamment de fonds pour les payer comme elle l'avait prévu. Si  $F_i^{t-1}$  ne suffit pas pour payer les  $n'_i$  employés, alors l'entreprise répartit ce fonds en parts égales, ce qui représente un salaire « rationné »  $\bar{s}'_i$  :

$$\bar{s}'_i = F_i^{t-1} / n'_i.$$

*b) Le processus d'interaction*

Au tout début du processus, les prix, salaires, nombre d'employés et nombre d'acheteurs de chaque entreprise sont fixés de façon aléatoire dans un certain intervalle. La circulation de la monnaie s'effectue de la façon suivante : les entreprises avancent l'argent en payant les employés pour leur travail ; les employés peuvent alors acheter le bien de consommation, ce qui remplit les caisses des entreprises qui peuvent à nouveau payer leurs employés, et ainsi de suite.

Lorsque débute la période courante  $t$ , chaque entreprise  $E_i$  a fixé un prix de vente  $p'_t$  et un salaire  $s'_t$ . A partir de là, le processus se déroule en cinq phases successives par période :

**1<sup>re</sup> phase** : Chaque individu contacte deux entreprises, l'une dont il pourrait devenir client, et l'autre dont il pourrait devenir employé ; il compare les diverses possibilités qui lui sont ouvertes en tant que client et employé (il y en a quatre), puis il choisit de quelle entreprise il va devenir client et dans quelle entreprise il va travailler, de façon à obtenir l'utilité anticipée la plus élevée ; à l'issue de cette phase, chaque entreprise sait quels sont ses clients et ses employés.

**2<sup>e</sup> phase** : Étant donné le nombre de ses employés ( $n'_t$ ), chaque entreprise produit une quantité de biens de consommation déterminée par sa fonction de production ( $f_i(n'_t)$ ).

**3<sup>e</sup> phase** : Les entreprises versent les salaires à leurs employés ; deux cas peuvent se présenter :

- 1 / si les fonds obtenus grâce à la vente du bien pendant la période précédente sont suffisants ( $F_i^{t-1} \geq n'_t s'_t$ ), l'entreprise paye les salaires prévus et conserve pour la période suivante le surplus éventuel ;
- 2 / si les fonds sont insuffisants ( $F_i^{t-1} < n'_t s'_t$ ), l'entreprise ne peut payer les salaires prévus et verse à chaque employé un salaire rationné  $\bar{s}'_t = F_i^{t-1} / n'_t$ .

**4<sup>e</sup> phase** : Les individus achètent le bien de consommation dans l'entreprise où ils sont clients ; ici encore, deux cas peuvent se produire :

- 1 / si l'entreprise a produit suffisamment pour satisfaire la demande, chacun est servi au maximum de ce que lui permet son revenu ; on suppose que le bien de consommation est non durable : s'il y a des inventus, ils sont perdus ;
- 2 / si l'entreprise n'a pas produit suffisamment, elle divise sa production en parts respectant l'importance relative des demandes individuelles :

si on note  $D'_t$  la demande totale en provenance des clients,  $Q'_t$  la quantité totale produite par l'entreprise ( $Q'_t = f_i(n'_t)$ ), et  $c'_t$  la consommation maximale permise par sa contrainte de budget à l'individu  $I_i$ , alors ce dernier reçoit la part  $Q'_t$  ( $c'_t/D'_t$ ) ; les individus rationnés ne dépensent pas l'intégralité de leur revenu courant : ce qui leur reste est transmis à la période suivante où il pourra être utilisé.

**5<sup>e</sup> phase :** Compte tenu des fonds qu'elles viennent de récolter et de la demande  $D'_t$  à laquelle elles ont fait face, les entreprises fixent pour la période suivante ( $t+1$ ) un nouveau salaire  $s'_{t+1} = F'_t/n'_t$ , et un nouveau prix de vente  $p'_{t+1}$  ( $p'_{t+1} = p'_t + \Delta p$  si la demande a été excédentaire,  $p'_{t+1} = p'_t - \Delta p$  sinon).

#### *c) Résultats du modèle*

Le modèle a été programmé sur ordinateur avec 1 000 individus et 25 entreprises, sur 30 périodes. Les fonctions de production ont été choisies linéaires et toutes identiques ( $f(n) = kn$ ,  $k = 100$ ). Les résultats que l'on obtient sont satisfaisants (voir fig. 1) : on observe au début du processus une tendance à la diminution de l'utilité totale (somme des utilités individuelles) et une tendance à l'augmentation des profits ; par la suite, il s'opère une sorte de stabilisation des niveaux d'utilité globale et de profit total.

Il est intéressant de comparer l'évolution de ce système avec l'évolution d'un système « amorphe » dans lequel les agents ne modifient pas leur action en fonction des circonstances : les individus ne changent pas d'entreprise, les entreprises ne modifient pas les prix et les salaires. On s'aperçoit que le système « amorphe » n'est pas stable : la somme des utilités individuelles tend vers zéro, alors que les profits augmentent sans cesse. Ceci est dû au fait que dans ce cas les entreprises n'augmentent pas les salaires en fonction des fonds dont elles disposent. Ces fonds tendent donc à s'accumuler.

#### *d) Apports conceptuels du modèle*

Que nous apporte ce modèle élémentaire ? Quelques points méritent d'être soulignés.

1 / Le problème conceptuel de la convergence vers l'équilibre, qui se posait pour le paradigme néo-classique (voir la deuxième section), dispa-

rait. La théorie des prix est abordée ici du point de vue de la formation même des prix : les prix sont ce qu'ils sont parce que les entreprises l'ont décidé, et elles prennent leur décision en fonction de la situation dans laquelle elles se trouvent. Il est vrai que les règles de décision que je leur ai attribuées dans le modèle sont très grossières et mériteraient d'être affinées, mais l'essentiel, en l'occurrence, n'est pas là.

Je dois insister sur la différence qu'il y a entre cette conception de la formation des prix et la conception qui sous-tend par exemple le cas classique de l'équilibre de Cournot-Nash. Dans cet équilibre, il y a une forme d'« endogénéisation des décisions » en ce sens que chacune des deux entreprises fixe la quantité qu'elle va produire en considérant comme une donnée la quantité produite par l'autre entreprise. En déterminant sa

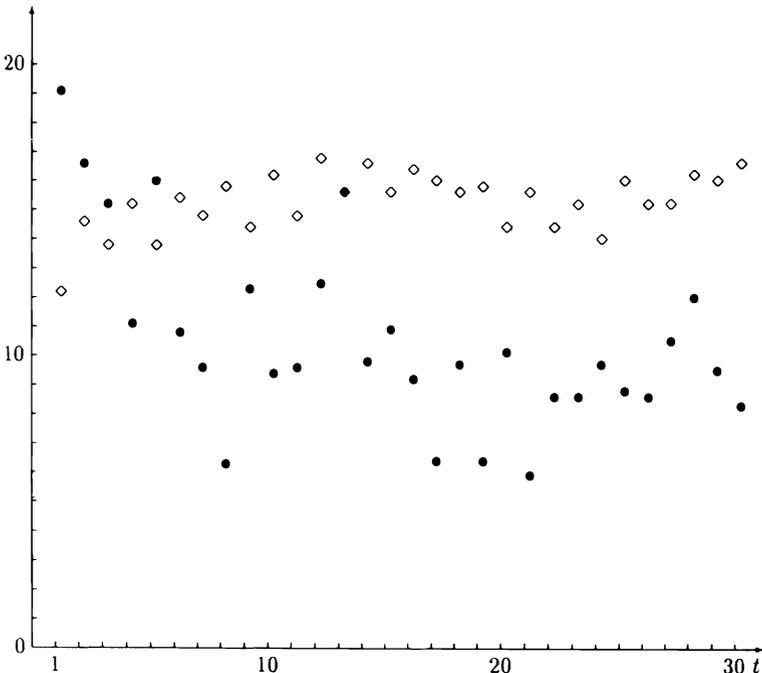


FIG. 1. — Résultats du modèle sur 30 périodes avec 1 000 individus et 25 entreprises. Les ronds (●) représentent  $U/100\ 000$ , où  $U$  est la somme totale des utilités individuelles. Les diamants (◇) représentent  $\pi/1\ 000$ , où  $\pi$  est la somme totale des profits.

quantité produite, chaque entreprise fixe du même coup son prix de vente par l'intermédiaire de la fonction de demande (supposée connue). Mais la quantité et le prix ainsi fixés dépendent de la quantité qui sera produite par l'autre entreprise, et cette dernière quantité est inconnue (d'un strict point de vue mathématique, elle peut prendre une infinité de valeurs). C'est l'hypothèse d'équilibre qui permet, grâce à la résolution d'un système d'équations simultanées, de déterminer un couple unique de quantités compatibles, et d'attribuer *de l'extérieur du système* une ligne d'action à chaque entreprise. L'équilibre de Cournot-Nash n'échappe donc en aucune façon au problème conceptuel de la convergence. Plus généralement, l'endogénéisation des prix telle qu'elle est conçue dans les modèles de concurrence imparfaite ne prend son sens que lorsqu'on la combine avec l'hypothèse d'équilibre et que l'on fait l'impasse sur la question de la convergence.

2 / Ce modèle illustre la notion d'action recadrée. A chaque période, les agents modifient le cadre de leur décision en fonction de ce qu'ils viennent d'apprendre. Le recadrage ne porte pas sur ce que j'ai appelé la structure des décisions (par exemple, le nombre de biens), mais seulement sur les paramètres des décisions (principalement les prix et les salaires).

Ce phénomène de recadrage signifie également que le modèle laisse une place à l'erreur : si un agent modifie son cadre de décision, cela veut bien dire qu'auparavant il ne possédait pas une information exacte ou complète sur la situation. Cette prise en compte de l'erreur me semble très importante car je crois que l'erreur est omniprésente dans la vie économique.

3 / La dernière remarque que je souhaiterais faire concerne la notion d'équilibre. On n'observe pas, dans ce modèle, de convergence vers l'équilibre. Les prix ne convergent pas, pas plus que les salaires. Ceci n'a rien d'étonnant. Dès lors que l'on modélise un processus sur la base de règles de décision individuelles données, il est probable que l'on n'atteindra jamais une position d'équilibre, c'est-à-dire la compatibilité *ex ante* des décisions prises par tous les agents lors d'une période donnée. A supposer même que le système atteigne cette position lors d'une période donnée, il risque d'en être rejeté dès la période suivante car les règles de comportement qui président au déroulement du processus ne seront pas adaptées au maintien dans une position d'équilibre. Il me semble que, dans l'hypothèse d'une forte division de la connaissance, la convergence

vers l'équilibre ne pourra survenir que si les règles de comportement des agents se modifient au cours du processus et en fonction de l'évolution de ce processus. Ceci paraît bien délicat à modéliser.

## CONCLUSION

Quelles sont les relations qu'entretiennent les deux paradigmes économiques évoqués dans cet article ? Ils appartiennent à la science sociale entendue comme une *praxéologie* au sens de von Mises<sup>1</sup>, c'est-à-dire comme une science de l'action humaine qui repose sur le postulat d'action intentionnelle (ou rationnelle) et qui étudie les effets d'interaction. Ils se différencient principalement par leur conception de l'action. Le paradigme néo-classique ne permet de concevoir que des actions optimisatrices entre lesquelles règne l'harmonie définie par l'équilibre. Le paradigme évolutionniste, nous l'avons vu, repose sur une conception beaucoup plus large de l'action humaine. Du point de vue du paradigme néo-classique, la convergence vers l'équilibre est en général tenue pour acquise, et le paradigme évolutionniste est implicitement considéré comme un complément. Mais, du point de vue du paradigme évolutionniste, la convergence vers l'équilibre est loin d'être tenue pour acquise, et dans ces conditions la relation avec le paradigme néo-classique a des chances d'être conçue, non pas en termes de complémentarité, mais plutôt en termes de concurrence.

1. L. von Mises, *ibid.*, chap. 2.

## *Individualisme méthodologique et économie de l'information*

### 1 | INTRODUCTION

Selon l'individualisme méthodologique, analyser une institution sociale revient à étudier les raisons individuelles expliquant pourquoi différents individus sont amenés à faire tel ou tel choix et la façon dont l'institution agrège ces actions individuelles<sup>1</sup>. Cette méthodologie caractérise notamment l'économie théorique contemporaine, et plus généralement l'ensemble des modèles formalisés développés dans les sciences sociales (économie, sociologie, science politique...). Un modèle formalisé commence toujours par une description précise de l'institution étudiée (marchés, institutions politiques, « jeux »...), c'est-à-dire des ensembles de choix individuels possibles et des règles définissant la compatibilité et les conséquences des différentes combinaisons d'actions individuelles. La description des préférences individuelles motivant les actions permet ensuite de décrire l'« équilibre » de l'institution étudiée, défini par la combinaison d'actions individuelles satisfaisant le principe de rationalité, c'est-à-dire les actions remplissant au mieux les objectifs individuels étant donné les règles de l'institution. Le moment essentiel de l'analyse consiste alors à étudier dans quelle mesure cette agrégation des rationalités individuelles est collectivement rationnelle, et donc dans quelle mesure l'institution étudiée est une « bonne » institution, c'est-à-dire une institution permettant de satisfaire de façon efficace les objectifs individuels.

1. Cf. par exemple R. Boudon (dir.), *Traité de sociologie*, PUF, 1992, p. 21-55.

Pourtant, malgré ce consensus de façade, la méthodologie des sciences sociales formalisées est traversée par des évolutions profondes. Le principal développement depuis la fin des années 1960, auquel se rattachent la plupart des contributions importantes des trente dernières années, est celui de l'économie de l'information. L'économie de l'information se caractérise non seulement la prise en compte du fait que les individus disposent en général d'une information imparfaite au moment de faire leurs choix, mais surtout par la modélisation explicite de la façon dont toutes ces informations individuelles imparfaites, les « croyances » individuelles, se forment et évoluent, et en particulier la façon dont l'institution étudiée permet de les transmettre et de les utiliser dans l'équilibre collectif.

L'objectif de ce texte est de montrer en quoi l'économie de l'information marque une étape essentielle dans le développement de l'individualisme méthodologique. L'économie de l'information permet de mieux comprendre et analyser les institutions et phénomènes traditionnellement étudiés par les sciences sociales, ou au moins d'enrichir le débat en formulant les différentes théories possibles et les prédictions testables permettant de les départager en des termes intelligibles par tous. L'économie de l'information permet en particulier de prendre rigoureusement en compte les phénomènes de croyances et d'apprentissages individuels et collectifs et d'explicitier le statut épistémologique du théoricien vis-à-vis des agents de ses modèles, et ainsi de combler certaines des insuffisances légitimement attribuées à l'individualisme méthodologique par ses critiques. Ce texte se propose de conduire cette démonstration en présentant brièvement les principales contributions de l'économie de l'information à l'analyse de trois questions centrales des sciences sociales : le fonctionnement du marché, la production de l'inégalité sociale et le rôle des institutions politiques.

## 2 | LE FONCTIONNEMENT DU MARCHÉ

Même s'il n'est pas toujours formulé explicitement en ces termes, le premier résultat de l'économie de l'information est incontestablement le résultat d'efficacité du marché du modèle Arrow-Debreu<sup>1</sup> : un planificateur centralisé qui disposerait d'une information exhaustive sur les tech-

1. G. Debreu, *Theory of Value*, Yale University Press, 1959.

niques permettant de produire tous les biens et services possibles et sur les préférences exactes de chaque consommateur ne pourrait pas réorganiser la production et la consommation de manière à ce que chacun soit mieux loti que dans la situation où chaque agent connaît uniquement ses propres préférences et techniques de production et où seul le système de prix permet de coordonner les actions individuelles.

Ce résultat est important, car il correspond à l'intuition commune : le marché ne sert à quelque chose que dans la mesure où il permet de coordonner l'utilisation efficace des informations dispersées localement parmi un grand nombre d'agents. L'avantage de la formulation que donne le modèle Arrow-Debreu de cet argument classique, aussi vieux sans doute que le marché lui-même, est qu'il permet d'analyser de façon rigoureuse les conditions exactes sous lesquelles ce résultat d'efficacité reste valable ou cesse de l'être, comme le montrent par exemple les travaux ultérieurs de Kenneth Arrow (cf. *infra*), par opposition aux discours se contentant d'affirmer l'efficacité présumée du marché, comme par exemple celui de Hayek<sup>1</sup>. De fait, depuis les années 1960, les recherches sur le fonctionnement du marché ont tenté de préciser les conditions de ce résultat d'« efficacité informationnelle » du système de prix. Certaines de ces recherches ont contribué à renforcer ce résultat, comme par exemple les travaux sur la transmission d'information par les prix et les anticipations rationnelles en situation où certains agents disposent d'informations<sup>2</sup>. Mais la plupart des travaux ont au contraire montré les limites de l'efficacité informationnelle du système de prix, limites qui peuvent expliquer l'émergence d'institutions telles que les firmes ou les systèmes « informels » de crédit ou d'assurance, et plus généralement des formes d'organisation économique permettant de dépasser les limites informationnelles du marché<sup>3</sup>. Lorsque les hypothèses du modèle Arrow-Debreu ne sont pas vérifiées, on peut souvent montrer que le marché laissé à lui-même ne conduit pas à une utilisation efficace de l'information disponible, par exemple en présence de rendements d'échelle croissants impliquant l'existence de secteurs non concurrentiels, ce qui peut justifier une « régulation » publique des entreprises opérant dans ces secteurs, qui doit elle-même être analysée en termes informationnels<sup>4</sup>; ou encore du fait des limites informationnelles des

1. F. Hayek, « The Use of Knowledge in Society », *American Economic Review*, 1945.

2. S. Grossman, « An Introduction to the Theory of Rational Expectations under Asymmetric Information », *Review of Economic Studies*, 154, p. 541-559.

3. K. J. Arrow, *The Limits of Organization*, Norton & Company, 1974.

4. J.-J. Laffont et J. Tirole, *A Theory of Incentives in Procurement and Regulation*, MIT Press, 1993.

marchés intertemporels<sup>1</sup> du fait de phénomènes pervers d'« anti-sélection » impliquant une séparation excessive sur les marchés d'assurance<sup>2</sup>, ou plus simplement de l'asymétrie d'information entre le vendeur et l'acheteur<sup>3</sup>, qui peuvent justifier des interventions publiques, en particulier dans le domaine des assurances sociales.

L'état présent des connaissances sur ces questions est loin d'être totalement satisfaisant, et les recherches continuent. Mais ces développements montrent toute l'utilité des concepts et des outils de l'économie de l'information : ils permettent d'explicitier la signification et les limites de l'efficacité informationnelle du marché, fondement essentiel des institutions contemporaines. En outre, ce premier domaine de développement et d'application de l'économie de l'information illustre un des principaux atouts de cette méthodologie : elle permet d'explicitier le statut du théoricien, qui ne dispose pas d'information ou de « rationalité » supérieures à celles des agents de ses modèles, et qui se contente d'évaluer les institutions agrégeant ces informations et rationalités individuelles. Cela permet de répondre à des critiques légitimes comme celle de Lucas, qui reprochait aux théories traditionnelles de reposer sur l'hypothèse de myopie des agents face au système de prix, et plus généralement à tous ceux qui reprochent à l'individualisme méthodologique et aux modèles formalisés le « naturalisme » avec lequel ils décrivent les « agents ».

### 3 | LA PRODUCTION DE L'INÉGALITÉ SOCIALE

L'analyse de l'inégalité est une autre question centrale que l'économie de l'information a permis de faire progresser de façon sensible. L'exemple du modèle de discrimination est particulièrement révélateur. Cette théorie, initialement développée par Arrow et Phelps dans le contexte de

1. R. Guesnerie, « The Arrow-Debreu Paradigm Faced with Modern Theories of Contracting : A Discussion of Selected Issues Involving Information and Time », in L. Werin et H. Wijkander (dir.), *Contract Economics*, Blackwell, 1992.
2. M. Rothschild et J. Stiglitz, « Equilibrium in Competitive Insurance Markets : an Essay on the Economics of Imperfect Information », *Quarterly Journal of Economics*, 90 (1976), p. 629-650 ; D. Henriot et J.-C. Rochet, « Équilibres et optima sur les marchés d'assurance », in *Mélanges économiques en l'honneur d'Edmond Malinvaud*, Economica, 1988.
3. K. J. Arrow, « Uncertainty and the Welfare Economics of Medical Care », *American Economic Review*, 53 (1963), p. 941-973 ; G. Akerlof, « The Market for Lemons : Qualitative Uncertainty and the Market Mechanism », *Quarterly Journal of Economics*, 89 (1970), p. 488-500.

la discrimination vis-à-vis de la minorité noire aux États-Unis<sup>1</sup>, peut également s'appliquer à tout autre groupe dont l'appartenance individuelle est observable pour l'employeur, comme les femmes, les basses castes en Inde, les chômeurs de longue durée, ou plus généralement toute origine sociale observable et susceptible d'engendrer des préjugés négatifs. L'idée de base de la théorie est très simple. Supposons que les employeurs, ou tout autre groupe ayant à décider de l'orientation des ressources humaines, anticipent que certains groupes sociaux ont objectivement moins de chances que les autres d'être suffisamment qualifiés pour remplir certains emplois ou certaines tâches exigeant une qualification élevée. Les employeurs observent imparfaitement le niveau exact de qualifications et de motivations des candidats à l'embauche, si bien qu'ils décident l'embauche pour un emploi qualifié sur la base de signaux imparfaits, comme le résultat d'un test, d'un entretien ou un *curriculum vitae*. Puisqu'ils anticipent que certains groupes ont *a priori* moins de chances que les autres d'avoir le capital humain nécessaire, ils prendront des membres de ces groupes uniquement si le résultat de leur test est exceptionnellement bon, c'est-à-dire qu'ils placeront la barre plus haut que pour les autres groupes. Quelle sera la réaction des groupes discriminés face à cette pratique des employeurs ? Puisque la probabilité d'être embauché sur un poste qualifié est faible, ils entreprendront les investissements nécessaires moins souvent que la moyenne. Autrement dit, leur comportement aura tendance à valider les anticipations des employeurs, à savoir que le niveau moyen de qualification de ce groupe sera effectivement inférieur aux autres. On peut ainsi montrer que, même si deux groupes, par exemple les Noirs et les Blancs, ont initialement exactement les mêmes capacités à acquérir une qualification élevée, et *a fortiori* si l'un des groupes a initialement des capacités légèrement inférieures du fait d'origines sociales plus modestes, alors l'interaction perverse entre les croyances et anticipations des employeurs et les comportements engendrés par ces anticipations peut produire une inégalité profonde et persistante des qualifications et des emplois obtenus par ces deux groupes. Cette inégalité est totalement inefficace et « inutile », puisqu'elle n'est fondée que sur un phénomène de « croyance autoréalisatrice » de la part des employeurs et que l'efficacité économique exigerait au contraire que des

1. K. Arrow, « The Theory of Discrimination », in O. Ashenfelter et A. Rees (dir.), *Discrimination on Labor Markets*, Princeton University Press, 1972 ; E. Phelps, « The Statistical Theory of Racism and Sexism », *American Economic Review*, 62 (1968), p. 659-661.

groupes aux capacités identiques effectuent les mêmes investissements et occupent les mêmes emplois.

Cette théorie est importante à plus d'un titre. D'abord, elle illustre une classe générale de phénomènes de « croyance autoréalisatrice » que l'économie de l'information permet d'étudier. De tels phénomènes ont été analysés à propos de questions économiques extrêmement diverses, mais l'étude de la formation des croyances au sujet de l'inégalité des caractéristiques individuelles a toujours occupé une position centrale, non seulement avec la théorie de la discrimination, mais également avec la théorie du *signalling*<sup>1</sup>, la théorie de la réputation<sup>2</sup> ou la théorie de l'apprentissage<sup>3</sup>.

