

## MODÈLE RATIONNEL OU MODÈLE ÉCONOMIQUE DE LA RATIONALITÉ ?

SUIVANT une représentation courante, la théorie économique serait aussi la théorie du choix rationnel pris en général. Elle se caractériserait non par les hypothèses spécifiques qui lui servent à décrire la rationalité, mais par la systémativité et la précision formelle qu'elle met dans cette description. Comme toute affirmation d'équivalence, cette thèse peut se lire dans deux sens. D'une part, elle pose que la théorie économique n'en appelle jamais, en droit tout au moins, à des prémisses explicatives dont le contenu déborde les conséquences logiques du principe de rationalité. Cette première partie de la thèse relève d'une discussion, qui ne sera pas traitée ici, sur la place exacte de l'individualisme méthodologique en économie<sup>1</sup>. Pour la plupart des économistes, cette question est au moins tranchée dans les faits : d'après une idée banale, le corpus doctrinal que désigne assez lâchement l'expression de « théorie économique » n'aurait besoin que d'un

J'ai tiré profit, pour l'élaboration des thèses de cet article, de très nombreuses discussions sur les problèmes de la rationalité poursuivies avec Alan Balloo, Alain Boyer, Pierre-André Chiappori, Bertrand Munier et Pietro Pecari. Ce texte a par ailleurs bénéficié des commentaires des membres du GRASCE et de P. A. Chiappori, qui a bien voulu relire la totalité du manuscrit. Que tous soient ici remerciés.

1. Défendre l'individualisme méthodologique, c'est aussi, classiquement, adhérer à un principe de rationalité : cette démarche revient à considérer comme donnée primaire, qui sert à expliquer, mais n'a pas elle-même à être analysée, la rationalité des comportements individuels. A l'inverse, on peut défendre une forme de principe de rationalité sans pour autant adhérer à l'individualisme méthodologique : c'est ce que propose en substance l'approche *fonctionnaliste* (bien définie dans Merton [1957] ; cf. aussi Harsanyi [1968] ; Watkins [1970]), qui prête directement une finalité aux tous. Cette approche n'a guère de représentants explicites en économie (si ce n'est, peut-être, certaines variantes du marxisme, cf. Elster [1979], p. 34-35), de sorte qu'une discussion sur l'universalité du principe de rationalité en économie porterait sans doute exclusivement sur l'individualisme méthodologique.

seul point de départ méthodologique — le principe de rationalité appliqué aux agents individuels. Si elle présente un intérêt épistémologique général, cette première partie de la thèse n'est sans doute pas de nature à retenir très longtemps l'attention des économistes. Il ne devrait pas en aller de même de la seconde information contenue dans la thèse. Celle-ci affirme aussi, d'autre part, que toute espèce de choix rationnel de l'individu est, en droit tout au moins, analysable par la théorie économique. Entre la conceptualisation par fins et moyens, qui est courante en sociologie et en philosophie, et l'analyse du choix par préférences et contraintes à laquelle procède l'économiste, il n'y aurait de distinction autre que linguistique. Le premier langage serait plus traditionnel, mais aussi plus imparfait que le second, sans avoir pour autant une extension différente : la théorie économique fournirait la traduction claire, appropriée au calcul, et en même temps rigoureusement fidèle, de ce que l'on appelle « rationalité instrumentale » ou modèle fins - moyens<sup>2</sup>. Cette réduction du choix rationnel de l'individu au modèle qu'en donne l'économie est, on va le voir, de nature à diviser les économistes plus vivement que ne le fait la querelle du holisme et de l'individualisme méthodologique.

Tout dépend, naturellement, du sens prêté aux termes : celui de « rationalité instrumentale » n'est pas essentiellement problématique à ce stade<sup>3</sup>, mais celui de « théorie économique », ou de « modèle économique » du choix, doit être rendu explicite. Il nous semble que la théorie de l'utilité espérée, redécouverte par Von Neuman et Morgenstern [1944], puis raffinée et axiomatisée par Savage [1954] et beaucoup d'autres à sa suite, a quelques bonnes raisons de passer pour *le* modèle que la théorie économique propose du choix rationnel. Tel est bien, au demeurant, le point de vue qu'ont endossé des théoriciens de la décision aussi différents que Simon [1966, 1979] ou Harsanyi [1976]. L'accord des deux auteurs sur cette assignation du modèle « économique » est d'autant plus significatif qu'ils en interprètent très différemment l'extension. Récemment, Simon pouvait écrire : « Ce que l'économie peut exporter n'est pas la rationalité, mais une forme très particulière et spécialisée de celle-ci — la rationalité du maximisateur d'utilité, et, encore, celle d'un maximisateur plutôt astucieux (a pretty smart one

2. Les deux occurrences les plus célèbres du modèle fins - moyens sont peut-être : Weber in ([1905], § 1 ; [1913], § 1 et 2 ; [1922-1972], p. 13) ; Parsons, in [1937], chap. II.

3. Il suffit, pour l'instant, de s'appuyer sur une formulation très générale du type : « Les agents disposent de leurs moyens d'une façon qui est appropriée à leurs fins et aux conditions dans lesquelles ils se trouvent. » Cette phrase correspond approximativement à la définition parsonnienne de l'acte ([1973], p. 58).

at that). » ([1979], p. 2.) Harsanyi, en revanche, maintient — avec quelques réserves relatives au statut de la théorie des jeux — le point de vue suivant lequel tout individu rationnel est aussi « bayésien », c'est-à-dire maximisateur d'utilité espérée<sup>4</sup>. Loin de constituer un exemple isolé, ces deux prises de positions nous semblent révéler un clivage de la discipline : tous les économistes, loin de là, ne partagent pas l'idée que « rationnel » veut dire « bayésien », même *en principe*, c'est-à-dire indépendamment des difficultés techniques — non négligeables — qu'il faut affronter lorsque l'on veut révéler effectivement les probabilités subjectives des individus. Ainsi, nous pouvons déceler deux positions clairement antagoniques à l'intérieur même de la discipline — celle du « réductionnisme » bayésien d'une part, celle du « dualisme » de l'autre.

Cet article se propose d'aborder quelques-uns des problèmes méthodologiques traditionnels de la rationalité sous la forme, peut-être moins traditionnelle, d'une critique de la position réductionniste qui repose presque exclusivement sur les idées de H. Simon. Pour mettre en évidence les difficultés de cette position, nous rappelons les concepts simoniens de rationalité « limitée » ou « procédurale » et tentons de montrer, par une argumentation informelle, qu'ils sont irréductibles au modèle de la théorie économique, sans qu'on puisse, par ailleurs, récuser leur pertinence pour l'analyse du comportement rationnel. Notre thèse est donc que le dualisme est inévitable et qu'il faut clairement distinguer le principe de rationalité lui-même, comme modèle *générique*, et les modèles divers, *spécifiques*, qui peuvent se réclamer de lui, mais non prétendre en épuiser le contenu, comme ceux de la rationalité limitée, ou, tout aussi bien, de l'utilité espérée. Cette thèse signifie que nous récusons l'implication : « rationnel » → « économique » ; elle est neutre

4. Dans un premier temps, Harsanyi reconnaît que « en dépit de ses mérites, le concept de rationalité développé par la théorie bayésienne demande à être à nouveau généralisé, car il ne couvre pas adéquatement les comportements rationnels en situations de jeu ». Mais, dans un second temps, il ajoute : « On peut considérer les situations de jeu comme un cas particulier de l'incertitude, puisqu'en général aucun des joueurs ne peut prédire l'issue du jeu, ni même les probabilités associées aux différentes issues. » (1976, p. 96). On se trouverait donc dans le cas de l'« incertitude », par opposition au « risque », suivant la distinction célèbre de Knight [1921]. Or la conception subjectiviste des probabilités regarde comme probabilisables non seulement les situations de risque, mais aussi celles d'incertitude : le domaine d'application du calcul des probabilités serait celui du *pari*, qui peut porter sur des objets quelconques, et non pas le domaine restreint des *fréquences*, comme le prétend l'épistémologie objectiviste traditionnelle (cf. de Finetti [1937]). C'est l'épistémologie dont se réclame Harsanyi : c'est pourquoi il peut, dans un troisième temps, attribuer des probabilités *subjectives* aux agents placés en situation de jeu. Ce dernier procédé est à l'origine de la « tracing procedure » appliquée aux jeux non coopératifs à équilibres multiples (cf. *infra*, p. 19-20).

pour l'implication inverse. Nous ne toucherons donc pas, sauf par accident, à la question de l'individualisme méthodologique<sup>5</sup>.

La première partie de ce travail met en place les deux thèses antagoniques. Nous avons sélectionné trois variantes significatives du *réductionnisme* : la méthodologie aprioriste de von Mises ; celle de Becker et de son école, qui se prétend, au contraire, empiriste ; enfin, et surtout, parce qu'elle est sans doute la plus représentative des options choisies par les économistes, la position « bayésienne » incarnée par Savage, qui s'articule essentiellement grâce à une distribution du positif et du normatif, et dont nous verrons qu'elle est, en quelque sorte, à mi-chemin des deux premières. Pour le *dualisme*, plus brièvement traité dans cette partie puisqu'il constitue la thèse de l'article en général, nous invoquons Allais, Simon, et certains aspects de la théorie des jeux.

La deuxième partie développe une argumentation dualiste à partir des idées de H. Simon. Comme celles-ci ne se sont pas encore déposées dans une théorie au sens propre du terme, mais s'expriment plutôt dans un ensemble de modèles particuliers, nous nous sommes efforcé, dans un premier temps, de rassembler diverses indications importantes de Simon sur son approche de la rationalité : le résultat tient dans une présentation stylisée de cette approche à partir de trois principes de modélisation — la méthode des buts intermédiaires (« subgoals »), le principe de satisfaction et le principe de recherche (« search »). Ce schéma nous permet de revenir sur la question, toujours obscure, des liens entre rationalité « limitée » ou « procédurale » et théorie « classique » (c'est-à-dire, en fait, bayésienne) de la rationalité : nous plaçons la cause de l'*irréductibilité* avec des arguments dont quelques-uns, peut-être, sont nouveaux. Cette position conforte évidemment celle du dualisme.

La dernière partie (en forme de conclusion) revient à des considérations de méthodologie générale, qui seront là esquissées plutôt que développées. Nous y défendons les avantages d'une formulation *triviale* du principe de rationalité, qui tiennent compte de la dénivellation entre l'approche générique du social par la rationalité et les théories ou modèles spécifiques qui s'en réclament (théorie de l'utilité espérée,

5. Sur l'individualisme méthodologique dans les sciences sociales, on se reportera à Boudon [1977] et Boyer [1983], notamment les pages 189-249 qui présentent de façon détaillée le débat anglo-saxon sur le problème. Une application conséquente de l'individualisme méthodologique à l'économie obligerait à réévaluer entièrement certaines procédures traditionnelles : non seulement le raisonnement marshallien en termes d'agent représentatif, mais aussi le choix, comme unités d'analyse micro-économique, d'entités composites (la firme, le ménage), que l'on devrait en toute rigueur analyser à partir d'interactions individuelles ; l'article de F. Bourguignon, dans ce numéro, illustre la difficulté de ce dernier problème.

théorie des jeux, modèles simoniens). Par ailleurs, nous proposons, sans fournir naturellement de réponse définitive, d'ouvrir l'inventaire des justifications traditionnelles du principe de rationalité, et notamment de reconsidérer la fameuse « méthode zéro » de Popper et Weber, à la lumière du dualisme radical défendu dans cet article.

### DUALISTES ET REDUCTIONNISTES DANS LA THEORIE DE LA RATIONALITE

Historiquement, la position réductionniste semble remonter à la variante autrichienne de l'économie néo-classique. Elle apparaît avec une netteté particulière chez von Mises, qui a pu écrire : « L'économie subjective moderne a transformé la théorie des prix de marché en une théorie générale du choix humain. » ([1949], p. 3.) Pour l'auteur de *Human action*, la théorie subjective de la valeur avait fondé deux sciences à la fois : la praxéologie, science du choix individuel en général <sup>6</sup>, et la catallactique, science de la coordination des comportements analysés par la praxéologie <sup>7</sup>. Bien qu'il n'ait connu qu'un état assez primitif de la discipline, von Mises anticipe directement sur la thèse méthodologique qui est celle de nombreux auteurs de l'ère post-savagienne : l'économie ayant résolu de manière satisfaisante la question du choix rationnel individuel, il importe désormais de privilégier soit l'étude de la *coordination* — le rapport entre micro-économie et macro-économie —, soit l'extension du paradigme à des zones du comportement *individuel* qui, pour des raisons historiques, sont encore livrées aux méthodologies holistiques.

La question du dualisme et du réductionnisme n'en était pas une pour les auteurs autrichiens, et il serait vain de rechercher dans leurs écrits une argumentation directe en faveur de la seconde thèse. Toute-

6. Selon von Mises, le choix individuel *humain* est aussi, par définition, un choix rationnel : « L'action humaine est par nécessité toujours rationnelle. L'expression « action rationnelle » est donc un pléonasme et il convient de l'éviter. » (1949, p. 18.) Non content d'assimiler modèle économique de la rationalité et modèle rationnel en général, von Mises identifie donc tous les choix à des choix rationnels. Cette seconde assimilation — facultative pour le point de vue réductionniste — est aujourd'hui défendue par Becker et son école.

7. Strictement parlant, « catallactique » veut dire science de l'échange ; récemment [1976], Hicks a employé le terme dans ce sens, plus restreint que celui de von Mises.

fois, il peut être intéressant de rappeler quel était le mode de raisonnement classique de cette école sur les problèmes de la rationalité : ceux-ci, disent Menger ou von Mises, peuvent se traiter entièrement a priori, par référence seulement à la signification des concepts choisis comme point de départ. Pour von Mises, le marginalisme tout entier découle de l'examen du concept d'action humaine, analytiquement identifié à ceux d'action « rationnelle », « délibérée », et « consciente » ([1949], chap. I et IV). Il n'est pas besoin de faire la moindre hypothèse empirique pour construire la théorie économique du choix. La nature exacte de cet apriorisme économique est très obscure et a souvent été discutée dans la littérature<sup>8</sup>. Nous n'avons pas l'intention d'y revenir ici, mais il peut être utile de souligner ce point : si l'on est un aprioriste radical, au sens de von Mises et Menger, il faut renoncer à vouloir tester les lois économiques. Puisque les lois de la praxéologie sont vraies a priori (indépendamment de toute expérience) et que celles de la catallactique sont censées en découler logiquement, l'ensemble de la théorie économique bénéficie de la même certitude que son point de départ. Il serait vain, par exemple, de faire des tests statistiques sur la loi de demande<sup>9</sup>.

On peut douter que les économistes contemporains veuillent fonder le réductionnisme sur un déni aussi radical de l'expérience. Or il existe une variante moins extrême de l'apriorisme qui peut sembler attirante. La façon dont von Mises ou d'autres ont voulu dériver « logiquement » la loi des utilités décroissantes à partir de la simple considération du choix rationnel est assurément inacceptable<sup>10</sup> ; on peut alors vouloir distinguer entre la notion pure du choix et les *conditions* dans lesquelles s'exerce celui-ci, afin de restreindre la validité de la déduction a priori à celle-là seulement : les conditions, en revanche, seront empiriques. Ainsi, on considérera comme vrai, a priori, que chaque individu possède une fonction d'utilité, en laissant à l'expérience le soin de fixer le signe

8. Récemment en français : Meidinger [1978], Mongin ([1983], chap. I).

9. Tout test réalisé à l'aide de la théorie économique n'est pas vain pour autant : il peut porter sur les *conditions initiales* de l'explicans. Mais aucune loi n'a à être testée, puisqu'elle est non empirique ; ainsi la loi de demande selon Menger ([1883], p. 57-59).

10. Cf. [1949], p. 124, où il est dit que la loi de l'utilité marginale décroissante n'a rien à voir avec la loi psychologique de Gossen et résulte de considérations purement praxéologiques : cette loi « est formelle et a priori, et ne dépend d'aucune expérience ». Sur ce point, comme sur beaucoup d'autres, « Human Action » ne va guère au-delà d'une pétition de principe. On peut noter que Robbins a voulu traiter dans le même esprit — à partir de considérations a priori sur le sens des termes — la loi des rendements marginaux décroissants ([1932], p. 77-78). Sa tentative, plus discursive que celle de von Mises, n'en tourne pas moins court (Agassi, [1971]).

des dérivées secondes. Historiquement, cette position ne paraît pas très éloignée de celle de l'apriorisme anglo-saxon, qui sut combiner le principe autrichien de l'analyse a priori avec un certain empirisme hérité de la tradition millienne <sup>11</sup>.

Pour assouplir encore les thèses de l'apriorisme originel, on peut vouloir réintroduire une autre distinction que récusait explicitement von Mises : celle du normatif et du positif. L'auteur de *Human Action* contestait qu'il y eût un sens quelconque à parler d'action irrationnelle :

L'homme erre souvent dans le choix et l'application de ses moyens. Une action qui n'est pas appropriée à sa fin trompe son attente. Elle est *contraire à son propos, mais rationnelle* (contrary to purpose, but rational) — en d'autres termes, elle est le résultat d'une délibération raisonnable, quoique fautive, et constitue une tentative pour atteindre un certain but, quoique aussi bien une tentative inefficace. ([1949], p. 20.)

La radicalisme des premiers Autrichiens s'exprime non seulement par leur refus de tout empirisme, mais aussi par un certain utopisme : aucune action n'est irrationnelle, elle peut et doit toujours être reconstruite comme conforme à ses buts, relativement à l'information subjective de l'agent. On peut, là encore, récuser cette dimension de la théorie sans la sacrifier tout à fait. Il nous semble que telle est, précisément, la position d'un auteur comme Savage [1954]. Pour lui, l'existence d'une fonction d'utilité von Neuman-Morgenstern découle analytiquement du concept de la rationalité, mais non pas de l'action en général, étant donné que la rationalité a un sens normatif : toute action ne s'y conforme pas immédiatement, bien que, pour un sujet conscient de ce qu'il fait, elle doive s'y conformer <sup>12</sup>. Afin de ne pas introduire d'exigences arbitraires sur les normes, Savage propose, en substance, de recourir au test suivant : un postulat de la théorie est acceptable au point de vue normatif si, lorsqu'il est spontanément violé dans le comportement et qu'on le fait remarquer au sujet, celui-ci modifie sa décision initiale pour la rendre compatible avec le postulat ([1954], p. 7). Cette méthode de *révélation des normes* joue un rôle essentiel,

11. Knight, mieux que Robbins, pourrait bien représenter cette attitude ([1921], [1946]).

12. A propos du postulat P1, qui requiert l'existence d'un préordre total sur l'ensemble des actions, Savage écrit : « Le principal usage que je ferais de P1, et d'ailleurs des autres postulats, est normatif : il me sert à raffiner (« police ») mes décisions dans le sens d'une cohérence plus grande et, lorsque cela est possible, à faire dépendre des décisions compliquées de décisions plus simples. » ([1954], p. 20.)

chez Savage et les bayésiens en général, dans la défense du fameux postulat P 2, ou « principe de la chose sûre », dont la signification pour la rationalité semble tout d'abord indécise<sup>13</sup>.

