

Mikaël Cozic

Anti-réalisme, rationalité limitée et théorie expérimentale de la décision

Résumé. *L'objet de cet article est de présenter et discuter les objections d'inspiration anti-réaliste (ou instrumentaliste) qui sont adressées au programme de modélisation de la rationalité limitée dans les sciences de la décision, et en particulier dans les sciences économiques. Après avoir reconstruit ces objections dans leurs diverses variantes, nous proposons un argument qui entend contrer de manière constructive la variante que nous appelons comportementale, selon laquelle les modèles de décision ne doivent être évalués qu'en fonction de leur capacité à décrire les comportements individuels. Cet argument repose sur les avancées récentes de la théorie de la décision expérimentale et fait appel à la prise en compte de données cognitives.*

Mots-clés. *Anti-réalisme – Cognition – Instrumentalisme – Rationalité limitée – Théorie de la décision expérimentale*

Abstract. *This paper discusses anti-realist (or instrumentalist) objections against the bounded rationality research program in decision sciences. After a reconstruction of several versions of these objections, we expose an argument against the so-called behavioral version of the anti-realist objection according to which decision models should be assessed exclusively on the basis of behavioral data. Our argument relies on recent advances in experimental decision theory, especially on works that collect and exploit cognitive data.*

Keywords. *Anti-realism – Bounded rationality – Cognition – Experimental decision theory – Instrumentalism*

Introduction

Le développement sans précédent de la théorie de la décision et de la théorie des jeux comportementales (voir respectivement Kahneman & Tversky,

2000, et Camerer, 2003) et, plus récemment, de la neuroéconomie (voir Camerer, Loewenstein & Prelec, 2005), auquel nous assistons aujourd'hui, relance inévitablement les débats méthodologiques sur le statut des modèles de décision individuelle et stratégique (voir, respectivement, Gul & Pesendorfer, 2005; Guala, 2006) qui sont au fondement de l'économie contemporaine. Il est aisé de voir pourquoi. Une grande partie de ces recherches procèdent en effet d'une forme ou d'une autre de *révisionnisme empirique*, démarche que l'on peut logiquement décomposer en deux étapes. Ces recherches mettent en avant des classes de situations où les modèles classiques de décisions semblent *empiriquement* inadéquats. Elles proposent ensuite de réagir à ces inadéquations empiriques en *révisant* les modèles classiques de décision.¹ Le révisionnisme empirique soulève par conséquent deux types de questions méthodologiques, chacune correspondant à l'une de ces deux étapes.

L'affirmation d'inadéquation empirique soulève la question de la *testabilité* des modèles de décision: ces modèles sont-ils testables? Si oui, par quels types de données empiriques? Quelles sont les principales hypothèses auxiliaires en jeu dans la mise en relation entre modèles et données empiriques? La réaction révisionniste soulève ensuite une autre question: quelle est la part de l'adéquation empirique dans l'évaluation globale des modèles de décision: l'inadéquation empirique d'un modèle de décision constitue-t-elle une raison suffisante pour le rejeter? Toutes les familles de données empiriques ont-elles un pouvoir égal de remise en question des modèles de décision?

Nous ne proposerons pas ici, loin s'en faut, de réponse générale à ces deux familles (étroitement liées) de questions. Pour le dire autrement, nous ne proposerons pas une analyse méthodologique générale du révisionnisme empirique. Il nous semble que le domaine est à la fois trop hétérogène et trop récent pour que l'on procède à une telle analyse. Nous allons donc nous concentrer sur une partie relativement circonscrite du domaine, qui correspond à l'un des courants les plus anciens et les plus influents du révisionnisme empirique: le programme de modélisation de la rationalité limitée (que nous abrègerons par la suite PMRL). Notre étude comporte une seconde restriction importante: nous ne dessinerons pas un paysage exhaustif des débats méthodologiques qui peuvent entourer le PMRL.² Nous allons nous restreindre à la discussion d'une objection importante que l'on peut adresser au PMRL: *l'objection anti-réaliste*. Importante, cette objection l'est en un double sens. D'abord, nous le verrons, parce qu'elle comporte une part de vérité non négligeable. Ensuite (et surtout) parce que nous pensons qu'elle reflète une conception méthodologique répandue (même si elle reste souvent implicite) dans les sciences de la décision. Nous étayerons plus loin cette conjecture. Un mot, pour terminer, sur le rôle de la théorie

de la décision expérimentale dans notre argumentation. L'une des caractéristiques importantes de cette argumentation est qu'elle invoque de manière cruciale des résultats et des méthodes de la théorie de la décision expérimentale. La méthodologie expérimentale permet en effet d'introduire avec rigueur les *données cognitives* qui jouent un rôle de premier plan dans notre argument.