Ensuite, ces théories de l'inégalité illustrent la capacité de l'individualisme méthodologique à analyser rigoureusement l'importance des phénomènes de croyances pour la production de l'inégalité sociale et donc pour les outils adaptés à la redistribution. En effet, dans le cadre du résultat d'efficacité informationnelle du modèle Arrow-Debreu, la seule intervention publique légitime est une redistribution au moyen d'impôts et de transferts fiscaux n'interférant pas avec le fonctionnement efficace du système de prix, et dont l'ampleur est d'ailleurs limitée par les capacités informationnelles imparfaites de la puissance publique à conditionner les transferts sur les caractéristiques individuelles pertinentes<sup>4</sup>. Au contraire, si l'inégalité sociale est due, au moins en partie, aux phénomènes pervers du type de ceux décrits par la théorie de la discrimination, alors d'autres outils de redistribution sont requis, comme par exemple l'emploi de lois antidiscrimination ou même de quotas, dont l'efficacité informationnelle doit être étudiée avec la même rigueur<sup>5</sup>.

Cette capacité de l'économie de l'information à élargir ainsi le champ d'analyse de l'individualisme méthodologique est essentielle, car un reproche souvent fait à cette méthodologie est précisément son incapacité à analyser de tels processus de production « sociale » de l'inégalité. Par exemple, la théorie de la discrimination n'est pas sans similarité avec

1. M. Spence, « Job Market Signalling », *Quarterly Journal of Economics*, 87 (1973), p. 355-374.
2. D. Kreps et R. Wilson, « Reputation and Incomplete Information », *Journal of Economic Theory*, 27 (1982), p. 253-279.
3. T. Piketty, « Social Mobility and Redistributive Politics », *Quarterly Journal of Economics*, 110 (1995), p. 551-584.
4. J. Mirrlees, « An Exploration in the Theory of Optimal Income Taxation », *Review of Economic Studies* 38 (1971), p. 175-208 ; et « Notes on Welfare Economics, Information and Uncertainty », in Balch, McFadden et Wu (dir.), *Essays on Economic Behavior under Uncertainty* (1974).
5. S. Coate et G. Loury, « Will Affirmative-Action Policies Eliminate Negative Stereotypes ? », *American Economic Review*, 83 (1993), p. 1220-1240.

certaines théories sociologiques, selon lesquelles l'inégalité est souvent le produit d'un « discours dominant », qui, en insistant sur les faibles chances d'ascension sociale des membres de certains groupes défavorisés, finit par les décourager et par s'autoréaliser<sup>1</sup>, et ces théories sociologiques se sont souvent opposées aux théories issues de l'individualisme méthodologique insistant sur l'inégalité des motivations individuelles, comme celle de Boudon<sup>2</sup>. L'économie de l'information permet d'exprimer ces oppositions dans un langage commun, ce qui doit permettre de mieux cerner leur nature. Cette même opposition se retrouve d'ailleurs au sein même de l'économie formalisée, avec le conflit entre la théorie dite « statistique » de la discrimination fondée sur les croyances autoréalisatrices et les théories issues de l'école de Chicago, insistant sur les « préférences » individuelles intangibles engendrant la discrimination<sup>3</sup>, et plus généralement sur l'inégalité inévitable des caractéristiques et des motivations individuelles qui sont au fondement de l'inégalité sociale.

#### 4 | LE RÔLE DES INSTITUTIONS POLITIQUES

Ce conflit entre une approche « informationnelle » en termes de « croyances » et une approche en termes de « préférences individuelles » se retrouve dans de très nombreux domaines. Il s'exprime de la façon particulièrement transparente dans l'analyse des institutions politiques, et notamment des processus de vote et de la règle de la majorité. La première tentative de modélisation formalisée dans ce domaine, qui est également la première véritable formalisation de toutes les sciences sociales, a été apportée par le livre classique de Condorcet<sup>4</sup>. Condorcet se situait dans une perspective résolument « informationnelle », en décrivant la fonction des institutions politiques comme un problème d'agrégation de l'information dispersée dans la société afin d'aboutir au choix collectif de la « bonne » politique, supposée être la même pour tous. Le « théorème du jury » de Condorcet montre que, si chaque citoyen-électeur a une pro-

1. P. Bourdieu et J.-C. Passeron, *Les héritiers*, Les Éditions de Minuit, 1964 ; et *La reproduction*, Les Éditions de Minuit, 1970.
2. R. Boudon, *L'inégalité des chances*, Armand Colin, 1973
3. G. Becker, *The Economics of Discrimination*, The University of Chicago Press, 1971.
4. Condorcet, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, 1785.

babilité supérieure à 50 % de connaître la « bonne » politique, alors cette dernière sera correctement choisie avec une probabilité avoisinant les 100 % si la règle de la majorité est appliquée et si la taille de l'électorat est suffisamment importante. Ce résultat d'efficacité informationnelle du vote doit être considéré comme l'équivalent du résultat d'efficacité informationnelle du marché de Arrow-Debreu : il apporte la justification élémentaire d'une institution sociale fondamentale, et surtout il fournit le cadre conceptuel permettant d'approfondir la signification et les limites de ce résultat, comme Condorcet avait d'ailleurs commencé à le faire en analysant l'efficacité de procédures de vote plus complexes mettant en jeu plusieurs étapes ou la démocratie indirecte.

Pourtant, ce cadre conceptuel a été rejeté lors du développement des sciences politiques formalisées dans les années 1950-1960 aux États-Unis. Les fondateurs de l'école dite du *public choice*<sup>1</sup> se proposaient d'analyser la politique comme un conflit d'intérêts entre des acteurs politiques poursuivant rationnellement leurs objectifs, et voyaient dans l'approche informationnelle de Condorcet et sa notion de « bonne » politique identique pour tous des postulats bien peu prometteurs. De la même façon que les théoriciens de l'inégalité sociale de Chicago insistent sur l'importance de l'inégalité des « préférences », expression inaliénable et intangible des particularités individuelles qu'une logique totalitaire voudrait nier et transformer par la force, les théoriciens du *public choice* se méfient d'une approche informationnelle de la politique et de sa notion de « bonne » politique, dans laquelle ils pensent reconnaître la dérive totalitaire de la Révolution française, et préfèrent décrire l'hétérogénéité des attitudes politiques individuelles comme résultant d'un conflit irréductible de préférences et d'intérêts, dont les institutions politiques doivent simplement modérer les manifestations inefficaces. Comme toujours, les conflits méthodologiques reposent sur des conflits intellectuels et politiques fondamentaux, et ce au sein même de l'individualisme méthodologique.

L'approche *public choice* décrivant la politique comme un affrontement de préférences et de rationalités individuelles contradictoires est restée très largement dominante dans le domaine des sciences politiques formalisées<sup>2</sup>, ce qui explique sans doute que la formalisation ne se soit pas encore imposée en science politique comme elle l'a fait en économie. La vision *public choice* de la politique et du rôle de ses institutions est en effet

1. A. Downs, *An Economic Theory of Democracy*, Harper & Collins, 1957 ; J. Buchanan et G. Tullock, *The Calculus of Consent*, The University of Michigan Press, 1962.
2. Cf. D. Mueller, *Public Choice II*, Cambridge University Press, 1989.

condamnée à faire de grands écarts entre le mélange complexe d'intérêts et de croyances que constituent toujours les attitudes politiques individuelles, à commencer par celles des théoriciens eux-mêmes, et les comportements des agents des modèles proposés.

Depuis le début des années 1990, l'influence de l'économie de l'information sur les sciences politiques et le retour à la problématique formulée par Condorcet commencent cependant à se faire sentir. Le lien entre le théorème du jury de Condorcet et ses idées sur les « cycles de majorité », qui étaient les seules à avoir été retenues par l'école du *public choice*, a été remarquablement clarifié par les travaux de Peyton Young<sup>1</sup>. L'étude des institutions politiques du point de vue de leur capacité à agréger l'information disponible a été étendue de l'analyse du vote proprement dit à celle de très nombreuses autres institutions, comme les sondages<sup>2</sup>, les partis politiques<sup>3</sup>, les parlements<sup>4</sup>, le débat politique et la communication stratégique<sup>5</sup>, les manifestations<sup>6</sup>. De très nombreux autres travaux sont actuellement en cours et approfondissent ces analyses, par exemple en comparant l'efficacité informationnelle de différents systèmes électoraux et de différentes combinaisons constitutionnelles. Il est probable que cette approche deviendra bientôt dominante et permettra de renouveler l'analyse normative des institutions politiques. Parallèlement, le développement de grandes enquêtes d'opinion de plus en plus fines permet maintenant de faire le lien entre les théories informationnelles de la formation des attitudes politiques et les études empiriques de ces dernières, et ces études confirment largement la nécessité de dépasser l'approche *public choice* pour mieux comprendre la nature des attitudes politiques individuelles et donc le rôle des institutions politiques, comme par exemple sur la question des attitudes face à l'inégalité et à la redistribution<sup>7</sup>.

1. P. Young, « Condorcet's Theory of Voting », *American Political Science Review*, 82 (1988), p. 1231-1244 ; et « Optimal Voting Rules », *Journal of Economic Perspectives*, 9 (1995), p. 51-64.
2. R. McKelvey et P. Ordeshook, « Information and the Democratic Ideal », *Journal of Politics*, 1986.
3. J. Banks, *Signalling Models in Political Science*, Harwood Academic Publishers, 1991.
4. K. Krehbiel, *Information and Legislative Organization*, Michigan University Press, 1991.
5. D. Austen-Smith, « Strategic Models of Talk in Political Decision Making », *International Political Science Review*, 1992.
6. S. Lohmann, « Information Aggregation through Costly Political Action », *American Economic Review*, 84 (1994), p. 518-530.
7. T. Piketty, *ibid.*, et « Mobilité économique et attitudes politiques face à la redistribution », Document de travail du CEPREMAP, n° 9603, 1996.



## *Rationalité cognitive et théorie des jeux*

Depuis vingt ans, la théorie des jeux met de plus en plus l'accent sur la rationalité cognitive des joueurs, qui traduit leur manière d'appréhender et de représenter leur environnement naturel et surtout social. Plus précisément, elle a successivement étudié les informations qu'ils reçoivent sur leur environnement, les anticipations qu'ils projettent sur cet environnement, enfin les croyances qui modélisent l'environnement. Les croyances des joueurs sont elles-mêmes examinées quant à leur structure formelle au niveau individuel, leur forme hiérarchique dans un contexte collectif et leur processus de révision au cours du temps. Parallèlement aux préférences et aux opportunités, les croyances des joueurs sont ainsi devenues un facteur explicatif essentiel tant de leurs comportements que de leurs interactions.

La typologie classique de l'information, qui caractérise un jeu aux yeux du modélisateur (§ 1), peut se transcrire simplement en une typologie de l'incertitude ressentie par les joueurs (§ 2). La structure individuelle des croyances, qui s'exprime dans un cadre ensembliste ou probabiliste (§ 3), trouve son application dans la valeur de l'information, parfois négative dans un jeu (§ 4). La hiérarchie collective des savoirs croisés, qui culmine dans la notion de connaissance commune (§ 5), fournit une justification cognitive aux équilibres, soutenus par les seuls raisonnements des joueurs (§ 6). La révision individuelle des croyances, dont les règles s'adaptent aux contextes variés de révision (§ 7), autorise un traitement stratégique de l'information, dont la diffusion est contrôlée par les joueurs (§ 8). La révision collective des croyances, qui incite à une forme de consensus (§ 9), fournit une justification dynamique aux équilibres, résultant de la convergence de processus d'apprentissage (§ 10).

## I | TYPES D'INFORMATION DU MODÉLISATEUR

Dans tout jeu, le modélisateur considère deux types d'acteurs qui, au cours d'interactions stratégiques directes non médiatisées par des institutions, adoptent des modes de comportement profondément différents. Un acteur passif, la nature, qui symbolise les influences physiques sur le jeu, revêt un état engendré par une règle automatique, généralement de nature aléatoire. Des acteurs actifs, les joueurs, qui sont les véritables centres de décision autonomes du jeu, posent chacun une action résultant d'un processus rationnel de délibération mentale. Tout joueur se trouve ainsi plongé dans un environnement formé conjointement des autres joueurs et de la nature, et subit des conséquences résultant non seulement de sa propre action, mais de celles des autres et de l'état de la nature.

La rationalité de chaque joueur s'appuie sur des caractéristiques individuelles de trois natures : ses opportunités, ses préférences et ses croyances, qui interviennent à deux moments successifs de la délibération. La « rationalité cognitive » traduit la manière dont il transforme ses informations sur le passé en anticipations sur le futur, en s'appuyant sur ses croyances concernant le fonctionnement de son environnement. La « rationalité instrumentale » traduit la manière dont il sélectionne une action à mettre en œuvre à partir de ses anticipations, en confrontant ses opportunités d'action et ses préférences définies sur leurs conséquences. La version forte de ces deux formes de rationalité affirme respectivement que le joueur prévoit parfaitement l'avenir de son environnement et qu'il maximise son espérance d'utilité sous certaines contraintes.

L'« information structurelle », qui décrit la structure du jeu, se compose, pour la nature, de sa règle d'engendrement des états, et pour chaque joueur, tant de ses caractéristiques exogènes (opportunités, préférences, croyances *a priori*) que de ses formes de rationalité cognitive et instrumentale. Cette information est immuable au cours du jeu, mais, contrairement aux opportunités et aux préférences des joueurs qui sont supposées exogènes et stables au cours du temps, les croyances des joueurs vont se modifier de façon endogène en fonction des messages reçus. Cette information est cachée aux yeux du modélisateur, dans la mesure où il s'agit pour l'essentiel d'états mentaux des joueurs. Il peut néanmoins révéler ces derniers dans des circonstances favorables à partir de leurs manifestations visibles ou pour le moins en postuler certaines propriétés remarquables.

L'« information factuelle », qui résume le déroulement du jeu, se compose des états que prend la nature et des actions qu'exécutent les joueurs, et a donc une nature profondément endogène. Cette information se concrétise progressivement au cours du jeu supposé dynamique, et si elle est rétrospective en ce qui concerne les actions et états passés, elle est prospective en ce qui concerne les actions et états futurs. Cette information est naturellement supposée observable par le modélisateur quant à sa partie rétrospective, mais reste virtuelle quant à sa partie prospective. Cette dernière fait justement l'objet de prévisions du modélisateur qu'il pourra confronter aux réalisations correspondantes pour tester les hypothèses de comportement et d'interaction qu'il a été amené à faire.

## 2 | TYPES D'INCERTITUDE DU JOUEUR

L'« incertitude structurelle » du joueur concerne les formes de rationalité des autres joueurs, mais surtout leurs caractéristiques exogènes (situation d'« information incomplète »). Elle est traitée en considérant que les caractéristiques de tout joueur peuvent être résumées en un « type », étant entendu que chaque joueur connaît son propre type, mais seulement l'ensemble des types possibles des autres joueurs. De plus, elle est supposée probabilisable, au sens où la nature est censée tirer au sort, au début du jeu, un type pour chacun des joueurs, selon une distribution de probabilités conjointe, connue de tous les joueurs. Si, dans un jeu statique, cette distribution résume ce que sait un joueur, dans un jeu dynamique, cette distribution peut être affinée au cours du jeu en fonction de l'observation des actions d'autrui.

L'« incertitude rétrospective » du joueur concerne les états passés de la nature, mais surtout les actions passées des autres joueurs (situation d'« information imparfaite »). Elle est traitée, dans l'arbre d'un jeu dynamique, en considérant que les nœuds auxquels le joueur est susceptible de se trouver et entre lesquels il ne peut discriminer par défaut d'observation sont réunis en un « ensemble d'information ». Là encore, elle est souvent supposée probabilisable au sens où le joueur concerné va projeter une distribution de probabilités purement personnelle sur les nœuds de son ensemble d'information. Cette distribution n'est élaborée qu'une seule fois par le joueur en fonction de sa connaissance du type des autres joueurs, qui les incite à privilégier telles actions plutôt que telles autres.

L'«incertitude prospective» du joueur porte moins sur les états futurs de la nature (qui joue généralement en début de jeu) que sur les actions futures des autres joueurs (situation de manque d'«information stratégique»). Elle est traitée en considérant que le joueur s'efforce de prévoir les actions de l'adversaire, ou même sa stratégie, à savoir l'action qu'il envisage de mettre en œuvre en tout ensemble d'information auquel il est susceptible de se trouver. Cette prévision est une fois de plus supposée probabilisée, une distribution de probabilités sur les stratégies futures d'autrui définissant sa «stratégie mixte» anticipée. Celle-ci résulte d'un processus mental par lequel chaque joueur simule le processus de délibération des autres, compte tenu de l'ensemble des informations qu'il possède sur eux.

En résumé, chaque joueur utilise son information rétrospective pour réduire son incertitude structurelle, en particulier en révélant progressivement le type d'autrui à travers les actions qu'il exécute. Puis, chaque joueur s'appuie sur son information structurelle pour réduire son incertitude prospective, en anticipant l'action d'autrui à partir du type présumé d'autrui, mais aussi de sa propre action. Ainsi, l'action de chacun dépendant de celle des autres (à travers ses préférences et ses opportunités), chaque joueur est engagé dans un système d'anticipations croisées sur leurs actions respectives (j'anticipe que tu anticipes que j'anticipe...). Ce système d'anticipations croisées va lui-même reposer sur un système de savoirs croisés sur leurs types respectifs (je sais que tu sais que je sais...).

### 3 | STRUCTURE INDIVIDUELLE DE CONNAISSANCE

L'incertitude d'un joueur est formalisée en considérant qu'il envisage un ensemble de «mondes possibles», chaque monde traduisant une modalité possible de tous les éléments du jeu qui sont entachés d'incertitude. L'incertitude est «ensembliste» quand le joueur ne peut discriminer entre les mondes d'un «événement», à savoir un sous-ensemble de mondes ; l'incertitude est «probabiliste» quand un joueur est capable de définir une distribution de probabilités sur l'ensemble des mondes possibles. De plus, l'incertitude que le joueur affronte est généralement conditionnelle au «monde réel», connu du seul modélisateur ; une «relation d'accessibilité» associe dès lors à tout monde réel envisagé par le modélisateur l'événement ou la distribution correspondante du joueur.

Pour tout joueur, il est possible de considérer une structure hiérarchique de croyances, faisant alterner des niveaux de nature ensembliste et de nature probabiliste. Dans une structure biprobabiliste, un joueur a une « incertitude » probabiliste de premier niveau sur un phénomène aléatoire et une « ambiguïté » probabiliste de second niveau sur son évaluation de premier niveau. Pour une fonction de croyances au sens de Dempster-Shafer, qui est une généralisation non additive d'une distribution de probabilités, le joueur considère soit une distribution de probabilités sur des événements de base, soit une famille de distributions de probabilités sur les mondes. Dans une structure de Kripke, introduite en « logique épistémique », le joueur considère une hiérarchie infinie de niveaux ensemblistes, autosimilaire à chaque niveau.

Les croyances d'un joueur peuvent ainsi être axiomatisées (et même géométrisées) au même titre que ses préférences, la forme la plus usuelle combinant une double structure. D'une part, le joueur possède une distribution de probabilités *a priori* sur les mondes possibles, de nature objective ou subjective ; d'autre part, le joueur possède une « structure d'information » ensembliste sur les mondes, formée des événements accessibles à partir de chaque monde. A toute représentation « sémantique », définie comme une relation d'accessibilité sur des mondes possibles, on peut alors associer une représentation « syntaxique », définie par un opérateur de connaissance sur des propositions. La règle de passage indique qu'un joueur connaît une proposition (assimilée à un événement) en un monde si cette proposition est vraie en tous les mondes accessibles à partir de ce monde.

Si la relation d'accessibilité engendre une partition d'événements, l'opérateur de connaissance associé obéit à des propriétés remarquables. Il s'agit de l'omniscience logique (le joueur connaît toutes les conséquences de son savoir), de la vérité (ce que sait un joueur est vrai), de l'introspection positive (le joueur sait ce qu'il sait) et de l'introspection négative (le joueur sait ce qu'il ne sait pas). Cependant, la relation d'accessibilité peut être moins restrictive, ce qui correspond à des formes variées de rationalité cognitive limitée. Ainsi, un affaiblissement de la propriété d'omniscience logique traduit des capacités limitées de traitement de l'information, alors qu'un affaiblissement de la propriété de vérité transforme la « connaissance » (toujours vraie) en « croyance » (susceptible d'être fausse).

## 4 | VALEUR INDIVIDUELLE DE L'INFORMATION

En théorie des jeux, un « jeu bayésien » est caractérisé par un espace spécifique de mondes possibles sur lequel sont définies les croyances des joueurs, à savoir une distribution de probabilités et une structure d'information partitionnelle. La stratégie d'un joueur est définie comme l'action conditionnelle qu'il met en œuvre en chacun des mondes possibles, étant entendu qu'il applique la même action en deux mondes indiscernables. L'utilité moyenne du joueur en tout monde, pour des stratégies données des autres joueurs, n'est autre que l'espérance de l'utilité qu'il obtient en tous les mondes accessibles à partir de ce monde. Un équilibre bayésien est alors un équilibre de Nash du jeu élargi, à savoir un ensemble de stratégies des joueurs tel qu'aucun joueur n'a intérêt à en dévier unilatéralement pour une autre stratégie.

Lorsque l'espace des mondes possibles est celui des états de la nature (qui incluent éventuellement les types des joueurs), la structure d'information du joueur exprime le message-événement qu'il reçoit en chaque monde réel envisageable. Les structures d'information peuvent alors être ordonnées selon le contenu en information du message qu'elles véhiculent, une structure étant plus fine qu'une autre si tout domaine d'accessibilité de la première est inclus dans un domaine de la seconde. Les structures extrêmes correspondent à l'incertitude totale (en tout monde, tous les autres mondes sont accessibles) et à la certitude (en tout monde, seul ce monde est accessible). Enfin, la valeur de l'information apportée par une structure d'information plus fine qu'une autre n'est autre que la différence d'utilité moyenne obtenue par le décideur entre les deux structures.

En se restreignant au choix individuel en incertitude, un résultat fondamental affirme que, si la structure d'information d'un décideur est partitionnelle, la valeur de l'information est toujours positive. C'est dire que le décideur ne peut que voir son utilité moyenne s'améliorer s'il reçoit une information (supposé vraie) qui vient rendre sa partition d'information plus fine. Ce résultat est maintenu pour une structure d'information partitionnelle, lorsqu'un message lui-même probabiliste (il exprime une probabilité conditionnelle au monde réel) est plus fin qu'un autre (au sens où il en constitue un « mélange probabiliste »). Cependant, ce résultat ne tient plus si la structure d'information ensembliste du décideur cesse

d'être partitionnelle, en particulier lorsque l'axiome d'introspection positive cesse d'être vérifié.

En théorie des jeux proprement dite, le résultat fondamental cesse généralement d'être valide et peut conduire à une valeur de l'information négative, même lorsque les joueurs ont une structure d'information partitionnelle. De façon triviale, le dévoilement d'une information que possède un joueur et non l'autre, peut se révéler défavorable au premier et favorable au second, par exemple en ce qui concerne le coût de production d'un producteur sur un marché duopolistique. De façon plus subtile, la diffusion d'une information inconnue des joueurs (ou même connue de l'un) peut se révéler nuisible simultanément pour les deux joueurs, par exemple en ce qui concerne le type d'un monopole affrontant de façon répétitive un entrant potentiel sur un marché. Contrairement à l'intuition, il est alors impossible pour les joueurs de revenir à une situation meilleure en ignorant l'information, car chacun sait que l'autre la possède et agit en conséquence.