Comme Savage admet par ailleurs que la théorie de l'utilité espérée a un contenu empirique important — ses articles de 1948 et 1952 écrits avec Friedman visaient à déterminer par l'expérience la forme typique de la fonction d'utilité von Neuman-Morgenstern —, sa position nous semble assez représentative des deux atténuations qu'appelle, simultanément, la conception de von Mises. Plus généralement, les défenseurs modernes du réductionnisme bayésien adhèrent, le plus souvent, à cette position médiane : la théorie de l'utilité espérée a certain contenu a priori, dans la mesure où ses postulats (y compris le fatidique « principe de la chose sûre ») découlent analytiquement de la notion de rationalité ; elle est en même temps empirique dans la mesure où les postulats n'imposent que des contraintes formelles sur le comportement, laissant indéterminée la nature exacte de la fonction d'utilité<sup>14</sup> ; enfin, elle est principalement normative, ce qui ne veut pas dire qu'avec quelque précaution, on ne puisse la faire servir à la prédiction des comportements effectifs, et, donc, à un usage scientifique<sup>15</sup>. Ces affirmations successives paraissent correspondre assez bien aux vues de Harsanyi [1976], qui, tout comme Savage, ne voit aucune contradiction entre la normativité de la théorie et le fait qu'elle permette de dériver des prédictions scientifiques.

Avant de présenter le point de vue dualiste, il faut peut-être signaler une argumentation possible en faveur du réductionnisme qui est moins répandue que celle-ci, mais néanmoins significative. C'est celle de Becker, qui a pu écrire :

13. Cf. [1954], p. 21-24 et 102-104, où Savage se sert de la méthode ci-dessus (appliquée à lui-même !) pour récuser le « paradoxe d'Allais ». Le « principe de la chose sûre » affirme en substance que, si un agent ne préfère pas une action  $f$  à une action  $g$  et qu'il est indifférent entre  $f$  et  $g$  lorsque l'événement B se produit, alors il est vrai qu'il ne préfère pas  $f$  à  $g$  lorsque non B se produit. Il s'agit essentiellement d'un principe de décomposabilité, qui facilite la comparaison des actions deux à deux : si l'on veut comparer  $f$  et  $g$ , il suffira de considérer le sous-ensemble d'états du monde sur lequel l'évaluation des deux actions diverge. Malgré les apparences, ce principe n'est ni tautologique (ce que reconnaît Savage), ni trivial (ce qu'il affirme, en dépit des critiques avancées par Allais, sur la base, notamment, du fameux « paradoxe »).

14. Selon Friedman-Savage [1948], le discrédit dans lequel avait sombré la théorie de l'utilité espérée avant sa réinvention par von Neuman et Morgenstern vient de ce que les économistes, Marshall, par exemple, ne distinguaient pas bien la théorie elle-même de certaines hypothèses contingentes sur l'utilité marginale de la monnaie.

15. Cf. Savage [1954], p. 20 ; Friedman-Savage [1952], p. 463. Nous revenons plus bas sur la question de l'usage prédictif des théories normatives.

L'approche économique est assez compréhensive pour s'appliquer à l'ensemble du comportement humain, que ce comportement fasse intervenir des prix monétaires ou des prix fantômes supposés, des décisions répétées ou des décisions rares, des décisions importantes ou des décisions mineures, des fins émotives ou mécaniques, des riches ou des pauvres, des hommes ou des femmes, des gens intelligents ou stupides... ([1976], p. 8.)

L'esprit de provocation a peut-être nui à la réputation scientifique de l'École de Chicago : celle-ci défend une position méthodologique au moins aussi articulée et cohérente que celle des Autrichiens (dont elle s'inspire) ou que le bayésianisme (auquel elle emprunte ses outils analytiques les plus importants). Par rapport à ces deux écoles, on peut analyser — sommairement — ainsi la conception beckerienne de « l'approche économique ». D'une part, Becker récuse tout à fait l'idée de construire une science économique a priori. On ne retrouve donc pas, chez lui, l'équivalent du « principe économique » dont F. Knight, à Chicago, professait le caractère analytiquement nécessaire (par exemple, [1946]). L'approche économique du mariage et du crime se justifiera uniquement par la fécondité de ses prédictions empiriques : prédictions dont il nous est dit qu'elles sont nombreuses, souvent insuitées (lorsque les théories sociologiques avec lesquelles l'approche est en concurrence ne parviennent pas à les dériver), et largement conformes aux données<sup>16</sup>. Ainsi, Becker semble aller plus loin que les bayésiens dans le rejet des analyses a priori. Par ailleurs, il se sépare d'eux et rejoint von Mises sur ce point décisif : il n'y a pas lieu de distinguer entre action et action rationnelle, entre positif et normatif, puisque toute action humaine, lorsqu'elle est convenablement analysée, apparaît comme rationnelle. « L'approche économique du comportement humain » offre une systématisation presque monomaniaque de cette idée venue de la tradition autrichienne, grâce notamment (mais pas exclusivement) aux possibilités techniques ouvertes par la maximisation de l'utilité espérée<sup>17</sup>.

16. Ainsi, « l'approche économique » permettrait de rendre compte du fait, jusqu'alors inexpliqué, que les individus à haut revenu se marient plus jeunes et divorcent moins fréquemment que les autres ([1976], p. 11). Pour d'autres exemples, voir [1976], *passim* et [1979]. Becker insiste sur le fait qu'aucune de ces prédictions n'est analytiquement vraie. Il s'oppose donc tout à fait à la méthodologie aprioriste de l'école autrichienne.

17. Les théories du mariage, du crime, ou de l'accoutumance à la drogue, reposent évidemment sur des considérations d'utilité espérée. Un autre procédé caractéristique consiste à modifier l'espace de départ de la fonction d'utilité, par exemple en la faisant dépendre du temps, comme dans la « nouvelle théorie du consommateur », [1976], p. 91 et suiv., p. 134 et suiv.

Chez les économistes, la conception dualiste semble s'être affirmée d'abord polémiquement contre l'universalité prétendue du modèle bayésien, quitte à se transformer ensuite en théorie positive. Ce point vaudrait pour des théoriciens de la décision individuelle, comme Simon et Allais, ainsi que, avec beaucoup plus de nuances, pour certains spécialistes de théorie des jeux. En laissant celle-ci de côté pour l'instant, et au risque de forcer quelque peu les positions, nous pouvons tenter de caractériser ainsi l'apport de la théorie de la décision à la thèse dualiste.

LE MODÈLE DE L'UTILITÉ ESPÉRÉE NE PEUT SE DONNER POUR UNE THÉORIE POSITIVE UNIVERSELLE DU CHOIX

Les expériences rapportées par Edwards, et plus récemment par Kahneman et Tversky, révèlent que les sujets ne traitent pas l'information nouvelle qui leur est fournie suivant la méthode du théorème de Bayes, soit qu'ils accordent trop d'importance à l'information nouvelle, soit, au contraire, qu'ils hésitent à réviser leurs probabilités. Une expérience aussi simple que celle du choix binaire (« binary choice experiment »), analysée par Simon, montre que de nombreux sujets s'obstinent à prêter une structure au hasard, au lieu de faire les calculs élémentaires d'espérance mathématique que l'on attendrait d'eux<sup>18</sup>.

Ces résultats paraissent très défavorables à la théorie de l'utilité espérée, mais on peut, dans une certaine mesure, y parer : la théorie a les moyens de réintégrer des contre-exemples importants en s'adjoignant des hypothèses appropriées sur la forme des fonctions d'utilité<sup>19</sup> ou sur la nature des processus d'apprentissage sous-jacents aux expé-

18. Pour une revue de ces résultats et une bibliographie, cf. Simon [1966], p. 5-8 ; [1978], p. 9 ; [1979], p. 506-507. Voici comment se présente l'expérience du *choix binaire* : on demande au sujet, dans une série d'épreuves successives, de choisir entre deux symboles, par exemple « plus » et « moins » ; chacun de ses choix fait l'objet d'un jugement de la part de l'expérimentateur (celui-ci lui dit s'il a fait un « bon » ou un « mauvais » choix, voire lui attribue une récompense dans le cas d'un « bon » choix). La distribution des approbations et des improbations de l'expérimentateur sur la séquence des réponses varie suivant le type d'expériences. Dans la plus simple, les réponses sont jugées « bonnes » ou « mauvaises » aléatoirement, suivant des proportions fixes et connues du sujet : par exemple, on lui dira que « moins » est récompensé deux fois plus, en moyenne, que « plus ». On a alors la surprise de constater que de nombreux sujets construisent des séquences présentant une fréquence relative de deux « moins » pour un « plus ». Cf. Simon [1966].

19. Simon donne un exemple de cette stratégie ([1978], p. 1, note 8). Elle avait été inaugurée dans les articles célèbres de Friedman et Savage [1948] et [1952], que l'on peut juger bien peu conformes à la méthodologie falsificationniste dont ils se réclament pourtant (Watkins [1970] ; Mongin [1983]).

riences<sup>20</sup>. On est alors renvoyé à un problème classique de la philosophie des sciences : comme il est possible, avec un minimum d'ingéniosité mathématique, de sauver n'importe quelle théorie apparemment démentie par l'expérience, quand choisira-t-on de regarder un contre-exemple comme dirimant ? Des auteurs, tels Simon ou Allais, récuse ces stratégies de récupération et considèrent la théorie de l'utilité espérée comme empiriquement réfutée<sup>21</sup>. Ceci conforte pour le moins l'idée anti-réductionniste qu'il doit exister une *autre* théorie positive du choix.

Certains défenseurs notoires du bayésianisme, comme Harsanyi, approuveraient sans doute, au nom de considérations de méthodologie générale (cf. [1976], p. 122 et 1960), le rejet des stratégies « immunisantes » qui permettent de sauver, en toutes circonstances, une théorie en difficulté. Ils rappelleraient peut-être que, même réfutée, une théorie scientifique continue souvent, en l'absence d'une meilleure, à présenter un intérêt *positif*. Mais, surtout, ils argueraient du caractère *normatif* de la théorie considérée. En ce sens, des expériences du type Kahneman et Tversky peuvent être très gênantes pour la version beckerienne du réductionnisme, mais ne semblent pas devoir menacer celle de Savage ou de Harsanyi.

LE MODÈLE DE L'UTILITÉ ESPÉRÉE NE CONSTITUE PAS UNE REPRÉSENTATION NORMATIVE INDISPUTABLE DU CHOIX RATIONNEL

Cette affirmation cruciale, qui revient à remettre en cause le réductionnisme dans son ensemble, peut s'autoriser de deux types de considérations. Tout d'abord, certaines expériences du type précédent sont difficiles à analyser comme des échecs de la rationalité individuelle, qui mettraient seulement en cause l'adéquation des comportements à une norme. L'intérêt exceptionnel du « paradoxe d'Allais », que B. Munier discute dans ce même numéro, tient à cette difficulté. Savage considère qu'il ne met pas en cause la valeur normative du « principe de la chose sûre » parce que, dit-il, il s'est soumis lui-même à la méthode

20. L'idée que les agents font l'apprentissage de leurs propres préférences remonte à Stigler et Becker [1977].

21. Voici comment Simon résume les implications théoriques des résultats précédents : « Même pour des situations relativement simples, le comportement des sujets humains s'est en général considérablement éloigné (de la prédiction faite à l'aide de la théorie) ; [1978] ; « Sur la base de ces données et d'autres, il semble inévitable de conclure que la théorie de l'utilité espérée ne fournit pas une bonne prédiction, même par approximation, du comportement effectif » [1979]. Allais, de même, regarde la théorie comme empiriquement réfutée ([1953] ; [1979], p. 553-554).

de révélation des normes précédemment définies (voir note 13) : on peut juger cette récusation du paradoxe quelque peu cavalière. M. Allais n'est pas du tout mal justifié à dire, comme il le fait dans son article célèbre de 1953, que les sujets qui choisissent « paradoxalement » dans son expérience ne sont pas irrationnels suivant la définition intuitive de ce mot, mais seulement « très prudents »<sup>22</sup>. Sa thèse générale est que la définition « néo-bernoullienne » de la rationalité par la maximisation de l'utilité espérée repose sur des postulats trop déterminés. Sous le couvert d'une théorie universelle de la cohérence des choix, Savage aurait introduit dans ses postulats des spécifications discutables, comme le « principe de la chose sûre », qui conduisent à rejeter dans l'irrationnel certains comportements auxquels on ne peut, normativement, rien reprocher.

Les modèles de rationalité « limitée » ou « procédurale » de H. Simon permettent de mettre en cause, par un autre biais, l'exclusivité *normative* du modèle économique du choix. Nous reviendrons longuement sur ce point dans la deuxième partie.

IL CONVIENT DE DISTINGUER ENTRE LE MODÈLE GÉNÉRIQUE DE LA RATIONALITÉ ET LES MODÈLES SPÉCIFIQUES, COMME CELUI DE L'UTILITÉ ESPÉRÉE

Cette conclusion découle immédiatement du paragraphe ci-dessus. Il faut bien voir qu'elle n'est qu'une conséquence minimale des objections précédemment rapportées. Certains spécialistes de la décision iraient jusqu'à nier que la maximisation de l'utilité espérée ait un intérêt quelconque pour une théorie de la rationalité. M. Allais semble se rapprocher de ce point de vue dans sa polémique contre l'« école anglo-américaine ». La raison en est peut-être qu'il cherche à formaliser directement le concept *générique* du choix rationnel<sup>23</sup>. La position d'un auteur comme H. Simon est moins radicale. S'il reconnaît les

22. Cf. aussi [1952], in [1979] (notamment p. 88-90). Le mémoire rédigé par Allais en 1952 est malheureusement aujourd'hui indisponible en français.

23. Voici comment Allais, en 1953, définit l'homme rationnel : celui-ci « poursuit des fins cohérentes avec elles-mêmes, emploie des moyens appropriés aux fins poursuivies » (p. 518), ce qui impliquerait trois « conditions de cohérence » : 1. le champ du choix est ordonné ; 2. si deux perspectives aléatoires (« loteries » au sens de Savage) sont telles que la première comporte dans tous les cas possibles des gains plus élevés que la seconde, la première sera préférée à la seconde ; (« axiome de la préférence absolue ») ; 3. l'homme rationnel considère des probabilités objectives seulement. On voit que, dans cette axiomatique informelle, ne figure pas le postulat P 2 de Savage. — La définition générale de l'homme rationnel est ici irréprochable, mais on pourrait discuter le fait qu'elle *implique* les conditions (1), (2), (3). Celles-ci, en réalité, correspondent déjà à un modèle *spécifique* de la rationalité, même s'il est moins restrictif que celui de Savage.

déficiences de ce qu'il appelle la « théorie classique » du choix, il ne la disqualifie pas tout à fait, pour autant, comme théorie normative (mais non comme description) de la rationalité. Sa thèse est au fond qu'il faut distinguer la notion très générale, et même triviale, de rationalité comme adaptation des moyens disponibles aux fins, et différents modèles fondés sur des hypothèses particulières relatives à l'interprétation des termes « fins », « moyens », « adaptation »<sup>24</sup>. Nous y adhérons tout à fait dans cet article, dont le but n'est pas de réfuter, mais de relativiser, le modèle économique de l'homme rationnel. La difficulté de cette thèse est qu'elle oblige à produire une taxinomie cohérente des divers modèles *spécifiques* — utilité espérée, théorie des jeux, rationalité « limitée » ou « procédurale » — qui peuvent se réclamer du principe *générique* de rationalité.

Avant d'exposer en détail les notions de rationalité « limitée » ou « procédurale », il convient peut-être d'indiquer sommairement comment le débat entre réductionnisme et dualisme peut affecter la théorie des jeux. Il apparaîtra, nous semble-t-il, dans certaines situations « paradoxales » d'interaction stratégique du type analysé par H. Moulin dans ce numéro. Ces situations ont ceci de commun que les concepts de stratégie et d'équilibre ordinairement utilisés par la théorie des jeux ne permettent pas de déterminer une *notion univoque de décision rationnelle* du joueur, ou, s'ils le permettent, se heurtent à certaines objections intuitives quant à la notion de rationalité utilisée. L'exemple le plus significatif pour notre propos est sans doute celui des jeux non coopératifs à points d'équilibre multiples, dont Harsanyi dit qu'ils sont les plus intéressants pour l'application de la théorie des jeux [1976].

Celle-ci, jusqu'à présent, n'a pas donné le moyen de privilégier un équilibre plutôt qu'un autre comme *la* solution naturelle du jeu ; la définition de ce que peut être, dans ces conditions, une stratégie rationnelle des joueurs demeure tout à fait indéterminée. Harsanyi a tenté de réduire à la fois l'indétermination du concept de stratégie rationnelle et la multiplicité des équilibres en introduisant le point de vue bayésien dans le traitement de ces jeux. À côté des distributions de probabilité *objectives* qui décrivent les différentes stratégies mixtes à sa disposition, chaque joueur  $i$  est caractérisé par une distribution  $p_i = (p^1_i, p^2_i, \dots, p^{K_i}_i)$ , telle que chaque composante  $p^k_i, k = 1, 2, \dots, K_i$ , « est la valeur de probabilité *subjective* initiale que, par hypothèse, chaque

24. Cf. Simon [1964] « La rationalité caractérise un style de comportement qui est approprié à la réalisation de buts donnés, dans les limites qu'imposent des conditions données. » Cette formule est reprise dans [1972], p. 161.

joueur  $j$  ( $j \neq i$ ) attribuée à l'utilisation, par le joueur  $i$ , de sa  $k^e$  stratégie pure disponible » ([1976], p. 113). S'il y a  $n$  joueurs, on définira  $p = (p_1, \dots, p_{i-1}, p_i, p_{i+1}, \dots, p_n)$  comme le  $n$ -uplet de ces distributions de probabilité subjectives. La procédure inventée par Harsanyi (« tracing procedure ») consiste alors grossièrement en ceci : la stratégie rationnelle *initiale* du joueur  $i$  est la meilleure réponse  $s_i$  qu'il associe au  $(n - 1)$ -uplet  $(p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_n)$  ; comme le point  $s = (s_1, s_2, \dots, s_n)$  n'est généralement pas un point d'équilibre, les joueurs devront réviser leurs stratégies ; Harsanyi décrit alors une procédure de révisions successives qui converge vers l'un des équilibres du jeu<sup>25</sup>. Cette méthode, on le voit, est typiquement réductionniste : elle postule que la maximisation de l'utilité espérée, relativement à des probabilités subjectives, s'applique aussi au contexte d'incertitude *stratégique* qui vient d'être défini. Les théoriciens des jeux qui récusent la méthode le feront alors ou bien sur la base du point de vue dualiste (même s'il n'a pas été encore découvert, il existerait un concept de rationalité proprement stratégique qui s'applique au contexte), ou bien suivant l'idée qu'une approche par la rationalité, qu'elle soit stratégique ou bayésienne, rencontre une limite intrinsèque avec les jeux non coopératifs à équilibres multiples (il faudrait alors en appeler à des considérations extra-rationnelles, c'est-à-dire causales, pour déterminer la solution du jeu)<sup>26</sup>.