Nous démarrons avec une reconstruction du programme de modélisation de la rationalité limitée (Section 1). La Section 2 est ensuite consacrée à l'anti-réalisme dans les sciences de la décision et à la distinction entre sa variante comportementale et sa variante collective. La Section 3 introduit l'objection au programme de modélisation de la rationalité limitée fondée sur la variante comportementale de l'anti-réalisme. Les Sections 4 à 6 sont consacrées à la critique de cette objection. La Section 4 distingue les principales stratégies de réponse tandis que les Sections 5 et 6 étayent successivement les deux prémisses de l'argument que nous favorisons.

1. Rationalité limitée et modèles classiques de choix

Cinquante ans après l'article fondateur d'Herbert Simon, 'A behavioral model of rational choice' (Simon, 1955), la rationalité limitée est constamment invoquée dans le champ des sciences de la décision. En témoigne, par exemple, le discours que Daniel Kahneman a prononcé lors de la remise du Prix Nobel d'Economie, intitulé 'Maps of bounded rationality' (Kahneman, 2003). Une certaine *confusion sémantique* accompagne cette popularité: l'expression 'rationalité limitée' recouvre aujourd'hui des phénomènes, des modèles et des domaines de recherches passablement hétérogènes.

Dans une première acception, la rationalité limitée se confond avec les écarts entre le comportement effectif des agents et les prédictions que font les modèles classiques de choix. Ces modèles étant censés codifier des principes de rationalité 'parfaite', on désigne naturellement les déviations ou les anomalies relatives à ces modèles par le terme de 'rationalité limitée'. Dans ce cas, la modélisation de la rationalité limitée se confond avec le projet général d'élaboration de modèles de décision plus adéquats empiriquement que ne le sont les modèles classiques.³

Nous retiendrons toutefois une autre acception, qui est plus restreinte et qui correspond plus étroitement au programme original d'H. Simon. Pour Simon, en effet, la rationalité limitée évoque les *limitations cognitives* qui affectent la formation des états mentaux et la prise de décision des agents. Dans ce cas, la modélisation de la rationalité limitée est la recherche de modèles de décision susceptibles de prendre en compte ces limitations cognitives.

Le projet de modélisation de la rationalité limitée s'enracine dans une critique des *modèles classiques de choix*, qui servent notamment de socle à l'économie théorique. Si l'on reprend la tripartition usuelle entre les situations de certitude, de risque et d'incertitude (Luce & Raiffa, 1985), alors les trois principaux modèles classiques sont les suivants:

1. Le modèle classique de choix en certitude part d'un ensemble A d'actions réalisables. Le choix est certain dans la mesure où l'agent est supposé connaître avec certitude la conséquence de chacune des actions réalisables. Les valeurs de l'agent sont représentées par une relation de préférence \geq sur A qu'on suppose en général réflexive, transitive et complète. Le modèle se résout d'ordinaire par maximisation: l'agent est supposé choisir l'une des actions $a^* \in A$ qui est \geq -maximale.
2. Le modèle classique de choix en situation de risque est le modèle d'espérance d'utilité axiomatisé pour la première fois par (von Neumann & Morgenstern, 1944). Dans ce cadre, les conséquences des actions de l'agent sont supposées dépendre de l'état de l'environnement, état qui, dans le cas général, est ignoré de l'agent. Les actions de l'agent sont alors conçues comme des fonctions qui associent à tout état de l'environnement possible $s \in S$ l'une des conséquences $c \in C$. Le choix est risqué au sens où une distribution de probabilité (dite parfois 'objective') P sur S est donnée à l'agent. Chaque action réalisable induit alors une distribution de probabilité ou *loterie* sur l'ensemble C des conséquences. Le modèle d'espérance d'utilité suppose que les valeurs de l'agent relatives aux conséquences possibles puissent être représentées par une fonction d'utilité à valeur réelle $u: C \rightarrow \mathbf{R}$ et que l'agent choisit l'une des actions dont l'espérance d'utilité (relativement à $u(\cdot)$) est maximale. Le théorème de représentation associé au modèle d'espérance d'utilité implique un certain nombre d'axiomes concernant les préférences de l'agent sur les actions réalisables, axiomes qui garantissent que ces préférences puissent être représentées par l'espérance d'une fonction d'utilité.⁴
3. Le modèle classique de choix en situation d'incertitude, où l'on ne suppose plus qu'une distribution de probabilité exogène est disponible, est celui de l'espérance subjective d'utilité tel qu'il a été axiomatisé pour la première fois par L. Savage (1954). Le modèle suppose que les croyances de l'agent sur l'environnement sont représentées par une distribution de probabilité subjective sur S ; et que l'agent choisit l'une des actions dont l'espérance subjective d'utilité (de nouveau relativement à une fonction d'utilité $u(\cdot)$ sur C) est maximale. Le théorème de représentation associé au modèle d'espérance subjective d'utilité implique des axiomes concernant les préférences de l'agent sur les actions réalisables