## 5 | HIÉRARCHIE COLLECTIVE DE CONNAISSANCE

En syntaxe, en s'appuyant sur les opérateurs individuels de connaissance, il est possible de définir des degrés successifs dans la distribution des connaissances (croisées) entre plusieurs joueurs. Un énoncé est de « connaissance diluée » si aucun joueur ne le connaît à lui seul, mais la mise en commun de leurs connaissances locales permettrait d'en déduire l'énoncé en question. Un énoncé est de « connaissance partagée à l'ordre  $k$  » si chaque joueur le connaît, sait que les autres le connaissent et ainsi de suite jusqu'au niveau  $k$  de connaissances croisées. Un énoncé est de « connaissance commune » s'il est de connaissance partagée à l'ordre infini, cette définition hiérarchique reposant ainsi sur la connaissance d'une infinité d'énoncés.

L'opérateur de connaissance commune obéit lui-même à certaines des propriétés remarquables introduites pour les opérateurs de connaissance individuels et desquelles elles se déduisent. L'opérateur de connaissance commune peut par ailleurs recevoir une définition circulaire, à savoir qu'un énoncé est de connaissance commune s'il est de connaissance partagée à l'ordre un à la fois qu'il est vrai et qu'il est de connaissance commune ; la définition circulaire implique la définition hiérarchi-

que, mais non l'inverse. La notion de connaissance commune d'un événement est enfin souvent considérée comme impliquée par le fait que cet événement est public, à savoir que chaque joueur constate son occurrence sous le regard des autres joueurs qui en sont aussi témoins.

En sémantique, dans un cadre ensembliste, les notions de connaissance partagée et de connaissance commune ont une contrepartie sous forme d'une relation d'accessibilité déduite des relations d'accessibilité des joueurs. En particulier, si les joueurs sont munis de partitions individuelles, la partition définissant la connaissance commune est la partition la plus fine des partitions plus grossières que ces partitions individuelles. Une difficulté surgit cependant du fait que les relations d'accessibilité (en particulier partitionnelles) doivent elles-mêmes être de connaissance commune, sinon en un sens technique, du moins en un sens plus informel. Ceci est empiriquement possible dans des situations suffisamment simples, mais s'avère théoriquement plus problématique, même en élargissant convenablement l'espace des mondes possibles.

Toujours en sémantique, mais dans un cadre probabiliste, la notion de connaissances croisées a une contrepartie sous forme de hiérarchie probabiliste de connaissances, toujours assise à sa base sur les mondes possibles. A chaque niveau, chaque joueur a une distribution de probabilités sur les distributions de probabilités des autres joueurs, l'ensemble étant soumis à des conditions de cohérence inter-niveaux. La notion de connaissance commune s'en déduit lorsqu'un joueur affecte une probabilité 1 à chaque niveau impliqué par la connaissance d'un événement de base ; une notion affaiblie d' $\epsilon$ -connaissance commune se contente d'affecter une probabilité  $1-\epsilon$  à chaque niveau. Enfin, sous certaines conditions, on peut montrer que la connaissance par un joueur de la hiérarchie infinie de croyances n'apporte rien de plus que la hiérarchie elle-même.

## 6 | JUSTIFICATION COGNITIVE DES ÉQUILIBRES

En théorie des jeux, les notions usuelles d'équilibre sont définies comme des états stationnaires du jeu, à savoir des états qui restent inchangés en l'absence de modification de l'environnement du jeu. Plus précisément, l'équilibre de Nash est un état résumé par les stratégies des joueurs et caractérisé par le fait qu'aucun joueur n'a intérêt à en dévier unilatéralement, une fois qu'il est établi. Cependant, le processus par

lequel on peut atteindre un équilibre n'est pas précisé, sauf à introduire une entité fictive externe au jeu, qui calcule un équilibre et le propose aux joueurs, qui ont intérêt à s'y conformer. Plus précisément, le « régulateur nashien » cherche un point fixe des fonctions de meilleure réponse des joueurs, puis suggère aux joueurs les stratégies correspondantes, qu'ils ont intérêt à adopter.

En réaction, la « justification cognitive des équilibres » vise à fonder les notions d'équilibre sur les seuls raisonnements des joueurs, c'est-à-dire à en faire des « équilibres de croyances » dans la tête des joueurs. Pour ce faire, en sus des états de la nature et des caractéristiques exogènes des joueurs, les stratégies endogènes des joueurs sont elles-mêmes introduites dans l'espace des mondes possibles. Les joueurs sont toujours supposés rationnels bayésiens, à savoir qu'ils maximisent leur espérance d'utilité par rapport à leur espace d'incertitude élargi, ce qui pose problème compte tenu du caractère désormais endogène des mondes possibles. Outre l'hypothèse drastique qu'il est de connaissance commune que les joueurs sont rationnels bayésiens, on introduit des hypothèses supplémentaires desquelles on infère diverses notions d'équilibre plus larges que l'équilibre de Nash.

Si l'on fait l'hypothèse supplémentaire qu'il est de connaissance commune que les joueurs jouent de façon indépendante, on obtient un « équilibre rationalisable ». Dans cet équilibre, toute stratégie d'un joueur est une meilleure réponse à des stratégies anticipées des autres, elles-mêmes évaluées comme meilleures réponses à des stratégies anticipées des tiers, et ainsi de suite jusqu'à boucler à un certain niveau fini. Si l'on fait l'hypothèse supplémentaire qu'il est de connaissance commune que les joueurs ont la même distribution de probabilités *a priori*, on obtient un « équilibre corrélé ». Dans cet équilibre, un « corrélateur » choisit une issue du jeu en conformité avec une distribution de probabilités adéquate sur ces issues, puis indique à chaque joueur la stratégie correspondante, que ce dernier a intérêt à respecter si les autres respectent la leur.

Si la connaissance commune de la rationalité des joueurs permet ainsi, dans un jeu statique, de singulariser une notion d'équilibre, le problème de la sélection d'un équilibre d'un certain type est plus délicat. Il peut être résolu par la donnée des distributions de probabilités *a priori* des joueurs, ou plus généralement par l'introduction de « conventions » qui sont de connaissance commune, bien que l'origine de ces conventions ne soit pas précisée. Par ailleurs, dans un jeu dynamique, la connaissance commune de la rationalité des joueurs peut conduire à des situations paradoxales, comme le « paradoxe de la rétroduction » (*backward induction paradox*). Ce

paradoxe indique qu'un joueur, qui se contente de raisonner par rétroduction à partir du futur, peut se trouver dans des situations incompatibles avec ses hypothèses, et doit donc les réviser par référence au passé.

## 7 | RÉVISION INDIVIDUELLE DES CROYANCES

Sous forme syntaxique, le processus de révision des croyances suppose que les joueurs disposent d'une connaissance initiale et reçoivent un message, ces deux éléments pouvant s'avérer soit compatibles soit contradictoires au sens logique. Un « principe d'homogénéité » postule dès lors que le joueur recherche une connaissance finale qui soit cohérente et de même type que la connaissance initiale (le message étant d'un type souvent plus simple). Un « principe de priorité » affirme, en particulier en cas de contradiction, que le message est supposé vrai alors que la connaissance initiale est sujette à caution (plutôt que de les traiter sur un pied d'égalité). Un « principe d'économie » stipule enfin que, si la connaissance initiale peut ne pas être modifiée, elle reste inchangée, et si elle doit être modifiée, cette modification se fait de façon « minimale ».

Sous forme sémantique, si la connaissance initiale est ensembliste comme le message, et s'ils ne sont pas contradictoires, la connaissance finale est simplement symbolisée par les mondes à l'intersection des deux ensembles. En revanche, en cas de contradiction, la révision va retenir les mondes du message les plus « proches » de la connaissance initiale (ce qui revient, en syntaxe, à conserver les propositions les plus « épistémiquement enracinées »). Si la connaissance initiale est probabiliste et le message toujours ensembliste, les deux n'étant pas contradictoires, la « révision bayésienne » effectue une homothétie des probabilités des mondes qui appartiennent au message. En revanche, en cas de contradiction, la « révision lewisienne » (*imaging*) consiste encore à reporter la probabilité de chaque monde hors message sur le monde le plus proche à l'intérieur du message.

Les méthodes de révision peuvent être étendues, pour une connaissance initiale de structure donnée, à des messages non ensemblistes, par exemple des messages de nature aléatoire (règle de Jeffrey). Elles peuvent aussi être généralisées à des connaissances initiales exprimées sous forme de structures hiérarchiques quelconques, en particulier à des fonctions de croyances au sens de Dempster-Shafer (règle de Dempster, règle de

Fagin-Halpern). Toutes ces méthodes de révision obéissent ou non à des propriétés dynamiques associées à la réception d'une succession de messages. On peut citer la propriété de « commutativité » (la connaissance finale ne dépend pas de l'ordre dans lequel sont arrivés deux messages successifs) et la propriété d'« idempotence » (la connaissance finale reste inchangée lorsqu'un même message est répété plusieurs fois).

Les méthodes de révision sont souvent multiples, pour une même structure de la connaissance initiale et du message, et doivent être adaptées au « contexte » de la révision, lui-même lié à la nature du message (contexte illustré par une urne contenant au départ des boules bleues, blanches et rouges). L'« actualisation » (*updating*) intervient lorsque le message annonce que le monde réel a changé (les boules rouges ont été retirées de l'urne). La « rectification » (*revising*) intervient lorsque le message annonce une « erreur » dans la perception du monde réel (l'urne ne contient pas en réalité de boules rouges). La « focalisation » (*focusing*) intervient lorsque les mondes sont eux-mêmes formés de populations d'objets formant une population globale dans laquelle on a effectué un tirage (la boule tirée de l'urne n'est pas rouge).

## 8 | UTILISATION STRATÉGIQUE DE L'INFORMATION

En théorie des jeux, les diverses formes de révision peuvent se rencontrer lorsqu'un joueur reçoit un message spécifique, à savoir l'observation d'une action d'autrui (ou plus rarement l'observation d'une intention d'action ou d'une anticipation). Une actualisation devient nécessaire, car le joueur apprend que le monde réel a changé sous l'impulsion de l'action de l'adversaire, en particulier lorsque cette action elle-même est perçue de façon bruitée. Une rectification peut intervenir si l'action observée est en contradiction avec l'ensemble d'hypothèses que fait le joueur sur son adversaire, en particulier lorsque l'action se situe hors de la trajectoire d'équilibre prévue. Une focalisation s'impose enfin si l'action observée révèle partiellement les informations privées détenues par l'adversaire, en particulier des informations sur les états de la nature ou sur le type du joueur.

De fait, le joueur reçoit son message de façon plus ou moins volontariste, ce qui l'amène à réaliser un compromis entre exploration d'une information nouvelle et exploitation de l'information ancienne. L'expérimentation passive correspond à une situation où l'information est sim-

plement le sous-produit du déroulement normal du jeu, c'est-à-dire qu'elle est acquise à un coût nul, mais à un rythme imposé. L'expérimentation active résulte d'une configuration où l'information est délibérément recherchée en déplaçant l'action usuelle pour tester la réaction de l'adversaire, au prix d'une perte d'utilité locale. L'expérimentation pure renvoie à la possibilité pour un joueur d'acquérir directement une information exogène, à un coût défini, par une « action d'information » qui n'a pas d'autre impact.

En sens inverse, si un joueur a conscience que son action est susceptible de révéler l'information privée qu'il détient, il va tenter de lui faire jouer un rôle stratégique, conditionnellement à l'usage que son adversaire peut en faire. Si l'information révélée lui est favorable, il va tenter de la rendre aussi transparente et univoque que possible, particulièrement si l'action présente est susceptible de transmettre des intentions d'action pour le futur. Si l'information révélée lui est défavorable, il va tenter de la masquer par un décalage volontaire de son action, en arbitrant toutefois entre le coût de l'information révélée et l'avantage intrinsèque de l'action. Dans ce dernier cas, il peut même aller plus loin et tenter de faire révéler par l'autre une information fausse, qui va dans le sens de son propre intérêt, bien qu'il sache que l'autre envisage cette éventualité.

C'est ainsi qu'un joueur peut procéder à un « bluff » en faisant croire qu'il a une information différente de la vraie sur un état de la nature (en particulier pour un joueur de poker sur les cartes qu'il détient). De même, un joueur peut se construire une « réputation » en faisant croire qu'il est d'un type différent de son type réel (en particulier pour un entrepreneur sur son attitude dure ou molle en affaires). Selon les intérêts respectifs des joueurs, les équilibres obtenus dans des « jeux de signalement » peuvent alors être classés en révéléteurs, non révéléteurs ou semi-révéléteurs. De plus, dans un mécanisme révéléteur, la relation que postule chaque joueur entre l'information et l'action de son adversaire (pour inférer la première à partir de la seconde) apparaît comme un prototype de théorie autoréalisatrice.

## 9 | AJUSTEMENT COLLECTIF DES CROYANCES

On considère un événement sur lequel deux joueurs ont même probabilité *a priori*, mais des partitions d'information différentes traduisant des informations préalables différentes, et en conséquence des probabili-

tés *a posteriori* différentes. Dans un processus de communication séquentiel, à chaque période, les joueurs affichent leur dernière évaluation de l'événement et tiennent compte de ce message pour réviser leur propre évaluation. Un théorème fondamental affirme alors que le processus ne peut converger que vers une évaluation commune, à savoir que les joueurs ne peuvent « s'accorder à ne pas s'accorder » (*agree to disagree*) si leurs évaluations sont de connaissance commune. Tout se passe comme si chacun révélait implicitement les informations privées détenues par l'autre, si bien que leurs évaluations finales coïncident, mais non forcément leurs partitions d'information finales.

Plus généralement, on considère des joueurs qui possèdent des croyances asymétriques et en déduisent des actions obéissant au « principe de la chose sûre » (si une même action est associée à deux événements disjoints, la même action est associée à leur réunion). Dans un processus de communication séquentiel, à chaque période, les joueurs affichent leurs actions et tiennent compte de ce message pour réviser leurs croyances en conséquence. Le « théorème de mise en accord » (*agreement theorem*) affirme alors que les actions successives ne peuvent converger que vers des actions finales qui sont justifiables sur la base de croyances symétriques (et qui s'avèrent donc identiques si leurs fonctions de comportement sont elles-mêmes symétriques). Ce théorème suppose que les structures d'information sont partitionnelles et n'est plus valable pour des structures trop affaiblies ; de même, les actions doivent accéder au statut de connaissance commune.

Une première illustration concerne le « problème des trois chapeaux » où trois joueurs ont à deviner la couleur (rouge ou bleue) du chapeau qu'ils ont sur leur tête, sachant qu'ils voient le chapeau des autres, mais non le leur. A partir d'une annonce initiale se constituant en connaissance commune (par ex., « il existe un chapeau rouge ») et à des tops successifs, chaque joueur doit déclarer s'il connaît la couleur de son chapeau. Dans ce jeu, après trois tops au plus, les joueurs connaissent la couleur de leur chapeau (ils la connaissent simultanément au troisième top exactement si tous les chapeaux sont rouges). La raison en est que le nombre de mondes possibles (les distributions possibles de couleurs) étant fini et égal à  $k$ , tout événement de connaissance partagée par les joueurs à l'ordre  $k$  devient connaissance commune.

Une seconde illustration concerne le « problème des deux généraux » où deux généraux alliés doivent se coordonner sur le choix d'une tactique (attaquer ou non), chaque tactique étant adaptée à un état de l'adversaire, observable par le seul premier général. Si l'état est favorable, le premier

général envoie un message d'attaque à l'autre (qui a une probabilité  $\epsilon$  de se perdre) ; si le second général reçoit le message, il envoie un contre-message de confirmation (qui a une probabilité  $\epsilon$  de se perdre), et ainsi de suite. Dans ce jeu, les deux généraux n'attaqueront jamais car ils n'auront jamais la garantie que la coordination sera parfaite et le risque ainsi encouru est trop important. La raison en est que le nombre de mondes possibles (le nombre d'aller-retour de messages) étant infini, un événement peut être de connaissance partagée à tout ordre  $k$  sans jamais devenir de connaissance commune.

## 10 | JUSTIFICATION DYNAMIQUE DES ÉQUILIBRES

La justification cognitive des équilibres faisait appel à des joueurs hyper-rationnels, capables de simuler le comportement de leurs adversaires de manière quasi parfaite dans un « processus éducatif » se déroulant instantanément. Une autre manière de voir consiste à remplacer le travail mental des anticipations croisées à des niveaux successifs par le travail effectif de la temporalité séquentielle sur des périodes successives. La « justification dynamique des équilibres » fait ainsi appel à des joueurs ayant une rationalité cognitive et/ou instrumentale limitée, engagés dans un « processus d'apprentissage » prenant en compte les leçons du passé. De fait, il est utile de distinguer deux types d'apprentissage selon l'étendue du champ des informations que récoltent les joueurs et la sophistication du traitement des informations qu'ils réalisent.

Dans un « apprentissage épistémique », chaque joueur optimise sa stratégie à partir d'une anticipation de la stratégie d'autrui, dérivée de croyances sur le comportement d'autrui, elles-mêmes révisées en fonction de l'observation des actions passées d'autrui. Dans une première variante, chaque joueur a des croyances sur le type d'autrui qu'il révise après observation de ses actions ; dans une seconde variante, chaque joueur a directement un schéma d'anticipation de la stratégie d'autrui à partir de l'observation de ses actions. Le processus de révision des croyances induit en général un processus collectif dynamique, stochastique et non stationnaire, qui converge vers un « équilibre de Nash subjectif » ou « équilibre auto-confirmant » (*self-confirming equilibrium*). Notion d'équilibre plus faible que l'équilibre de Nash, l'équilibre auto-confirmant

est un état tel qu'aucun joueur n'a intérêt à en dévier, compte tenu des informations dont il dispose.

Dans un « apprentissage comportemental », chaque joueur joue une stratégie avec une probabilité qui est proportionnelle à un index résumant l'utilité obtenue dans le passé avec chacune des stratégies mises en œuvre. Dans une première variante, le joueur n'observe que l'utilité de ses propres actions, puis inhibe les actions défavorables et renforce les actions favorables ; dans une seconde variante, le joueur observe les utilités obtenues par les autres joueurs avec leurs actions, puis imite les joueurs qui ont réussi. Le mécanisme de renforcement des stratégies induit là encore un processus dynamique stochastique, qui converge éventuellement vers une classe d'équilibres encore mal caractérisée. On peut toutefois remarquer que ce mécanisme est isomorphe à un mécanisme évolutionnaire, concernant cette fois des populations de joueurs à stratégies fixées, mais qui se reproduisent proportionnellement à l'utilité qu'ils retirent de leurs interactions.

Si l'expérience commune accumulée par les joueurs permet ainsi, dans un contexte dynamique, de singulariser une notion d'équilibre, le problème de la sélection d'un équilibre particulier est lui aussi relativement simple. En effet, celui-ci résulte, outre des règles de révision ou de renforcement et des facteurs aléatoires, des conditions initiales de l'apprentissage, à savoir des croyances *a priori* sur l'action des autres ou sur l'utilité de ses propres actions. Il faut cependant observer que la coordination des joueurs sur un équilibre peut là encore reposer sur des dispositifs collectifs exogènes, qui atténuent l'ambition d'individualisme méthodologique soutenue par la théorie des jeux. Il s'agit, en apprentissage épistémique, de contraintes de compatibilité sur les croyances *a priori* concernant les stratégies d'autrui, et, en apprentissage comportemental, de conditions de précoordination des stratégies expérimentées.

\*

Poursuivant un programme de recherches cognitiviste, la théorie des jeux s'est tout naturellement rapprochée de certaines sciences cognitives, qui ont abordé les mêmes questions à partir de présupposés différents. La structure et la révision des croyances sont formalisées à l'aide de la logique épistémique, en particulier en ce qui concerne l'introduction de la connaissance commune et du raisonnement contrefactuel. Les comportements et les équilibres afférents sont examinés conjointement avec l'intelligence artificielle, en particulier en ce qui concerne l'étude de la cognition distribuée et des mécanismes d'apprentissage. Enfin, ces processus

théoriques sont soumis à des tests expérimentaux dans le sillage de la psychologie cognitive, en particulier en ce qui concerne la validation des inférences probabilistes et des principes décisionnels.

Il apparaît néanmoins, au regard d'approches plus psychosociologiques, que la prise en compte des croyances des acteurs dans la théorie des jeux souffre encore de faiblesses plus ou moins faciles à combler. Tout d'abord, les croyances sont formées d'un simple stock d'informations élémentaires de même poids, sans considérer l'existence de modèles mentaux plus structurés dans leur contenu et plus relativisés dans leur crédibilité. Ensuite, les informations et les croyances ont un sens univoque et identique pour tous les acteurs, sans envisager de difficultés de compréhension pour chaque joueur et de divergences d'interprétation de l'un à l'autre. Enfin, les informations nouvelles sont directement assimilables par l'acteur et s'empilent de façon cumulative, sans évoquer des problèmes de filtrage sélectif des informations ou de restructuration discontinue des croyances.

## *Science ou poésie de l'action ? L'espace poppérien du raisonnement historique*

« La poésie est un genre plus philosophique et d'un caractère plus élevé que l'histoire » (*Poétique*, 1451 b5).