25. Nous nous appuyons sur Harsanyi [1976], p. 113-115. Pour plus de détails, cf. Harsanyi [1975].

26. Pour cette deuxième branche de l'alternative, cf. Elster [1979], p. 156 : « Afin de transformer la situation en une situation à sortie unique (single-exit situation), nous devons ajouter certaines hypothèses causales aux hypothèses purement intentionnelles, de façon qu'il soit possible de prédire laquelle des options restantes sera effectivement réalisée. » Choisir cette voie revient à reconnaître que l'individualisme méthodologique dans sa formulation traditionnelle comporte une limite inhérente : il ne serait pas capable, même en principe, de fonder une explication *complète* de certaines situations sociales. On est alors conduit à un problème méthodologique difficile : comment faut-il coordonner l'explication intentionnelle, fondée sur le principe de la rationalité, et l'explication causale — par la convention ou « l'habitus », par exemple — qui l'ignore ? Pour une première approche, très générale, de ce problème, cf. Toulmin [1970].

Il faut noter que les problèmes méthodologiques que nous venons de soulever à propos des jeux non coopératifs à équilibres multiples ne sont pas du tout limités à cet exemple. La théorie des jeux répétés conduit généralement à retenir comme équilibres une infinité de  $n$ -uplets de stratégies, même après élimination des équilibres de Nash instables, c'est-à-dire après restriction de l'ensemble des équilibres aux seuls équilibres *parfaits*. La méthode qui consiste à surajouter des déterminations causales aux déterminations intentionnelles ou stratégiques afin de spécifier la solution du jeu apparaît alors particulièrement tentante (pour un exemple, cf. Chiappori et Mongin, à paraître).

**L'APPROCHE DE LA RATIONALITE « LIMITEE »  
OU « PROCEDURALE »  
COMPAREE A L'APPROCHE ECONOMIQUE**

**Généralités**

L'argumentation que nous proposons en faveur du dualisme repose sur un examen de l'approche simonienne de la rationalité « limitée » ou « procédurale ». Elle s'articule ainsi : cette approche satisfait à la définition générale de la rationalité ; elle s'exprime dans un projet de modélisation qui est irréductible au formalisme de l'utilité espérée ; elle permet de rendre compte d'une situation, au moins, que le modèle classique des économistes ne parvient pas à traiter de façon satisfaisante. Pris ensemble, les trois points impliquent que l'on doit faire place à cette approche à l'intérieur d'une théorie générale de la rationalité : le réductionnisme est intenable. L'essentiel de nos analyses est consacré à la deuxième étape du raisonnement : la nature exacte des rapports qu'entretiennent les propositions de Simon et le modèle classique de la rationalité est restée, jusqu'à présent, très obscure.

La raison en est sans doute que l'approche de la rationalité « limitée » ou « procédurale » n'a pas fait l'objet d'une axiomatisation en bonne et due forme. Les quelques formalismes dont nous avons eu connaissance ne nous semblent pas rendre justice à l'intérêt des idées de H. Simon : nous nous en expliquerons. La dissymétrie entre l'axiomatisation élégante de Savage et des propositions qui, pour l'essentiel, sont exprimées dans la langue vulgaire fera douter beaucoup d'économistes qu'il y ait matière à comparaison. Mais leur position ne serait pas épistémologiquement justifiée. L'absence d'une axiomatisation convenable des idées de rationalité « limitée » ou « procédurale » renvoie à deux aspects distincts : d'une part, l'éclatement de l'approche en une pluralité de modèles que relie entre eux certains principes communs, mais qui sont souvent très dissemblables ; d'autre part, le caractère hybride de ces modèles, généralement peu formalisés, et plus faciles à transposer en langage informatique qu'à traduire intégralement en symboles mathématiques. Seul le premier aspect est significatif d'un point de vue épistémologique. Un modèle d'économie dit « mathématique » n'est, aux yeux du logicien, que médiocrement formalisé. Toutes les sciences empiriques connaissent des gradations dans la formalisation — l'important est que la combinaison de langue symbolique et de

langue vulgaire dans un modèle ou une théorie, en premier lieu, n'empêche pas de tester la cohérence logique des propositions rassemblées, en second lieu, ne les rende pas si équivoques qu'on ne puisse en dériver des prédictions empiriques significatives. Les modèles de Simon passent le test sur ces deux plans ; beaucoup, d'ailleurs, sont, épistémologiquement parlants, de même nature que les modèles « classiques » auxquels ils s'opposent<sup>27</sup>. Le véritable problème est que l'approche n'a, jusqu'à présent, donné lieu qu'à une pluralité de *modèles*, regroupés autour d'une heuristique commune, mais dont on ne peut dire qu'ils s'ordonnent en une *théorie*. L'approche de l'utilité espérée a, en revanche, tous les caractères que l'on peut prêter à une théorie scientifique : distinction claire d'une syntaxe et d'une sémantique (rendue possible par une axiomatique) ; généralité considérable (en droit indéfinie) ; souplesse d'emploi (tous les axiomes ne sont pas simultanément mobilisés dans chaque modèle particulier)<sup>28</sup>.

Simon a évidemment pris la mesure de cette difficulté. En 1963, passant en revue les rapports de l'économie et de la psychologie, il soulignait les résultats désastreux qu'obtient le modèle classique dans la confrontation aux données sur les comportements individuels ; mais il reconnaissait aussi bien que, dans la plupart des secteurs d'étude, « nous n'avons malheureusement, pour remplacer le modèle classique,

27. Soit, par exemple, l'étude récente de Ijiri et Simon [1977] où la courbe de coût à long terme des entreprises est expliquée à l'aide d'un modèle de rationalité limitée et de certaines hypothèses sur la distribution des tailles et les taux de croissance des entreprises. Il s'agit d'un travail de modélisation au sens tout à fait traditionnel, où des hypothèses générales (rationalité limitée) reçoivent une forme particulière, s'adjoignent des conditions initiales (distribution des tailles et des taux de croissance) et servent à prédire un certain résultat empirique (forme de la courbe de coût à long terme). L'appréciation de la *cohérence logique* et de la *testabilité* de l'explicans poserait, ici, des problèmes de même nature, exactement, que l'appréciation d'un modèle « classique » de la recherche économique ordinaire. Les travaux d'intelligence artificielle, en revanche, soulèvent des problèmes d'appréciation distincts, dans la mesure où cohérence et testabilité s'éprouvent indirectement, en quelque sorte, par la simulation informatique. Ce type d'application empirique appelle sans doute l'élaboration d'une épistémologie spécifique, dont il n'existe encore, çà et là, que des ébauches partielles (la plus significative, peut-être, est due à Simon lui-même : [1969], chap. I).

28. Très sommairement, une approche par « modèles » préserve plus nettement l'individualité du phénomène qu'une approche « théorique » au sens fort, puisque le modèle s'applique *en propre* au phénomène, alors que la théorie est en droit d'application universelle et regarde celui-ci comme un *cas particulier*. En fait, la distinction ne peut être que de degrés, non de nature. Pariente a élaboré une distinction voisine de celle-ci entre « connaissance par modèle » et « connaissance par système » [1973]. Lorsqu'ils ne procèdent pas d'une théorie préalable, les modèles comportent toujours un certain élément d'*ad hocité*, comme le souligne Simon dans la citation du paragraphe suivant.

qu'une masse de spécifications empiriques qui ne donnent pas lieu à des prédictions précises (sharp) en dehors des données auxquelles elles-mêmes s'appliquent » ([1963], p. 735). Ceci pourrait se paraphraser : il est facile de faire mieux, pour la prédiction ou l'explication des comportements individuels, que la théorie classique, mais on ne retrouve pas, avec des modèles mieux corroborés, son remarquable pouvoir d'unification. Aujourd'hui, Simon accorderait sans doute moins d'importance à cet aspect de la comparaison : ses articles les plus récents ([1972], [1978], [1979]) suggèrent qu'il assumerait, dans une certaine mesure, l'éclatement de son approche, en invoquant les particularités du domaine — l'hétérogénéité des choix humains — et, plus encore peut-être, celle de la technique employée — la simulation informatique, qui rend généralement inutile une formalisation intégrale des modèles.

Ces préliminaires visaient à dissocier le vrai du faux problème que pose la comparaison entre l'approche de Simon et celle des économistes. Nous aurons à tenir compte sans cesse de la dénivellation méthodologique entre une théorie, au sens traditionnel, et des modèles qui sont particuliers dans leur objet (le choix d'un coup aux échecs, le comportement de prix d'une entreprise) et ont souvent fait leurs preuves grâce à leur incorporation à des artefacts (ce qui est un mode très particulier du test). Ce problème nous a conduit à ne pas aborder immédiatement la relation entre Simon et les classiques, qui fait l'objet véritable de cette section. Nous la réservons à une rubrique ultérieure, après avoir essayé de reconstruire l'approche simonienne telle qu'elle résulte des éléments de modélisation proposés depuis 1947.

Le choix de cette méthode se trouve conforté par une autre considération : les caractéristiques explicites que Simon a données de sa conception sont relativement équivoques. Il emploie généralement les trois expressions de « théorie comportementale », de « rationalité limitée » et de « rationalité procédurale » ; il peut être utile de s'arrêter un instant sur les difficultés d'interprétation qu'elles soulèvent.

« Behavioral », à vrai dire, n'est guère gênant. L'approche simonienne dans son ensemble tombe sous cette étiquette (cf. [1955] ; [1979], p. 496), mais elle ne lui est évidemment pas spécifique : elle a été revendiquée dans d'autres tentatives hétérodoxes — qui communiquent sans doute avec les travaux de Simon — comme celle de Cyert et March [1963]. Par ce terme, il faut entendre non pas un modèle bien défini, ni même une classe de modèles, mais une direction méthodologique : sont « behaviorial » les recherches sur les comportements économiques qui se prêtent explicitement à des tests menés sur les agents *individuels*

de décision, et récuse donc le principe méthodologique friedmanien du « comme si » tel qu'il est généralement entendu<sup>29</sup>.

Pour décrire ce qui est spécifique dans son projet, Simon a employé avec une assez grande régularité l'expression de « rationalité limitée » : « limited rationality » ([1955], p. 225), puis « bounded rationality » (par exemple [1972], [1979]). Cette expression a une valeur polémique contre la théorie classique, mais elle est équivoque : elle peut renvoyer soit à l'idée d'une rationalité *affaiblie* par rapport à la rationalité achevée dont ferait preuve le maximisateur d'utilité espérée, soit à celle d'une rationalité *achevée*, sur le même pied que la rationalité du modèle classique, mais qui, contrairement à celle-ci, tiendrait compte des limitations effectives dans les *moyens* du choix, notamment les moyens de calcul.

Enfin, Simon a été plusieurs fois tenté de caractériser sa conception comme « procédurale », par opposition à l'idée « substantielle » (« substantive ») de la rationalité qui inspirerait les économistes. Le principe de cette dichotomie est apparemment le suivant :

( $\alpha$ ) Les modèles de rationalité « procédurale » s'intéressent à la rationalité du *processus* du choix ; ceux de la rationalité « substantielle » privilégient la rationalité des *résultats* du choix<sup>30</sup>.

( $\beta$ ) Est « procédurale » la conception qui met l'accent sur l'aspect *délibératif* de la décision, c'est-à-dire qui insiste sur le fait que les conditions du choix, du côté des fins aussi bien que des moyens, ne sont

29. Cf [1979], p. 435-496. On interprète généralement le principe friedmanien du « as if » comme limitant le test des hypothèses économiques à une confrontation avec des données *agrégées* (dans la prédiction macro-économique, ou, au minimum, la prédiction de marché, faite sur un agent représentatif). Telle est, en tout cas, l'interprétation de Simon : [1966], p. 2 ; [1979], p. 494-495. En fait, sous cette forme, ce serait Becker ([1962], in [1976], chap. VIII), plutôt que Friedman, qui a fixé le sens du « comme si », cf. Mongin [1982] et [1983].

Le mot « behavioral » a peut-être des connotations malheureuses : il n'y a rien de commun entre l'approche d'un auteur comme Simon et le courant *behavioriste* en psychologie expérimentale, puisque celle-là met toujours l'accent sur les caractéristiques internes des sujets de décision et rejette expressément le principe de la « boîte noire ». En fait, c'est plutôt l'approche ordinaire du « comme si » qui participe de la méthodologie behavioriste, par l'usage régulier qu'elle fait des « boîtes noires ».

30. Cf. 1964 : « La rationalité renvoie tantôt aux processus de choix... tantôt aux choix eux-mêmes. L'usage ancien du mot en psychologie, en logique et en morale privilégie le premier aspect, tandis que l'économie et la sociologie s'intéressent principalement au second. » Et, plus récemment : « Nous devons rendre compte non seulement de la *rationalité substantielle* — du degré auquel le sujet choisit des actions appropriées —, mais aussi de la *rationalité procédurale* — de l'efficacité, relativement au pouvoir et aux limites cognitives de l'homme, des procédures qu'il utilise pour choisir ses actions » ([1978], p. 9).

pas données au décideur, mais font l'objet d'une recherche. Est « substantielle » la conception qui regarde les conditions du choix comme fixées, et identifie donc la décision à la simple application d'un critère d'évaluation *donné* à un ensemble d'actions possibles également *donné* <sup>31</sup>.

Cette opposition du « procédural » et du « substantiel » est aussi équivoque parce qu'elle donne à penser — plus nettement dans la définition  $\alpha$ ) que dans le commentaire  $\beta$ ) — qu'une comparaison entre théorie classique et approche simonienne est dénuée d'objet : l'une et l'autre auraient des compétences différentes. Mais s'il en allait ainsi, les polémiques de Simon contre la trivialité, ou le manque d'intérêt pratique, du modèle de l'utilité espérée perdrait toute signification. Par ailleurs, il reste à comprendre le lien que Simon envisage entre les deux aspects de rationalité limitée, ou avec des moyens limités, et de rationalité procédurale : s'appliquent-ils à deux classes de modèles ou à l'approche dans son ensemble ?

Pour définir celle-ci, il vaut donc mieux esquisser une caractérisation d'ensemble.

### **Caractérisation de l'approche**

#### LE MODÈLE ORGANISATIONNEL ET LA NOTION DE « SUBGOALS »

Il n'est pas indifférent que les recherches de Simon sur la décision aient eu pour point de départ historique l'organisation, et non pas l'agent individuel. C'est en étudiant l'administration municipale qu'il a été frappé de l'inadéquation des concepts marginalistes : dans le meilleur des cas, ceux-ci pouvaient prétendre décrire les décisions d'une unité interne à l'organisation, par exemple, le choix des fournisseurs par tel ou tel bureau de l'administration communale ; mais ils étaient manifestement inadéquats lorsqu'il fallait comprendre une décision prise par l'organisation elle-même, par exemple celle de construire une piscine, conformément aux recommandations du bureau des sports, plutôt qu'un théâtre, comme le demanderait le bureau de la culture. L'égali-sation à la marge des coûts et des avantages perd ici toute valeur

31. Cf [1976], p. 131-137, où Simon reprend le concept, emprunté à la philosophie classique, de « délibération » (« Le comportement est rationnel au sens procédural quand il résulte d'une délibération appropriée »), et l'analyse à partir de l'idée moderne de « processus de recherche ».

descriptive, pour la simple raison que les données du problème d'optimisation ne sont pas définies. Simon a donc été conduit à distinguer le mode de réalisation des buts inférieurs intermédiaires (« subgoals »), qui peut, dans certains cas, s'analyser d'après les modèles économiques de la rationalité, et les règles de fixation de ces buts inférieurs, au niveau de l'organisation elle-même, qui échappent complètement à ces modèles <sup>32</sup>.

Transposée à l'individu, cette description sommaire — nous ne prétendons pas rendre justice à la richesse de la conception simonienne de l'organisation — a des implications intéressantes. Tout d'abord, elle suggère que la représentation du choix doit faire appel à des modèles multiples et coordonnés entre eux : ils correspondent aux différentes procédures de décision qui coexistent dans un même sujet et auxquelles celui-ci recourra préférentiellement suivant les problèmes à traiter. L'exemple de l'organisation a donc orienté Simon vers le dualisme sous cette forme radicale : un même sujet empirique de décision tombe généralement sous plusieurs modèles décisionnels. D'autre part, la façon dont l'administration municipale *hiérarchise* ses décisions suggère une modalité possible de la décision individuelle, inédite pour les économistes, bien que les sciences sociales l'aient connue depuis toujours. On peut la désigner par l'expression de « méthode des buts intermédiaires », quand on se place au point de vue de l'agent, et celle « d'analyse fonctionnelle », quand on la prête non pas à l'agent lui-même, mais au sociologue ou à l'économiste.

La méthode des buts intermédiaires consiste simplement à assigner des buts pour lesquels on cherchera tour à tour les moyens les plus appropriés ; une fois ces moyens caractérisés, on les regardera à leur tour comme des buts (« subgoals ») dont il s'agira de déterminer les moyens de réalisation ; et ainsi de suite. Cette méthode familière correspond très largement à la vision traditionnelle de la délibération : « On ne délibère que sur les moyens. » (Aristote.) L'influence de la théorie marginaliste, ou plus généralement néo-classique, du choix en a fait

32. Cf. Simon [1974]. Il peut être intéressant de mettre en parallèle la façon dont Arrow [1974] et Simon, respectivement, introduisent le concept d'*organisation* dans la théorie économique. Pour ces deux auteurs, les organisations apparaissent comme des moyens de pallier les défaillances de la rationalité des choix — ce qui explique que les versions élémentaires de la théorie néo-classique en ignorent purement et simplement l'existence. Mais, pour Arrow, il s'agit d'un *défaillance* de la rationalité *collective* (en fait, du système de prix), alors que, pour Simon, une organisation est l'une des réponses que l'on peut donner au problème du choix *individuel* ([1979], p. 501). Dans les termes de von Mises, Arrow ne voit que l'aspect catallactique de l'organisation, alors que Simon en propose également une analyse praxéologique.

oublier l'importance. Du point de vue de cette théorie, il est évidemment sous-optimal de *fixer* un but et de délibérer *ensuite* sur les moyens. L'intérêt d'une approche en termes de « préférences » et de « contraintes », plutôt que de « fins » et de « moyens », vient de ce qu'elle permet d'*endogénéiser les fins au même titre que les moyens*. Cette considération importante a conduit Harsanyi à proclamer, d'une manière typiquement réductionniste, que l'approche contraintes/préférences était « plus générale » que celle de la fin et des moyens ([1976], p. 94). Conformément à son option dualiste, Simon rétorquerait sans doute qu'elle ne peut être la plus générale, puisqu'elle n'est pas toujours applicable : le sujet de décision ne dispose pas toujours d'un préordre complet sur ses actions (ou sur les conséquences de ses actions). Dans le cas d'information imparfaite sur les préférences, la méthode des buts intermédiaires est l'un des procédés *rationnels* par lesquels on parvient à une décision<sup>33</sup>. En dépit du réductionnisme, le sens commun et l'économie appliquée (la politique économique notamment) ne l'ont, heureusement, jamais ignoré.