(assimilées ici à des fonctions de S dans C), axiomes qui garantissent qu'il existe une unique distribution de probabilité P sur S et une fonction d'utilité $u(\cdot)$ sur C (unique à une transformation affine positive près) telles que les préférences peuvent être représentées par l'espérance subjective d'utilité calculée à l'aide de P et $u(\cdot)$.

La liste des modèles de choix classiques que nous venons de donner n'est évidemment pas exhaustive. Ils incluent les *spécifications* des modèles qui viennent d'être esquissés – par exemple ceux qui précisent la forme mathématique des fonctions d'utilité; mais également leurs *extensions*, – par exemple les extensions aux choix dynamiques. Par ailleurs, nous incluons dans la même famille les modèles de la théorie des jeux quand elle applique les concepts d'équilibres ordinaires: équilibre de Nash pour les jeux sous forme normale et équilibre parfait en sous-jeux pour les jeux sous forme extensive.

La définition de ce que nous entendons par 'modèles classiques de choix' serait bien incomplète si nous ne précisions pas, en outre, quelle interprétation accompagne ces modèles. On sait en effet que les modèles de choix font l'objet de différentes interprétations dont les deux plus importantes sont l'interprétation *descriptive*, suivant laquelle ils servent à prédire et expliquer le comportement des agents, et l'interprétation *normative*, suivant laquelle ils apparaissent comme l'expression de principes de rationalité décisionnelle. De nombreux philosophes ont contesté le bien-fondé d'une telle distinction, mettant en avant la perméabilité du descriptif au normatif (et réciproquement). Nous ne nous engagerons cependant pas dans cette discussion: notre discussion portera essentiellement sur la première de ces deux interprétations, l'interprétation descriptive, sans préjuger de la question des liens avec la seconde. Car c'est d'abord du point de vue descriptif que les tenants du programme de modélisation de la rationalité limitée (PMRL) estiment que les modèles classiques souffrent d'un défaut majeur.

C'est un fait que les agents auxquels on applique les modèles de choix subissent des limitations cognitives de différentes sortes: notre attention, nos capacités mémorielles, nos capacités de détection des régularités, nos capacités calculatoires sont limitées. Bien sûr, certaines limitations sont passagères et peuvent être levées quand le besoin s'en fait vraiment sentir ou tout simplement dans de 'bonnes conditions' épistémiques. Mais la plupart de ces limitations sont structurelles et robustes. L'intuition fondamentale des tenants de la rationalité limitée est que ces limitations ont une influence cruciale sur la prise de décision et qu'il faut les prendre en compte dans les modèles de choix. Pour être plus précis, on peut distinguer deux conséquences distinctes qui découlent de l'existence de nos limitations cognitives.

La première est qu'elles rendent les modèles classiques que nous avons brièvement présentés descriptivement inadéquats. La seconde est que, en réponse à cette inadéquation descriptive, il faut réélaborer des modèles de choix. Ces deux conséquences font de la rationalité limitée une forme de révisionnisme empirique.

2. L'anti-réalisme

Si l'existence de limitations cognitives est difficilement contestable, les conséquences qu'en tirent les tenants du PMRL ne vont pas de soi. Dans ce qui suit, nous allons nous concentrer sur l'une des familles d'objections méthodologiques au PMRL, famille d'inspiration anti-réaliste au sens où l'on emploie désormais ce terme en philosophie des sciences.