Dans ce texte fameux, Aristote dévalorise le récit historique au profit du récit de fiction. Son geste est semble-t-il dirigé d'une part contre Hérodote et Thucydide<sup>1</sup>, d'autre part contre Platon. Celui-ci répudiait la poésie au motif qu'elle n'était que simulacre et, comme toute *mimésis*, « accomplissait son œuvre (*ergon*) loin de la vérité » (*Rép.* X, 603 a 12). Contre eux, le Stagirite « joue » en quelque sorte Euripide<sup>2</sup>. Il n'est certes pas question de dire que la poésie est plus philosophique que la géométrie : en général, Aristote ne *renverse* pas les hiérarchies platoniciennes, mais il réhabilite le terme dévalorisé par Platon, par exemple la rhétorique et la poésie, en montrant qu'elles sont des pratiques réglées, susceptibles de tomber sous la juridiction du philosophe<sup>3</sup>. La *mimésis* n'est pas condamnable, puisqu'elle est naturelle (*kata phusin*) et qu'elle est la condition de tout apprentissage. Elle n'est pas simulacre ou copie de copie, puisqu'il

1. Lequel prétendait que son œuvre avait des leçons universelles à délivrer (*katēma eis aei*) parce que des « faits analogues » à ceux qu'il rapportait ne sauraient manquer de revenir (*Guerre du Péloponnèse*, I, 22).
2. Éventuellement soumis à la critique (1461 b 22), mais au nom des normes de la tragédie, et non de considérations extrinsèques au genre.
3. Ainsi, l'apport propre qu'Aristote revendique pour sa *Rhétorique* est d'avoir montré que le cœur de cet art n'est autre que le raisonnement, qu'il soit déductif, l'*enthymème* (syllogisme à partir de vraisemblances et de signes), ou inductif, l'*exemple* (passage du singulier au singulier sans mention de l'universel implicite). En revanche, il n'a pas cru nécessaire de constituer une *technè historikè*. C'est d'autant plus étrange que Thucydide emploie aussi le terme même de *tekmerion* (preuve sensible, signe nécessaire) et qu'Aristote ne cesse de donner des exemples historiques d'« exemples » rhétoriques : Pisistrate, etc. Par ailleurs, les signes non réciproques avec la cause sont pour le Stagirite des lieux de *sophismes* : il n'a pas semble-t-il de théorie du « faisceau d'indices convergents ».

n'existe pas de Formes originaires séparées. Au contraire, elle est, comme la métaphore, dotée d'une authentique valeur cognitive, en tant qu'elle permet une certaine connaissance des formes (sensibles). Elle est plus proche que l'histoire d'une science des mœurs, au sens d'une théorie de l'action (*praxis*) doublée d'une psychologie des types de caractères (*éthoi*), telles qu'on en trouve la mise en œuvre dans les *Éthiques* du Maître du Lycée. Œdipe nous en apprend plus sur l'homme qu'Alcibiade<sup>1</sup>. Le terme *poiësis*, même s'il n'a pas dans ce contexte le sens général de fabrication, de production d'une œuvre séparée, a tout de même un contenu plus large que le terme actuel, puisqu'il désigne plus ou moins ce que nous appelons la fiction littéraire, à commencer par l'épopée et surtout le théâtre (le roman n'existait pas en tant que tel). Aristote précise bien que le critère de démarcation entre la poésie et l'histoire n'est pas la versification : les *Enquêtes* d'Hérodote versifiées demeureraient de l'histoire, non de la poésie<sup>2</sup>. Mais il n'en reste pas moins que l'histoire est « moins philosophique » que la poésie. Je me permettrai dans ce qui suit de commenter librement les arguments d'Aristote :

1 / La poésie est plus générale (*katholou*) que l'histoire, laquelle, comme la sensation et l'action, a toujours pour objet le particulier. Or il n'est de science que du général<sup>3</sup>, même si toute substance première est individuelle et concrète. Telle est l'« aporie » aristotélicienne de la connaissance<sup>4</sup>. Notons que la généralité dont est capable la fiction est « statistique » et de l'ordre du « vraisemblable », non de l'ordre de la science *proprio sensu*. Mais le vraisemblable (ce qui arrive la plupart du temps) est le substitut sublunaire de la nécessité.

2 / La poésie est modalement plus riche que la description historique : elle narre des événements *possibles*, n'étant pas contrainte par le seul

1. L'allusion à Alcibiade paraît viser plutôt Thucydide que Platon. Mais il faut reconnaître qu'Aristote ne cite pas l'historien dans la *Poétique*. Le problème est d'autant plus délicat qu'il est évidemment loin de mépriser le « matériau » empirique offert par l'histoire : sa *Constitution des Athéniens* en témoigne, ainsi que le livre V de la *Politique*.
2. *Don Juan* n'est pas moins « poétique » que *Le Misanthrope*.
3. *Seconds Analytiques*, 87 b 39 ; on note que, curieusement, le magnifique contrefactuel qui suit (« Si nous étions sur la Lune, nous ne verrions pas la cause de l'éclipse ») paraît être contredit un peu plus loin : « Si nous étions sur la Lune, le fait et le pourquoi (*dia ti*) seraient évidents. » Toute la difficulté de la théorie aristotélicienne du rapport de la sensation et de l'universel se livre peut-être en cette apparente contradiction.
4. *Métaphysique*, Z, 15, et M, 10, qui paraît introduire une solution ambiguë : « En un sens la science est universelle, en un autre sens elle ne l'est pas. »

réel. Elle permet donc de rendre mieux compte de ce dont les hommes sont capables, et ce dans des situations plus favorables à l'expression de leurs formes d'excellence. La fiction permet un meilleur accès à ce que l'on peut appeler la « situation humaine », pour reprendre une expression de M. Scheler.

3 / La poésie est un genre plus *spoudaïon*, plus noble, plus grave, plus profond (Aristote paraît penser d'abord à la tragédie<sup>1</sup> et à l'épopée, même s'il cite la comédie deux lignes plus loin). Elle peut se permettre de *styliser* les événements, en éliminant le futile, l'accidentel, l'anecdotique. Et ses héros sont d'une autre stature que les individus concrets, humains, trop humains. En peignant l'homme « meilleur » qu'il n'est, le poète épique et le tragique s'approchent plus de sa fin, de son essence, que ne le fait l'historien.

4 / La poésie peut constituer de belles totalités intégrées : il y a *unité* du *muthos*, de l'intrigue<sup>2</sup>, de l'action (non du héros, puisque, depuis Eschyle, le *dialogue* entre au moins deux héros est devenu la règle). A l'opposé, l'histoire est désordonnée, inachevée. Cette unité structurale (ou mieux « structuration », *sustasis*) du récit poétique lui donne en quelque sorte une consistance ontologique dont le récit historique est dépourvu. De même que le bonheur ne peut être attribué à un homme qu'après sa mort (cf. Solon et Crésus), de même la limitation et l'ordonnement de l'action tragique permettent de la juger comme une totalité achevée. On sait le parti que l'esthétique classique, singulièrement dans la France du Grand Siècle, tirera de cette valorisation de l'organicité de l'œuvre. Il faudra attendre le siècle suivant pour que l'histoire « laïque » (non sacrée) soit réhabilitée (Vico, Voltaire, Hume).

5 / La poésie, mise en spectacle public, a une fonction sociale cathartique<sup>3</sup> (pour ainsi dire thérapeutique) (1449 b 27) que n'aurait pas la lecture des événements historiques, comme si Aristote avait entrevu que l'histoire thucydidiennne, laïcisée et rationaliste, risquait de désenchanter le monde<sup>4</sup>. La fiction permet d'agir sur les passions dangereuses de manière

1. En 1449 b 24, la tragédie est définie comme « imitation d'une action de caractère élevé » (*mimésis praxeôs spoudaïas*). Le *spoudaïos*, dans l'*Éthique à Nicomaque*, c'est l'homme vertueux.

2. Cf. les belles analyses de P. Ricœur in *Temps et récit*, I, 2, Seuil, 1983.

3. Voir M. Revaut d'Allonnes, *D'un mort à l'autre*, Seuil, 1989, p. 66 sq.

4. En fait, Aristote ne prend guère en compte explicitement la dimension sacrée du tragique, mais il est vrai que Thucydide lui-même insiste (contre Hérodote) sur le fait que « l'absence de merveilleux dans (mes) récits les rendra peut-être moins agréables à entendre » (*loc. cit.*).

plus efficace que le récit historique. Le discours aux Méliens dans Thucydide n'a guère de quoi susciter l'enthousiasme, encore que la terreur et la pitié y trouveraient sans doute leur compte, mais sans pour autant provoquer la « purgation » souhaitée par Aristote. La transfiguration poétique opère une thérapie politique socialement unifiante que l'histoire est incapable d'accomplir.

Laissant de côté cette question des fonctions morales et politiques de la poésie et de l'histoire, reprenons les quatre arguments « épistémologiques » que nous avons attribués au Stagirite :

1 / *La généralité*. Il est clair que l'enquête historique en tant que telle s'intéresse au réel, en ses déterminations aussi bien récurrentes que singulières. Pour Cournot, comme pour Windelband et Popper, tout objet du monde peut se prêter à une approche singularisante, historisante, à la fois de nature descriptive et de nature explicative. Même la physique, en tant que cosmogonie, vise à rendre compte de l'histoire de notre Univers, et la théorie de l'évolution est d'abord une théorie historique. Toute science empirique a pour visée ultime le singulier (le réel) par la construction de modèles. G.-G. Granger, dans son classique *Pensée formelle et sciences de l'homme*<sup>1</sup>, voit là une aporie transcendante du discours de l'historien, qui vise à constituer une science de l'individuel sans modélisation ni médiation pratico-clinique à son objet. En revanche, Hempel, Hayek ou Popper soulignent que même l'approche historique opère avec des modèles (singuliers)<sup>2</sup>, par abstraction et sélection, et qu'elle présuppose nécessairement l'usage de « généralités » du sens commun ou tirées des sciences sociales théoriques (dont la psychologie). Le nominalisme radical est impossible, et tout chercheur doit pouvoir faire librement usage d'universels. Ces présuppositions universelles sont souvent (mais pas toujours) : 1 / vagues, 2 / triviales, 3 / implicites dans les concepts utilisés : le raisonnement de l'historien est fortement enthymématique, au sens moderne de raisonnement auquel manque au moins une prémisse, considérée comme allant de

1. Aubier-Montaigne, 1969. Cf. aussi le bel ouvrage de J.-C. Pariente, *Le Langage et l'individuel*, Colin, 1973 ; et Alain Boyer, *L'Explication en histoire*, Presses Universitaires de Lille, 1992, part. II, chap. 3.

2. Cf. Popper, « La logique des sciences sociales », in Adorno-Popper, *De Vienne à Francfort, La querelle allemande des sciences sociales*, Complexe, 1979, thèses 25 et 26, avec l'exemple – malheureusement trivial ! –, de Charlemagne, et *La quête inachevée*, Agora, § 24, à propos des « modèles singuliers qui sont les hypothèses singulières (en principe testables) de l'histoire » : cette petite phrase trop elliptique va à l'encontre de l'interprétation donnée par J.-C. Passeron de la théorie de la réfutabilité, in *Le raisonnement sociologique. L'espace non poppérien du raisonnement naturel*, Nathan, 1991 (voir plus loin).

soi. Or toute culture se dit par ce qu'elle ne dit pas, autrement dit par ce qu'elle considère comme connu de tous (*Common Knowledge*) et donc impropre à être sans cesse explicité<sup>1</sup>. L'historien peut avoir à déterminer (conjecturalement) ces prémisses manquantes, alors même que son propre discours utilise des *topoi* éventuellement différents. Ces universels implicites du sens commun peuvent devenir à leur tour des objets des sciences sociales théoriques, lesquelles ont charge de préciser et de « détrivialiser » les présupposés en question. Un historien peut affirmer sans autre forme de procès que Louis XIV devint impopulaire du fait des augmentations d'impôt qu'il dut effectuer à la fin de son règne pour financer ses guerres. Malgré la critique de W. Dray<sup>2</sup>, il me semble que cette explication est enthymématique, et qu'elle présuppose une vérité du type « la plupart du temps, l'augmentation des impôts est inversement proportionnelle à la popularité du gouvernement », ce qui n'est pas toujours vrai. Or le quanteur aristotélicien (« la plupart du temps », « normalement ») est vague, et il ne nous donne pas le moyen terme, la cause des exceptions. Il me semble qu'un politologue pourrait vouloir préciser le présupposé, en transformant « la plupart du temps » en « sauf si », suivi d'une liste finie de cas dûment analysés. (Les clauses *ceteris paribus*, on le sait, empoisonnent la question de la testabilité des généralisations empiriques en sciences sociales, singulièrement en économie et en psychologie.) A quelles conditions une telle généralisation ne vaut-elle pas ? La science peut ainsi chercher non à rompre de manière altière avec le sens commun, ce qu'elle ne doit faire que si elle a les moyens de le réfuter (localement, au coup par coup), mais à le préciser, à en déployer les conditions d'applicabilité, quitte à passer pour cela à un niveau supérieur de formalisation. L'historien pourra alors se réappropriier la « loi », qui lui permettra de repérer plus aisément des différences pertinentes de situations. En dernière analyse, ces généralisations empiriques devraient pouvoir être déduites de « modèles situationnels abstraits » faisant appel, comme le voulait d'ailleurs W. Dray, fidèle à la tradition compréhensiviste, et en particulier à Collingwood<sup>3</sup>, à un « principe de rationalité » des agents<sup>4</sup>.

1. Cf. mon article « Cela va sans le dire. Éloge de l'enthymème », in Boyer et Vignaux éd., *Argumentation et rhétorique, Hermès*, 15, 1995.

2. Voir *L'explication en histoire, op. cit.*, part. 1.

3. *The Idea of History*, Oxford, 1946 (un « classique », non traduit en français), et les remarques de Popper, in *La connaissance objective*, Aubier, 1991, chap. IV, en part. § 11 (« La compréhension (herméneutique) dans les Humanités »).

4. Cf. Popper, « La rationalité et le statut du principe de rationalité », in *Hommage à Rueff*, Payot, 1967. Voir aussi le livre savant d'E. Picavet, *Choix rationnel et vie publique*, PUF, 1996.

2 / *Quant aux modalités.* L'historien s'intéresse effectivement à ce qui est réellement arrivé, et singulièrement à ce qui a produit des effets (*Wirklichkeit*) à moyen ou long terme (qui s'intéresserait à Raspoutine, si l'hémophilie de l'héritier de l'empire n'avait pas eu d'effets possibles sur le comportement de la famille du tsar, et donc sur les événements de 1917 ?). Contrairement au romancier, l'historien ne s'intéresse pas en tant que tel à des mondes seulement possibles, alors que la démarche généralisante, loin de pouvoir se ramener à la terne « induction », consiste souvent à lever des conditions (abstraire, c'est soustraire), comme en mathématiques, où même les conditions de toute expérience concrète sont levées. L'historien, qui ne peut pas ne pas sélectionner, ne peut cependant omettre un facteur que s'il est prêt, le cas échéant, à en montrer la non-pertinence pour le problème qu'il s'est donné.

En revanche, il n'est pas exact que l'historien puisse se passer tout à fait de raisonnements contrefactuels : ils ont une valeur heuristique, et, accessoirement, persuasive. Weber et Aron ont insisté sur le rôle de ce type de raisonnements dans la recherche des causes (Que se serait-il passé si... ?). Mais, comme « avec des si on peut mettre Paris en bouteille », il conviendra de faire un usage prudent de ce type de raisonnement, sauf à tomber dans la fiction pure<sup>1</sup>. Par ailleurs, lorsque l'on a affaire à des agents délibérant avant d'agir, il est clair qu'on ne peut les comprendre qu'en restituant autant que faire se peut leurs constructions imaginaires des scénarios possibles de leur point de vue (limité). Le possible ici agit dans le réel, en tant qu'il est envisagé, comme attracteur ou repoussoir<sup>2</sup>. L'heuristique anti-historiciste équivaut à se doter d'une « opération historique », qui consiste à revenir par la pensée au moment où rien n'est joué, où le futur est encore contingent, afin de « rendre au passé l'incertitude de l'avenir », selon la belle formule d'Aron<sup>3</sup>. Le problème herméneutique est ici particulièrement délicat, puisque mis à part la possibilité inéliminable d'une histoire du présent (de Thucydide à la « sociologie empirique » actuelle), l'historien connaît souvent la « fin » de l'histoire (par exemple celle de l'Empire romain). Il doit faire comme s'il ne la connaissait pas, tout en la connaissant, puisque c'est de cette connaissance quasi « divine » qu'il tire son savoir de l'importance et de la signification (*des significations*) de la situation qu'il décrit et explique. L'importance d'un événe-

1. *L'explication en histoire*, part. 5, chap. 3 (« Les sentiers qui bifurquent »).

2. Cf. Popper, *Un univers de propensions*, L'Éclat, 1992.

3. *Introduction à la philosophie de l'histoire*, rééd. Tel, Gallimard, 1981 [texte de 1938].

ment lui est conférée après coup, par ses conséquences, la plupart non voulues. Le sens transcende l'intention.

La modalité du *nécessaire* a-t-elle une place dans le raisonnement de l'historien ? Sans doute pas au sens où Marx parlait de la « nécessité de fer » des lois du capitalisme (et de sa nécessaire autodestruction), ce qui est de l'historicisme (« Lois de l'Histoire »). La contingence et le hasard (au sens cournotien<sup>1</sup> et au sens de Poincaré)<sup>2</sup>, la pluralité des chaînes causales, l'importance des effets inattendus des actes intentionnels sont au contraire des notions indispensables pour penser une histoire pluraliste et post-historiciste. Néanmoins, une sorte de quasi-nécessité peut avoir un rôle dans les explications historiques : il s'agit des « logiques situationnelles fortes », qui entraînent presque nécessairement telle ou telle conséquence dramatique (où l'on retrouve la tragédie, sans avoir à supposer que l'Histoire en sa totalité serait tragique). Car un historien ne peut pas ne pas tenter d'expliquer : même un récit donne un ordre, suggère un enchaînement des événements qui fait signe vers la causalité, alors qu'un romancier peut jouer sur des éléments qu'il maintient sciemment dans une zone d'ombre propre à exciter l'imagination du lecteur, le laissant dans l'obligation de remplir les « blancs », non tant par les « trivialités » dont il a été question plus haut que par des interprétations des intentions ou des motivations des personnages, qu'il aurait peut-être tort de chercher à expliciter totalement. L'historien, comme tout scientifique, cherche à réduire l'apparente indétermination des faits, sans qu'il ait pour autant à présupposer quelque thèse déterministe.

La causalité demeure un concept philosophique, au moins au sens où il est à la fois obscur et indispensable (la science cherche le pourquoi). Mettre en évidence des conditions quasi nécessaires d'un événement est possible, les conditions suffisantes, que Popper, comme à sa suite Hempel, paraît identifier aux causes, demeurant seulement à l'horizon de la recherche : elles ne comprennent rien de moins que la totalité « laplacienne » des antécédents d'un événement. Il me semble que l'historien peut se contenter d'énoncer des « conditions favorisantes » des événements, eu égard à un état du monde qu'il suppose donné. (L'univers historique est éminemment un « univers de propensions ».) L'expli-

1. Rencontre de séries causales indépendantes, qui simule la finalité, au sens d'Aristote (*Physique*, II).
2. Très petites causes, très grands effets (sensibilité aux conditions initiales) ; on pourrait parler d'« effets Cléopâtre » ou d'« effets Cromwell », en hommage à Blaise Pascal. Ce principe va à l'encontre de la thèse (acceptée par Descartes) de la proportionnalité de la cause et de l'effet et peut-être aussi du principe de continuité leibnizien.

cation de cet état du monde est toujours exigible en droit, mais « c'est une autre histoire ».

3 / *La stylisation par élimination.* L'École des Annales a définitivement montré qu'il convenait de se méfier de toute hiérarchisation aristocratique *a priori* des événements en événements « nobles » et événements « vulgaires ». La prise en compte des attitudes quotidiennes et des mœurs (Montesquieu étant sur ce point le « grand ancêtre ») amène nécessairement à un « élargissement du questionnaire » dont L. Febvre et P. Veyne ont admirablement parlé. La faiblesse logique de la coupure administrative entre « sciences sociales » (empiriques) et « histoire » devient d'autant plus patente. Même dans le champ politique, l'importance donnée à l'agrégation de multiples « petites décisions », localement rationnelles (intelligibles) mais globalement « aveugles », relativise celle des « grands hommes », même si elle ne l'abolit pas<sup>1</sup>. Par ailleurs, l'historien ne peut négliger les séries causales qui « ne mènent nulle part », pas plus que le paléontologue ne saurait négliger l'Homme de Néandertal. On oublie souvent ce que l'on pourrait appeler un second type d'« effet Cournot » : il y a des séries causales qui se perdent dans les sables, et qui rendent paradoxalement « le passé plus difficile à connaître que l'avenir ». Mais tout cela n'empêche évidemment pas que, sous contrainte de factualité, si l'on peut dire, l'historien ne puisse être un grand styliste : simplement, cette détermination esthétique n'est pas essentielle à l'objectivité de son œuvre. Il est clair en revanche que le travail historique est pour partie un travail de sélection, certaines périodes historiques présentant paradoxalement « trop » d'informations. Comme dans tous les domaines, la pertinence d'un document est liée au problème que l'on se pose : si l'historien néglige telle ou telle information, il doit pouvoir expliquer pourquoi. Toute théorie contient par sa problématique des critères internes de pertinence (il n'existe peut-être pas d'épistémologie générale de la pertinence). Il n'en reste pas moins que la question de la pertinence d'une « donnée » (forcément interprétée et soumise à la critique) est une question objective, intersubjectivement discutable, comme dans toute science.

4 / *La stylisation par totalisation.* Il y a bien dans le travail historique une dimension de mise en intrigue (Veyne, Ricœur). Il faut d'abord dresser le « paysage situationnel », dans ses dimensions géographiques (parti-

1. Popper se réfère à *Guerre et paix*, dans *Misère de l'historicisme*, Agora, 1988 [1944], pour illustrer ce point (« la logique des événements », in § 31).

nelles), économiques, institutionnelles, psychologiques, événementielles, etc. L'historien construit son récit, il fait des choix narratifs. Le principe de sélection repose, on l'a dit, sur la notion de problème, le méta-problème de l'historien se ramenant souvent à celui de savoir quelle était la « situation de problèmes » des agents en question. Collingwood semble avoir vu là une aporie : si les solutions tentées par les agents ont échoué, il ne serait plus possible de les comprendre, puisque l'on ne comprend une question que grâce à sa solution, et une solution qu'en tant que réponse à une question. Cercle herméneutique. Les deux domaines pratiqués par Collingwood en histoire, l'archéologie et l'histoire des idées philosophiques et scientifiques, pourraient paraître lui donner raison. Mais il faut résister à ce scepticisme<sup>1</sup> : la reconstruction des problèmes à partir des tentatives de solution est un processus « par essais et erreurs », par approximations successives visant à produire une sorte d'équilibre entre nos hypothèses et les données, et non une vision immédiate du rapport vécu entre question et réponse, d'autant que certains acteurs évaluent mal la réelle portée du problème dans lequel ils sont « embarqués ». En ce cas, l'historien est paradoxalement mieux placé que l'acteur pour le comprendre, sans que l'on ait à utiliser des hypothèses coûteuses sur l'inconscient des acteurs. La distance est aussi bénéfique que la proximité pour entendre.

Il me paraît nécessaire d'insister un peu lourdement sur ceci que « construire » les faits ne signifie pas les *créer*. Toute version forte du « constructionnisme » est délirante : autant accorder au révisionniste que les chambres à gaz (ne) sont (que) des « construits historiques » ! Celui qui, sous prétexte de non-naïveté, méprise l'idée même de « fait historique » est sur la pente de l'idéalisme fantaisiste, tout comme celui qui prétend que les faits objectivés par le savant dans son laboratoire ne sont que des « artefacts ». Que tout fait soit « interprété à la lumière d'une théorie » (Popper) n'implique pas que tout n'est qu'interprétation. Il y a des interprétations *fausses* des faits (et donc des vraies, même si nous ne savons pas avec certitude quelles elles sont). La valeur de vérité d'une proposition claire et distincte ne dépend pas de nous, elle est immaîtrisable, inaccessible à la décision humaine<sup>2</sup>. Barguigner sur cette thèse paraît dangereux à tous points de vue. En ce sens, l'historien est soumis, ou

1. Cf. J. Watkins, « Imperfect Rationality », in Borger & Cioffi (eds), *Explanation in the Behavioural Sciences*, Cambridge, 1970. Cet article, où l'on trouve notamment une critique détaillée de la théorie classique de la décision, mériterait d'être plus connu.
2. Ce que nous décidons (rationnellement), c'est d'*accepter* (« jusqu'à preuve du contraire ») une hypothèse comme étant, à la lumière du débat critique, la meilleure candidate à la vérité.

plutôt se soumet librement à une contrainte de véridicité factuelle qui le spécifie suffisamment par rapport au poète, même si l'on peut accorder à ce dernier, par-delà sa liberté de « variation » imaginaire, une visée plus générale de vérité : Aristote eût sans doute été enchanté par *L'Avare*, en y trouvant plus de « vérité » que dans une description sociologique (historique) détaillée. Mais l'historien ne travaille pas à ce niveau d'abstraction là.