Simon appelle « analyse fonctionnelle » la méthode des buts intermédiaires quand elle est considérée du point de vue de l'économiste ou du sociologue : « Des comportements sont fonctionnels s'ils contribuent à la réalisation de certains buts » (sous-entendu : supposés donnés ([1978], p. 3). Cette détermination provient de la sociologie et de l'anthropologie, où le raisonnement fonctionnel (en ce sens) est très répandu — mais pas toujours, loin de là, réservé aux individus. L'approche qu'on appelle « fonctionnaliste » consiste le plus souvent à poser que des buts sont *donnés* (plus ou moins vaguement : le bien de la société, le maintien de la classe dominante) et qu'à ces buts donnés et considérés isolément les uns des autres, les moyens vont s'ajuster (telle institution, tel comportement de classe)<sup>34</sup>. L'une des thèses épistémologiques intéressantes de Simon est que la méthode archaïque de l'« analyse fonctionnelle » devrait suffire à traiter de nombreux problèmes, économiques ou non, auxquels il est maintenant courant d'appliquer

33. Ce qui est gênant dans l'hypothèse traditionnelle, ce n'est pas tant l'axiome de transitivité que celui de *complétude*. L'évaluation est possible pour des « grappes » d'actions que relie la transitivité, mais toutes les actions ne sont pas commensurables. On peut analyser cette situation comme un cas d'information imparfaite, et justifier rationnellement la méthode des subgoals par les coûts de recherche ou de transaction — infinis — que représenterait le passage du préordre partiel à un préordre total.

34. L'idée marxiste des « contradictions » d'une société illustre, dans un contexte holistique, le conflit des buts intermédiaires qui est inhérent à l'approche des « subgoals » : pour Marx, le but intermédiaire « réaliser de la plus-value relative » entre en conflit avec l'autre but intermédiaire du capital, « réaliser de la plus-value absolue ».

le calcul marginal ([1978], p. 3-6). Pour expliquer le fait que deux individus puissent entrer dans la relation d'employeur à employé, il n'est sans doute pas nécessaire de décrire des ajustements marginaux de coûts présents et d'avantages subjectivement anticipés ; on peut aisément se contenter d'« un argument qualitatif simple suivant lequel un contrat d'emploi peut représenter une manière fonctionnelle (« raisonnable ») de faire face à certains types d'incertitude » ([1978], p. 5). Cette position, qui paraîtra sans doute obscurantiste à certains, tire évidemment sa force d'une constatation que Simon laisse implicite : parmi les modèles marginalistes de la relation d'emploi, nombreux sont ceux qui, comme l'analyse fonctionnelle des sociologues, ne visent pas à dériver de conclusions autres que qualitatives. On ne peut rejeter aussi rapidement, ou du moins sur la même base épistémologique, des modèles qui prétendent endogénéiser le taux de salaire et la durée de travail fixés dans le contrat. Par ailleurs, comme il est bien connu depuis Samuelson, certaines conclusions *qualitatives* demandent, pour être établies, la comparaison *quantitative* d'effets de direction contraire.

La discussion précédente donne peut-être à penser que Simon regarde comme peu problématique le transfert des concepts entre la théorie de l'organisation et celle de la décision individuelle. Il est vrai que, à la suite de Marschak [1955], il refuse parfois de distinguer entre les sujets, individuel ou collectif, dont il décrit le processus de décision : « Une théorie du comportement rationnel peut avoir affaire à la rationalité des individus ou à celle des organisations. En fait, les deux corpus théoriques ne sont pas complètement distincts. » ([1972], p. 161.) Cette attitude peut s'autoriser de l'idée intéressante, mais problématique, suivant laquelle les problèmes de choix seraient au moins partiellement indépendants de la nature du sujet : il existerait une théorie générale de la rationalité des organismes qui recouperait des théories spécifiques, comme celle de l'organisation ou du choix individuel<sup>35</sup>. Sous cette forme, l'attitude de Simon se heurterait aux objections de l'individualisme méthodologique. Pour éviter un débat sur ce point, ici hors de propos, il vaut mieux rendre indépendantes l'approche simonienne des choix individuels et l'idée d'une théorie de la rationalité des organismes, en considérant que le modèle organisationnel n'a, par rapport à l'individu, qu'une valeur heuristique. Si la méthode des « subgoals » a pu se transposer de l'administration municipale à

35. Ceci paraît implicite dans [1979], p. 501, où la liste des procédures de la rationalité « limitée » ne distingue pas entre celles qui sont propres à l'organisation et celles qui sont communes à l'individu et à l'organisation.

l'individu, cela n'est pas dû à l'identité des problèmes qui se posent à l'une et l'autre, mais seulement à une situation de fait, qui n'allait pas de soi avant que la méthode fût testée sur le cas de l'individu.

#### LE PRINCIPE DE SATISFACTION

A côté de la méthode qui consiste à « remplacer des buts abstraits et globaux par des buts inférieurs tangibles dont la réalisation se prête à l'observation et à la mesure », Simon attache une grande importance à celle qui définit « des choix satisfaisants au lieu de choix optimaux » ([1979], p. 501). De fait, les économistes connaissent Simon principalement comme le théoricien du « satisficing », dont il a défendu avec succès l'application aux décisions d'entreprises<sup>36</sup>.

Le principe de satisfaction peut se formuler simplement ainsi : est choisie non pas l'action la meilleure relativement aux conditions objectives et subjectives de la décision, mais une action satisfaisante, c'est-à-dire dont l'évaluation est supérieure au niveau d'aspiration, ou seuil de satisfaction, du sujet. Plus précisément, un modèle qui fait intervenir le principe de satisfaction devra spécifier les deux mécanismes suivants :

a) un mécanisme de comparaison entre les actions et le niveau d'aspiration, qui assure un classement séquentiel des actions en « satisfaisantes » et « non satisfaisantes » ;

b) un mécanisme d'endogonisation partielle du niveau d'aspiration qui décrit comment celui-ci s'ajuste lorsque le mécanisme de comparaison précédent n'engendre pas immédiatement de solution satisfaisante.

Cette définition (qui correspond à [1979], p. 503) appelle trois types de remarques. Tout d'abord, l'aspect *séquentiel* des règles de « satisficing » est tout à fait fondamental : le sujet n'attribue pas une évaluation à chaque action possible, mais il considère successivement  $a_1, a_2 \dots$  qu'il évalue et compare au seuil de satisfaction  $k$  ; il arrête sa recherche dès qu'il rencontre une action  $a_s$  dont l'évaluation égale ou excède  $k$ .

36. Sur les liens entre le « satisficing » et les règles de « mark-up » employées dans les entreprises, cf. [1966], p. 11.

Le mot « satisficing » n'apparaît pas au vocabulaire de l'anglais courant ; il s'agit d'un terme écossais d'emploi archaïque que Simon a remis en usage. J. L. Le Moigne propose la traduction : principe du « satisfecit » ([1969], édit. française [1974], p. 82).

L'aspect séquentiel n'est pas impliqué logiquement dans l'idée de la satisfaction : on pourrait concevoir, à la place de (a), un mécanisme *globalisant* qui consisterait à évaluer toutes les actions possibles et à définir ensuite un sous-ensemble d'actions satisfaisantes dont l'une serait finalement sélectionnée d'après un critère quelconque. Naturellement, un tel mécanisme n'a aucune vraisemblance psychologique, mais il est a priori concevable<sup>37</sup>. Ce n'est donc pas tant le principe de satisfaction en lui-même, que sa version séquentielle, dont H. Simon fait une pièce constante de ses propositions et de sa critique de la théorie classique<sup>38</sup>.

Cette première remarque en appelle une seconde sur les rapports de l'optimisation et de la satisfaction. Il y a deux aspects à ce problème, l'opposition entre « séquentiel » et « globalisant » d'une part, la forme de la fonction d'évaluation de l'autre. Si l'on oublie l'aspect séquentiel, le « satisficing » est un cas particulier d'optimisation : les solutions satisfaisantes résultent de la maximisation d'une fonction d'évaluation à deux valeurs, 0 (insatisfaisant) et 1 (satisfaisant). Le remplacement d'une fonction d'évaluation complexe par une fonction à deux valeurs peut représenter une simplification du calcul importante, et, dans son premier modèle, Simon semble avoir vu là un des enjeux des modèles de satisfaction ([1955], p. 246-247). Mais, en réalité, la psychologie empirique montre que le gain réalisé du côté de l'évaluation est bien moins significatif que l'économie d'information liée au caractère séquentiel du choix : le véritable intérêt d'une règle de « satisficing » par rapport à la règle d'optimisation, est qu'elle dispense d'explorer la totalité de l'ensemble des actions possibles<sup>39</sup>. On peut exprimer cette différence de la manière suivante : l'optimisation sépare le moment de l'évaluation et celui du choix (on évalue tous les éléments de l'espace considéré, puis on choisit ceux dont l'évaluation est maximale) ; une règle de « satisficing » au sens précédent consiste en une succession de couples d'évaluation et de choix (chaque action, tour à tour, est évaluée,

37. Formellement parlant, il revient à effacer la distinction entre respect des *contraintes* et réalisation de la fonction *objectif* : en programmation linéaire, il s'agirait de remplacer la condition d'optimisation par une inégalité supplémentaire.

38. Cf. « Dans la plupart des modèles globaux du choix rationnel, toutes les variantes (alternatives) sont évaluées avant que le choix soit fait. Dans les processus de décision humains, les alternatives sont souvent examinées de façon séquentielle. » ([1955], p. 252).

39. Ce point peut se constater sur l'exemple du jeu d'échecs : ignorant en quoi consiste le couple de stratégies minimax qui doit correspondre, suivant le théorème de Zermelo, à chaque configuration du jeu, les joueurs se contentent de critères intuitifs d'évaluation des *positions* auxquelles mènent leurs coups ; la façon dont est construite cette fonction d'utilité implicite est moins importante que le nombre de positions futures que le joueur parvient à analyser.

puis soumise au choix, c'est-à-dire rejetée ou acceptée ; le processus cesse à la première action acceptée).

La troisième remarque concerne le mécanisme d'endogénéisation (*b*). Tout comme (*a*), il est avéré par la psychologie empirique, par exemple les travaux de K. Lewin. Ceux-ci ont montré que les niveaux d'aspiration avaient tendance à s'ajuster de manière à faire apparaître des solutions : si la valeur  $k$  est trop élevée initialement, elle tendra à s'abaisser (cf. [1955], p. 253 ; [1979], p. 503). Cette considération fait apparaître la relation entre satisfaction et optimisation sous un angle nouveau : si le niveau de satisfaction s'ajuste en direction du possible, le processus *dynamique* de « satisficing » pourra fort bien converger vers l'optimum ([1966], p. 10). Ce point est appelé à jouer un rôle non négligeable dans l'échange d'arguments entre Simon et les défenseurs de la théorie classique.

Pour préciser la définition abstraite du principe de satisfaction, nous utiliserons le formalisme de Simon en 1955. Il appelle  $A$  l'ensemble des actions,  $S$  l'ensemble des conséquences et  $V : S \rightarrow \mathbb{R}$  une fonction d'évaluation des conséquences. L'information des agents est décrite par une correspondance qui, à chaque  $a \in A$ , associe un sous-ensemble  $S_a$  de  $S$ . On peut alors concevoir la règle de décision suivante : examiner tour à tour  $a_1, \dots, a_i, \dots$ , en calculant à chaque étape  $V(S)$  pour tout  $s \in S_{a_i}$  ; choisir la première action  $a_s$  telle que :  $V(s) \geq k$  pour tout  $s \in S_{a_s}$ .<sup>40</sup>

Une telle règle décrit un mécanisme séquentiel du type (*a*) ci-dessus. Si l'on tente de la situer par rapport à la théorie bayésienne de la décision, on s'apercevra qu'elle partage avec elle un présupposé important : l'évaluation ne porte pas directement sur les actions, mais sur les conséquences ; elle « remonte » des conséquences aux actions. Cette hypothèse est naturelle dans une théorie de la rationalité, mais il n'est pas sûr qu'il faille lui donner un caractère absolu<sup>41</sup>. D'autre part, et ceci est plus important, Simon innove par une conception hétérodoxe de l'information de l'agent. Il faut en préciser le rapport exact avec la représentation savagienne de l'incertitude.

40. La règle que propose Simon ([1955], p. 248 et 253) nous semble plutôt se rattacher à la « méthode des subgoals » qu'au principe de satisfaction. C'est pourquoi nous ne la reproduisons pas ici.

41. On peut se demander si les travaux de Simon sur l'intelligence artificielle ne conduisent pas à remettre en cause cette hypothèse apparemment naturelle. La force d'un *coup*, aux échecs, s'apprécie dans une certaine mesure indépendamment de ses conséquences : comment formaliser, en termes d'évaluation des conséquences, la notion intuitive qu'a le joueur de « reprendre l'initiative » ? Cf. Simon [1972], p. 169, où est présentée une distinction intéressante entre « coups » et « stratégies » aux échecs ; et Newell, Shaw, Simon [1962], qui traitent en détail des heuristiques concrètes du joueur d'échec.

Savage formalise l'idée d'action en la définissant comme une application de l'ensemble des états du monde dans l'ensemble des conséquences, une  $a : E \rightarrow S$  ([1954], p. 14 ; cf. Radner [1972]). Ceci veut dire que l'agent est « incertain » des conséquences de son action : il peut associer à celle-ci un sous-ensemble de  $S$  (soit  $a(E)$ , l'image par la fonction-action de l'ensemble des états du monde), et non pas un point unique de  $S$ , comme le permettrait l'information parfaite. Cette incertitude a ceci de remarquable qu'elle disparaît dès que l'agent connaît l'état du monde réalisé  $e$  : le sous-ensemble associé  $a(E)$  se contracte alors en un point  $a(e)$ . Chez Simon, l'ensemble  $E$  des états du monde n'apparaît nulle part. Il est seulement dit qu'à chaque action  $a$  (regardée maintenant comme un point dans un espace  $A$ , non plus comme une fonction) est associé un sous-ensemble  $S_a$ . Rien n'interdit que cette correspondance grossière  $a \rightarrow S_a$  soit « raffinée » de la façon que précise Savage ; mais rien ne l'impose non plus. Le refus de Simon, à première vue surprenant, d'introduire un espace des états du monde lui permet de se situer à un niveau de généralité plus grand que Savage. Il évite d'identifier, comme le fait celui-ci, l'incertitude de l'agent à une incertitude *externe*. Il se peut que l'agent simonien associe  $S_a = a(E)$  à son action parce qu'il ignore quel est l'état du monde réalisé et n'ignore rien d'autre ; mais il se peut aussi bien qu'il connaisse le  $e$  pertinent et soit incertain parce qu'il ne sait pas calculer les conséquences de ses actions.

Le modèle de 1955 ne comporte pas de mécanisme explicite du type (b). Parmi les tentatives faites pour endogénéiser le niveau d'aspiration, il faut citer celle de Radner [1974], mais il nous semble qu'elle manque partiellement son but. L'article de Radner formalise la situation suivante : un organisme fait alterner les périodes de recherche et les périodes de repos suivant l'écart qui apparaît à chaque moment  $t$  entre sa performance  $U(t)$  et un certain niveau de satisfaction ; celui-ci est endogène parce qu'il dépend du niveau  $u$  atteint à la fin de la dernière période de recherche. La valeur  $u$  restera fixe pendant toute la durée de la période de recherche qui s'achève seulement lorsque  $U(t)$  atteint ou dépasse de nouveau  $u$ . Le modèle de Radner appartient en fait à la classe des modèles *adaptatifs*, dans lesquels certains paramètres de la décision varient dans le temps, en fonction de leurs propres valeurs passées. Il ne formalise pas l'idée suivant laquelle le niveau de satisfaction peut se modifier en fonction des performances courantes et de la durée de la recherche, pour ajuster, en quelque sorte, le désirable au possible. C'est, en fait, aux modèles inspirés des idées de

K. Lewin sur les niveaux d'aspiration, c'est-à-dire à la psychologie empirique, qu'il conviendrait d'emprunter les outils nécessaires à une endogénéisation pertinente du seuil de satisfaction.

#### LE PRINCIPE DE RECHERCHE

Simon a parfois mis en parallèle l'idée de « search » et celle de « satisficing » (par exemple [1979]). Il nous semble que l'on peut définir un principe de recherche sur le même pied que la méthode des buts intermédiaires ou le principe de satisfaction, et complémentaire d'eux. Pour le décrire précisément, le plus simple est d'en caractériser directement les rapports avec B, le principe de satisfaction. Nous laissons de côté, désormais, la « méthode des subgoals », à la fois parce qu'elle est moins intéressante que le « satisficing » et que le type de recherche qu'elle fait intervenir est élémentaire : identification des buts, puis recherche des moyens, transformation des moyens, une fois identifiés, en buts intermédiaires, etc. <sup>42</sup>.

Les rapports du « search » et du « satisficing » sont doubles. Une idée de recherche apparaît déjà à l'intérieur du principe de satisfaction, puisque celui-ci n'a de sens que lorsque l'ensemble des *actions* possibles (« alternatives ») n'est pas entièrement donné au sujet. Mais le principe de recherche est aussi distinct de B, car, pour Simon, le « search » doit porter aussi sur l'ensemble des *conséquences* attachées aux actions que l'individu envisage tour à tour. Cette distinction est, par exemple, implicite in [1979], p. 500-501, où Simon décrit ainsi les deux points de théorie qui le séparent le plus nettement de l'approche classique : « les options de recherche doivent être recherchées (the alternatives of search had to be sought out) » *et* « les conséquences du choix d'une option particulière sont connues de façon imparfaite seulement ».

Le formalisme précédent peut aider à faire ressortir la différence des deux types de « search ». Le premier ( $R_1$ ) consiste à explorer séquentiellement A, en prenant la correspondance des *a* avec  $S_a$  comme donnée. Le second ( $R_2$ ) consiste, pour un *a* donné, à « raffiner » la correspondance avec  $S_a$ , voire, dans certains cas, à la transformer en fonction. Simon décrit ainsi ce processus :

42. Les trois aspects A, B, C se trouvent réunis dans l'approche globale du « General Problem Solving » en intelligence artificielle. Développée à partir de 1957 par Newell, Shaw et Simon, cette approche ne vise pas moins qu'« à modéliser la plupart des caractéristiques de la capacité qu'a l'homme de résoudre des problèmes » (Simon [1969] ; édit. franç., p. 86). Les résultats importants qu'elle a permis d'atteindre, notamment sur l'exemple privilégié du jeu d'échecs, sont consignés dans Newell et Simon [1972], et, plus récemment, dans Simon [1979 a].