Commençons par caractériser l'anti-réalisme. En partant du célèbre essai de Milton Friedman (1953), nous donnerons une reconstruction dont l'objectif sera plus de proposer un idéal-type commode de l'anti-réalisme que de viser à la rigueur exégétique. L'une des premières affirmations contenues dans Friedman (1953: 7) est la suivante:

L'objectif ultime d'une science positive est de développer une 'théorie' ou une 'hypothèse' qui produise des prédictions valides et signifiantes (i.e. qui ne se réduisent pas à des truisimes) à propos de phénomènes qui n'ont pas encore été observés.

La notion de prédiction que Friedman mobilise ici est à la fois épistémique et temporelle: P est la prédiction d'une certaine théorie à un moment t si P est une proposition empirique dont la valeur de vérité n'est pas encore connue par le modélisateur en t . Nous considérerons plus généralement les conséquences observationnelles d'une théorie: ce qu'elle dit à propos de phénomènes observables. On peut alors s'appuyer sur le concept d'*adéquation empirique* telle qu'il est employé par B. van Fraassen dans *The scientific image* (1980): une théorie est empiriquement adéquate si l'ensemble de ses conséquences observationnelles sont correctes, c'est-à-dire si 'ce qu'elle dit des choses et événements observables dans ce monde est vrai – précisément si elle "sauve les phénomènes"' (cf. Van Fraassen, 1980: 12). On s'autorisera à parler de *degré* d'adéquation empirique d'une théorie, par quoi il faut entendre le degré de correction des conséquences observationnelles de cette théorie. Si l'on s'appuie sur la notion d'adéquation empirique (plutôt que sur celle de prédiction), alors on obtient comme reconstruction de l'affirmation de Friedman la thèse suivante:

(ARG) *l'objectif d'une science positive est de développer une théorie empiriquement adéquate.*

L'un des aspects les plus intéressants de la position de Friedman est qu'il restreint d'une manière très particulière cette première affirmation. Il affirme en effet par la suite que:

... une théorie doit être jugée par son pouvoir prédictif *relativement à la classe de phénomènes qu'elle est faite pour 'expliquer'*. (1953: 8, italiques de l'auteur)

Bien sûr, on peut entendre cette seconde affirmation comme une simple reformulation de la première. Ce serait aller trop vite. Friedman semble en effet mettre en avant le fait que, dans l'évaluation de théories comme, par exemple, les théories économiques, le domaine entier d'application de la théorie n'est pas nécessairement pertinent. Autrement dit, ce n'est pas l'adéquation empirique globale qui est le bon étalon d'évaluation d'une théorie, mais son adéquation par rapport à un sous-domaine d'application de la théorie que l'on pourrait appeler son 'domaine-cible'.⁵ Il faut donc soigneusement distinguer entre

1. l'*anti-réalisme général*, qui suppose qu'il existe une distinction relativement claire (bien qu'elle puisse être vague) entre ce qui est observable et ce qui ne l'est pas et qui assigne comme objectif principal à une théorie scientifique d'avoir des conséquences observationnelles adéquates.⁶
2. l'*anti-réalisme ciblé* qui suppose en outre qu'il existe un certain domaine-cible, disons D^* , et qui assigne comme objectif principal à une théorie scientifique d'avoir des conséquences observationnelles adéquates relativement à son domaine-cible D^* .

Ce qui distingue les deux formes d'anti-réalisme, c'est leur attitude concernant l'adéquation empirique d'une théorie relativement aux données qui sont *exclues* du domaine-cible, disons D^* . Pour l'anti-réaliste général, une inadéquation empirique relativement à D^* est un mauvais point pour la théorie; pour l'anti-réaliste ciblé, une telle inadéquation est *non pertinente* et ne doit pas intervenir dans l'évaluation de la théorie.

Une fois ce cadre épistémologique posé, revenons aux modèles de décision: quelle forme peut revêtir, dans ce cas, l'anti-réalisme ciblé? Il y a en fait deux grandes réponses possibles à cette question, bien que la distinction entre les deux réponses passe généralement inaperçue.

(2a) Une première variante de l'anti-réalisme ciblé, la **variante comportementale**, consiste à soutenir que le domaine-cible des modèles de décision est exclusivement constitué par les comportements des agents. Dans ce cas, un modèle de décision ne doit être évalué que par son adéquation comportementale, c'est-à-dire par sa capacité à décrire correctement les comportements individuels. En revanche, les modèles de décision n'ont à donner de

description satisfaisante ni des états mentaux ni des processus cognitifs à l'œuvre dans la prise de décision. Pour employer la terminologie mise en place ci-dessus, la cognition est exclue du domaine-cible. Typiquement, pour un partisan de la variante comportementale, peu importe que la maximisation d'une fonction d'utilité soit un processus cognitif plausible ou non, tant que le comportement des agents obéit aux propriétés comportementales déduites du modèle.