La critique des témoignages et des documents est de l'ordre de la déontologie minimale de l'historien. Il lui faut des données *indépendantes* et *convergentes*. En ce cas, il peut conjecturer qu'il est improbable que cette convergence soit une pure coïncidence. Comme le faisait remarquer finement M. Bloch<sup>1</sup>, un excès de convergence dans le détail peut amener l'historien, en utilisant le même argument de l'« improbabilité des coïncidences »<sup>2</sup>, à soupçonner une contrefaçon, puisqu'il est improbable que des documents indépendants concordent sur des détails accidentels et que les déformations dues au temps aillent exactement dans le même sens (un falsificateur doué peut lui compliquer la tâche en introduisant de manière aléatoire de telles variations dans ses fabrications : « Élémentaire, mon cher Watson ! »). Il est improbable que le hasard simule la finalité dans le détail (et dans un temps très court). A condition de ne pas passer par-dessus bord l'idéal de vérité, l'usage de l'argumentation par l'improbabilité des coïncidences me paraît indispensable pour lutter contre les effets délétères du « vérificationnisme », lorsque celui-ci sert à de mauvaises gens pour soutenir leur mauvaise cause. Sans le recours à cet argument, tout événement historique devient également douteux.

Par-delà la question de la fidélité critique par rapport aux faits, demeure-t-il un espace pour quelque chose comme l'idéal d'unité organique du récit ? Lorsque le processus décrit est achevé, non dans ses conséquences, ce qui n'est pas la règle, mais en quelque sorte dans sa « substance », au sens où la féodalité en tant que système ou « paysage situationnel » est « chose du passé », le récit peut se déployer comme un drame, avec sa fin « naturelle », au demeurant peut-être due à des intrications de tendances « lourdes » et de « concours de circonstances » complexes. Mais si le processus en tant que tel est encore « vivant » (la laïcité, l'Islam, etc.), l'historien ne peut semble-t-il que faire appel à des « interprétations » non testables, caractérisées par Popper comme des « cristalli-

1. In *Apologie de l'histoire*, Colin, 1941 (rééd. 1964).

2. Cf. Boyer, « Microcosme et Malin Génie », *Les années trente. La formalisation*, F. Nef et D. Vernant (éd.), Vrin, 1998.

sations de points de vue » sur l'histoire<sup>1</sup>, à condition de les expliciter et de ne pas faire comme si elles étaient des hypothèses confirmées par les faits qu'elles subsument. L'histoire de la civilisation européenne vue sous le point de vue de la lutte des classes, ou sous le point de vue de l'émergence de la démocratie, voilà des exemples de telles interprétations. Bien sûr, de tels points de vue sont également nécessaires lorsque l'on analyse des « totalités achevées », mais il me semble que leur nécessité se fait encore plus sentir dans le cas philosophiquement plus troublant de l'étude d'« histoires en cours ».

Plus généralement, le récit historique est soumis à ce que l'on pourrait appeler *la tentation de la poésie*, concurrentement avec *la tentation de la science*. La tentation poétique peut être de nature globale ou locale : globale, si l'historien se fait philosophe de l'Histoire, processus dont il croit pouvoir déchiffrer l'énigme fondamentale, et donc éventuellement le sens futur : tels sont l'« historicisme » au sens de Popper, la « sociodicée » au sens de Raymond Aron, les « Grands Récits », au sens de J.-F. Lyotard ; locale, s'il considère que son objet privilégié est constitué d'événements eux-mêmes sublimes, au sens kantien<sup>2</sup>, ce qui l'amènera à se concentrer sur des séries bien unifiées d'événements à teneur poétique, épique ou tragique (la Révolution et l'Empire, par exemple) : l'histoire ici imite à son tour la poésie, avec des catégories comme le héros, le coup de théâtre, la péripétie, le dénouement, la conséquence involontaire (*hétérotélie*)<sup>3</sup>, avec des contraintes d'unité de temps et d'action (voire de lieu, laquelle n'est pas explicite dans Aristote). Du reste, si l'on s'avise du fait que nombre d'acteurs politiques ne sont pas sans avoir été influencés par leur propre tendance à s'identifier à des « postures » mythico-poétiques, il n'y a rien d'étonnant à constater que l'histoire peut imiter la poésie. Par ailleurs, si la poésie *dramatique* est une « imitation de l'action », l'histoire « poétique » se concentrera sur le domaine par excellence de la *praxis*, à savoir la politique, y compris sa « continuation par d'autres moyens », la guerre. Dans les deux cas, on peut dire que le défi aristotélicien est relevé : l'histoire est au moins aussi philosophique que la poésie, et elle est elle-même poésie. Or, le double mouvement qui paraît caractériser l'his-

1. *The Open Society*, Routledge & Kegan Paul, 1962 [1945], chap. 25 (« Has History any meaning ? »). Cette thèse demanderait élucidation ; les analyses de R. Aron sont beaucoup plus riches (voir les Appendices à *L'introduction à la philosophie de l'Histoire*, Gallimard, rééd. Tel), mais elles vont dans le même sens.
2. Sur l'importance de l'exemple, voir *Critique de la faculté de juger*, § 32 (Aubier, trad. A. Renaut, p. 267).
3. Au sens de J. Monnerot, *Les lois du tragique*, Gallimard, 1969.

toriographie moderne se manifeste d'une part par la méfiance à l'égard de toute « philosophie de l'Histoire », de toute dramaturgie de l'histoire globale<sup>1</sup>, d'autre part par la relativisation de l'importance de l'histoire « événementielle » (Braudel). La belle linéarité de l'Histoire et sa temporalité monophonique se voient discréditées, au profit d'enchevêtrements de temporalisations multiples, non nécessairement convergentes ; ce qui d'ailleurs est compatible avec une réévaluation de l'événement, sensible en France depuis une vingtaine d'années. Comme on le sait, il ne s'agit sans doute plus du même « événement ».

A propos de la « scientificité » éventuelle de l'histoire et des sciences sociales, qu'on me permette de faire quelques remarques sur la question de la « falsifiabilité » ou testabilité<sup>2</sup>. J.-C. Passeron explique longuement, dans l'ouvrage cité plus haut, en quoi les sciences sociales définiraient un « espace non poppérien », marqué par l'usage nécessaire de noms propres, alors que les hypothèses scientifiques au sens de Popper seraient obligatoirement des universelles au sens strict, définies précisément dans la *Logique de la découverte scientifique* (*Logik der Forschung*, 1934), laquelle se voulait d'ailleurs *explicitement* une analyse des « méthodes des sciences de la nature » (*Naturwissenschaften*)<sup>3</sup>. Le style de J.-C. Passeron est très polémique et ironique. On est tenté de dire : « Pourquoi tant d'agressivité ? »... Après tout, les propositions avancées par un philosophe ne sont pas à prendre ou à laisser, mais à discuter, et, éventuellement, à réfuter. Or, on peut montrer que la polémique de M. Passeron *porte à faux*. Il est inexact que des hypothèses singulières ne puissent être par construction réfutées empiriquement : les « conditions initiales » de toute explication scientifique au sens de Popper (et à sa suite de Hempel) *sont* (ou doivent être) testables, même si la question de savoir si la réfutation porte sur elles ou sur la théorie se pose toujours (problème de Duhem). Elles doivent être indépendamment testables (non *ad hoc*), testables par d'autres énoncés que l'*explanandum*. On peut toujours « éviter » une réfutation. Popper propose de considérer comme ne respectant pas les « règles du jeu de la

1. Popper définit l'historicisme – concept qu'il prit soin de distinguer de celui d'historisme (*Historismus*), avant que L. Strauss ne le définisse tout autrement –, comme l'idée selon laquelle l'Histoire aurait une « intrigue » (*a plot*). Il insiste sur ses origines religieuses.
2. On ne présuppose pas dans ce qui suit que le concept de falsifiabilité soit sans problèmes, mais seulement qu'il convient de ne pas en voir là où il n'y en a pas. Popper a dit beaucoup de choses fausses, et s'est souvent trompé, y compris de cibles ; ce n'est pas une raison pour lui faire dire ce qu'il n'a pas dit.
3. *Le raisonnement sociologique*, p. 378 : « Seuls les énoncés universels au sens strict répondent aux exigences de la mécanique falsificatrice » : cela est très contestable. Si je dis « Il ne pleuvra pas à Paris le 31 décembre 1999 », une telle prédiction singulière est falsifiable.

science » les stratégies d'évitement systématique de la mise en cause empirique, ou « stratégies immunisantes ». Mais cela ne veut pas dire que tout ce qui n'est pas testable soit sophistique ou pseudo-scientifique. Nous pouvons difficilement nous passer de « points de vue » ou d'« interprétations ». Cela dit, l'historien, même s'il est guidé (comme d'ailleurs le physicien) par des idées métaphysiques, peut tenter de rendre ses hypothèses plus (intersubjectivement) contrôlables. Il convient de contester plusieurs erreurs communes quant à la testabilité des hypothèses historiques (sans parler de celle-ci : croire que « testabilité » implique « universalité au sens strict »)<sup>1</sup> : 1 / croire que le test des hypothèses historiques passe par leur capacité prédictive au sens obvie (prédire l'avenir à long terme) ; il ne s'agit selon Popper que d'un avatar de l'*historicisme* ; 2 / croire que « testable » implique « mathématiquement formulé », alors que la mathématisation n'est sur ce point qu'un moyen d'augmenter l'informativité (et donc la testabilité) des hypothèses, sans être un réquisit nécessaire de l'« empiricité » ; 3 / croire que tester veut dire nécessairement expérimenter, comme si les hypothèses astronomiques n'étaient pas testables observationnellement ; 4 / croire que le test des hypothèses exige une base de données neutres et évidentes. La testabilité des hypothèses singulières est évidemment plus faible au sens logique que celle des hypothèses universelles : elles ont moins de contenu. Par ailleurs, certaines hypothèses contiennent à la fois des concepts universels et des noms propres. Elles sont *stricto sensu* non vérifiables, mais peuvent être réfutables. L'hypothèse « Toutes les orbites des planètes sont elliptiques » est plus testable que « La planète Mars a une orbite elliptique », puisqu'elle prédit beaucoup plus, en prenant plus de risques, et en étant potentiellement bien plus informative. C'est pourquoi Kepler ne s'est pas contenté de la seconde proposition, et qu'il l'a *généralisée* pour obtenir sa loi, contrôlée ensuite sur les autres planètes. En revanche, l'énoncé « Il existe (quelque part dans l'espace-temps) une planète qui a une orbite elliptique (ou triangulaire) » n'est pas en tant que tel réfutable, et demeure « métaphysique »<sup>2</sup>, comme « Il existe une pierre philosophale ». C'est une hypothèse

1. En fait, même les théories physiques ne peuvent se passer de tout nom propre, dès lors qu'elles ont un aspect historique, comme l'astronomie, ou qu'on désire les appliquer ou les tester (*Logique de la découverte scientifique*, § 14, en particulier n. 2, p. 64, sur les « déterminations (spatiales, temporelles ou autres) » comme « principes d'individuation »). Sur l'usage des particuliers égocentriques tels qu'« ici » et « maintenant » dans les sciences, cf. la réponse de Popper à Grünbaum in *The Philosophy of K. Popper*, Schülpp (éd.), La Salle, 1974, vol. II, p. 1143.
2. Il est de plus non vérifiable, à cause de l'universel « ellipse ». « Il existe (quelque part dans l'espace-temps) un cheval ailé » est en revanche unilatéralement décidable, parce que vérifiable et non réfutable par l'expérience.

« purement existentielle », logiquement très faible, même si elle peut être la *conséquence* d'une hypothèse forte et testable : *l'empiricité n'est pas une propriété déductivement transmissible*. Les thèses métaphysiques *proprio sensu* sont souvent selon Popper de la forme « Tout et quelque chose » (*All-and-some*) (« Tout événement a une cause »), ce qui les rend à la fois non vérifiables (quantificateur universel) et non falsifiables (quantificateur existentiel), mais n'implique pas qu'elles ne puissent jouer un rôle heuristique fécond (« Programmes métaphysiques de recherche »), ni qu'elles soient inaccessibles à la critique rationnelle. On peut d'ailleurs tenter, à la manière kantienne qu'adoptait Popper dans *Logique de la découverte scientifique (LDS)*, de les transformer en règles de méthode, ou principes régulateurs. Quoi qu'il en soit, les hypothèses *singulières* peuvent être testables (à condition que leurs opérateurs d'individuation permettent d'identifier spatio-temporellement leurs objets). Les « énoncés de base » (énoncés tests) eux-mêmes sont des énoncés *singuliers* « facilement testables intersubjectivement » (*LDS*, § 29). « Facilement » ici est à entendre en un sens pragmatique : des énoncés sur la valeur de vérité desquels lesquels l'accord peut se faire « facilement », au rebours des énoncés plus théoriques. Sans une telle « facilité », la science serait impossible, ce serait « l'échec du langage comme moyen de communication universel, une nouvelle Tour de Babel »<sup>1</sup>. Par ailleurs, Popper critique l'*historicisme*, pas l'idée de « connaissance historique », qu'au contraire il défend *explicitement*. Les hypothèses de l'historien, comme celles du linguiste, de l'ethnologue, de l'économiste, du psychologue ou du sociologue<sup>2</sup>, *peuvent* donc être en principe testables. Ce qui ne veut pas dire que certains points de vue ou principes de sélection ne soient pas nécessaires, on l'a dit : c'est ce que Popper appelle des « interprétations », qu'il ne récuse pas en bloc pour « historicisme », contrairement à ce qu'affirme J.-C. Passeron<sup>3</sup>. Celui-ci a raison lorsqu'il pointe une difficulté dans l'exigence poppérienne de répétabilité intersubjective de l'expérience, exigence qui vise à éliminer la présence d'effets « occultes ». Mais la non-reproductibilité des événements histori-

1. *LDS*, § 29. Il est intéressant de noter que Popper utilisera plus tard la même métaphore en un tout autre sens : « Si la Tour de Babel n'avait pas existé, il nous faudrait l'inventer » (*Conjectures et réfutations*, Payot, 1985, chap. 17, § 4). Il s'agit alors de défendre l'idée que *la valeur d'une discussion dépend de la variété des points de vue en concurrence*. Toute la question du pluralisme théorique et pratique se ramène à celle du « degré de babélisation » souhaitable.

2. Cf. R. Boudon, *La logique du social*, Pluriel.

3. *Op. cit.*, p. 191 : une telle erreur de lecture conforte l'idée de Popper selon laquelle on ne lit jamais un texte sans préjugés, sans « anticipations ». Par ailleurs, ce n'est pas la mise sous forme spinoziste/witgensteinienne, avec des propositions numérotées et des scolies, mais sans démonstration, qui peut à elle seule donner de la rigueur argumentative à la critique.

ques n'implique pas la non-reproductibilité des examens critiques des documents (les observations astronomiques sont également intersubjectivement reproductibles en ce sens seulement). Que l'on puisse tenter de montrer qu'en sciences sociales la part relative de l'interprétation (par rapport à celle des hypothèses testables) est « plus grande » que dans les sciences de la nature, Popper n'en disconvierait pas : la testabilité n'en demeurerait pas moins un *desideratum* méthodologique universel, et les « stratégies immunisantes », grâce auxquelles on change subrepticement le sens de la théorie en la protégeant de toute expérience négative possible, sont à proscrire.

On peut citer de nombreux exemples. Ainsi, la recherche en archéologie présente certaines similitudes avec les sciences de la nature, et une découverte est souvent utilisée pour réfuter des hypothèses admises. Les thèses de Max Weber sur l'origine du capitalisme ne sont pas « à prendre ou à laisser », et elles ont été critiquées *empiriquement*. L'histoire quantitative a depuis quarante ans habitude l'historien, tels l'économiste et le sociologue, à tester ses hypothèses de manière assez rigoureuse. Critiquant avec verve les statistiques de l'Anglais G. King à propos des différences entre les économies anglaise et française dans la deuxième moitié du XVII<sup>e</sup> siècle, E. LeRoy Ladurie écrivait il y a près de trente ans : « Les historiens sont tatillons ; ils veulent savoir, connaître ces “méchants faits”, qui détruisent les belles hypothèses. »<sup>1</sup> Comme déclaration « non poppérienne », on fait mieux... En histoire politique, l'ouverture des archives de Moscou est en train de nous montrer que certaines de nos croyances sur l'histoire du communisme sont à réviser (même si, bien entendu, il n'existe rien de tel que des « données brutes », non soumises à la critique, ce qui est une banalité de base). Les analyses des structures sociales<sup>2</sup>, rapports de propriété, rapports de force, rapports de coordination, d'échanges, sont des domaines où il ne paraît en rien impossible d'obtenir des hypothèses contrôlables indépendamment de ce qu'elles expliquent, testables. L'histoire des sciences, qui bien entendu requiert tout particulièrement des interprétations philosophiques, ne passe pas pour un domaine

1. « Les comptes fantastiques de Gregory King », in *Le territoire de l'historien*, Tel-Gallimard, 1997, p. 259.
2. Voir la belle analyse que donne K. Pomian des travaux de P. Toubert sur les « structures du Latium médiéval », in « L'histoire des structures », *La nouvelle histoire*, Retz, 1978. L'histoire structurale, à condition qu'elle ne fasse pas fi des *acteurs*, en ne connaissant que le *système*, pour reprendre les heureux termes de Crozier et Friedberg, est compatible avec la théorie de l'« analyse situationnelle ». Sur cette notion, voir *L'explication en histoire*, *op. cit.*, et « Loi, rationalité et situation », in *Introduction à la lecture de Karl Popper*, Presses de l'École normale supérieure, 1994.

où il soit transcendentalement impossible de contrôler ses dires par la découverte de documents nouveaux. Même l'histoire des mentalités offre de quoi alimenter la « machinerie falsificatrice », selon l'expression polémique de J.-C. Passeron. L'hypothèse de M. Vovelle, selon laquelle la déchristianisation *précède* la Révolution française, paraît ainsi relativement contrôlable : on peut imaginer des travaux empiriques montrant que ce qui vaut pour telle province ne vaut pas pour les autres, ce qui amènerait à relativiser les généralisations (légitimes) auxquelles se livre l'historien lorsqu'il suppose que son échantillon est suffisamment représentatif. La communauté des historiens ou celle des *social scientists* peut en ce sens fonctionner comme une communauté scientifique, avec ses propres méthodes de contrôle des énoncés, même s'il paraît difficile en ces domaines d'atteindre des résultats aussi impressionnants par l'audace de leurs prédictions que ceux que l'on obtient en physique, et si la polémique passionnée y est encore plus forte que dans les sciences « dures ». Croire cependant que ces dernières se développent dans le silence des passions serait une grossière erreur. L'idée que l'objectivité scientifique serait due à la neutralité des savants est un des mythes que Popper s'est toujours efforcé de dénoncer, arguant (dès 1944) que ce qu'on oublie la sociologie de la science, c'est la... sociologie de la science, le caractère institutionnel de l'objectivité et du progrès scientifiques. Mais rien ne s'oppose du point de vue de la méthode à ce que des normes de type scientifique concernant la mise à l'épreuve critique et publique des hypothèses puissent réguler la recherche en sciences sociales et en histoire. Où sont là les « interdits » poppériens ? Le seul qu'on puisse formuler serait du type : « Imaginez des hypothèses audacieuses et concurrentes et essayez de les exposer le plus possible à la critique et à l'épreuve de l'expérience ! », ce qui en toute rigueur est moins un interdit qu'une suggestion ou une provocation.

La reconstitution des modèles situationnels des agents eux-mêmes, de leurs problèmes, de leurs buts et de leurs intentions, est à dire vrai très délicate, mais elle est souvent nécessaire. L'explication causale comprend presque toujours au moins un moment de compréhension. On augmente la rigueur (et donc la vulnérabilité) des hypothèses en se donnant pour règle un « principe de rationalité », dont la formulation ne laisse d'ailleurs pas d'être problématique : elle pourrait constituer l'un des problèmes principaux des sciences sociales théoriques. Il paraît de bonne méthode de partir d'une rationalité forte<sup>1</sup>, et de l'affaiblir si besoin est, en tenant compte des évidentes limites cognitives des

1. Voir « égoïste » : cf. P. Van Parijs, *Le modèle économique et ses rivaux*, Droz, 1990.

acteurs : il est irréaliste de faire comme si un individu pouvait connaître la totalité (infinie) des conséquences de ses croyances et des scénarios d'action qu'il envisage, sans parler de ceux qu'il n'imagine même pas. Mais il paraît également nécessaire de postuler, en l'absence d'indices du contraire, une cohérence minimale de l'acteur et une sorte de stabilité de ses objectifs centraux : reconstruire son « plan de vie », sa hiérarchie mouvante d'objectifs, de désirs et d'intentions plus ou moins conscientes, telle est l'une des tâches de l'historien (même quand il s'intéresse à des « acteurs anonymes »), qui ne peut, me semble-t-il, éviter de présupposer quelque chose comme une thèse de l'identité personnelle, ou de la *subjectivité* des actes intentionnels. Une histoire radicalement behavioriste est absurde. Si un objectif partiel paraît changer, il convient de déterminer conjecturalement un méta-objectif qui permette de rendre compréhensible le fait que tel changement dans l'environnement situationnel de l'agent (ou plutôt dans sa représentation subjective de son environnement, laquelle peut être erronée ou partielle) ait provoqué telle modification apparente de l'objectif de premier ordre.

Ces *logiques situationnelles*, nécessairement stylisées par l'historien, mais sous contrainte de conformité (approchée) aux faits, peuvent imiter le drame, comme la nature imite l'art, en ce qu'elles peuvent exhiber des impasses tragiques, des dilemmes, des retournements de situation, etc. A cela est liée la possibilité heuristique de sympathiser avec les acteurs : sympathie à entendre en un sens large, qui recouvrirait les deux aspects de la *katharsis* : la terreur (*phobos*) et la pitié (*éléos*). Cet outil heuristique n'est cependant pas suffisant, et il vaut mieux être en mesure d'exhiber des données indépendantes, voire une relecture enrichissante de données « connues » : une théorie selon Popper est fortement corroborée si elle amène à réviser les données qu'elle avait pour tâche d'expliquer (ce qui incidemment montre qu'elle n'a pas pu être « induite » de ces mêmes données)<sup>1</sup>.