Un trait de réalisme que nous pourrions souhaiter d'introduire consiste à abandonner l'hypothèse que la relation entre  $A$  et les sous-ensembles de  $S$  soit connue à l'avance, alors même que  $V(S)$  le serait. Dans un cas extrême, on partira d'une situation où chaque élément  $a$  est associé à la totalité de l'ensemble  $S$ . Ensuite, on introduira dans le processus de décision des étapes de collecte de l'information qui ont pour effet de rendre plus précise la correspondance entre les  $a$  et des sous-ensembles différenciés de  $S$ . Si ce processus de collecte de l'information est coûteux, alors l'un des aspects de la décision sera la détermination du degré auquel la correspondance doit être raffinée. ([1955], p. 248.)

Les processus  $R_1$  et  $R_2$  ont-ils des équivalents dans la conception savagienne de la décision ?  $R_1$  se laisse facilement transposer : il suffira de poser que l'espace  $A^s = \{a\}$  des actions <sup>43</sup> n'est pas donné à l'agent, mais exploré séquentiellement.  $R_2$  n'a pas d'équivalent véritable. Plus exactement, Savage ne peut considérer, comme nous l'avons déjà suggéré, qu'un cas spécial de  $R_2$  : l'agent commence par faire une recherche sur l'ensemble des états du monde  $E$  ; celle-ci lui permet de restreindre son application  $a : E \rightarrow S$  en une application  $a|E_1$  définie par  $E_1 \subset E$  ; on a alors  $a|E_1(E_1) \subseteq a(E)$  avec parfois l'inclusion stricte ; et ainsi de suite pour  $E_2, E_3, \dots$  <sup>44</sup>. En d'autres termes, la réduction du sous-ensemble associé provient *uniquement* d'une réduction de l'incertitude externe.

En revanche — et cela ne saurait surprendre — les processus  $R_1$  et  $R_2$  se laissent immédiatement traduire dans la terminologie des informaticiens. Simon décrit ainsi les constituants d'un programme joueur d'échecs ; des « générateurs de coups », qui fournissent l'équivalent de l'ensemble  $A$  ; des « évaluateurs », correspondant à la fonction  $V(S)$  ; enfin des *règles d'arrêt*, destinées, respectivement, « à arrêter l'exploration des continuations aux positions stables qui peuvent être évaluées suivant des critères statiques » (=  $R_2$ ) et « à arrêter l'ensemble du processus par la sélection d'un mouvement satisfaisant » (=  $R_1$ ) <sup>45</sup>.

Le problème que soulève le principe de recherche tient essentiellement à la formalisation de  $R_2$ , et, plus encore, de l'interaction de  $R_1$  avec  $R_2$ . L'exemple du jeu d'échecs pourra servir à donner la mesure

43.  $A^s$  est un espace d'actions au sens de Savage, c'est-à-dire d'applications de  $E$  dans  $S$ . Nous le considérons comme un sous-ensemble de l'espace des actions au sens de Simon, en redéfinissant ainsi ce dernier : une action est une correspondance qui, à chaque  $e \in E$ , associe un sous-ensemble  $S_a$  de  $S$ .

44. Cf. la notion de « structure d'information » chez Marschak et Radner [1972].

45. Simon [1972], p. 168. Sur la distinction générateur/évaluateur en intelligence artificielle, cf. aussi Newell, Shaw et Simon [1962] ; cet article ne comporte pas encore la distinction des deux « règles d'arrêt ».

de la difficulté. Les premiers ordinateurs joueurs d'échecs étaient fondés sur la puissance de calcul ; même en explorant un nombre immense de continuations par rapport aux possibilités de l'intelligence humaine (une centaine de continuations pour chaque coup chez les grands maîtres), ils étaient peu performants. Le programme de Newell, Shaw et Simon [1958] a été l'un des premiers à intégrer directement certaines caractéristiques réalistes du joueur humain <sup>46</sup>. Un des obstacles les plus difficiles, à ce point, que semblent avoir rencontrés les spécialistes d'intelligence artificielle concerne la *flexibilité* des processus de recherche humains. Le joueur d'échecs explore les continuations à des profondeurs très diverses ; pour certaines séquences, il se contentera d'une hypothèse portant sur une ou deux réponses successives de l'adversaire ; pour d'autres — celles qui comportent des « continuations forcées » ou qui lui semblent particulièrement prometteuses —, il analysera plusieurs branches d'un arbre comportant huit, dix réponses enchaînées. Cette flexibilité embarrassante pour les informaticiens <sup>47</sup> peut s'analyser comme un processus  $R_2$  dans lequel les deux aspects de calcul des conséquences et d'évaluation des positions interfèrent constamment. Si une continuation donnée  $s$  est fortement valorisée, dans un sens ou dans un autre, le joueur tendra vraisemblablement :  $\alpha$ ) à l'explorer sur une profondeur accrue ;  $\beta$ ) à explorer les configurations « voisines ».  $\alpha$ ) et  $\beta$ ) sont deux illustrations de ce que peut être, concrètement, le « raffinement » de la correspondance  $a \rightarrow S_a$ . Le point important, et difficile à modéliser, est que ce processus est déterminé, en chacune de ses étapes, par une évaluation des conséquences provisoirement constatées, évaluation qui peut reposer sur la fonction d'évaluation terminale  $V$  ( ), mais aussi bien sur des fonctions d'évaluation spécifiques distinctes d'elle <sup>48</sup>.

46. Voici, très schématiquement, comment Simon analyse cette involution. Après avoir tablé, sans succès véritable, sur la supériorité matérielle de la machine, l'intelligence artificielle s'est efforcée de reconstituer ce fait psychologique : dans le choix de ses coups, l'homme compte moins sur la mémoire à court terme (le nombre de continuations qu'il peut analyser) que sur l'appréciation globale des positions et le savoir acquis, c'est-à-dire la mémoire à long terme. On a pu montrer que les grands maîtres pouvaient « reconnaître » des dizaines de milliers de configurations du jeu, certaines limitées à deux pions, d'autres étendues à l'ensemble de l'échiquier — ce qui leur permet d'économiser les ressources (rares) de leur mémoire à court terme. Les programmes contemporains parviennent, dans une certaine mesure, à tirer parti à la fois de la puissance de la machine et de l'humanisation des procédures : « L'intelligence artificielle gardera vraisemblablement cette forme hybride pendant quelque temps. » ([1978 a], p. 497.)

47. Pour une présentation très concrète du problème de la flexibilité, appuyée sur l'exemple du jeu d'échecs et du « backgammon », cf. Berliner [1980].

48. Sur ce dernier point, cf. Newell, Shaw et Simon [1962].

## Relations entre les deux approches « simonienne » et « classique »

Nous examinerons trois manières de caractériser ces relations, avant de défendre le point de vue de l'irréductibilité réciproque : la distinction du court et du long termes ; celle de la nature interne ou externe des coûts de recherche ; enfin, celle du positif et du normatif.

### LA DISTINCTION DU COURT TERME ET DU LONG TERME

On peut se demander si les modèles de rationalité « limitée » ou « procédurale » ne convergent pas, dans des conditions ordinaires, vers des modèles optimisateurs traditionnels. Le problème a été soulevé notamment à propos du principe de satisfaction B, dont l'énoncé suggère deux mécanismes plausibles :

(M 1) Les données psychologiques sur le comportement individuel montrent que les aspirations tendent à s'ajuster sur le possible. Ainsi, d'après un argument courant, le niveau d'aspiration et le maximum que l'individu est susceptible d'atteindre se rapprochent à long terme. ([1966], p. 10.)

(M 2) A supposer que certaines firmes adoptent un principe de satisfaction (*satisfied*) elles devraient perdre progressivement du terrain par rapport aux entreprises maximisatrices, qui, faisant des profits plus élevés, grandiront plus vite que les autres. (*Ibid*).

La plausibilité de ces deux mécanismes a conduit Simon à justifier ainsi son approche : *même si* (M 1) et (M 2) sont avérés, ils ne sont guère significatifs, empiriquement, dans un monde en changement incessant. Tel est le sens de l'apologue du bol de mélasse que l'on peut lire dans l'article de 1966 : supposons que l'on s'intéresse au comportement d'un liquide visqueux — la mélasse — contenu dans un bol. Si, à la suite d'un choc initial, le bol est maintenu stable *et* que l'on se préoccupe seulement des positions d'équilibre, la loi de la gravitation (ici, la minimisation de la hauteur du centre de gravité du liquide) est la seule information théorique nécessaire. Si le bol est agité sans cesse *ou* que l'on s'intéresse aussi aux positions de court terme, il faut faire intervenir une théorie plus compréhensive que la loi de la gravitation, puisque des paramètres internes à l'organisme (viscosité de la mélasse) deviendront alors pertinents. La morale de cet apologue est double :

— une conception de la rationalité qui tienne compte des paramètres *internes* de la décision est indispensable, puisque, comme on le sait depuis Keynes, le court terme n'est pas exactement dénué d'importance ;

— quand bien même on ne s'intéresserait qu'au long terme, la variabilité de l'environnement social fait douter que les mécanismes (M 1) et (M 2) décrivent autre chose que des positions de référence théoriques ; l'approche de la rationalité limitée serait *aussi* nécessaire pour une théorie réaliste du long terme.

*Mutatis mutandis*, ces deux points apparaissent souvent, en macro-économie, dans les justifications méthodologiques que les keynésiens donnent de l'approche des « déséquilibres », mais ils ne sont pas toujours très clairement distingués l'un de l'autre. Par ailleurs, ils figurent souvent dans la littérature critique consacrée à la théorie de la sélection naturelle économique défendue par Alchian [1950] et Friedman ([1953], p. 22-23), dont (M 2) représente une version particulière<sup>49</sup>. Il serait intéressant de voir ce que deviennent les deux objections de Simon lorsqu'on revient à la source, c'est-à-dire à la théorie *biologique* de la sélection naturelle. La première n'aurait alors guère de sens : plus nettement que les sciences sociales, les théories biologiques peuvent revendiquer le fait de ne s'intéresser qu'aux positions de long terme<sup>50</sup>. La seconde objection demeurerait, en revanche, sous la forme d'un problème empirique. Comme l'écrit Elster :

Il n'y a pas d'objection logique qu'on puisse opposer à l'idée d'un monde où le taux de variation de l'environnement relativement au taux d'apparition des mutations est si élevé que la plupart des organismes sont, la plupart du temps, mal adaptés les uns aux autres et à leur environnement. ([1979], p. 7.)

C'est au défenseur de la théorie néo-darwinienne de faire la preuve que l'environnement des espèces animales change suffisamment lentement pour que les mécanismes sélectionnistes aient le temps de les porter à un degré d'adaptation convenable.

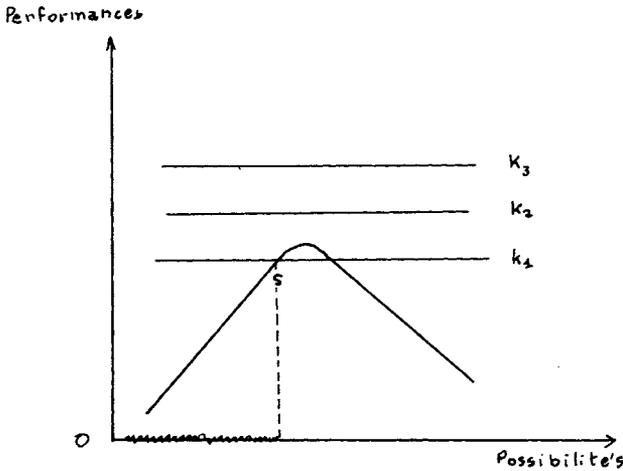
49. La théorie de la sélection naturelle économique (SNE) affirme que, quelle que soit la manière concrète dont les entreprises décident — « satisficing », optimisation ou tirage au sort des décisions —, l'élimination de certaines d'entre elles par la concurrence fera que l'économiste peut prédire les décisions de la branche à l'aide de la théorie optimisatrice traditionnelle. (M2) est plus particulier, puisqu'il concerne cette modalité concrète de la décision qu'est le « satisficing ». Comme il n'existe pas de facteur équivalent à la concurrence dans la théorie du consommateur, le mécanisme (M2) ne peut régler les rapports entre satisfaction et optimisation, dans le meilleur des cas, qu'à propos de l'*entreprise*.

50. Un biologiste qui applique la théorie des jeux à l'étude des populations animales a pu avancer une thèse provocante : *puisque ils sont des concepts d'équilibre, c'est-à-dire de long terme*, les concepts de la théorie des jeux sont plus immédiatement applicables à l'étude des populations animales qu'à celle des comportements humains (Maynard-Smith [1982], p. VII).

Les objections précédentes sur la *pertinence* des mécanismes (M 1) et (M 2) suffisent, aux yeux de Simon, à faire perdre tout intérêt autre que formel à ce qu'on pourrait appeler la thèse de convergence. Du coup, il ne s'est pas interrogé longuement sur les objections *logiques* éventuelles que l'on pourrait *aussi* élever contre (M 1) et (M 2). Cette attitude, qui peut passer soit pour bienveillante, soit pour méprisante — on en jugera par le ton de son article de 1979 —, nous semble présenter certains inconvénients pour l'approche de H. Simon. Car si les mécanismes envisagés sont logiquement admissibles, cette approche ne pourra peut-être pas dénier à la théorie traditionnelle une certaine valeur normative. Il s'agirait d'une *normativité au sens faible* — appréciée du point de vue de l'économiste, et non pas de l'agent : la théorie optimisatrice, dans cette optique, ne constituerait pas un modèle à donner en exemple à l'individu ; mais pour celui qui l'étudie, elle constituerait une référence théorique indispensable. Appliquée à l'équilibre de marché, cette idée a une longue histoire dans la littérature économique (par exemple Knight [1921]). On devrait certainement la prendre en considération s'il s'avérait qu'elle s'applique aussi à l'optimisation, et aux rapports entre optimisation et satisfaction. De ce point de vue, le désintérêt relatif de Simon pour des modèles du type (M 1) ou (M 2) peut l'amener à devoir faire plus de concessions qu'il ne l'envisageait initialement aux mérites de l'approche « classique » ; à tout le moins, il peut être conduit à s'engager dans un débat complexe sur l'antériorité logique d'une approche en termes de positions d'équilibre simplifiées par rapport à des approches plus concrètes.

Par chance, ce débat incommode peut être évité complètement : les mécanismes (M 1) et (M 2) appellent une série d'objections a priori. La thèse de la convergence est non seulement dénuée de pertinence, mais aussi logiquement récusée.

Concernant (M 1), il paraît évident que seules des hypothèses très restrictives peuvent imposer la convergence du modèle de satisfaction vers un modèle optimisateur. La raison principale en est que le mécanisme du « satisficing » est, par nature, séquentiel. Plaçons-nous dans le cas, apparemment favorable à la thèse de convergence, d'une révision du seuil à la baisse, en l'absence d'une solution satisfaisante antérieure. Si le processus de révision du seuil n'est pas continu, mais discret, il est presque certain que son ajustement ne conduira pas à retenir l'optimum comme première solution satisfaisante : dans le schéma ci-dessous, l'existence d'un « pas » de révision  $k_n - k_{n-1}$  et le caractère séquentiel de l'exploration des possibles (menée, par hypo-



thèse, à partir de l'origine) impliquent que le premier choix satisfaisant, S, est sous-optimal.

Le mécanisme (M2) soulève des problèmes plus intéressants, mais il est aussi logiquement douteux : on peut lui opposer tous les arguments qu'appelle l'utilisation hâtive, par Alchian et Friedman, de la théorie de la Sélection naturelle économique (SNE). Ces arguments remontent, pour l'essentiel, aux travaux de Winter ([1964], [1971]), [1975]), dont une citation permettra peut-être de faire saisir l'intention exacte :

Même s'il s'avérait que l'argument par la sélection ne conforte pas vraiment une théorie particulière de l'entreprise, il n'en reste pas moins que l'hypothèse suivant laquelle les entreprises doivent, à long terme, éviter de faire des profits négatifs afin de survivre a nécessairement ... *quelque* conséquence relativement au type de comportement qui se trouve être viable. ([1964], p. 236.)

Autrement dit, Winter ne remet pas en cause l'analogie biologique dont procède l'idée de la SNE. On pourrait, certes, faire ressortir le caractère bien approximatif de la relation qui unit l'hérédité aux règles de décisions, les mutations génétiques aux innovations et la survie à la réalisation d'un certain objectif de profits ; d'autres ont insisté sur la

fragilité logique de cette triple identification<sup>51</sup>. Winter, pour sa part, ne récuse pas le projet de faire une théorie de la SNE : acceptant l'analogie, il en explore les conséquences sur différents modèles de l'entreprise. Ses conclusions sont, dans l'ensemble, très défavorables à la thèse de l'optimisation et relativement favorables à la conception simonienne du « satisficing ».

Dans son modèle de 1964, Winter montre que l'idée intuitive selon laquelle « seuls survivent les optimisateurs » repose sur des conditions *nécessaires* particulièrement contraignantes : outre la condition ordinaire de stabilité de l'environnement, il faut imposer des mécanismes spéciaux d'entrée et de sortie des entreprises, notamment une condition de survie pour les entreprises quasi optimisatrices qui réaliseraient « temporairement » des profits négatifs, ainsi qu'une hypothèse de rendements d'échelle constants<sup>52</sup>. Par ailleurs, la convergence vers l'optimisation ne semble assurée en général que si l'on exclut l'apparition dans l'économie, au-delà de la période initiale, d'un nombre important d'entreprises sous-optimisatrices<sup>53</sup>. Quant au modèle simonien du « satisficing », Winter suggère, sans le montrer en détail, qu'à mettre les choses au pis, la théorie de la SNE ne lui est *pas plus défavorable* qu'elle ne l'est à la théorie classique ([1964], p. 264). Le modèle de P. A. Chiappori présenté dans ce numéro permet de préciser cette affirmation : il met en évidence des conditions *suffisantes*, relativement générales et plausibles, sous lesquelles, dans le long terme et à environnement inchangé, la proportion de « satisficers » ne s'annule pas.

51. Penrose [1952] suggère que le principe héréditaire (le gène) n'a peut-être pas d'équivalent en théorie de l'entreprise ; en tout cas, la règle de décision (« forme d'organisation » dans la théorie de Winter) est sans doute le moins mauvais candidat. S'ils ne discutent pas l'analogie de la mutation et d'une innovation (au sens d'une règle de décision modifiée), Elster [1979] et Chiappori (ce numéro) insistent sur l'étroitesse du stock de mutations potentielles dont dispose la population des entreprises, au contraire des populations animales. Par ailleurs, le problème des liens entre le caractère adaptatif des règles de décision et l'*adoption* aveugle, par le milieu, des règles qui réussissent, demeure entier (cf. Elster [1979], p. 1-4 et 9-12). On peut y voir une objection supplémentaire à la transposition des concepts de principe héréditaire ou de gène. Une quatrième analogie apparaît dans la version la plus strictement néo-darwinienne de la thèse de Friedman-Alchian : l'*avantage comparatif* a pour répondant la capacité différentielle des entreprises à s'accroître. Ce n'est pas l'aspect le moins discuté de cette thèse (cf. [Winter], 1971, p. 246).

52. Cette dernière hypothèse vise à éviter qu'une entreprise faiblement performante puisse coexister avec une entreprise quasi optimisatrice, chacune opérant sur des portions différentes de sa fonction de production. Cf. Winter [1964], p. 242-245.