Prenons comme exemple le modèle de choix certain. Il suppose que l'agent choisisse, parmi les actions qui s'offrent à lui, celle dont la conséquence est maximale relativement à une relation de préférence réflexive, transitive et complète. La théorie économique, surtout celle du consommateur, a exploré les implications comportementales d'un tel modèle lorsque l'on fait varier les actions offertes à l'agent (les ensembles de choix) (cf. par exemple Sen, 1971). Les choix d'un agent qui obéit au modèle de choix certain satisfont par exemple la Propriété α : si l'ensemble de choix X est inclus dans l'ensemble de choix Y et si l'action x appartient à X alors qu'elle est choisie par l'agent parmi les éléments de Y , alors x est choisie parmi les éléments de X . Dans ce contexte, la variante (2a) (comportementale) de l'anti-réalisme ciblé consiste à affirmer qu'il n'est pas pertinent de se demander si la maximisation d'une relation de préférence est plausible cognitivement; seule importe la question de savoir si les choix des agents se conforment à la Propriété α – et aux autres conséquences comportementales du modèle.

(2b) Il y a une seconde variante de l'anti-réalisme ciblé qui est concevable, et c'est bien souvent celle qui est en jeu dans les discussions méthodologiques en économie. C'est la **variante collective** selon laquelle le domaine-cible des modèles de décision n'est constitué ni par les processus cognitifs des individus, ni par leurs comportements, mais par les phénomènes collectifs.⁷ Typiquement, pour un partisan de la variante collective, peu importe que le comportement des consommateurs se conforme aux propriétés impliquées par le modèle de choix certain tant que, par exemple, le modèle permet de décrire correctement l'évolution de la demande agrégée pour un bien en fonction de son prix. On trouve par exemple chez F. Machlup (1967) – dont les positions méthodologiques ne sont pas sans rapport avec celle de Friedman – une expression de cette variante collective dans sa discussion méthodologique de la théorie de la firme:

L'empiriste a tendance à vérifier les actions déduites à partir de la théorie en les testant par rapport au comportement individuel, alors que la théorie ne sert qu'à expliquer et prédire le comportement de masse. (Machlup, 1967: 6)

Quand je dis qu'il y a une grande confusion à propos des objectifs de la 'théorie de la firme' mobilisé dans la théorie traditionnelle des prix, je veux dire ceci: contrairement à ce que beaucoup croient, le modèle de la firme dans cette théorie n'est pas destiné à expliquer et prédire le comportement des firmes réelles. En réalité, le modèle est destiné à expliquer et prédire les changements dans les prix observés ... en fonction de changements dans certaines conditions (taux de rémunération, taux d'intérêts, droits de douane, etc.). (Machlup, 1967: 9)

Il est important de remarquer que⁸ dans ces deux citations on peut repérer deux types de ciblage passablement distincts: un ciblage 'agrégatif' selon lequel les données empiriques cibles sont des données agrégées, et un ciblage 'comparatif' selon lequel les données cibles sont des données relatives aux changements dans la valeur de certaines variables en fonction de changements dans la valeur d'autres variables. En microéconomie, le ciblage comparatif revient à ne retenir que les conséquences observationnelles de la statique comparative (par opposition à la statique pure).⁹ Le fait de combiner le ciblage agrégatif avec le ciblage comparatif se comprend dans la mesure où c'est typiquement dans la statique comparative que se révèle le pouvoir prédictif du modèle. Mais, en ce qui nous concerne, c'est bien sûr le ciblage agrégatif qui correspond conceptuellement à la variante collective de l'anti-réalisme.

3. L'objection anti-réaliste

La distinction entre les deux variantes de l'anti-réalisme ciblé, la variante comportementale et la variante collective, est de première importance pour le révisionnisme empirique en général, et pour le PMRL en particulier. Commençons par celui-ci.