L'environnement situationnel est constitué de tout ce qui est pertinent pour les agents en question, autrement dit tout ce qui se rapporte à une structuration problématique particulière : des éléments physiques et climatiques (toute histoire est géographique), avec leurs contraintes et leurs « espaces de liberté », les champs d'action possible dont ils définissent les contours ; culturels, aussi bien institutionnels que traditionnels ; stratégiques, mettant face à face des acteurs en interaction coopérative ou conflictuelle. Seuls les individus agissent, mais toujours dans des condi-

1. Cf. *La connaissance objective*, Aubier, 1991, chap. V.

tions données, qu'ils n'inventent ni ne dominent souverainement, mais qu'ils peuvent modifier, volontairement ou non (c'est en ce sens que les « individualistes méthodologiques » peuvent, comme Popper, se réclamer de Marx, sans avoir à hypostasier et à privilégier par principe les rapports de domination). Dans certains cas, l'ensemble des stratégies possibles est en quelque sorte donné, et l'acteur n'est qu'un sélectionneur, optimal ou sous-optimal (*satisficer* simonien), et son comportement peut être conjecturalement prédit sur la base de l'analyse situationnelle (« animée », comme dit Popper, par un principe de rationalité). Dans d'autres cas, l'ensemble des solutions possibles n'est pas donné, ou pas entièrement, ce qui amène certains acteurs à inventer, à innover : invention qui peut dans certains cas être retenue et imitée, plus ou moins consciemment, par d'autres. L'invention institutionnelle, sur laquelle insiste à juste titre Castoriadis, est ainsi l'une des sources essentielles du devenir historique. Or, dans ce cas, l'analyse situationnelle ne peut qu'après coup reconstituer la rationalité de la solution, sans pouvoir la prédire : il me semble qu'il y a là une *limite a priori de toute logique situationnelle* et, avec elle, de toute théorie normative de la décision<sup>1</sup>. Que l'acteur ne soit pas seulement décideur mais aussi créateur de « voies de sortie » des situations suggère la thèse d'une transcendance inévitable de l'historique sur le théorique. Aucune théorie de la décision, malgré les avancées dues à la théorie simonienne de l'heuristique<sup>2</sup>, ne peut prédire une innovation réelle, en histoire de l'art comme en histoire des sciences, en histoire des techniques comme en histoire politique. Il est trop facile d'être « sage après coup » (*to be wise after the event*), selon l'expression du poppérien J. Agassi, à propos de l'histoire des sciences. La « solution » gaullienne au problème de la défaite militaire de Juin 40 peut être reconstruite, elle n'était pas donnée au sens où les matrices de la théorie des jeux présentent l'espace des stratégies possibles ; on peut peut-être construire cette matrice après coup, en intégrant la stratégie « Londres », mais rien ne permet de déduire effectivement cette stratégie à partir d'une analyse situationnelle pure et simple, alors que le « coup » imaginaire de De Gaulle a consisté précisément à inventer une solution inouïe : et s' « il est beau de tenter des choses inouïes » (Corneille), on ne s'étonnera pas que l'histoire puisse être plus poétique que la science. *Il ne s'agit pas de choisir entre des options données, mais d'en générer de nouvelles*. On peut reconstruire la rationalité de la démarche de Descartes, et

1. Cf. B. Saint-Sernin, *Les mathématiques de la décision*, PUF, 1973, p. 320 (« Limites de l'usage des calculs et subjectivité »).

2. Cf. « D'où viennent les idées justes ? », in *Introduction à la lecture de Karl Popper*.

la logique de son argumentation<sup>1</sup>, il n'en demeure pas moins que l'invention cartésienne n'est pas réductible au choix d'une stratégie philosophique *donnée*. L'invention est affaire d'audace et d'imagination critique, et les sciences cognitives paraissent encore loin d'être en mesure de rendre compte de ces processus complexes de la pensée imaginative et créatrice, où la présence de mécanismes spécifiques de production d'essais et de contrôle des erreurs est assez plausible.

Bien entendu, tout n'est pas innovation dans l'action, ce qui d'ailleurs ne manquerait pas d'avoir des effets pervers sur la possibilité même de la précompréhension intersubjective anticipatrice, fondement des sentiments de confiance (et de méfiance) qui permettent l'action<sup>2</sup>. Autrement dit, la structure de base traditionnelle de l'action sociale est constituée de représentations collectives relativement stables (à environnement situationnel donné, répétitif ou évoluant très lentement), représentations en général et légitimement non problématisées (d'où leur apparence de routines non conscientes, alors même qu'elles sont acquises par inculcation ou par mimétisme). En l'absence d'un *habitus* critique (ou « tradition de second ordre »), ces normes peuvent se « pétrifier » sans garder nécessairement la trace de leur originelle rationalité, et donc devenir inertielles (résistantes au changement). R. Boudon a parfaitement raison de soutenir que bien des comportements apparemment « irrationnels » peuvent être expliqués du point de vue de la rationalité située de l'acteur, ce qui est d'ailleurs assez proche du point de vue de Popper. Mais il est rationnel qu'il y ait des normes non problématisées, que l'acteur ne calcule pas à chaque fois la décision la plus rationnelle et qu'il suive donc des règles, éventuellement sous-optimales. Il est compréhensible que certains individus particulièrement intégrés à un groupe préfèrent la sécurité de la norme, même si le paysage situationnel qui lui donnait sens a changé, à l'aventure incertaine de la rupture. Il me semble qu'on ne peut exclure *a priori* la possibilité de normes suivies simplement parce qu'elles sont des normes « inertielles », ce qui permettrait de justifier partiellement la classification quaternaire de Max Weber. Le rationalisme méthodologique encourage simplement l'historien à ne pas recourir en premier lieu à la catégorie des comportements « traditionnels » (et à celle des comportements passionnels). Nier l'importance dans toute action des dimensions normative et émotionnelle est cependant impossible, la psychologie

1. Voir A. Bouvier, *L'argumentation philosophique*, PUF, 1995.

2. Cf. Popper, « Vers une théorie rationaliste de la tradition », in *Conjectures et réfutations*, Payot, 1985.

cognitive tendant du reste à corroborer l'idée selon laquelle aucun comportement intelligent ne s'opère dans le silence total des passions.

Il n'existe pas de méthode mécanique pour reconstruire les croyances et les intentions des autres, juste une heuristique projective et donc conjecturale (« Qu'est-ce que je ferais à sa place ? »). L'historien est donc légitimement prudent sur ce point, accusant celui qui s'autoriserait à reconstruire des pans entiers de monologues intérieurs de « faire du roman ». Cette prudence est d'autant plus justifiée que les manières dont une reconstruction des croyances est effectuée ne sont pas en général substituables l'une à l'autre, même si elles sont logiquement équivalentes. Il s'agit du fait bien connu de l'opacité référentielle, selon l'heureuse expression de Quine : De « Œdipe savait qu'il avait tué Laïos » et de « Laïos était son père », aucune personne raisonnable ne déduirait « Œdipe savait qu'il avait tué son père » (avant qu'il ne le découvre avec horreur). Les motivations que l'on est amené à attribuer aux agents sont également une source de perplexité non éliminable. En dehors des motivations « éternelles » de l'animal humain (peur, faim, désir sexuel, protection des proches), vraies « la plupart du temps », et qui, même si elles sont déclinées de manière diverse par toutes les cultures, n'en demeurent pas moins accessibles, par simple introspection, c'est à l'histoire elle-même de faire la carte des « mentalités » d'une époque donnée, par exemple des mentalités religieuses. On peut rappeler ici que, selon Popper, la psychologie ne saurait être autonome : la structure motivationnelle des acteurs dépend de leur environnement situationnel, lequel peut même contraindre les individus à restreindre ou à modifier leurs motivations. Ainsi, une institution comme le marché peut amener, sous peine d'élimination, les agents à se comporter « comme si » ils étaient mus simplement par l'appât du gain ; une structure du type « dilemme des prisonniers » (modèle de situation contraignante abstraite) peut conduire les agents à ne pas coopérer (au moins dans un jeu non répété), alors même qu'ils n'ont pas de « pulsions » agressives ; une structure institutionnelle comme la démocratie, ainsi que l'avait fait remarquer (polémiquement) Platon, avant Tocqueville, peut induire une psychologie particulière, etc. La structure des incitations à agir est variable, et sa mise en évidence permet de tracer le cadre dans lequel est possible l'assignation de désirs et de croyances aux agents. Sans connaître les règles d'un jeu, on ne peut bien évidemment rien comprendre au comportement d'un agent : il paraît « absurde ».

Comme l'avait brillamment souligné Raymond Aron, la connaissance historique est en un sens un cas particulier de la connaissance d'autrui, avec ceci de singulier qu'il faut dans le cas de l'histoire conjecturer les

intentions sans pouvoir tester nos conjectures en interrogeant l'autre ou en observant ses réactions, ce qui du reste ne supprime pas l'incertitude et l'ambiguïté de la compréhension (mensonges, rationalisations, stylisation), mais les limite. Il n'y a pas de raison d'en déduire que des hypothèses portant sur les intuitions et les représentations d'autrui soient par essence inaccessibles à l'examen. Une telle pratique demande seulement beaucoup d'imagination et de sagacité critique. Après avoir commencé avec une parole d'Aristote, terminons en citant Hobbes :

« Quoique les mots soient les signes que nous avons des opinions & des intentions, comme ils sont si souvent équivoques suivant la diversité de la texture du discours & celle des personnes en la compagnie desquelles ces mots se débitent, & comme pour nous en faire démêler le vrai sens il faut voir celui qui parle, être témoin de ses actions & conjecturer ses intentions, il s'ensuit qu'il doit être extrêmement difficile de découvrir les opinions & le vrai sens de ceux qui ont vécu longtemps avant nous, & qui ne nous ont point laissé que leurs ouvrages pour nous en instruire, vu que nous ne pouvons les entendre qu'à l'aide de l'histoire, par le moyen de laquelle nous suppléerons, peut-être, au défaut des circonstances passées, mais non sans beaucoup de sagacité. »<sup>1</sup>

1. *De la nature humaine*, rééd. Vrin, 1971 (trad. du baron d'Holbach), p. 160.



## *Complétude et incomplétude de l'action*

### INTRODUCTION

L'idée que l'action peut être complète, au sens où l'agent en inspire le projet et en assume toutes les péripéties, même quand il ne les a pas voulues, n'est pas familière aux Anciens ; Hannah Arendt en attribue la paternité à saint Augustin. Analysant les opacités du vouloir, celui-ci, tout en mettant au jour la scission de la volonté, déclare l'homme responsable de ses entraînements et de ses passions. Mais Augustin, tout en formant l'idée d'une action entièrement habitée et assumée par son auteur, est aussi le théoricien de la « causalité déficiente » : le métal dont nous sommes faits comporte des pailles indétectables, que seule l'action révèle. Elle est donc toujours un apprentissage de la faiblesse, alors que, justement, elle exige de l'habileté et de la force. Cette contrariété pourrait engendrer un dilemme. Or ce n'est pas le cas. Selon saint Augustin, l'action conduit l'agent à accepter – et même à revendiquer – la dualité de son expérience : il est faible, les circonstances peuvent le ballotter, le réduire, le pousser à renoncer ou à trahir ; toutefois, l'action suscite en lui une instance invulnérable, sorte de « chambre de veille », pourrait-on dire, où l'esprit – même aux instants de découragement et de peur – se tient en quelque sorte au sec. Ces suppositions reposent pour Augustin sur sa foi ; mais leur évidence anthropologique n'en est pas tributaire.

Il ne faut pourtant pas imaginer, inhérent à l'agent, une sorte d'être intérieur à son être, même si la métaphore de l'« homme intérieur » sem-

ble le suggérer. On ne cesse, en effet, de psalmodier, après Husserl, le fameux : « *Noli ire foras, in teipsum redi, in interiore homine habitat veritas.* »<sup>1</sup> Mais on omet la suite, qui dit pourtant l'essentiel : « *et si tuam naturam mutabilem inveneris, transcede et teipsum* »; et si tu trouves ta nature changeante, alors surmonte-toi, toi aussi ». Une telle opération – qu'un examen trop rapide fait appeler « mystique » – est liée, pour saint Augustin, à la rationalité de l'action. Il ajoute, en effet : « *Sed memento cum te transcendis, ratiocinantem animam te transcendere. Illuc ergo tende, unde ipsum lumen rationis accenditur.* » Tel est le projet que nous formons, aujourd'hui encore, quand nous méditons sur la rationalité de l'action : tendre « vers ce point où la lumière de la raison elle-même s'embrace ».

Une fois encore, il ne semble pas que ce soit la foi de saint Augustin qui confère un contenu – réel pour certains, illusoire pour d'autres – à son analyse : il nous prescrit de chercher comment la lumière de la raison illumine ou non ceux qui agissent. C'est le même esprit qui, aujourd'hui encore, anime les théoriciens de la décision : ils savent qu'on ne greffe pas une lumière d'emprunt sur l'esprit obscurci : les modèles de la décision rationnelle sont aussi difficiles à incorporer du dehors à des agents que de leur greffer un foie ou un cœur. Les phénomènes de rejet, quoique différents, sont tout aussi importants. En effet, la raison, au sens de la reconnaissance et de l'acceptation lucides des liens nécessaires entre les choses, heurte trop nos intérêts, nos désirs, nos peurs pour ne pas susciter en nous une hostilité compacte, quoique déguisée. Il est facile d'être à la fois rationaliste et naïf ; il est plus malaisé de croire en la raison, quand on mesure combien il est difficile – et même héroïque – d'agir selon la raison.

C'est pourquoi, au XVII<sup>e</sup> siècle, des penseurs comme Bossuet et Fénelon, responsables des princes de France, insistent sur la nécessité, pour l'homme d'action, de réduire son moi, au sens du « *transcede et teipsum* » augustinien. Bossuet note : « L'esprit du prince doit être une glace nette et unie, où tout ce qui vient, de quelque côté que ce soit, est représenté comme il est, selon la vérité. Il est dans un parfait équilibre ; il ne se détourne ni à droite ni à gauche. »<sup>2</sup>

Tel est le climat intellectuel et spirituel dans lequel est née la « Géométrie du hasard », dont Pascal dit avec exaltation, dans son « Adresse à la très célèbre Académie parisienne » : « *Sic matheseos demonstrationes cum aleae incertitudinem jungendo, et quae contraria videntur conciliando, ab utraque*

1. Saint Augustin, *De Vera Religione*, IV, XXXIX, 72.

2. Bossuet, *Politique tirée des propres paroles de l'Écriture sainte, Œuvres choisies*, éd. Lanson, 1897, p. 254.

*nominationem suam accipiens, stupendum hunc titulum jure sibi arrogat: aleae Geometria.* »<sup>1</sup>

Ces allusions à des textes anciens n'ont pas d'autre intérêt que de mettre en relief quelques caractères paradoxaux de l'action, qui se heurte, comme le signale Pascal, à la nécessité de « joindre les démonstrations des mathématiques aux incertitudes du hasard » et de « concilier ce qui paraît contraire ». Du même coup, il n'est pas facile de dire ce que l'on peut mettre sous l'expression d'« action complète », puisque l'agent est affecté, du fait de sa condition d'homme, d'une labilité essentielle : ses résolutions les plus fermes rencontrent l'incertitude et le risque ; toute action élémentaire est prise dans un réseau d'interactions qui en modifie la portée et le sens.

## 1 | LA COMPLÉTUDE INTERNE DE L'ACTION

Et pourtant, les hommes qui ont réfléchi sur l'action, et qui ont eu la possibilité de comparer entre elles des actions appartenant à une même série et accomplies par le même agent, signalent une singularité : il n'y aurait pas apprentissage de l'action, amélioration d'une action sur l'autre. L'action, pourrait-on dire, est « complète » en ce sens que l'agent ne note pas un perfectionnement de ses actions, lié à l'expérience.

Cournot et Taine, en réfléchissant tous les deux sur les guerres de la Révolution et de l'Empire, le premier en peignant le maréchal Gouvion Saint-Cyr, le second Napoléon, sont frappés par la constatation identique de leurs héros, qui n'ont pas du tout le sentiment qu'ils aient été de meilleurs chefs de guerre à leur dixième ou à leur vingtième bataille qu'à la première.

Dans la notice biographique du maréchal Gouvion Saint-Cyr, qui figure au début des *Mémoires* de ce dernier, Cournot observe, à propos de la première bataille, la reprise du camp de Nothweiler, le 14 septembre 1793, livrée par Gouvion Saint-Cyr à la tête de la division de Lembach : « On trouve déjà, dans cette première affaire, les traits caractéristiques de son génie guerrier, l'art de faire valoir de faibles moyens et la

1. « Ainsi, joignant les démonstrations des mathématiques à l'incertitude du hasard, et conciliant ce qui paraissait contraire, elle tire son nom des deux et s'arroge à juste titre le nom stupéfiant de "Géométrie du hasard" » (Pascal, *Œuvres*, Pléiade, p. 74).

sagacité dans l'appréciation des causes morales. Aussi l'estimait-il à l'égal de ses plus belles, et généralement il pensait que les hommes se montraient au début de leur carrière militaire tels qu'ils devaient toujours rester. » Il souligne l'aspect de jeu de stratégie de l'action militaire : « A compter de ce jour, les montagnes des Vosges devinrent le théâtre habituel de ses combinaisons : "Saint-Cyr joue aux échecs", disaient ses camarades, quand on entendait le canon de ce côté<sup>1</sup> ; [...] » Cournot, qui fut l'un des premiers à établir que le calcul mathématique peut se substituer au calcul machiavélien dans la conception d'une stratégie, ajoute plus loin : « C'est une chose digne de remarque que le général qui peut-être a fait la guerre avec le plus de calcul, n'avait aucune teinture des sciences exactes, dont on suppose communément que l'étude doit être une partie intégrante de l'éducation militaire. »<sup>2</sup> Et il poursuit : « D'ailleurs, on l'entendait dire souvent qu'il y avait plus à profiter dans l'observation des ruses guerrières des animaux que dans l'étude de la plupart des traités de didactique... »<sup>3</sup> Ici, l'action est complète au sens où Buffon, à propos de la génération, parle de « moule intérieur » : comme si l'homme de guerre découvrait d'instinct le moule dans lequel les actes ou les pensées qui constituent les parties – ou les organes – d'une action doivent être coulés pour que cette dernière arrive à un état de perfection momentanée mais réelle.

De son côté, Napoléon confiait à Las Cases : « C'est une chose bien remarquable, que le nombre de grands généraux qui ont surgi tout à coup dans la Révolution. Pichegru, Kléber, Masséna, Marceau, Desaix, Hoche, etc. ; et presque tous de simples soldats ; mais aussi, là semblent s'être épuisés les efforts de la nature ; elle n'a plus rien produit depuis, je veux dire au moins d'une telle force. C'est qu'à cette époque tout fut donné au concours parmi trente millions d'hommes, et la nature doit prendre ses droits ; tandis que plus tard on était rentré dans les bornes plus resserrées de l'ordre et de la société [...] » Et Napoléon ajoute : « Une autre chose non moins remarquable, c'est l'extrême jeunesse de plusieurs de ces généraux qui semblent sortir tout faits des mains de la nature. »<sup>4</sup>

Ce thème de la complétude de l'action – due à l'instinct ou à la grâce – est une idée que les Romantiques allemands ont développée : ainsi Hölderlin dans *Hyperion* ou Kleist dans *Sur le théâtre de marionnettes*.

1. Cournot, « Notice biographique du maréchal Gouvion Saint-Cyr », in maréchal Gouvion Saint-Cyr, *Mémoires pour servir à l'histoire du Consulat et de l'Empire*, p. 9.

2. *Ibid.*, p. 96.

3. *Ibid.*, p. 98-99.

4. Las Cases, *Mémorial de Sainte-Hélène*, 10 mai 1816.

Racontant l'étrange invitation reçue par un jeune voyageur de se battre à l'épée contre un ours que possédait son hôte, Kleist lui fait dire : « Non seulement l'ours paraît tous mes coups, comme le premier bretteur du monde, mais (chose en quoi nul bretteur au monde ne l'eût imité), il ne se prêtait même pas aux feintes : debout, me fixant dans les yeux comme s'il avait pu lire dans mon âme, la patte levée prête au coup, si mes attaques n'étaient que simulées, il ne bougeait pas. »<sup>1</sup> L'artiste en tire des réflexions sur la complétude organique de l'action : « Nous voyons que, dans le monde organique, plus obscure et plus faible est la réflexion, d'autant plus rayonnante et souveraine s'étend la grâce. »<sup>2</sup> Et Kleist retrouve l'image platonicienne de la marionnette du début des *Lois*, et du dieu : « De même la grâce, quand la connaissance est pour ainsi dire passée par un infini, est de nouveau là ; de sorte qu'elle apparaît en sa plus grande pureté dans cette conformation humaine du corps qui, ou bien n'a aucune conscience, ou bien a une conscience infinie, c'est-à-dire dans le mannequin ou dans le dieu. »<sup>3</sup>

Les derniers mots du récit montrent que, pour Kleist, l'idéal de la complétude subjective – mais aussi organique – de l'action a quelque chose d'inaccessible :

« En sorte, dis-je un peu rêveur, qu'il nous faudrait de nouveau manger du fruit de l'arbre de la connaissance, pour retomber dans l'état d'innocence ?

— Sans nul doute, répondit-il ; c'est le dernier chapitre de l'histoire du monde. »<sup>4</sup>

Entre l'ours de Kleist et les animaux de Gouvion Saint-Cyr, il y a plus d'un trait commun : l'action – ou, plus exactement, sa partie incessible qu'est l'exécution – est avant tout biologique ou organique ; ensuite, il y a dans l'action une forme de grâce qui implique un passage par l'infini, c'est-à-dire une sorte d'exténuation, ou de mise entre parenthèses de son « moi » par l'agent. L'ours de Kleist, pourrait-on dire, est leibnizien : attentif aux mille messages ténus que constituent les « petites perceptions », il pare les coups et évite les feintes ; il bénéficie d'une grâce intersubjective, il opère à l'intérieur du réseau d'un *vinculum substantiale* naturel, qui relie, par exemple dans l'assaut, les deux adversaires. Quel

1. Heinrich von Kleist, *Über das Marionetten Theater, Sur le théâtre de marionnettes*, trad. franç. et présentation par Roger Munier, La Traversière, 1981, p. 63.

2. *Ibid.*, p. 63.

3. *Ibid.*, p. 65.

4. *Ibid.*, p. 65-66.

homme d'action n'a pas la nostalgie de tels instants parfaits, où, sous l'effet d'une grâce d'état, il ferait jeu égal avec un animal sauvage ?

Mais ce type de complétude subjective devient impraticable, dès que l'hésitation ou la peur figent l'agent. Nos actions sont « incomplètes » en ce qu'elles ne partent pas d'un lieu intérieur qui serait la pièce de commandement. Nous ne décidons pas, parce que nous nous arrangeons pour que nos actions n'aient rien de décisif. On retrouve ici, scruté avec le microscope des Moralistes, le *causas habent deficientes* de saint Augustin. Ici, incomplétude est synonyme de défaillance ou, plutôt, de déficience. A quoi tient cette déficience, au-delà des accidents, des « chutes », aurait dit Camus ? A une raison liée à la constitution de l'être humain : pour agir, il faut rassembler ses forces ; mais les forces de l'individu ne se gardent pas quand celui-ci les entrepose dans son « moi » ; elles se décomposent, se délitent, se perdent. Aucun principe de conservation ne joue dans l'ordre psychique : quelle que soit la résolution affichée par un homme, on peut toujours essayer de le retourner, de l'abattre, de l'effrayer. C'est le ressort de la propagande, et l'un des leviers de l'action. D'où le conseil de saint Augustin, répétons-le : « *Et si tuam naturam mutabilem inveneris, transcede et teipsum.* »

Pour remédier à la déficience naturelle de l'homme, les moralistes – et Descartes ne fait pas autre chose – cherchent dans l'action elle-même les remèdes aux déficiences de l'action, en faisant fond sur un trait positif de la nature humaine, à savoir que « l'action elle-même, comme le dit Simone Weil, constitue le plus puissant des appels à l'action et le stimulant le plus irrésistible »<sup>1</sup>.

Jusqu'ici, nous avons réfléchi sur le caractère de complétude ou d'incomplétude de l'action en partant d'un constat banal : une action ne prend corps que si des individus acceptent de s'engager, de revendiquer la paternité d'une entreprise, d'assumer des échecs, parfois la mort de certains de leurs compagnons, etc. De fait, quand l'action touche à la vie et à la mort, le plus difficile n'est pas d'imaginer ce qu'il faudrait faire ni même de l'accomplir, mais c'est d'assumer un malheur qui, dans une certaine mesure, est inséparable de l'action. Simone Weil, dans *L'Iliade ou le poème de la force*, a mis en évidence ce fait ; et les tragédies du XX<sup>e</sup> siècle nous ont donné d'amples occasions de vérifier la réalité du malheur, au cœur de l'action humaine, y compris quand elle a pour objet de le conjurer. Ainsi s'explique le paradoxe central de l'action : puisque, si l'on en croit saint Augustin, la volonté est à l'œuvre même dans les entraînements et les passions, nous ne cessons d'agir ; inversement, puisque nous sommes

1. Simone Weil, *Écrits de Londres*, Gallimard, p. 114.

incapables de nous lancer entièrement dans ce que nous décidons, nous découvrons la rareté de l'action.