53. « Toutes les entreprises nées après  $t_0$  ont des formes d'organisation du type maximisateur » (*ibid.*, p. 243). On peut se demander s'il s'agit véritablement d'une condition nécessaire.

Ces résultats ne constituent assurément que des applications particulières de la théorie de la SNE : comme théorie générale, celle-ci est encore à naître. Suivant l'intérêt que l'on prête aux objections avancées contre la précarité de l'analogie, on peut approuver ou non les efforts entrepris par Winter et d'autres pour aller au-delà des modèles spéciaux<sup>54</sup>. Pour notre propos, en tout cas, ceux-ci suffisent amplement. Ils établissent que (M 2) ne constitue pas une loi vraie universellement, ni même un mécanisme « plausible » en un sens lâche<sup>55</sup>.

#### LA DISTINCTION DES CONTRAINTES INTERNES ET EXTERNES

Suivant ce point de vue, la spécificité de l'approche simonienne viendrait de ce qu'elle intègre des coûts internes, comme ceux qui s'attachent à une exploration de l'ensemble des actions et de l'ensemble des conséquences. Voici comment Simon a pu définir la rationalité limitée (bounded) : « On peut désigner comme théories de la rationalité limitée celles qui incorporent des contraintes sur la capacité de l'acteur à traiter l'information » ([1972]). Suivant cette caractérisation, la différence entre modèles classiques et modèles de rationalité limitée tiendrait finalement à ceci : les premiers font coïncider l'énoncé des contraintes avec la description de l'environnement, l'énoncé des préférences avec la description du sujet ; les seconds refusent d'identifier la dichotomie contraintes/préférences à la dichotomie environnement/sujet, puisqu'ils créent la catégorie nouvelle de *contraintes internes*.

Une telle manière de présenter le débat a pour défaut de renvoyer à une version aujourd'hui dépassée de la théorie classique. Les modèles de recherches nés de l'article célèbre de Stigler [1961], voire les développements que Becker [1976] a donnés à la micro-économie traditionnelle, devraient permettre à celle-ci de tenir compte des contraintes

54. Cf. Winter [1975].

55. Rétrospectivement, on ne saurait s'étonner de cette conclusion négative. La lecture des articles classiques de Friedman et d'Alchian montre qu'il leur manquait des distinctions essentielles à une théorisation, même rudimentaire, de la SNE. Par exemple, ils semblent mêler constamment deux analyses de ce que doit être l'analogie, en économie, du gène (ou du principe héréditaire) : il s'agit tantôt de l'action particulière décidée à chaque instant, tantôt de la  *règle de décision*  (« forme d'organisation » dans la terminologie de Winter), qui conduit l'entreprise à engager telle ou telle action. On perçoit intuitivement que la première version (qui inclut, par exemple, le cas où les décisions d'entreprises sont purement aléatoires : Friedman [1953], p. 35) est plus difficile à défendre que la seconde : cf. [Winter], 1971, p. 244. Il serait facile de montrer, d'autre part, que l'article d'Alchian identifie incorrectement le raisonnement mené sur les cohortes d'individus d'âge identique et l'analyse en coupe de l'ensemble de la population totale à une date considérée.

internes à l'agent. Lévy-Garboua a très bien résumé l'esprit de ces travaux en proposant (i) de distinguer la « cohérence » d'un comportement et son « optimalité » relativement à l'environnement ; (ii) d'analyser cette « cohérence » à l'aide des outils de la théorie traditionnelle : un comportement sub-optimal relativement à un environnement donné est « cohérent » (c'est-à-dire rationnel) si l'on parvient à le décrire comme *optimal relativement à un ensemble plus large de contraintes*, incluant des coûts de recherche internes<sup>56</sup>. Cette double prise de parti soulève un problème délicat. Il n'y a guère de doute que Simon accepterait volontiers (i), la distinction entre « cohérence » et « rationalité ». Il approuverait probablement aussi cette manière de la justifier : « Si les économistes ont pu considérer la formation du choix comme une activité non coûteuse, c'est qu'ils assimilaient l'expérience et l'action, et ignoraient le rôle des ressources humaines<sup>57</sup> » A cet égard, les travaux de Stigler, Becker, Lévy-Garboua, ressortissent à la catégorie des modèles de rationalité limitée. Dans son article de 1972, Simon le concédait facilement à propos de Stigler<sup>58</sup>. Mais s'il en va ainsi, les modèles simoniens sont-ils vraiment indispensables ? De deux choses l'une, en réalité : ou bien Simon accepte la parenté de ses modèles et de l'approche classique du « search », et il lui sera difficile de faire obstacle au réductionnisme bayésien ; ou bien il maintient fermement le point de vue dualiste, et il doit alors expliciter ce qui est inadéquat dans la manière classique de représenter la rationalité limitée.

Simon a choisi sans équivoque la dernière branche de l'alternative lorsque, récemment, il s'est prononcé contre la conception post-stiglierienne du « search » :

Stigler a versé le vin neuf de la théorie de la recherche dans les vieilles outres de la maximisation de l'utilité, puisque le coût de la recherche est égalisé avec son rendement marginal... Mais la maximisation de l'utilité n'était en rien essentielle au schéma de la recherche — heureusement d'ailleurs, car elle aurait exigé que le décideur fût capable d'estimer les coûts et rendements marginaux de la recherche dans un contexte de décision qui était déjà trop complexe pour que s'y applique une rationalité globale. ([1979], p. 503.)

Cette prise de position vaut, plus généralement, contre la thèse (ii) de Lévy-Garboua. Il faut tenter de l'analyser précisément.

56. Lévy-Garboua [1983]. Ces deux thèses méthodologiques sont illustrées par un modèle de formation des choix dans lequel l'agent découvre séquentiellement ses propres préférences (Lévy-Garboua [1979]).

57. Lévy-Garboua [1979], p. 101.

58. Simon [1972], p. 163.

Le modèle initial de Stigler décrivait la façon dont un consommateur décide d'acheter une unité d'un bien sur un marché caractérisé par des prix multiples. Par hypothèse, le consommateur connaît la fonction de répartition des prix  $F(p)$  ; chaque observation  $k$ , par exemple à cause de la distance entre les magasins, présente un coût  $c_k > 0$ ,  $\forall k$ . On sait que la règle initialement retenue par Stigler — qui consiste à fixer au début de la recherche le nombre d'observations à faire — est en réalité sous-optimale. On a démontré que, sous les conditions de l'article de 1961, qui sont d'ailleurs relativement naturelles, il est optimal, pour le consommateur, de définir après chaque observation une *valeur de seuil*  $r_k$  en dessous de laquelle il décidera d'acheter (éventuellement, en retournant dans un magasin déjà visité)<sup>59</sup>. Les  $r_k$  sont en général variables avec  $k$  (ils sont constants lorsque  $c_k = c$ ,  $\forall k$ ) ; en outre, ils sont toujours strictement supérieurs au prix le plus faible auquel  $F(p)$  assigne une valeur positive, ce qui veut dire qu'il est optimal de ne pas acquérir une information parfaite<sup>60</sup>.

Ces résultats nous semblent d'un très grand intérêt parce qu'ils mettent le dualisme dans une position inconmode : voici un modèle dont l'inspiration est typiquement classique, mais qui permet de retrouver, apparemment sans difficulté, les deux principes simoniens de « satisficing » et de « search » ; il résout même de façon élégante la question de l'endogénéisation du seuil de satisfaction. Plus précisément, que deviennent, dans ce nouveau contexte, les processus de recherche  $R_1$  et  $R_2$  précédemment définis ?

Si l'on associe à chaque prix  $p_k$  successivement découvert par l'agent *l'action* : « acheter au prix  $p_k$  », on peut interpréter la recherche du consommateur stiglierien comme un processus du type  $R_1$ . Si, au contraire, on choisit de regarder l'action du consommateur comme fixée (= acheter une unité d'un produit donné), la recherche portera non plus sur l'espace des actions, mais sur celui des *conséquences* (= les coûts impliqués par son achat), et l'on a affaire à un processus du type  $R_2$ . Dans l'un et l'autre cas, la théorie de l'échantillonnage optimal permet d'expliquer l'arrêt du processus : dans la première interprétation, elle permet d'expliquer le seuil  $r_k$ , en le faisant dépendre des

59. Cf. Marciano [1982], qui fait une bonne revue des résultats obtenus à partir d'un modèle de Stigler. Rappelons les conditions posées dans ce modèle : le consommateur achète une unité et une seule du produit ; il connaît la loi  $F(p)$  ; il peut, à tout instant, acheter au prix le plus bas rencontré jusqu'alors (hypothèse dite du « perfect recall ») ; il est neutre par rapport au risque ; le coût moyen des observations est non décroissant ( $c_{k+1} \geq c_k$ ) et  $c_1 > 0$ . La théorie du « search » s'est particulièrement attachée à généraliser les deux premières conditions.

60. Marciano [1982], p. 19.

coûts de recherche et des réalisations précédentes ; dans la seconde interprétation, elle endogénéise la nature exacte de la transformation subie, pour une  $a$  donnée, par le sous-ensemble  $S_a$  initialement associé à cette action<sup>61</sup>. Ainsi,  $R_1$  et  $R_2$  se laissent apparemment formaliser dans un contexte typiquement classique. La théorie de l'échantillonnage optimal peut même prétendre à une généralité plus grande que le modèle simonien : elle montre que certaines méthodes de révision du seuil, et non pas d'autres, certaines manières de « raffiner » l'information initiale, et non pas d'autres, sont appropriées<sup>62</sup>. La conception simonienne de la rationalité limitée ne serait rien d'autre que l'heuristique des modèles post-stiglériens.

En réalité, cette variante (sophistiquée) de la thèse réductionniste est inadmissible pour trois ordres de raisons.

α) Simon reproche tout d'abord aux modèles classiques du « search » de poser une exigence irréaliste sur l'information initiale de l'agent. Dans le cas particulier de Stigler, le consommateur doit connaître la fonction de répartition  $F(p)$  et le coût  $c_k$  de chaque observation. Mais ces données sont généralement inaccessibles. Il en résulte que le modèle de Stigler n'a pas d'intérêt en dehors d'une approche explicite en termes de « comme si »<sup>63</sup>. Plus gravement peut-être, il est trivial : tout le problème de la théorie de la décision consiste à montrer comment les individus procèdent à des choix rationnels alors même qu'ils disposent d'une information initiale très faible. Présupposer que le consommateur a déjà estimé une distribution de probabilité et des coûts de recherche revient, bien loin de faire progresser la théorie de la décision, à en supprimer l'objet<sup>64</sup>.

β) Une autre objection porte sur la complexité des calculs d'optimisation impliqués dans les modèles classiques. A supposer même que l'on concède à Stigler ses hypothèses initiales, le consommateur devrait calculer effectivement les valeurs de seuil  $r_k$ . Simon doute qu'il soit en mesure de le faire. Contrairement à ce que l'on croit parfois, une telle objection ne procède pas d'un empirisme naïf. Simon ne cherche pas essentiellement à rappeler, après beaucoup d'autres que, « dans le monde réel », consommateurs et chefs d'entreprise ne savent pas

61. C'est évidemment le support de la distribution de  $F(p)$  qui joue ici le rôle de  $S_a$ .

62. Il n'est pas toujours vrai qu'un modèle de recherche avec coûts de recherche et optimisation conduise à définir un seuil de satisfaction, même variable ; cf. Radner, 1974.

63. Cf. Winter [1964], p. 250.

64. Cf. Simon [1955], p. 254 ; [1978], p. 10 ; [1979], p. 497-498.

dérivée des lagrangiens. Il défend plutôt une thèse méthodologique : un économiste doit respecter, dans la modélisation, certaines règles de *cohérence entre les capacités qu'il prête aux agents et l'état actuel de la science*. Il n'est pas cohérent de supposer, chez un consommateur, l'aptitude à déterminer instantanément et sans coût une solution que le modélisateur lui-même ne sait pas toujours calculer, même en s'aidant des plus gros ordinateurs. Bref, chaque fois que l'optimisation représente un coût non négligeable, eu égard aux techniques actuelles du calcul, il faut tenir compte, de quelque manière, de ce coût <sup>65</sup>. Et les modèles post-stiglériens accroissent sensiblement les difficultés de calcul par rapport aux modèles optimisateurs de certitude parfaite, qui, déjà, ne sont pas toujours aisément calculables <sup>66</sup>.

γ) Il nous semble que l'on peut avancer une ultime objection, valable, celle-ci, non pas contre la littérature du « search » en tant que telle, mais contre sa possible transposition à des processus généraux du type  $R_1$  et  $R_2$ . Le propre d'un modèle comme celui de Stigler est qu'il représente un agent préalablement doté de certaines caractéristiques internes, bien définies et affecté par une incertitude dont la source est uniquement *externe*. Si l'on veut décrire un processus de recherche au sens général où Simon l'entend, c'est-à-dire portant aussi bien sur des caractéristiques *internes* que sur des états de l'environnement, une transposition mécanique du modèle de Stigler rencontre une objection de circularité. Supposons que pour une action donnée  $\bar{a}$ , un agent procède à une recherche  $R_{\bar{a}}$ , c'est-à-dire « raffine » progressivement le sous-ensemble de conséquences associé  $S_{\bar{a}}$ . Si l'on veut endogénéiser le sous-ensemble  $S^*_{\bar{a}}$ , éventuellement réduit à un point unique, sur lequel  $R_2$  s'arrêtera, et que l'on recoure aux outils de la littérature du « search », on devra supposer donnés, ou calculables par l'agent, les coûts et gains marginaux espérés en chacune des étapes du processus de réduction du sous-ensemble associé. Par exemple, si l'agent est parvenu au sous-ensemble  $S'_{\bar{a}}$ , il devra connaître le gain attendu d'une réduction supplémentaire à  $S''_{\bar{a}} \subset S'_{\bar{a}}$ . Mais pour calculer ce gain espéré, il faudrait qu'il ait *déjà* évalué, à l'aide de sa fonction d'estimation  $V(\cdot)$ , les conséquences contenues dans  $S'_{\bar{a}} - S''_{\bar{a}}$  : cela supposerait qu'il connaisse *déjà* ces conséquences, alors que, par hypothèse, elles lui sont inconnues. Il y aurait donc circularité à appliquer le calcul marginal ordinaire à

65. Cf. Simon [1978], p. 11-12 ; [1978 a].

66. L'introduction de l'incertitude, en particulier, complique nécessairement les calculs ; c'est l'une des objections que Simon avance contre l'usage de la théorie de l'utilité espérée aussi bien que de la théorie des jeux ([1955], p. 246 ; [1972] ; [1978] et [1979]).

l'endogénéisation du point d'arrêt d'un processus général de type  $R_2$ . On peut, à l'évidence, construire un raisonnement analogue sur un processus  $R_1$ .

LA DISTINCTION DU POSITIF ET DU NORMATIF

Nous avons déjà considéré ci-dessus (p. 40) l'un des sens du mot « norme » : la position d'équilibre ou de long terme d'un processus. Il s'agira maintenant de normativité au sens fort : on dira qu'un modèle est normatif non seulement s'il constitue une référence théorique utile à l'observateur, mais s'il représente un processus de décision qui peut être donné en exemple à l'agent. C'est le sens de ce mot que Savage et Harsanyi ont en vue lorsqu'ils défendent le point de vue réductionniste (cf. p. 15).

Beaucoup d'auteurs chercheront, justement, à transposer aux modèles de Simon l'attitude que Savage croit pouvoir adopter face au paradoxe d'Allais : il s'agirait, encore une fois, de ces curiosités empiriques non probantes, qui témoignent, au mieux, de l'irrationalité effective des choix, et non pas de l'existence d'un substitut à la théorie de l'utilité espérée. Quant on en vient à la question de savoir ce que doit être un comportement, celle-ci demeure irremplaçable.

Il y a deux manières de défendre cette position, l'une fruste, l'autre sophistiquée, et aucune des deux ne nous semble probante. La première consiste à faire disparaître la considération des coûts de recherche de la définition normative de la rationalité. C'est la présence de ces coûts internes, et elle seule, qui justifie l'usage de procédures séquentielles manifestement sous-optimales comme la méthode des buts intermédiaires ou celle des seuils de satisfaction : le point important tient dans le « et elle seule ». Si l'on peut faire abstraction de ces coûts lorsqu'on définit la norme, l'approche simonienne ne décrit pas seulement une *rationalité avec des moyens limités*, mais une *rationalité limitée* au sens strict. Il n'est pas indifférent que Simon lui-même, dans ses premiers travaux, ait été attiré par cette position <sup>67</sup>.

67. Cf. : « C'est précisément à cause de ces limitations sur les connaissances et les capacités de l'organisme que les modèles de rationalité moins globaux (less global models of rationality) présentés ici sont significatifs et utiles. La question de savoir comment on se conduit "rationnellement", étant donné ces limitations, est distincte de la question de savoir comment on pourrait améliorer les capacités de l'organisme afin de rendre possible une action qui serait plus "rationnelle", vue depuis les hauteurs d'un modèle plus complet (judged from the mountain-top of a more complete top). » ([1955], p. 254.)

Aujourd'hui, Simon la récuserait sans doute tout à fait. Il rappellerait cette considération simple et, nous semble-t-il, indiscutable, qu'une norme qui est totalement (ou même seulement : très généralement) irréaliste est une fiction dénuée d'intérêt. Des auteurs qui, dans leurs travaux, ont privilégié l'approche classique, comme Radner, se retrouveraient aisément avec lui sur ce point : « L'existence de limitations (dans les capacités cognitives) pose de sérieux problèmes au concept du choix rationnel formulé en termes de comportement optimisateur. Un comportement qui n'est pas même *réalisable* (feasible) peut difficilement passer pour *rationnel* » (Radner [1972], p. 17). De là une deuxième argumentation, sophistiquée celle-ci, en faveur du caractère normatif des modèles économiques traditionnels : ceux-ci sont normatifs justement parce qu'ils tiennent compte, dans les contraintes de la décision, des coûts *internes* qui contribuent à en restreindre la liberté, à côté et en sus des contraintes imposées par l'environnement.

Mais contre cette argumentation sophistiquée, on ne peut que reprendre les objections détaillées ci-dessus (p. 46-47) sur le traitement, par la théorie classique, de ces coûts internes. Soit, par exemple, la théorie de l'échantillonnage optimal : il est vrai qu'elle permet, en un sens, de dire ce qu'est un « bon » processus de « satisficing ». Mais cette normativité est usurpée, car elle repose sur l'hypothèse d'une capacité cognitive des agents supérieure à ce qui est réalisable. L'argument de Radner contre la variante fruste de la thèse de normativité s'applique aussi bien à sa variante sophistiquée.

En résumé, l'approche de Simon ne se ramène ni à une conception de la rationalité à court terme, ni à une approche strictement descriptive, par opposition à une théorie classique constituée en norme <sup>68</sup>. Il n'est pas non plus vrai que le traitement simonien des contraintes internes à la décision soit un cas particulier d'une théorie classique élargie, susceptible d'intégrer les principes de recherche et de satisfaction. En bref, les deux approches mises en relation dans cette section sont irréductibles l'une à l'autre.