La variante comportementale permet de construire directement l'*objection anti-réaliste* contre le PMRL. En effet, dans cette variante, l'anti-réalisme ciblé peut reconnaître sans difficulté que les modèles classiques de choix ne sont pas adéquats du point de vue cognitif: s'ils ne donnent pas une description satisfaisante des états mentaux des agents ni des processus cognitifs mis en oeuvre dans la prise de décision, cela n'a pas d'importance. Pour le dire autrement: tout ce que l'on peut et doit exiger des modèles classiques de choix, c'est que les agents se comportent *comme si* leur comportement procédait de la maximisation d'une fonction d'utilité. De ce point de vue, la critique que les tenants du PMRL adressent aux modèles classiques est naïve: elle consiste à prendre à la lettre le critère de maximisation et les déterminants cognitifs sur lesquels il se base.

Il faut immédiatement ajouter que la variante comportementale de l'anti-réalisme ciblé ne permet pas de fonder une objection *générale* vis-à-vis du révisionnisme empirique. Considérons par exemple les modèles concurrents du modèle d'espérance d'utilité qui sont élaborés pour les situations de risques (modèle à utilité dépendante du rang, *prospect theory*, etc.) et d'incertitude (modèle *multiprior*, etc.).¹⁰ Ces travaux sont en général motivés par la mise en évidence de violations empiriques de certains des axiomes correspondants. C'est le cas avec le paradoxe d'Allais (1953) pour le risque, et avec celui d'Ellsberg (1961) pour l'incertitude. L'important pour la discussion qui nous occupe est que l'on peut concevoir ces violations comme des inadéquations *comportementales* du modèle d'espérance d'utilité. Les modèles en question cherchent avant tout à surmonter ces inadéquations comportementales, et non pour décrire de manière plus satisfaisante le type de processus délibératif que suivent les agents dans les situations risquées ou incertaines. Cette affirmation exige toutefois deux réserves. Tout d'abord, on pourrait objecter que la *prospect theory* de Kahneman & Tversky (1979) comporte une phase préliminaire dite d'*editing* durant laquelle le décideur est censé simplifier les informations dont il dispose à propos des options qui s'offrent à lui, de manière à pouvoir les évaluer par la suite. A cette objection il faut répondre que, précisément, D. Kahneman et A. Tversky ne donnent guère de détail sur la phase d'*editing*. Leur modèle ne porte que sur la seconde phase, qui est la contrepartie exacte de l'espérance d'utilité dans la perspective classique. On pourrait ensuite faire valoir que, en dissociant l'aversion face au risque et l'utilité marginale décroissante, certains de ces modèles constituent un authentique progrès psychologique.¹¹ C'est un point que nous reconnaissons: nous ne nions pas que les modèles concurrents au modèle d'espérance d'utilité accomplissent certains progrès du point de vue du réalisme psychologique. Mais ces modèles n'ont certainement pas pour objectif ni pour vertu de mieux représenter la délibération d'un agent qui fait face à un problème de décision complexe du point de vue cognitif. Il y a là une différence très importante avec les modèles conçus ou espérés par les tenants du PMRL.

Il résulte de ce qui précède que ces modèles nouveaux de choix en situations risquées ou incertaines sont à l'abri d'objection fondées sur la variante comportementale de l'anti-réalisme ciblé. En revanche, ils ne sont en principe pas à l'abri d'objections fondées sur la variante collective.

Pour terminer la présentation de l'objection anti-réaliste au PMRL, je voudrais montrer que, pour ce qui est des sciences économiques en tous cas, et la variante comportementale de l'anti-réalisme, et l'objection anti-réaliste sont répandues.

Commençons par la variante comportementale de l'anti-réalisme. Quand, par exemple, en théorie du consommateur il est question de tester le modèle du consommateur (une spécification du modèle de choix certain esquissé plus haut), les données prises en compte sont les données comportementales. Comme le rappelle P.-A. Chiappori:

[Dans la théorie du consommateur] l'explication des comportements repose sur un concept de base, celui de préférence ou de fonction d'utilité. Cependant, ce concept ne renvoie à aucune réalité directement observable; on admet en général que seuls les *comportements* induits (demandes, offres de travail ...) peuvent faire l'objet d'analyses empiriques. (Chiappori, 1990: 1006)¹²

Pour s'en convaincre, on peut considérer un manuel de microéconomie comme celui d'H. Varian (Varian, 1992).¹³