Sans le dire, nous faisons comme si nous admettions le postulat que la nature humaine est stable, que nos descriptions ont une valeur intemporelle. Examinons donc si la complétude de l'action est liée à la croyance en la stabilité de la nature humaine. Malebranche pensait avoir établi qu'on ne peut pas constituer une science de l'homme, comme on a fait une physique ou une biologie, car, observait-il, nous n'avons pas d'archétype de notre être. Si ce constat est juste, il laisse le choix entre plusieurs attitudes : suivre Montaigne et les moralistes français, c'est-à-dire consentir à la précarité de l'homme, sans renoncer à chercher dans les maximes ou les règles de la conduite le substitut d'invariants inexistantes dans la nature ; réfléchir sur l'action proprement dite, et voir si elle ne comporte pas une sorte de stabilité – et peut-être même de perfection – interne ; ou, enfin, adopter la position de Goethe, de Schleiermacher, de Dilthey, de Simmel et de Max Weber, mais aussi de Freud. Celle-ci garde la double idée goethéenne du *Bildungsroman* et des *Urtypen*, équivalents humains de l'*Urpflanze* ; et elle essaie d'articuler Forme [Types idéaux] et histoire [au sens d'expérience vécue ou *Erlebnis*].

Montaigne et les moralistes français ne postulent pas la stabilité de la nature humaine. Bien au contraire, ils peignent un homme inconstant, fragile, précaire. Ils ne voient pas là un obstacle à la description de la condition humaine ni à l'énoncé d'universaux concrets. Montaigne va plus loin, quand il se gausse d'Épicure postulant la robustesse de l'appareil sensoriel de l'homme. Pourtant, il y a bien, dans l'enquête sur l'homme des moralistes, la recherche d'invariants. Ces invariants, à leurs yeux, sont de deux types : d'abord, des principes d'explication de la variabilité humaine – par la diversité des circonstances où se passe l'existence ; par l'instabilité de l'humeur et le jeu des passions ; par le hasard ; en ce sens, la stabilité n'apparaît pas comme un postulat nécessaire. C'est dans cette tradition, en effet, qu'a été conduit l'appropriation mathématique du hasard ; en second lieu, des maximes de l'action, comme on le voit dans la correspondance entre Pascal et Fermat sur les jeux de hasard, dans les *Pensées* de Pascal relatives au pari ou dans les réflexions sur les jeux où intervient l'habileté des joueurs.

Ainsi, dans la tradition des moralistes lecteurs de Montaigne et de Pascal, la stabilité de la nature humaine n'est pas postulée : la variabilité de la nature humaine étant constatée, les invariants sont cherchés dans le cœur, c'est-à-dire dans la liberté ou la volonté d'une part, dans les règles conventionnelles [du jeu ou du droit] de l'autre.

Dans la perspective dramatique – celle de la tragédie grecque, du théâtre de Shakespeare et du roman anglais –, l'accent est mis sur l'action. Celle-ci, note Aristote, importe plus que le tempérament ou le caractère des agents. En effet, c'est l'action qui révèle à chacun son caractère. La question se pose alors de savoir si les structures de l'action restent inchangées, si divers que soient, pris individuellement, les agents ou les décideurs. A cette question, je suis, pour ma part, tenté de répondre par l'affirmative. Il y a une sorte de stabilité paradoxale de l'action. Alors qu'on attribue à l'action le pouvoir de faire surgir dans le monde – organique comme historique – ce qui est nouveau, celle-ci présente, aux yeux de l'observateur, deux propriétés curieuses : l'action est rare ; et l'action est stable. Dans ce cas, la complétude de l'action désigne la conformité de l'action concrète à un type idéal, ou, comme nous l'avons dit, à un « moule intérieur ».

Cournot découvre une telle conviction chez le maréchal Gouvion Saint-Cyr : « Intimement pénétré des difficultés de son art, il le plaçait sans hésiter au-dessus de tous les autres, et il ne pensait pas que ni l'étude, ni une longue expérience pussent ajouter beaucoup aux qualités guerrières que la nature donne seule. »<sup>1</sup> Et Napoléon, nous l'avons vu, pense de même. Ainsi, l'action réussie a une perfection interne telle que ni l'agent ni les observateurs ne peuvent se figurer qu'elle aurait pu être différente de ce qu'elle a été.

Du grand portrait de Napoléon Bonaparte qu'il a brossé dans *Les origines de la France contemporaine*, Taine tire les mêmes conclusions, mais là où Cournot parle de « nature », il parle d'« imagination constructive ». Évoquant les décisions de son héros, il note : « Avant d'agir, il a choisi son plan, et, s'il a choisi ce plan, c'est entre plusieurs autres. Derrière chaque combinaison adoptée, on entrevoit la foule des combinaisons rejetées... »<sup>2</sup> Il ajoute : « Certainement, parmi ses diverses facultés, si grandes qu'elles soient, celle-ci, l'*imagination constructive*, est la plus forte. »<sup>3</sup> De ces pages étonnantes sur la psychologie de Napoléon, on retire deux impressions opposées : d'un côté, comme celui-ci le dit lui-même, il « travaille toujours » ; de l'autre, il atteint dans l'action, d'emblée, une espèce de perfection, comme si, justement, l'action ne s'apprenait pas graduellement. « Sa mémoire topographique et son imagination géographique des contrées, des lieux, du terrain et des obstacles aboutissent à une vision interne qu'il

1. Cournot, « Notice biographique du maréchal Gouvion Saint-Cyr », in maréchal Gouvion Saint-Cyr, *Mémoires pour servir à l'histoire du Consulat et de l'Empire*, p. CVI.

2. Hippolyte Taine, *Les origines de la France contemporaine (1875-1893)*, Robert Laffont, coll. « Bouquins », t. 2, p. 393.

3. *Ibid.*, p. 394.

évoque à volonté et qui, après plusieurs années, ressuscite en lui aussi fraîche qu'au premier jour. Son calcul des distances, des marches et des manœuvres est une opération mathématique si rigoureuse que plusieurs fois, à deux ou trois cents lieues de distance, sa prévision militaire, antérieure de deux mois, de quatre mois, s'accomplit presque au jour fixé, précisément à la place dite. Ajoutez une dernière faculté, la plus rare de toutes ; car, si sa prévision s'accomplit, c'est que, comme les célèbres joueurs d'échecs, il a évalué juste, outre le jeu mécanique des pièces, le caractère et le talent de l'adversaire, "sondé son tirant d'eau", deviné ses fautes probables ; au calcul des quantités et des probabilités physiques, il a joint le calcul des quantités et des probabilités morales, et il s'est montré grand psychologue autant que stratéliste accompli. »<sup>1</sup>

Ces deux exemples, portant sur la même époque, et concernant des hommes dont l'un au moins fut extraordinaire, fournissent, par un effet de grossissement, une image plus palpable de l'action proprement dite. La peinture de Napoléon par Taine a d'ailleurs tellement frappé Duhem qu'il en a étendu les résultats à la psychologie de l'invention scientifique<sup>2</sup>.

De ces regards sur l'action, on retire l'impression que celle-ci, paradoxalement, est comme immobile, en raison même de sa perfection interne (*innere Vollkommenheit*, comme le dit Einstein des théories de principe). Cette unité entre la décision et l'exécution – que les auteurs et les observateurs des « grandes actions » ont souvent notée – a été décrite avec exactitude par Hölderlin dans *Hyperion* :

« Et, après un silence, Hypérior de demander :

— Y a-t-il place pour la satisfaction entre la décision et l'exécution ?

— C'est le repos du héros, répond Diotime. Il y a des décisions qui sont, à la fois, comme des paroles divines, ordre et accomplissement : la tienne est de celles-là. »<sup>3</sup>

La décision parfaite apparaît à son auteur comme à la fois libre et nécessaire, spontanée et dictée, surgie du dedans et pourtant impersonnelle.

Citons encore Napoléon : il aurait objecté – comme le fera Bergson –, à cette idée romantique de complétude interne de l'action, qu'il est exceptionnel que celle-ci sourde de notre moi entier : « On vous dit encore que, *quand on connaît le caractère d'un homme, on a la clef de sa conduite* ; c'est faux : tel fait une mauvaise action, qui est foncièrement hon-

1. *Ibid.*, p. 387-388.

2. Pierre Duhem, *La théorie physique. Son objet - sa structure* [1906], Vrin, 1981, p. 81 sq.

3. Hölderlin, *Hyperion*, Insel Verlag, 1979, t. 1, p. 113, trad. Pierre Bertaux.

nête homme ; tel fait une méchanceté sans être méchant. C'est que presque jamais l'homme n'agit par acte naturel de son caractère, mais par une passion secrète du moment, réfugiée, cachée dans les derniers replis du cœur. Autre erreur quand on vous dit que *le visage est le miroir de l'âme*. Le vrai est que l'homme est très difficile à connaître, et que, pour ne pas se tromper, il faut ne le juger que sur ses actions ; et encore faudrait-il que ce fût celles du moment, et seulement pour le moment. »<sup>1</sup>

La première singularité de l'action, c'est donc qu'elle soit si peu répandue. Alors que, dans *L'évolution créatrice*, Bergson déclare : « Il n'y a pas de choses, il n'y a que des actions »<sup>2</sup>, l'action complète ou parfaite, dans laquelle la décision et l'exécution se fondent, semble si rare qu'il y a probablement bien des hommes qui, de leur vie entière, n'ont pas accompli une seule action.

En effet, l'action introduit une rupture dans le cours du temps et dans la destinée d'un être. En idée, agir veut dire quitter le monde dans lequel on vivait jusque-là et se quitter soi-même. Il y a, dans le principe même de l'action, sacrifice et promesse : même si on est surchargé d'habitudes, de tâches, d'opinions, on se déclare prêt à les congédier, à les sacrifier. Cet élément sacrificiel, ou du moins ascétique, est dominant dans la conception chrétienne de l'action ; mais il a peut-être un caractère universel, même si nous croyons que la plupart des changements auxquels sont soumis les vivants sont graduels. En revanche, l'impossibilité d'oublier, de se détacher, de contracter d'autres liens, en un mot, la fixité structurelle ou le caractère répétitif du comportement entraînent une rareté de l'action.

Cette rareté de l'action, en effet, tient d'abord à des causes internes. Elle peut aussi résulter d'un manque d'occasions. Comme le note La Bruyère : « Le génie et les grands talents manquent souvent, aussi, les seules occasions : tels peuvent être loués de ce qu'ils ont fait, et tels de ce qu'ils auraient fait. »<sup>3</sup> Nous n'envisageons ici que les entraves internes à l'action.

Au XVII<sup>e</sup> siècle, l'étude de ces entraves internes conduit Fénelon et Bossuet à la conclusion que l'action, pour arriver à l'état le moins imparfait que la condition humaine l'autorise, doit s'accompagner, nous l'avons vu, d'un amenuisement délibéré du « moi ». Par là, l'une des causes majeures de la déficience de l'action, à savoir la fragilité du « moi », se trouve réduite. En effet, comme Bossuet le précise, « le contraire d'agir par raison, c'est agir par passion ou par humeur. Agir par humeur, ainsi

1. Las Cases, *Mémorial de Sainte-Hélène*, 20 mai 1816.

2. Bergson, *L'évolution créatrice* (1907), PUF, coll. « Quadrige », 1981, p. 249.

3. La Bruyère, *Les caractères*, « Du mérite personnel », Garnier-Flammarion, 1965, p. 103.

qu'agissait Saül contre David, ou poussé par sa jalousie, ou possédé par sa mélancolie noire, entraîne toute sorte d'irrégularité, d'inconstance, d'inégalité, de bizarrerie, d'injustice, d'étourdissement dans la conduite »<sup>1</sup>.

Par conséquent, il y a bien ici recherche d'invariants de l'action : ceux-ci se trouvent au-delà de l'agitation et de l'humeur, soit dans une « nature » qui donnerait originairement à l'homme d'action des dispositions particulières, soit dans une « raison » qui le rendrait capable de suspendre les effets perturbateurs de la nature.

De toute manière, ce qui fait la stabilité et la perfection interne de l'action, qu'elle soit vue par Bossuet ou par Taine, c'est l'ajustement quasi miraculeux de la décision aux circonstances, que cet ajustement soit l'œuvre du jugement ou l'effet de l'intuition. Ce qui est « stable », ce n'est pas un élément répétable ou généralisable, c'est la rencontre singulière d'un homme et d'un problème. Bref, l'action fait surgir, quand elle est parfaite, une singularité et un universel. Bien sûr, dans la mesure où les hommes participent de la raison, la décision, ainsi entendue, n'est ni opaque ni inintelligible. Quoique unique, elle est intrinsèquement stable, au sens où, filtrée par une sorte de mutation intérieure, elle ne procède plus alors de la partie de notre être dont Malebranche disait : « C'est que l'homme est une machine qui va comme on la pousse. C'est beaucoup plus le hasard que la Raison, qui le conduit. Tous vivent d'opinion. Tous agissent par imitation. Ils se font même un mérite de suivre ceux qui vont devant, sans savoir où. »<sup>2</sup>

En résumé, cette perspective – centrée sur l'action – prend à son compte les leçons des moralistes, mais elle s'attache à contourner l'obstacle de la variabilité de l'homme en construisant une doctrine de l'action qui minimise les effets de la précarité humaine. Par là, ce qu'on discerne, c'est un ensemble de conditions qui facilitent ou entravent la complétude interne ou subjective de l'action.

## 2 | L'ACTION FRAGMENTÉE

Il y a une seconde manière pour l'action d'être incomplète, non plus subjectivement, en raison des limites internes de l'agent, mais objectivement, en raison du caractère parcellaire ou éclaté du comportement des

1. *Ibid.*, p. 250.

2. Malebranche, *Entretiens sur la métaphysique et sur la religion* (V), A. Colin, 1922, p. 122.

individus et des groupes. Ce type d'incomplétude, virtuellement présent dans toutes les actions, est accentué par la structure des sociétés modernes. Pour s'en faire une idée, il suffit de comparer l'action administrative, telle qu'on la concevait en France il y a deux siècles et telle qu'elle se présente aujourd'hui.

a) *Le paradigme napoléonien*

Des livres qui peignent la nature et l'esprit de la décision administrative en France depuis la Révolution et surtout depuis l'Empire, *Les origines de la France contemporaine* d'Hippolyte Taine et *Des institutions d'instruction publique en France* d'A. A. Cournot sont parmi les plus remarquables.

Taine est un peintre de l'action ; les grands portraits qu'il a brossés de Robespierre, de Danton et de Napoléon mettent en lumière les conditions et la nature de la décision stratégique et politique. De son côté, Cournot, en étudiant la décision administrative, met en évidence plusieurs de ses traits modernes : à la suite de Bossuet mais dans une perspective de mathématicien, il établit que la compréhension de l'histoire est identique à la compréhension des jeux de stratégie, dans la mesure où interviennent dans les deux cas l'affrontement et la coopération des volontés d'une part, les rencontres fortuites de l'autre ; il remarque ensuite que l'administration résulte d'un mélange de mécanique et de bon sens ; il montre que les configurations que peut prendre une institution sont en nombre fini, comme les polyèdres réguliers dans le *Timée* (les divers baccalauréats, par exemple, forment un ensemble fini, composé de types peu nombreux) ; enfin, il propose une théorie des révolutions scientifiques, appelant *révolution* le passage d'un mode de description et d'intellection à un autre, c'est-à-dire un changement de clé. Or, en matière de décision administrative, deux siècles après l'instauration du paradigme propre à la France moderne, nous sommes en train de voir émerger un nouveau paradigme.

Trois expressions résument, d'une façon en apparence contradictoire, la figure du grand décideur administratif et politique, selon la tradition qui naît de la Révolution de 1789 et de l'Empire, même si, comme Tocqueville l'a démontré, ce type se forge sous l'Ancien Régime : c'est un « ingénieur politique », doué d'une « imagination constructive », comme le dit Taine, et il manifeste « le tact des circonstances », selon la formule de Mme de Staël. Celle-ci, dans le portrait qu'elle a brossé de Napoléon, insiste sur le caractère extraordinaire et presque monstrueux de l'homme qui, plus que tout autre, a inventé la conception et la pratique de l'admini-

nistration française. Il a ainsi forgé un modèle, un type idéal, qui, conforme dans son esprit à des traditions anciennes, innove pourtant de façon durable.

Ce modèle existe aujourd'hui encore, mais il a subi des transformations et des altérations. Pour les apprécier, il est bon de prendre à témoin A. A. Cournot, qui fut à la fois mathématicien et philosophe, et qui occupa, dans l'administration universitaire, des postes importants, comme inspecteur général des études et comme recteur.

Les analyses de Taine et de Cournot se conjuguent : ils portent la même appréciation sur le génie administratif de Napoléon, le même jugement sur le caractère fondateur de son œuvre. Il existerait donc un *paradigme* de l'action administrative qui donne à cette dernière sa pleine efficacité.

Quels en sont les traits ? Pour les connaître, on peut les prendre deux à deux :

1 / *Unité de la forme et maîtrise des contenus.* — Unité de style ou formelle de l'ensemble de l'administration, symbolisée, sur l'ensemble du territoire, par le préfet ; maîtrise, par le chef de d'administration, non seulement de la forme des procédures, mais du contenu de l'action. Un mot de Napoléon, rapporté par Taine, est à cet égard révélateur : « S'il n'y a personne pour faire de la poudre, je sais en fabriquer ; des affûts, je sais les construire ; s'il faut fondre des canons, je les ferai fondre ; les détails de la manœuvre, s'il faut les enseigner, je les enseignerai. »

Ces deux caractères, d'unité formelle d'une part, de maîtrise du contenu de l'autre, sont en tension, et ne peuvent se concilier que si les contenus ou ne sont pas trop techniques ou, tout en étant techniques, ne sont pas trop variables.

Au cours de l'évolution de l'administration française, le maintien simultané de ces deux exigences s'est révélé de plus en plus difficile. Dès lors, deux stratégies différentes ont été suivies : ou bien, on n'a maintenu que l'unité formelle de l'administration, sous la forme d'une réglementation plus précise et plus tatillonne, et sous la forme d'ingérences particulières de l'autorité supérieure, destinées seulement à signaler qu'elle existe ; mais aussi par la création de l'École nationale d'administration, qui repose sur le postulat de l'unité de l'action administrative, ou, si l'on veut être plus caustique, sur le mythe, comme dit Cournot de l'« évêque universel » ; ou bien, quand le domaine était trop technique, on a admis que, pour sauvegarder l'unité de la forme et du contenu, il fallait que l'autorité administrative fût exercée par des spécialistes : d'où l'existence des

grands corps techniques de l'État. Les exemples traditionnels sont connus (Mines, Ponts & Chaussées, Eaux & Forêts, Arsenaux, etc.), mais il en est de plus récents : CEA, Électricité de France, etc.

2 / *Variabilité des contenus de l'action et stabilité de son cadre.* — Au cours du temps, l'action administrative s'est diversifiée et étendue ; mais on s'est efforcé de maintenir aussi stables que possible son cadre et ses procédures.

Le respect simultané de ces deux exigences implique que l'évolution des contenus soit graduelle, sans ruptures brutales, et que les pratiques sociales qui forment le domaine où l'action administrative s'exerce soient elles-mêmes relativement fixes.

Il est clair que des évolutions considérables se sont produites, surtout depuis les années 1920-1930, car, même dans les pays démocratiques, les interventions de l'administration se sont diversifiées et amplifiées ; la Seconde Guerre mondiale a accéléré considérablement ce processus.

Du même coup, le maintien d'un cadre stable devient un défi, et bien souvent une fiction : un défi, car il est difficile d'imaginer un cadre de l'action administrative assez simple et assez souple pour qu'il reste stable malgré l'évolution des contenus (on y parvient si mal que la réglementation devient pléthorique, nuisant du même coup à la simplicité et à la compréhensibilité de l'action) ; une fiction, parce qu'on procède, en fait, comme si les pratiques étaient variables ou même indéterminées.

Une nouvelle figure, au sein de l'administration, apparaît, celle de l'expert ; et une nouvelle forme d'action administrative voit le jour : l'administration de mission. Au fond, c'est la figure de l'*ingénieur politique* qui s'estompe, au profit de celle de l'expert.

3 / *Distinction entre action administrative et action de gouvernement.* — Cournot remarque : « Autre chose est une institution de gouvernement, autre chose est une institution administrative. Nous vivons dans un temps où l'administration, prise dans son sens le plus général, va en se perfectionnant sans cesse, comme les sciences, comme l'industrie : car ce n'est qu'une science appliquée au mécanisme social, un agencement industriel de toutes les pièces, de tous les rouages de ce mécanisme. [...] Voilà le fait général, dominant ; et les conséquences à en tirer pour notre sujet se présentent d'elles-mêmes : car une institution telle que la nôtre [*s.e.* l'instruction publique] peut être considérée, tantôt, dans le sens le plus modeste, comme le mode d'organisation d'un service public, et à ce titre

comme une simple institution administrative, tantôt comme un moyen d'agir sur les mœurs, sur les idées des générations successives, ce qui évidemment est le caractère des institutions de gouvernement. »<sup>1</sup>

Comme il l'observe, « supposé que le régime ne change pas ou n'éprouve que des changements lents, on conçoit un état de choses où l'administration des intérêts et des services publics dépendrait à peine de la politique, où l'action du moteur politique, toujours nécessaire, se réduirait presque à donner cette chiquenaude de Descartes. »

Or plus la société devient mouvante ou fragile, plus il devient malaisé de délimiter ce qui incombe à l'administration et ce qui relève, comme le dit encore Cournot, des « institutions de gouvernement », qui agissent « sur les mœurs, sur les idées des générations successives ». Ainsi, comme le notait déjà Cournot, l'instruction publique ou, comme nous disons aujourd'hui, l'Éducation nationale comporte ces deux aspects : elle est « une branche importante des services publics » ou elle est un moyen privilégié de changer la société elle-même, d'influer sur le destin des citoyens, auquel cas elle est bien une « institution de gouvernement ». Les frontières, qui semblaient claires il n'y a pas si longtemps, sont devenues, au cours des vingt dernières années, de plus en plus indistinctes et poreuses.

4 / *Tension entre le caractère « scientifique » et le caractère « politique » de l'action administrative.* — Le paradigme napoléonien, comme le montrent aussi bien Taine que Cournot, tend à conférer à l'action administrative le statut d'une science appliquée, ou d'une industrie, sur laquelle règnent des sortes d'ingénieurs politiques, des spécialistes de la mécanique de l'action publique. La science de référence est toujours la mécanique, mais une mécanique appliquée, dont on ne doute pas qu'elle relève, en dernière instance, de la mécanique rationnelle. Les métaphores utilisées sont parlantes : « Nous vivons dans un temps où l'administration, prise dans son sens le plus général, va en se perfectionnant sans cesse, comme les sciences, comme l'industrie : car ce n'est qu'une science appliquée au mécanisme social, un agencement industriel de toutes les pièces, de tous les rouages de ce mécanisme. »

De même, nous l'avons vu, Taine décrit ainsi l'action de l'ingénieur politique : « Dans chacune des machines humaines qu'il construit et qu'il manie, il aperçoit d'un seul coup toutes les pièces, chacune à sa place et dans son office, les générateurs de la force, les organes de la transmission,

1. A. A. Cournot, *Des institutions d'instruction publique en France* [1864], Vrin, 1977, p. 334-335.

les engrenages superposés, les mouvements composants, la vitesse résultante, l'effet final et total, le rendement net ; jamais son regard ne demeure superficiel et sommaire : il plonge dans les angles obscurs et les derniers fonds, "par la précision technique de ses questions", avec une lucidité de spécialiste, et de cette façon, pour emprunter un mot des philosophes, l'idée chez lui se trouve *adéquate à son objet...* »

La métaphore de l'ingénieur politique implique que l'action a un caractère énergétiste et mécanique, et que son aspect « communicationnel » ou débattu est secondaire. Il y a des structures, des mécanismes, de l'énergie qui ressemble à celle des machines. L'énergie ou plutôt la canalisation de l'énergie appartient à l'autorité supérieure, les organes d'exécution n'ayant en principe que le privilège de la docilité. La réalité est plus subtile. Les contre-pouvoirs existent.