68. Cf. Simon [1979], p. 510 : « Si nous sommes intéressés par une théorie de la décision descriptive, voire même normative, il est devenu parfaitement clair que les théories classiques et néo-classiques ont été supplantées par une variante qui leur est supérieure : elle nous donne une approximation bien meilleure de ce qui arrive effectivement. » (Souligné par nous.)

## Justification du point de vue dualiste

Le but de cette section est de faire servir l'approche de Simon comme un contre-exemple significatif à la thèse réductionniste sur la rationalité : il s'agit de montrer qu'il existe une famille de modèles : ( $\alpha$ ) logiquement irréductibles à l'approche bayésienne ; ( $\beta$ ) portant néanmoins sur le choix rationnel ; ( $\gamma$ ) permettant de rendre compte d'une situation empirique de choix, au moins, que la modélisation classique ne parvient pas à traiter correctement. Prises ensemble, les trois conditions sont, à l'évidence, nécessaires et suffisantes pour qu'il y ait matière à parler de contre-exemple. Nous avons fait porter l'accent sur ( $\alpha$ ) qui est le seul point problématique. ( $\beta$ ) et ( $\gamma$ ) découlent en effet assez directement de considérations, qu'il faut développer maintenant, sur *l'échec de l'optimisation* <sup>69</sup>.

On se rappelle que Simon reproche aux modèles du « search » et, plus généralement, à toutes les variantes de la maximisation de l'utilité espérée, de négliger deux types de coûts : les coûts nécessaires à l'obtention d'une information (par exemple, la distribution  $F(p)$  chez Stigler) que cette approche considère, trivialement, comme donnée, mais dont il faudrait indiquer la genèse chez l'agent ; et les coûts de calcul au sens étroit, liés à la détermination d'une solution optimale, étant donné l'information exogène. Prendre en compte ces différents coûts conduira, en général, à déclasser l'action qui a été retenue comme optimale en leur absence. On doit donc distinguer deux niveaux logiques,  $N_0$  auquel les « coûts simoniens » sont considérés comme nuls, et  $N_1$  auquel ils sont pris en considération : une action optimale pour  $N_0$  ne le sera pas, en général, pour  $N_1$ .

La proposition des auteurs qui, avec Becker ou Lévy-Garboua, s'efforcent d'enrichir la théorie classique, en y faisant figurer des « coûts simoniens », revient essentiellement à redéfinir l'optimisation en  $N_1$  plutôt qu'en  $N_0$  : on élargit l'ensemble des contraintes sans renoncer, pour autant, à l'optimisation elle-même. Cette proposition néglige le fait qu'une telle optimisation « au niveau supérieur » implique, à son tour, des « coûts simoniens » : elle suppose donné un nouvel ensemble de paramètres, comme, par exemple, les coûts de recherche internes, que, trivialement, elle se dispense d'engendrer ; cette information exo-

69. Ce développement présente de manière informelle une argumentation de Chiappori et Mongin [1983 a]. L'idée que l'optimisation impliquerait une forme de régression à l'infini apparaît occasionnellement dans les travaux de Winter sur la sélection naturelle ([1964], [1971], [1975]).

gène une fois donnée, elle repose sur une détermination des solutions optimales qui sera presque certainement plus complexe que l'opération analogue du niveau  $N_0$ .

Si l'on devait prendre en compte ces nouveaux « coûts simoniens », le nouvel optimum n'en serait évidemment plus un ; il faut passer au niveau  $N_2$ , où ils seront, à leur tour, intégrés au calcul. Nous ne connaissons pas de travaux qui se soient encore situés à ce « deuxième niveau supérieur » ; mais il n'est pas improbable qu'ils apparaissent un jour et soient même salués par certains comme réalisant un progrès significatif par rapport aux modèles de la génération précédente. Pourtant, ils seraient exposés à une objection exactement similaire sur les « coûts simoniens » (d'ordre trois) qu'ils négligent.

On devine maintenant ce que nous entendons par « échec de l'optimisation ». De deux choses l'une, en effet : ou bien l'on parle d'une action « optimale », sans autre précision, mais alors il faut démontrer que cette action est optimale *pour tous les niveaux envisageables*, c'est-à-dire que la prise en compte de « coûts simoniens » appartenant à un niveau logique supérieur *quelconque* ne modifie pas l'identité de l'action optimale ; ou bien l'on choisit d'arrêter arbitrairement la régression des niveaux logiques à une valeur  $N_n$ , ce qui revient à négliger purement et simplement les « coûts simoniens » d'ordre  $n + 1$ ,  $n + 2$ , ...

Les deux branches de l'alternative sont également inconfortables. La première n'implique sans doute pas de contradiction logique intrinsèque : aucune considération a priori ne permet d'exclure qu'une action qui est solution d'un programme d'optimisation pour  $N_0$  puisse aussi l'être pour tous les niveaux supérieurs,  $N_i$ ,  $i = 1, 2, \dots$ <sup>70</sup>. La difficulté est en fait d'ordre empirique : il ne peut en aller ainsi pour toute valeur numérique de l'action et *tout* contexte de coût. Un contre-exemple fondé sur l'hypothèse de « coûts simoniens » croissant avec le niveau logique considéré (ce qui est raisonnable) suffit à établir ce point. Ainsi, la première branche de l'alternative ne permet pas de sauver l'optimisation *en général*. Il en est de même de la seconde branche, pour d'autres raisons. Exclure la considération des coûts d'ordre  $n + 1$  rend possible de caractériser une action comme optimale pour  $N_n$  ; mais, *en général*, elle n'est pas optimale pour les autres niveaux logiques, inférieurs ou supérieurs à  $N_n$ .

70. Contrairement à ce que suggère Winter [1975], p. 83 : « Il semblerait qu'une optimisation dont la portée s'étend à toutes ces considérations (de coût), y compris ses coûts propres, puisse comporter les difficultés logiques de l'auto-référence. »

La littérature économique courante a choisi très nettement la seconde branche de l'alternative. En fait, elle s'en est tenue jusqu'à présent à la considération des seuls niveaux  $N_0$  et  $N_1$ . S'élever dans la hiérarchie des niveaux logiques, nous l'avons suggéré, ne changerait rien à l'affaire. Les défenseurs de l'optimisation à l'ordre  $n$  n'en resteront pas moins, à leur insu, des défenseurs d'une *autre* théorie que l'optimisation à l'ordre  $n + 1$ . Quelle est cette autre théorie ? Une solution optimale à l'ordre  $N_n$  ne peut être, vue de  $N_{n+1}$ , que « satisfaisante ». Winter a été le premier, semble-t-il, à avancer cette idée provocante : « A un certain niveau d'analyse tout comportement dirigé vers un but (goal-seeking) est un comportement conforme au principe de satisfaction (satisficing). » ([1964], p. 264.)

Cette analyse de l'échec de l'optimisation a deux conséquences importantes pour notre propos. D'une part, elle montre qu'il doit exister *quelque* situation au moins pour laquelle l'optimisation échoue, ou plutôt ne sera, en fait, qu'une recherche déguisée de satisfaction. D'autre part, elle permet de rendre précise l'idée, déjà avancée à la fin de la rubrique précédente, que les modèles optimisateurs n'ont pas de privilège intrinsèque en matière de normativité. C'est un argument traditionnel des partisans de l'approche classique que « rationnel », au sens de « normativement rationnel », doit signifier : « optimisation de quelque variable »<sup>71</sup>. Or la discussion précédente rend impossible d'identifier ces deux concepts : elle pose un problème inédit, et sans doute, insoluble, au réductionnisme. Plus généralement, elle contribue à rendre incertaine la notion même de rationalité. Le lien entre optimisation et rationalité étant dénoué, nous sommes renvoyé, pour la définition de celle-ci, aux idées du sens commun : est rationnel un choix qui est approprié aux fins de l'individu et aux conditions dans lesquelles il se trouve. Et par rapport à une telle définition, les modèles simoniens sont aussi « rationnels » que ceux de l'approche classique.

71. Cf. Becker [1976], p. 5-8 ; Lévy-Garboua [1983].

## CONCLUSION

Le dualisme du modèle *générique* de la rationalité et de sa *spécification* « économique » étant établi, il faut s'efforcer d'en tirer les conséquences méthodologiques. Tout d'abord, quelle formulation convient-il de retenir pour le modèle rationnel lui-même ? Ce problème est moins important que celui de la justification du modèle rationnel : il nous semble qu'un grand nombre des arguments traditionnellement avancés en faveur du « principe de rationalité », de Max Weber à Robbins et Popper, présupposent une interprétation trop étroite de ce principe ; que, pour le dire schématiquement, elles sont mieux adaptées à la variante « économique » du principe lui-même. Il convient donc de reconsidérer ces justifications à la lumière de la position dualiste. Nous ne prétendons pas mener ici ce travail en général, mais seulement ouvrir la discussion sur l'un des arguments les mieux reçus en faveur de l'approche par la rationalité — l'idée que celle-ci permettra de faire apparaître la réalité des comportements *par différence* en quelque sorte, suivant ce que Weber nomme « méthode idéal-typique » et Popper, de façon plus suggestive, « méthode zéro ». Que reste-t-il de la « méthode zéro » une fois qu'on dissocie clairement la formulation générique du principe de rationalité et ses variantes spéciales, comme celle de la théorie économique, de la théorie des jeux et de la « rationalité limitée » ?

### Formulation d'un principe générique de rationalité

Deux candidats s'affrontent, qui ont chacun des titres bien établis dans la littérature méthodologique. Les formulations en termes de fins et de moyens ont un passé respectable, puisqu'elles remontent à l'analyse aristotélicienne du choix<sup>72</sup>. Et Weber a pu écrire cette phrase, que T. Parsons a ensuite placée en exergue de son ouvrage le plus célèbre : « Toute analyse réflexive (*denkende Besinnung*) concernant les éléments ultimes de l'activité humaine significative (*sinnvoll*) est

72. *Ethique à Nicomaque*, V, 3. Aristote distingue les actions volontaires et les choix, qui en sont des cas particuliers ; il identifie la sphère du choix à celle de la délibération, qu'il analyse dans des termes très proches de ceux que Simon emploiera pour décrire la méthode des « subgoals ». Aristote est non seulement l'ancêtre des formulations génériques en termes de fins et moyens, mais un théoricien de la rationalité « procédurale ».

tout d'abord liée aux catégories de la "fin" et des "moyens" »<sup>73</sup>. Pour être plus récents, puisqu'ils remontent approximativement à la fin du XIX<sup>e</sup> siècle, les modèles contraintes/préférences n'en ont pas moins créé à leur tour une tradition dans la philosophie de la rationalité : ces modèles ont cessé d'être l'apanage de l'économie ; un nombre croissant d'auteurs qui veulent traiter des problèmes généraux de la rationalité choisissent de s'y référer plutôt que d'utiliser l'antique conceptualisation par fins et moyens<sup>74</sup>. Y a-t-il quelque raison, à ce stade, de privilégier l'un des deux cas candidats ?

Nous avons rappelé l'argumentation de J. Harsanyi, suivant laquelle un modèle fins-moyens est *toujours* un cas particulier d'un modèle préférences-contraintes qui intégrerait le coût d'opportunité de la réalisation de chaque fin : fixer un but, et chercher ensuite les moyens de le réaliser, comme dans la méthode des « subgoals », c'est négliger le fait que la réalisation de ce but prédéterminé nuira à celle des autres buts ; c'est, dans les termes de la théorie économique, ignorer un coût d'opportunité, qu'un modèle préférences-contraintes fait entrer en ligne de compte, puisqu'il endogénéise les fins aussi bien que les moyens<sup>75</sup>. Cette objection n'a pas la portée que lui prête Harsanyi : elle concerne, en fait, deux modèles spécifiques, celui des « subgoals », d'une part, le modèle contraintes-préférences de la théorie économique traditionnelle, d'autre part. Mais il existe d'autres modèles fins-moyens que celui des « subgoals » (dont nous avons dit qu'il était particulièrement fruste : par exemple, il ne s'applique pas vraiment à la réflexion du joueur d'échecs), et d'autres modèles contraintes-préférences que celui de la théorie économique, des modèles où n'interviendrait pas la même possibilité indéfinie de comparaison et d'ajustement à la marge. A supposer, donc, qu'elle dise quelque chose sur la relation de ces deux modèles particuliers, l'objection de Harsanyi ne peut servir à justifier une formulation du *principe de rationalité lui-même* en termes de contraintes et préférences plutôt qu'en termes de fins et moyens.

Il y a, en fait, quelque raison de penser que la conceptualisation de Weber-Parsons est, sous un certain angle, la plus générale des deux : le problème de *tous* les modèles contraintes-préférences est qu'ils supposent, à des niveaux logiques variables suivant leur degré de sophistication, un ensemble de possibilités qui est *donné* à l'agent. Ceci les rattache sans équivoque possible à la classe des modèles de

73. Weber [1904], p. 149 ; trad. franç. 1965, p. 123 (nous la modifions pour « sinnvoll ». Parsons [1937].

74. Par exemple, Elster emploie une version modifiée de l'approche préférences-contraintes ([1979], p. 113-115).

75. Cf. *supra*, p. 29.

rationalité « substantielle ». Ils seront incapables de décrire de façon intéressante le processus de délibération ; ils ne pourront, au mieux, qu'en simuler les résultats. Une formulation du principe de rationalité en termes de fins et de moyens n'est pas sujette à une pareille limitation (à condition, naturellement, de ne pas l'identifier au modèle restrictif des « subgoals »). Il nous semble qu'elle préserve l'ambiguïté essentielle à la notion de choix rationnel : un choix rationnel l'est par son résultat, en ce sens qu'il débouche sur une *action* appropriée ; mais il l'est aussi par ce dont il provient, par le caractère approprié de la *délibération* qui le précède. La rationalité « substantive » privilégie unilatéralement le premier point de vue, la rationalité « procédurale » met en valeur le second. Les formulations du type parsonien ou webérien, par fins et moyens, peut-être à cause de leur vague intrinsèque, ont l'avantage de ne pas obliger à prendre parti et de préserver, ainsi, le caractère générique du principe de rationalité.

La trivialité de ces formulations apparaît finalement comme un avantage plutôt que comme un inconvénient. On peut la justifier par un autre biais. Les modèles de rationalité ont une finalité empirique et doivent faire l'objet de tests comparatifs. La meilleure stratégie, pour ce faire, semble être d'« immuniser », en quelque sorte, le principe de rationalité lui-même, et de faire porter le poids des réfutations éventuelles sur les aspects spécifiques des modèles. On fera servir les tests à rendre plus précises, et peut-être même solubles, des questions comme : l'agent procède-t-il séquentiellement ou globalement ? un allègement des coûts de calcul (apprécié par l'utilisation des ordinateurs) conduit-il à modifier substantiellement les décisions ? Cette méthode s'impose, pour des raisons assez évidentes, par rapport à celle que ferait porter le test *à la fois* sur le principe de rationalité et sur sa variante spécifique : une réfutation pourrait, alors, toujours s'interpréter comme portant sur le principe de rationalité lui-même ; de l'expérience, on ne tirerait plus que cette information indéterminée, inutilisable — les hommes ne sont pas rationnels <sup>76</sup>. Si, donc, il y a de bonnes raisons méthodologiques pour adopter une stratégie d'« immunisation » du principe de rationalité, il faut, aussi bien, choisir une formulation du principe cohérente avec cette stratégie. Or seule une formulation *triviale* peut convenir : il va de soi que, si le principe doit être protégé de la réfutation, il ne doit pas, non plus, être trop déterminé dans son contenu <sup>77</sup>. Il faut éviter par-dessus tout de faire bénéficier de la pro-

76. Cette argumentation est développée in Popper [1967], p. 147 ; Watkins [1970], p. 173-174 ; Boyer [1983], p. 313 et suiv.

77. Watkins, *ibid.*, a bien mis en valeur ce point.

tection une formulation qui serait déjà spécifique : on retomberait alors, après un détour compliqué, dans le réductionnisme <sup>78</sup>. Cette considération importante justifie que l'on donne au principe de rationalité une formulation d'une *trivialité étudiée*. Il nous semble que celle que nous avons transposée de Parsons peut faire l'affaire (cf. note 3. ci-dessus).

### Retour sur la « méthode zéro »

Parmi les justifications les mieux établies du principe de rationalité figure l'idée poppérienne de la « méthode zéro » ou du « point zéro ». Voici comment Popper la définit : c'est « la méthode qui consiste à élaborer un modèle à partir de l'hypothèse de rationalité complète (et peut-être aussi de l'hypothèse d'une information complète) de la part de tous les individus concernés, puis à estimer l'écart entre le comportement effectif et le comportement postulé par le modèle, en utilisant celui-ci, en quelque sorte, comme l'origine du système de coordonnées » ([1957], p. 141). La fécondité de la « méthode zéro », que Popper argumente en détail en la comparant à d'autres approches possibles du social ([1967] ; [1946], II, chap. 14), permet de justifier l'usage, dans les sciences sociales et historiques, d'un principe de rationalité.

Weber ne disait pas autre chose, encore qu'il ait formulé sa définition de la méthode, et son argumentation en faveur de la fécondité de son application, à l'aide d'un concept d'*idéal-type* qui n'est pas toujours très facile à interpréter. Grossièrement parlant, il considérait que les approches par la « Zweckrationalität » n'avaient pas de privilège inhérent en sciences sociales : elles ne sont pas fondées sur l'idée que, par essence, les hommes seraient rationnels. Mais elles se trouvent faire système, mieux que d'autres approches — par exemple celle qui supposerait l'irrationalité des comportements —, avec la méthode des idéaux-types <sup>79</sup>. Celle-ci revient à faire servir une représentation, logiquement cohérente

78. La formulation de Popper en [1967], p. 156 et suiv., nous paraît défectueuse. Tout en affirmant la trivialité nécessaire du principe de rationalité, Popper en donne un énoncé qui renvoie, en fait, à une conception « substantielle » de la rationalité. Ceci est d'autant plus étrange que la philosophie « faillibiliste » de Popper aurait pu le rapprocher de la conception simonienne de la rationalité procédurale ; cf. Boyer [1983], p. 351. Le paradoxe se dissipe si l'on songe que la méthodologie des sciences sociales esquissée (plutôt que développée) par Popper (in [1945], [1957], [1967]) révèle l'influence explicite de la théorie néo-classique dans sa version autrichienne.