Passons désormais à l'objection adressée au PMRL sur la base de l'anti-réalisme (dans sa variante comportementale). Elle se manifeste non seulement dans les manuels,¹⁴ mais également dans les débats actuels en méthodologie des sciences économiques, où l'on retrouve également des arguments que l'on peut assimiler à l'objection anti-réaliste. C'est le cas notamment dans l'essai retentissant de F. Gul et W. Pesendorfer (2005) pour une économie 'sans esprit' (*mindless*). Selon eux, les seules données à prendre en compte pour l'évaluation d'une théorie de la décision sont les données comportementales;¹⁵ les données cognitives, de quelque nature qu'elles soient, n'ont aucune place dans l'évaluation des modèles de choix.¹⁶ Il ne s'agit pas d'entreprendre une étude statistique sur la diffusion de l'anti-réalisme local et de l'objection anti-réaliste chez les économistes et les théoriciens de la décision. Les références qui précèdent suggèrent cependant à quel point il s'agit de conceptions et de préceptes méthodologiques répandus.

4. Défenses du PMRL

Après cette reconstruction du PMRL et des objections d'inspiration anti-réaliste qu'il soulève, nous allons désormais nous concentrer sur le premier type d'objection, celui qui est basé sur la variante comportementale: de quelle manière un tenant du PMRL peut-il répondre à ce type d'objection?

Il y a essentiellement deux réponses qui sont envisageables: l'une qui rejette radicalement les prémisses de l'objection anti-réaliste, l'autre qui est compatible avec ces prémisses mais en désaccord avec les conséquences qu'en tire l'objection. C'est en faveur de ce second type que nous allons argumenter.

La première réponse fait valoir qu'un modèle de décision ne doit pas seulement représenter les choix des agents, mais doit représenter *le processus de décision dans sa globalité*: aussi bien les processus délibératifs qui conduisent au comportement que le comportement lui-même. Mais cette première réponse rencontre la difficulté suivante, qui peut être exprimée sous la forme d'un *dilemme*. Le dilemme est structuré par la question de savoir s'il existe des données cognitives, c'est-à-dire des données empiriques (1) qui soient distinctes de l'observation des comportements 'bruts' et (2) qui puissent être connectées aux modèles de choix. (a) Si le défenseur du PMRL n'admet pas l'existence de telles données, alors sa défense repose entièrement sur l'idée qu'il existe réellement des processus et états cognitifs en amont des comportements qui doivent être 'pris au sérieux'. Mais en l'absence de données empiriques distinctes des comportements 'brut', on voit mal quelles pourraient être les conséquences méthodologiques d'une telle idée. (b) Supposons maintenant que le défenseur admette qu'il existe (en principe) des traces empiriques propres aux processus et états cognitifs (par opposition aux comportements 'bruts') alors cette réponse revient à nier la légitimité d'un ciblage des données empiriques – ou du moins la légitimité d'un ciblage purement comportemental des données empiriques. Mais pourquoi rejeter le ciblage comportemental? Le défenseur pourrait répondre en invoquant un principe méthodologique de *clotûre causale*: si les domaines D1 et D2 incluent chacun des conséquences observationnelles et si D1 est responsable causalement de D2, alors une théorie satisfaisante doit être empiriquement adéquate relativement à D2 *et* D1. Bien sûr, D1 serait ici l'ensemble des processus cognitifs délibératifs et D2 l'ensemble des comportements. Une autre façon de procéder pourrait consister à affirmer que si l'on ne tient pas compte de l'adéquation empirique relativement à D1, on se prive de la possibilité de fournir une explication satisfaisante des phénomènes du domaine D2. Ces deux arguments sont étroitement liés – autant que le sont en général causalité et explication. La difficulté de ces deux arguments, c'est qu'ils constituent une exigence épistémologique exorbitante, et qui paraît rien moins qu'absurde pour une science 'spéciale' comme l'est la théorie de la décision qui s'intéresse à un certain 'niveau' d'organisation et de description. *A contrario*, c'est dans la reconnaissance du fait que les sciences spéciales n'obéissent typiquement pas à ce genre d'exigence que nous voyons la part de vérité de l'objection anti-réaliste.

La discussion qui précède nous invite à considérer une autre voie argumentative: il nous semble en effet que l'on peut défendre le PMRL de manière, pour ainsi dire, plus économique, c'est-à-dire sans remettre en question l'idée que l'objectif principal des modèles de choix, dans leur

interprétation descriptive, soit de décrire correctement le comportement des agents. L'argument que nous proposons repose sur deux prémisses principales. La première (P1) est que, en raison des limitations cognitives des agents, il existe des situations de choix où les modèles classiques sont *comportementalement* inadéquats. La seconde (P2) est que la mise en adéquation *cognitive* (la capacité à décrire correctement les états mentaux et les processus cognitifs des agents) est un *moyen* privilégié pour corriger ces inadéquations comportementales. Nous allons désormais étayer successivement chacune de ces deux prémisses.