5 / *La nature et le style du commandement.* — Le modèle napoléonien est militaire et, si j'ose dire, « ignatien » : chacun, à sa place, doit exécuter les ordres, et tenir la position qu'il occupe, et à laquelle son *mérite* l'a fait accéder ; mais la soumission qui est attendue est *formelle* : elle ne met pas en cause le jugement intérieur ; elle n'implique pas une adhésion idéologique, mais plutôt une sorte de loyauté individuelle et aussi collective, au sein d'un corps de fonctionnaires donné.

On voit que deux formes de commandement existent dans un tel système : l'administratif, qui repose sur des règles stables et connues ; le politique, qui repose sur des règles singulières et liées aux circonstances. Taine et Cournot insistent, l'un comme l'autre, sur ce point : le premier cite Mme de Staël, observant que les discours de Napoléon « indiquaient le tact des circonstances » ; et le second évoque « le tact de l'homme d'État qui apprécie les circonstances, qui sent quand il faut gagner les esprits par la douceur, et quand il faut leur imposer par l'autorité ; qui se rend compte de ce qu'on peut obtenir de la multitude par la force matérielle et par la force morale, par la prudence et par l'audace » ; et il ajoute que « ce tact est tout autre chose que la science de l'économiste, de l'administrateur et du juge ».

En réalité, dans le modèle napoléonien, la distinction n'est pas aussi tranchée, et l'on constate que, dans un même cadre administratif, coexistent des styles très différents d'administration, liés à la personnalité des responsables d'une part, mais aussi au *génie des lieux* de l'autre.

b) *Émerge-t-il, aujourd'hui,  
un nouveau paradigme de l'action administrative ?*

Observons d'abord que le paradigme classique français comporte des traits simples et durables, qui ont des chances de subsister, même si un nouveau paradigme apparaît. Ces caractères touchent, pour les uns, à la nature du commandement et pour les autres à la nature des sociétés modernes.

1 / *Organes collectifs et responsabilité des individus.* — Cournot observait déjà que, dans l'administration comme ailleurs, il y a une « tentation de synodie »<sup>1</sup>, de gouvernement par des conseils et non plus par des individus. C'est un trait auquel l'esprit du système administratif français est, en principe, rebelle. Sous sa forme contemporaine, il est confronté à quelques problèmes majeurs : gestion des crises ; maîtrise des temporalités différentes ; intelligence des contre-pouvoirs, où les tensions entre individus et groupes se font sentir.

2 / *La gestion des crises.* — Le modèle français d'administration repose – comme sans doute les modèles étrangers comparables – sur un principe de spécialisation des domaines. Or le propre des crises est de faire sauter de telles distinctions et de rendre molles les commandes administratives habituelles. Dès lors, à l'interface entre les divers domaines ordinaires de compétence, du flou apparaît. De plus, la médiatisation des problèmes fait que le responsable administratif doit décider et informer de façon synchrone.

3 / *La maîtrise des temporalités différentes.* — Il est banal de le dire, mais la plupart des actes administratifs obéissent à un calendrier strict. Ils ont une périodicité propre, un rythme cyclique : les décisions administratives qui rythment le cycle annuel de la vie scolaire en sont un exemple. Chaque responsable doit agir à l'instant fixé – qui joue le rôle du *kairos* grec —, sous peine de perdre une année entière. Ce temps périodique est lui-même pris dans une temporalité irréversible, elle-même rythmée – ou chaotiquement ponctuée – par des événements politiques, sociaux, etc. Une des grandes difficultés de l'administration moderne est la réalisation d'une concordance entre ces temps multiples – pour les uns cycliques, pour les autres

1. Cournot, *Des institutions d'instruction publique en France*, p. 141.

linéaires, pour d'autres irréguliers – que la décision doit tenter de mettre en harmonie. Comme l'observait Bossuet : « Si toutes choses dépendent du temps, la science des temps est la vraie science des affaires... »<sup>1</sup>

4 / L'intelligence des contre-pouvoirs. — Pour tout décideur administratif, le pouvoir qui n'est pas le sien apparaît, à la limite, comme un contre-pouvoir. Non que les autres pouvoirs – qui sont aussi légitimes que le sien – lui semblent inamicaux, mais, dans la vie d'un organe administratif, plusieurs phénomènes interviennent : chacun défend ses compétences ; chacun dépend des autres pouvoirs et se trouve capable d'entraver ou de contrer les décisions des autres ; un mélange de coopération et de conflit, en un mot, marque, dans une démocratie (et sans doute, tout autant, dans un régime totalitaire) la vie administrative. C'est vrai, bien évidemment, dans les relations entre administration et syndicats. Mais c'est plus général.

De ces remarques, une conclusion évidente se dégage : dans les sociétés modernes, chaque agent n'agit jamais sur la totalité d'une entreprise, même quand il y participe librement. Comment, dans un tel contexte, retrouver alors un sentiment de complétude ? Sera-t-il uniquement subjectif ou existe-t-il des moyens – institutionnels, éthiques, scientifiques même – susceptibles de restituer à l'action une part de sa complétude rêvée ou attendue ?

### 3 | LE CONCEPT COSMIQUE D'ACTION

L'idée de la complétude de l'action est nécessaire en ce sens que l'action nous soustrait au lourd dépôt des habitudes incrustées. L'action, nous le sentons, devrait être comme la venue au monde d'un être nouveau ; la prononciation d'une parole qui serait indissolublement une vérité et une opération. Ce ne serait pas une complétude d'autonomie, de distance ou de retrait, mais une liberté d'avènement et d'adhésion. Le « moi », amenuisé, y serait à peine présent ; rien, pourtant, du contenu de l'action ne serait indifférencié. Nous avons tous la nostalgie d'une action qui, tout en récapitulant notre histoire, nous libérerait du passé, nous ferait entrer dans un avenir vierge.

1. Bossuet, Politique tirée des propres paroles de l'Écriture sainte, in *Extraits des œuvres diverses*, par G. Lanson, Delagrave, 1899, p. 252.

Il est facile de qualifier une telle aspiration d'illusoire ou de mythique ; en réalité, elle recèle une espérance qui, depuis toujours, habite l'action. Comme le dit Héraclite : « S'il n'espère pas, il ne trouvera pas l'inespéré, car il est hors de quête et sans accès. »<sup>1</sup> Une action n'est pas une œuvre ; son accomplissement l'inscrit dans les êtres et dans les choses ; son effet ne revient vers l'auteur que par accident ; la satisfaction est à trouver dans l'action elle-même, non dans ses fruits. Irréversible dans sa structure, l'action reste un événement, une singularité que le succès dissout aussi sûrement que l'échec. L'auteur d'une action n'atteint un état de complétude que s'il accepte entièrement, sans restriction mentale, de se livrer au génie de l'action, qui, comme la Nature, consomme les êtres pour accomplir ses desseins. L'action comporte donc inséparablement sacrifice et accomplissement : sa complétude ne va pas sans l'effacement de l'agent. Jean Cavallès a traduit ce constat, en confiant à Raymond Aron, en 1943, lors d'une mission à Londres : « Je suis spinoziste, je crois que nous saisissons partout du nécessaire. Nécessaires les enchaînements des mathématiques, nécessaires même les étapes de la science mathématique, nécessaire aussi cette lutte que nous menons. »

Cet amenuisement du « moi » est aussi pour l'agent une révélation : à l'instant où il s'insère librement dans le tissu de la nécessité, il découvre une sorte de joie ou de légèreté spinoziste. Il peut se dire : « *Sentimus experimurque nos aeternos esse.* » Et pourtant, cette joie est subordonnée à une combustion des scories internes, et, dans certains cas, au sacrifice.

Or, la peur de la destruction et de la mort est inhérente à l'être humain. Elle hante l'action. Non seulement nous avons peur de mourir, mais nous redoutons que les Érinyes se vengent sur nous d'avoir défié leur puissance. L'action tire sa complétude de la liberté qui la soutient. Mais agir librement, c'est s'arracher à un état ; que cet état soit de servitude ne change rien à l'affaire. Il faut que, dans l'instant, la négation soit entière et l'adhésion soit entière ; que je m'arrache à ce que j'étais et que je fasse être celui que je ne suis pas. En un sens, rien n'est plus naturel, s'il est vrai, comme le dit Héraclite, que nous vivons dans un monde « nouveau chaque jour ». Et pourtant, cette complétude sacrificielle nous effraie. C'est pourquoi, nous l'avons vu, il y a une rareté de l'action. Chacune de nos actions, si nous y sommes attentifs, a, dans son dessein et dans son inspiration, quelque chose de radical ; mais chacune de nos actions, par une sorte de déficience irréductible, reste comme à mi-par-

1. Héraclite, *Les fragments*, trad. franç. par Roger Munier, *Nouvelle Revue française*, n° 436, mai 1989, fragment 18.

cours. En effet, comme le dit un héros de Conrad, « l'homme est une grande chose, mais ce n'est pas un chef-d'œuvre »<sup>1</sup>.

Raisonnement ainsi, c'est faire de l'action humaine un élément des opérations de la nature et voir dans l'artifice l'invention du naturel. Dès lors, la complétude de l'action ne se définit pas dans le cadre de l'anthropologie mais dans celui de la cosmologie. Inscire l'action dans une philosophie de la nature est le rêve que poursuivent ceux qui veulent « naturaliser » la science, la psychologie, la morale, la culture. Par là, ils renouent avec la théologie naturelle ou avec la *Naturphilosophie* du Romantisme allemand. Un tel programme est plus austère qu'on ne le croit : il fait de la Nature l'arbitre de nos entreprises, de leur réussite, de leur échec et même de leur vérité.

Mais ce programme a aussi une ambiguïté foncière : car agir signifie-t-il entrer dans le flux de choses, se laisser brasser, rouler et porter par lui ? Ou bien, tout au contraire, accepter d'être étranger aux choses avant de décider d'entrer dans leurs flots ? L'action instaure et révèle, entre l'agent et la nature, un *dualisme opératoire*, et peut-être même ontologique. Agir, c'est faire entrer nos artifices parmi les productions de la nature, les rendre indiscernables des opérations de la nature. Comme l'écrivait Marcelin Berthelot en 1860 : « La Chimie crée son objet. Cette faculté créatrice, semblable à l'art lui-même, la distingue essentiellement des sciences naturelles et historiques. »

Toutefois, ce qui rend un tel résultat possible, c'est l'incomplétude de la nature : celle-ci ne produit pas la totalité des objets et des êtres qui seraient réalisables selon ses lois. Si bien que l'accomplissement même de l'action requiert une sorte d'inachèvement de la nature ; la capacité qu'elle nous laisse d'entremêler nos desseins à ses réalisations. Si les traces de l'action individuelle sur la nature en général s'effacent aussi vite que les pas du promeneur sur la grève, il n'en va plus de même des grandes entreprises humaines : non seulement elles laissent sur la nature une marque durable, mais elles sont susceptibles d'infléchir son cours. Comme si l'impact de l'action humaine sur la nature dépendait du caractère inachevé ou maniable de cette dernière. Comment, s'il en est ainsi, rendre l'action consonante au génie de la Nature ?

La réponse à cette question, c'est que nous devons nous doter d'un concept « cosmique » de l'action. Par « concept cosmique » ou *Weltbegriff* de l'action, nous entendons l'idée de l'action humaine envisagée sous le

1. Stein, dans *Lord Jim*.

double aspect de la « complétude » interne et de son inscription dans le réseau des causes naturelles ou artificielles.

L'action, nous l'avons vu, peut être examinée à un double point de vue : par rapport à l'agent, afin de jauger s'il a voulu et exécuté son action avec toute la plénitude et la liberté dont il était capable ; par rapport à la nature, afin d'évaluer si l'agent est parvenu à inscrire heureusement son projet dans le réseau complexe des causes.

Considérer ainsi l'action, c'est dire qu'on ne peut se contenter d'en élaborer un concept formel ou « scolastique », puisque l'essence de l'action dépend des lois, de la structure et de la nature du monde dans lequel elle s'accomplit.

Une action, par exemple, n'aura pas du tout la même structure selon que le monde où elle a lieu est l'univers tel que le conçoit Laplace ou l'univers tel que le conçoit Cournot. Dans le premier, la contingence est liée à l'ignorance ; dans le second, le hasard est un « fait naturel ».

Or il semble bien, sinon du point de vue ontologique et ultime, du moins du point de vue épistémique et empirique, que la Nature ressemble à l'univers de Cournot, formé de parties susceptibles de rester durablement autonomes, comme de se rencontrer à un instant qui est, pour les agents concernés, imprévisible. Dans une nature ainsi faite, même si le principe de causalité s'applique universellement, les relations causales changent selon les ordres : les lois de la géométrie ou de la cinématique ne sont pas celles de la physique ; les lois de la chimie sont radicalement différentes de celles de la mécanique newtonienne ; l'ordre biologique a sa spécificité, etc.

Dans un tel univers, l'action est rendue possible par le fait que la nature ne réalise pas tous les êtres, tous les processus ou tous les corps qui seraient réalisables selon ses lois ; en d'autres termes, elle n'est pas saturée, elle est incomplète.

Cette « incomplétude » de la nature ouvre à l'humanité un espace d'action. Pour Cournot, toutefois : « De roi de la création qu'il était ou croyait être, l'homme est monté ou descendu (comme il plaira de l'entendre) au rôle de concessionnaire d'une planète. [...] Il avait à faire valoir un domaine, il a une mine à exploiter : et ces quelques mots suffisent pour indiquer sous quelle face nouvelle vont désormais se présenter les plus graves problèmes de l'économie sociale, ainsi que les conditions de la vie historique des peuples. »<sup>1</sup> En opposant le domaine agricole à la mine,

1. Cournot, *Considérations sur la marche des idées et des événements dans les temps modernes*, p. 422.

Cournot suggère le caractère destructif de l'action humaine, comme l'a noté Georges Canguilhem<sup>1</sup>.

Si la nature est telle, et que les actions, la plupart du temps, mobilisent des enchaînements de causalité appartenant à deux ou plusieurs ordres de la réalité, alors l'agent doit avoir une perception des contraintes et des possibilités que présente l'insertion de ses actions dans ces divers ordres, pour en former, si possible, le « concept cosmique », c'est-à-dire discerner et organiser le cheminement enchevêtré de ses plans au sein de la nature. C'est une entreprise nécessaire et souvent irréalisable, comme la distinction stoïcienne entre les choses qui dépendent de nous et celles qui n'en dépendent pas l'exprime. Napoléon, abordant un jour cette question à Sainte-Hélène, observa : « Et chose bien remarquable, les obstacles qui m'ont fait échouer ne sont point venus des hommes ; ils sont venus des éléments : dans le Midi, c'est la mer qui m'a perdu ; et c'est l'incendie de Moscou, les glaces de l'hiver, qui m'ont perdu dans le Nord ; ainsi, l'eau, l'air et le feu, toute la nature, et rien que la nature, voilà quels ont été les ennemis [...] »<sup>2</sup>

Toute action, dès qu'elle cesse d'être un « jeu », c'est-à-dire une situation modélisable, pour devenir une entreprise qui met en branle des hommes, des dispositifs artificiels et la nature, acquiert une structure complexe qui fait d'elle, selon le mot de Simone Weil, une « composition sur plans multiples ». Du même coup, il y a inévitablement une tension entre la représentation mathématique et le cheminement causal de l'action. Il est rare, en effet, comme l'a montré Cournot, que le cheminement s'accomplisse de bout en bout dans la même région ou le même ordre de la nature.

Or, dès que le processus causal franchit la frontière d'un ordre – l'ordre mécanique ou chimique, par exemple – pour mettre en branle des événements ou des opérations de la nature inhérentes à un autre ordre – disons, biologique –, ce franchissement de l'interface entre deux ordres ne constitue pas, en général, un phénomène graduel ou continu : sans avoir, sur ces problèmes difficiles, l'information dont nous disposons aujourd'hui, Cournot, à propos de la discontinuité entre mécanique et chimie, remarque : « Ainsi, tandis que les physiciens ne tiraient de la mécanique newtonienne que des théories à beaucoup d'égards imparfaites, fragmentaires ou même incohérentes, les chimistes, en suivant d'au-

1. « Émile Littré, philosophe de la biologie et de la médecine », Colloque du Centenaire de la mort de Littré, *Revue de synthèse*, III, n° 106-108, avril-décembre 1982, p. 282.

2. Las Cases, *Mémorial de Sainte-Hélène*, 3 mars 1816.

tres voies, réussissaient à construire un système des mieux liés, une science autonome qui n'emprunte rien au mécanisme, et qui est même, autant qu'on en peut juger, irréductible à la mécanique ; [...] »<sup>1</sup> Poursuivant son analyse par une comparaison entre chimie minérale et chimie organique, il souligne : « La chimie organique se place, si l'on veut, aux confins des deux mondes inorganique et organique, mais sans cesser d'appartenir essentiellement au monde inorganique. »<sup>2</sup>

Tout serait simple et pourrait donner lieu sans peine à des prédictions scientifiques rigoureuses, si, à la jonction entre deux mondes, les relations causales restaient stables et transportaient leurs effets d'un monde à l'autre, sans altération ni variation de puissance. Or il n'en est rien : « Le monde physique [...] est comme la charpente qui supporte le monde de l'organisation et de la vie. Entre ces deux mondes existent des relations d'engrenage et de contact intime plutôt que de greffe ou de soudure ; et plus les observations se précisent scientifiquement, plus on est porté à croire que ce sont bien deux mondes distincts, ayant leurs lois propres, sans que l'on puisse concevoir le passage de l'un à l'autre par voie de développement graduel et de progrès continu. »<sup>3</sup> Cette remarque de Cournot n'a rien perdu de son actualité : quand on fait de l'ingénierie biologique, on tente de définir des règles de prudence, parce qu'on ne sait pas toujours si une manipulation ne va pas activer dans la nature une chaîne de causes et d'effets dont on ne soupçonne pas l'existence, car il s'agit de processus inactivés ou endormis.

C'est pourquoi, dans l'univers de Cournot, le hasard est un « fait naturel » et l'action, tout en étant, sous certaines conditions, modélisable mathématiquement, comporte, dès que sa réalisation implique des processus de causalité appartenant à des ordres ou à des mondes différents, des aléas qui ne tiennent pas à l'inexpérience ou aux erreurs des agents, mais au fait que la nature ne constitue pas un système d'un seul tenant mais qu'elle comporte des ordres, dont chacun est doté de structures et de lois propres.

A cet égard, la structure causale de l'action, entendue selon son concept « cosmique », ressemble à la structure logique de la théorie des choix multicritères : ici comme là, le problème est de trouver ou de construire une loi de composition, mettant en relation des ordres hétérogènes.

1. Cournot, *Matérialisme, vitalisme, rationalisme* [1875], Vrin, 1979, p. 20.

2. *Ibid.*, p. 21.

3. *Ibid.*, p. 53.

## CONCLUSION

Pour évaluer le degré de complétude d'une action, il faut donc recourir à deux étalons différents : l'un, pour apprécier sa complétude interne ; l'autre, pour discerner jusqu'à quel point un plan humain s'inscrit dans la nature. L'essence de l'action ne réside pas seulement dans l'attitude morale de l'agent ou dans la qualité technique de sa stratégie ; elle dépend principalement de l'accomplissement des projets et des plans au sein de la nature et de la société. Telle est l'évidence que souligne le concept cosmique de l'action.

Imprimé en France  
Imprimerie des Presses Universitaires de France  
73, avenue Ronsard, 41100 Vendôme  
Février 1998 — N° 44 705



Yves GRAFMEYER  
*Les gens de la banque*  
 Anne-Marie GUILLEMARD  
*Le déclin du social*  
 Wilhelm HENNIS  
*La problématique de Max Weber*  
 Monique HIRSCHHORN  
*L'ère des enseignants*  
 Albert O. HIRSCHMAN  
*Les passions et les intérêts. Justifications politiques du capitalisme avant son apogée*  
 Christopher JENCKS  
*L'inégalité. Influence de la famille et de l'école en Amérique*  
 Jean-Claude LAMBERTI  
*Tocqueville et les deux démocraties*  
 Dominique MARTIN  
*Démocratie industrielle*  
 Sylvie MESURE  
*Dilthey et la Fondation des Sciences historiques*  
 Sylvie MESURE (sous la direction de)  
*La rationalité des valeurs*  
 Wolfgang J. MOMMSEN  
*Max Weber et la politique allemande (1890-1920)*  
 Serge MOSCOVICI  
*Psychologie des minorités actives*  
 Mancur OLSON  
*Logique de l'action collective*  
 Jean-G. PADIOLEAU  
*L'État au concret*  
*« Le Monde » et le « Washington Post »*  
 Serge PAUGAM  
*La disqualification sociale - Essai sur la nouvelle pauvreté*  
 Patrick PHARO  
*Sociologie de l'esprit. Conceptualisation et vie sociale*  
 Antoine PROST  
*L'enseignement s'est-il démocratisé ?*  
 Dominique RAYNAUD  
*L'hypothèse d'Oxford. Essai sur les origines de la perspective*  
 Hélène RIFFAULT (sous la direction de)  
*Les valeurs des Français*  
 Bertrand SAINT-SERNIN, Emmanuel PICAVET,  
 Renaud FILLIEULE, Pierre DEMEULENAERE  
*Les modèles de l'action*  
 Max SCHELER  
*Problèmes de sociologie de la connaissance*  
 Georg SIMMEL  
*Sociologie et épistémologie*  
*Les problèmes de la philosophie de l'histoire*  
*Philosophie de l'argent*  
*Études sur les formes de la socialisation*  
 Werner SOMBART  
*Pourquoi le socialisme n'existe-t-il pas aux États-Unis ?*  
 Melford E. SPIRO  
*Culture et nature humaine*  
 Jean STOETZEL  
*Les valeurs du temps présent : une enquête européenne*  
 Abram de SWAAN  
*Sous l'aile protectrice de l'État*  
 Pierre-Éric TIXIER  
*Mutation ou déclin du syndicalisme ?*  
 Bernard VALADE  
*Pareto. La naissance d'une autre sociologie*

## ESSAIS

Maurice CUSSON  
*Le contrôle social du crime*  
 Olivier GALLAND (sous la direction de)  
*Le monde des étudiants*  
 José-Guilherme MERQUIOR  
*Foucault ou le nihilisme de la chaire*  
 Maurice de MONTMOLLIN  
*Le taylorisme à visage humain*  
 Antoinette CHAUVENET, Françoise ORLIC,  
 Georges BENGUIGUI  
*Le monde des surveillants de prison*  
 François LACASSE  
*Mythes, savoirs et décisions politiques*  
 Michel ORROT