79. Le texte décisif à cet égard est sans doute : [1917], p. 518 et suiv. (édit. franç., p. 464 et suiv.).

et empiriquement stylisée, des actions humaines, comme une norme de comparaison : on appréciera la réalité de ces actions *par différence* avec la norme. Le meilleur idéal-type est celui de la *rationalité absolue*, que Weber appelle « idéal-type de justesse (*Richtigkeistypus*) » : il suppose connues, de la part de l'agent, à la fois les conditions externes de son choix, c'est-à-dire, d'un côté, les moyens à sa disposition et les conditions exogènes de ce choix, d'un autre côté, les lois techniques qui lui permettent de faire l'usage le plus approprié de ses moyens<sup>80</sup>. Les irrationalités que fait apparaître l'application de ce premier idéal type doivent être réintégrées dans un second temps. De là un nouvel idéal-type, qui correspond à l'idée d'une *rationalité subjective*, ou rationalité par cohérence (Weber parle du « type de la rationalité subjective ») : on considérera l'action comme entièrement appropriée, mais relativement à une information imparfaite sur les conditions externes du choix et les lois pertinentes. Comme l'application de la méthode laissera en général substituer un résidu à ce stade encore, il faudra introduire une séquence d'idéals-types supplémentaires. Le principe de cette séquence est très simple : on postulera de moins en moins de rationalité de la part de l'agent, jusqu'à lui dénier finalement toute initiative ; le dernier résidu que produit cette méthode des différences se résorbe dans l'invocation d'une *causalité* mécanique. Certaines irrationalités irréductibles trouveront leur explication dans la psychologie ou la physiologie<sup>81</sup>.

Dans un langage différent, c'est, généralisée parce que rendue itérative, la « méthode zéro » définie par Popper. Le propre de cette méthodologie est qu'elle justifie simultanément deux notions de rationalité : le principe de rationalité en général (il est sous-jacent à chacun des idéaux-types énumérés par Weber) et une notion de rationalité « complète » (Popper) ou de « justesse » (Weber), par laquelle il nous est dit que les sciences sociales doivent toujours commencer. Sans être le moins du monde confondues, les deux notions sont étroitement solidaires dans leurs justifications : s'il y a lieu de privilégier, relativement à tout autre (par exemple aux explications causales), une approche par la rationalité *en général*, c'est parce qu'il est avantageux (informatif) de partir d'une notion *forte* de rationalité, quitte à descendre ensuite les degrés d'une échelle vers des notions affaiblies et plus réalistes.

80. Pour une définition du « *Richtigkeistypus* », [1913], p. 433 (édit. franç. ; p. 336).

81. Cf. [1913], p. 435, édit. franç., p. 339, où Weber présente une séquence orientée d'idéals types, du « type de la justesse » jusqu'à l'explication causale brute.

C'est là, précisément, que le bât blesse : que devient la justification wébéro-poppérienne du principe de rationalité, lorsque disparaît l'idée trop simple d'une hiérarchie des modèles de rationalité ? Au niveau de généralité où ils se situent, il n'est pas certain que Popper ou Weber aient eu en vue des modèles économiques bien déterminés<sup>82</sup>. Mais leur méthodologie présuppose, à tout le moins, que l'on puisse : *a*) donner un sens *normatif* déterminé à la rationalité, *b*) se servir de ce sens normatif pour *classer* les modèles en « plus » ou « moins » rationnels. Or la discussion des rapports entre l'approche optimisatrice de la théorie classique et l'approche simonienne de la rationalité limitée suggère que cela est impossible : chacune de ces deux approches a quelque raison de se présenter à la fois comme normative et positive ; il n'y a pas de sens à dire en général que l'une est « plus » rationnelle que l'autre. Comment pourrait-on, par ailleurs, situer l'approche de la théorie des jeux sur l'échelle wébérienne des idéaux-types ? Ses modèles sont-ils « plus » ou « moins » rationnels que ceux de l'utilité espérée ?

L'objection que nous élevons contre l'argumentation wébéro-poppérienne n'est peut-être valable que pour la science économique : il se trouve que les trois conceptions de la rationalité auxquelles l'économiste peut être tenté de faire appel sont très difficilement commensurables. L'application de la « méthode zéro » à d'autres disciplines n'est pas nécessairement exposée au même problème<sup>83</sup>. Mais le sociologue ou l'historien sont-ils mieux placés que l'économiste pour élaborer une hiérarchie d'idéaux-types rationnels ? Il y a lieu, croyons-nous, d'en douter, mais il ne peut s'agir de trancher la question dans le cadre de ce travail<sup>84</sup>.

A ce point de généralité, celui-ci se propose, à vrai dire, de soulever des problèmes plutôt que de les résoudre. Il peut être utile de récapituler, en les distinguant, les conclusions de cette enquête qui nous

82. Encore que Weber illustre constamment la méthode des idéaux-types par l'exemple de la théorie néo-classique autrichienne (cf. [1917], p. 523 et suiv., édit. franç., p. 471 et suiv.) et que Popper en ait subi l'influence. Le « Richtigkeitstypus » de Weber évoque naturellement un modèle d'optimisation avec information parfaite ; de même le « point zéro » de Popper, encore que celui-ci hésite sur le type d'information qu'il faut prêter à l'agent.

83. C'est ce que donnent à penser Watkins [1970] et Boyer [1983], qui proposent en substance de renoncer au « Richtigkeitstypus » et de faire jouer la méthode des idéaux-types à partir de la notion, moins contraignante, de rationalité subjective (d'information imparfaite). Cette proposition concerne l'histoire plutôt que les sciences sociales en général.

84. Il n'est pas indifférent de noter que deux sociologues, Crozier et Friedberg, semblent avoir opté *contre* la méthode idéal-typique à partir de considérations voisines de celles que nous venons d'esquisser ([1977], p. 277).

paraissent dans une certaine mesure établies, et les problèmes nouveaux sur lesquels débouchent ces conclusions :

— l'assimilation réductionniste de la théorie de la rationalité à une théorie économique de la rationalité, entendue au sens de l'utilité espérée, est strictement intenable ; cette thèse vaut contre les doctrines extrêmes de von Mises et de Becker, mais aussi contre le point de vue plus courant des bayésiens (Savage et Harsanyi) ;

— les modèles simoniens de rationalité « limitée » sont entièrement spécifiques, en ce sens qu'ils traitent des problèmes (ceux des choix rationnels faisant intervenir des coûts de décision internes) que la théorie de l'utilité espérée ne semble pas en mesure de traiter ; cette thèse a une portée polémique contre cette théorie, mais il nous semble plus important d'en voir la signification positive (l'axiomatisation rigoureuse des propositions de Simon apparaît comme une tâche urgente) ;

— puisqu'il est nécessaire de distinguer entre l'aspect générique (trivial) de la rationalité et les aspects spécifiques exploités dans telle ou telle famille de modèles, un inventaire de la méthodologie traditionnelle de la rationalité apparaît comme souhaitable ; des justifications du principe de rationalité élaborées en dehors de la perspective dualiste, comme la « méthode zéro » de Popper-Weber, pourraient s'avérer gravement inadéquates.

Octobre 1983

PHILIPPE MONGIN

CNRS et Ecole normale supérieure

### BIBLIOGRAPHIE

- [1950] AGASSI J., « Tautology and testability in economics », *Philosophy of the Social Sciences*, 1, p. 49-64.
- [1950] ALCHIAN A., « Uncertainty, evolution and economic theory », *Journal of Political Economy*, 58, p. 211-222.
- [1952] ALLAIS M., « Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine » in ALLAIS et HAGEN [1979], p. 27-145.

- [1953] ALLAIS M., « Le comportement de l'homme rationnel devant le risque - critique des postulats et axiomes de l'école américaine », *Econometrica*, 31, p. 503-546.
- [1979] ALLAIS M., HAGEN O. (sous la direction de), *Expected utility hypotheses and the Allais paradox*, Dordrecht, Reidel, Theory and Decision Library.
- [1972] ARISTOTE, *Ethique à Nicomaque*, trad. J. Tricot, Paris, Vrin.
- [1974] ARROW K. J., *Social choice and individual values*, New Haven, Yale University Press, Cowles Foundation Monograph, 2<sup>e</sup> édit. (1<sup>re</sup> édit., 1951).
- [1962] BECKER G. S., « Irrational behavior and economic theory » in BECKER [1976], p. 153-170.
- [1976] BECKER G. S., *The economic approach to human behavior*, Chicago, The University of Chicago Press.
- [1979] BECKER G. S., « Economic analysis and human behavior » in LÉVY-CARBOUA [1979], p. 7-24.
- [1980] BERLINER H., « L'ordinateur champion de "backgammon" », *Pour la Science*, août 1980, 34, p. 49-55.
- [1970] BORGER R., CIOFFI F. (sous la direction de), *Explanation in the behavioral sciences*, Cambridge, Cambridge University Press.
- [1977] BOUDON R., *Effets pervers et ordre social*, Paris, PUF.
- [1983] BOYER A., *L'explication en histoire. Essai sur des théories anglo-saxonnes de l'Histoire*, thèse pour le doctorat de troisième cycle, Université de Paris I-Panthéon-Sorbonne.
- [1983] CHIAPPORI P. A., MONGIN P., « Note sur une régression infinie dans la théorie classique de la décision », document de travail n° 78, Paris, Ecole normale supérieure, Laboratoire d'économie politique.
- CHIAPPORI P. A., MONGIN P., « Un modèle de politique monétaire avec aléas stratégiques », à paraître dans la *Revue économique*, 1984.
- [1967] CLAASSEN E. M. (sous la direction de), *Les fondements philosophiques des systèmes économiques. Textes de J. Rueff et essais rédigés en son honneur*, Paris, Payot.
- [1977] CROZIER M., FRIEDBERG E., *L'acteur et le système. Les contraintes de l'action collective*, Paris, Le Seuil.
- [1963] CYERT R. M., MARCH J. G., *A behavioral theory of the firm*, Englewood Cliffs, Prentice Hall.
- [1979] ELSTER J., *Ulysses and the Sirens. Studies in rationality and irrationality*, Cambridge, Cambridge University Press.
- [1937] FINETTI B. de, « La prévision, ses lois logiques, ses sources subjectives », *Annales de l'Institut Henri Poincaré*, t. VII, p. 1-68.
- [1953] FRIDMAN M., « The methodology of positive economics » in *Essays in positive economics*, Chicago, The University of Chicago Press, p. 3-46.
- [1948] FRIEDMAN M., SAVAGE L. J., « The utility analysis of choices involving risk », *Journal of Political Economy*, 56, p. 279-304.
- [1952] FRIEDMAN M., SAVAGE L. J., « The expected-utility hypothesis and the measurability of utility », *Journal of Political Economy*, 60, p. 463-474.
- [1960] HARSANYI J. C., « Popper's improbability criterion for the choice of scientific hypothesis » in Harsanyi [1976].
- [1968] HARSANYI J. C., « Individualistic and functionalistic explanations in the light of game theory : the example of social status » in I. LAKATOS et A. MUSGRAVE, *Problems in the philosophy of science*, Amsterdam, North Holland, p. 305-321.

- [1975] HARSANYI J. C., « The tracing procedure : a bayesian approach to defining a solution for n-person non-cooperative games », *International Journal of Game Theory*, 4, p. 61-94.
- [1976] HARSANYI J. C., *Essays on ethics, social behavior, and scientific explanation*, Dordrecht, Reidel, Theory and Decision Library.
- [1976] HICKS J. R., « "Revolution" in Economics » in LATSIS [1976], p. 207-218.
- [1977] IJRI Y., SIMON H. A., *Skew-distributions and the size of business firms*, Amsterdam.
- [1921] KNIGHT F., *Risk, uncertainty and profit*, rééd., Londres, London School of Economics, Series of Scarce Tracts in Economics and Political Science, 1957.
- [1946] KNIGHT F., « Immutable law in economics : its reality and limitations », *American Economic Review*, 36, Papers and Proceedings, p. 93-111.
- [1976] LATSIS S. J. (sous la direction de), *Method and appraisal in economics*, Cambridge, Cambridge University Press.
- [1979] LÉVY-GARBOUA L. (sous la direction de), *Sociological economics*, Londres, Sage Publications.
- [1979] LÉVY-GARBOUA L., « Perception and the formation of choice » in Lévy-Garboua [1979], p. 97-121.
- [1983] LÉVY-GARBOUA L., « L'économique et le rationnel », *L'Année sociologique*.
- [1982] MARCIANO M., « Acquisition optimale d'information par le consommateur sur le prix d'un bien homogène », note de recherche 82.01, GRASCE, Faculté d'économie appliquée, Aix-en-Provence.
- [1955] MARSCHAK J., « Elements for a theory of teams », *Management Science*, 1, p. 127-137.
- [1972] MARSCHAK J., RADNER R., « Economic theory of teams », Yale, Yale University Press.
- [1982] MAYNARD-SMITH J., « Evolution and the theory of games », Cambridge, Cambridge University Press.
- [1978] MEIDINGER C., « Apriorisme et théorie économique », *Revue économique*, 29, p. 261-290.
- [1983] MENCER C., *Untersuchungen über die Methode der Sozialwissenschaften und der politischen Oekonomie insbesondere*, Leipzig, von Duncker et Humblot.
- [1957] MERTON R. K., *Social theory and social structure*, Glencoe, The Free Press, 2<sup>e</sup> édit. (1<sup>re</sup> éd., 1953).
- [1949] MISES von L., *Human action. A treatise on economics*, New Haven, Yale University Press.
- [1982] MONGIN P., « Friedman et la thèse de l'irréalisme des suppositions ». Première partie : Le contexte épistémologique d'ensemble », *Conjoncture et analyse des déséquilibres*, document n° 55, Université de Paris I.
- [1983] MONGIN P., « Friedman et la thèse de l'irréalisme des suppositions ». Deuxième partie : les trois sens de la thèse friedmanienne », *Conjoncture et analyse des déséquilibres*, Université de Paris I (à paraître).
- [1944] NEUMAN von J., MORGENSTERN O., *Theory of games and economic behavior*, New York, Wiley (3<sup>e</sup> édit., 1953).
- [1958] NEWELL A., SHAW J. C., SIMON H. A., « Chess-playing programs and the problem of complexity », *Journal of Research and Development (IBM)*, 2, p. 320-335.

- [1962] NEWELL A., SHAW J. C., SIMON H. A., « The process of creative thinking » in H. E. GRUBER, G. TERRELL et M. WERTHEIMER, *Contemporary approaches to creative thinking*, Lieber-Atherton, p. 63-119 (repris in H. SIMON, [1979 a], p. 140-174).
- [1972] NEWELL A., SIMON H., *Human problem solving (a new approach to the psychology of human problem solving based on detailed analysis of human behavior and basic studies of the information processing systems)*, Englewood Cliffs, Prentice Hall.
- [1973] PARIENTE J. C., *Le langage et l'individuel*, Paris, Armand Colin.
- [1937] PARSONS T., *The structure of social action*, New York, The Free Press (2<sup>e</sup> édit., 1949).
- [1952] PENROSE E. T., « Biological analogies in the theory of the firm », *American Economic Review*, 62, p. 804-819.
- [1945] POPPER K. R., *The open society and its enemies*, 2, Londres, Routledge and Kegan Paul (4<sup>e</sup> édit., 1962).
- [1967] POPPER K. R., *The poverty of historicism*, Londres, Routledge and Kegan Paul (2<sup>e</sup> édit., 1960).
- [1967] POPPER K. R., « La rationalité et le statut du principe de rationalité » in E. M. CLAASSEN [1967], p. 142-150.
- [1972] RADNER R., « Normative theory of individual decision : an introduction » in R. RADNER et C. B. GUIRE [1972], chap. I, p. 1-18.
- [1974] RADNER R., *Satisficing*, communication faite au « Mathematical social science board colloquium on mathematical economics », Berkeley, University of California (août 1974).
- [1972] RADNER R., Mc GUIRE C. B. (sous la direction de), *Decision and organization. A volume in honor of Jacob Marschak*, Amsterdam, North-Holland.
- [1932] ROBBINS L., *An essay on the nature and significance of economic science*, Londres, Mac Millan (2<sup>e</sup> édit., 1935).
- [1954] SAVAGE L., *The foundations of statistics*, New York, Dover Publications (2<sup>e</sup> édit., 1972).
- [1947] SIMON H. A., *Administrative behavior*, New York, The Free Press (3<sup>e</sup> édit., 1976).
- [1955] SIMON H. A., « A Behavioral model of rational choice », *Quarterly Journal of Economics*, 69, p. 99-118.
- [1963] SIMON H. A., « Economics and psychology » in S. Koch, *Psychology. A study of a science*, t. VI, p. 685-723, New York, Mc Graw-Hill.
- [1964] SIMON H. A., « Rationality » in J. GOULD et W. L. KOLB, *A dictionary of the social sciences*, Glencoe, The Free Press, p. 573-574.
- [1966] SIMON H. A., « Theories of decision-making in economics and behavioral science » in *Surveys of economic theory*, Royal Economic Society and American Economic Association, Londres, MacMillan, t. III, chap. 1 (article d'abord publié in *American Economic Review*, 1959, 49, p. 253-283).
- [1969] SIMON H. A., *The sciences of the artificial*, Cambridge, MIT. Press (trad. franç. par J. L. Le Moigne, *La science des systèmes*, Paris, Epi, 1974).
- [1976] SIMON H. A., « From substantive to procedural rationality » in S. J. LATSKY [1976], p. 129-148.
- [1978] SIMON H. A., « Rationality as process and as product of thought », *American Economic Review*, 68, p. 1-16.

## Bibliographie

- [1978 a] SIMON H. A., « On how to decide what to do », *The Bell Journal of Economics*, 9, p. 494-507.
- [1979] SIMON H. A., « Rational decision making in business organizations », *American Economic Review*, 69, p. 494-513.
- [1979 a] SIMON H. A., *Models of thought*, New Haven, Yale University Press.
- [1983] SIMON H. A., *Models of bounded rationality*, 2 tomes, Cambridge, MIT Press. (Sauf indications contraires, ce recueil reprend les articles précédemment cités.)
- [1961] STIGLER G. J., « The economics of information », *Journal of Political Economy*, 69, p. 213-225.
- [1977] STIGLER G. J., BECKER G. S., « De gustibus non est disputandum », *American Economic Review*, 67.
- [1970] TOULMAIN S., « Reasons and causes » in R. BORGER et F. CIOFFI [1970], p. 1-26.
- [1970] WATKINS J., « Imperfect rationality » in R. BORGER et CIOFFI [1970], p. 167-217.
- [1904] WEBER M., « Die "Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis » in M. WEBER [1922 a].
- [1913] WEBER M., « Ueber einige Kategorien der verstehenden Soziologie », *ibid.*
- [1917] WEBER M., « Der Sinn der "Wertfreiheit" der soziologischen und ökonomischen Wissenschaften », *ibid.*
- [1922] WEBER M., *Wirtschaft und Gesellschaft*, Tübingen, Mohr (trad. franç. partielle, *Economie et société*, Paris, Plon, 1971).
- [1922 a] WEBER M., *Gesammelte Aufsätze zur Wissenschaftslehre*, Tübingen, Mohr (trad. franç. partielle de J. Freund, *Essais sur la théorie de la science*, Paris, Plon, 1965).
- [1964] WINTER S., « Economic "natural selection" and the theory of the firm », *Yale Economic essays*, 4, p. 225-272.
- [1971] WINTER S., « Satisficing, selection and the innovating remnant », *Quarterly Journal of Economics*, 85, p. 237-261.
- [1975] WINTER S., « Optimization and evolution in the theory of the firm » in R. H. DAY et T. GROVES, *Adaptive economic models*, New York, Academic Press, p. 73-118.