5. Inadéquations comportementales (P1)

Si les modèles classiques de choix étaient parfaitement adéquats au niveau comportemental, alors, du point de vue de la variante de l'anti-réalisme ciblé qui nous intéresse ici, il n'y aurait rien à redire à ces modèles et la critique des tenants du PMRL serait nulle et non avenue. La seule façon de contrer l'objection fondée sur l'anti-réalisme ciblé tout en préservant ses prémisses fondamentales consiste à mettre en avant des inadéquations comportementales. Cela signifie, de notre point de vue, que la question ne peut être tranchée indépendamment de la condition empirique des modèles de décision: il faut prendre en compte le degré d'adéquation empirique dont les modèles de choix sont *de facto* capables. Bien sûr, on connaît aujourd'hui, en particulier dans les situations risquées et incertaines, nombre d'inadéquations comportementales.¹⁷ Mais pour que la démonstration soit réellement satisfaisante, il faut se limiter à celles qui paraissent étroitement liées aux limitations cognitives des agents.

Que pourraient être de telles inadéquations comportementales? Ce pourrait être, par exemple, des situations de décision intuitivement difficiles du point de vue cognitif où l'agent n'obéit pas au modèle classique de décision correspondant. Ou, mieux, une classe de situations de choix telles que, *plus* la situation est difficile du point de vue cognitif, *moins* correcte est la prédiction comportementale du modèle classique. Récemment, C. N. McKinney et J. Van Huyck (2006, 2007) ont mené une étude expérimentale sur le jeu de Nim, un jeu à deux joueurs, à information parfaite et à somme nulle. Dans ce jeu, plusieurs pierres sont disposées sur plusieurs lignes (il peut y avoir plusieurs pierres par ligne) et, chacun son tour, les joueurs doivent ôter une ou plusieurs pierres de l'une (et de pas plus d'une) de ces lignes. Le gagnant est celui qui ôte la dernière pierre. Voici par exemple une représentation de la configuration de départ d'un jeu de Nim à 4 lignes et 11 pierres:

| | | | | |
|---|---|---|---|---|
| 1 | • | • | | |
| 2 | • | • | • | |
| 3 | • | • | | |
| 4 | • | • | • | • |

FIGURE 1
Un jeu de Nim

On peut démontrer qu'il existe deux configurations du jeu: dans la première, il existe une stratégie gagnante pour le premier joueur, dans la seconde une stratégie gagnante pour le second joueur. A la différence, par exemple, du jeu d'échecs, il existe un algorithme efficace (connu depuis le début du vingtième siècle)¹⁸ pour déterminer la stratégie gagnante d'un jeu donné. Par ailleurs, différentes mesures de la difficulté d'un jeu de Nim sont utilisables, comme le nombre de pierres (le rang) ou la longueur du chemin le plus court, mesures qui permettent de classer les jeux en fonction de leur difficulté cognitive. Si l'on considère les choix faits par les sujets lorsqu'ils jouent en premier et qu'il existe une stratégie gagnante pour eux, on se rend compte que la fréquence de victoire (la victoire est dans ce cas ce que prédit le modèle classique de décision) diminue rapidement avec l'augmentation de la difficulté du jeu – difficulté mesurée par l'un des critères précédemment évoqués. Les performances des sujets sont comprises entre l'issue rationnelle (la victoire) et les performances d'un joueur qui choisirait aléatoirement contre un programme qui met en oeuvre l'algorithme efficace à chaque fois qu'il a le trait. Quand le jeu est simple, les joueurs se rapprochent de l'issue rationnelle, mais très rapidement, leurs performances convergent vers celle du joueur 'aléatoire'. Par ailleurs, quand on fait jouer les sujets les uns contre les autres et que la complexité est élevée (relativement à des estimations empiriques préalables), alors la valeur du jeu n'est plus prédictive de l'issue du jeu: un joueur ne gagne pas significativement plus quand il possède (en principe) une stratégie dominante que quand c'est son adversaire qui en possède une.

Les auteurs interprètent ainsi leurs résultats: 'La valeur substantivement rationnelle des jeux étudiés dans cet article n'aide pas du tout à prédire la performance des sujets dans cette expérience.' Il est vrai que l'on suppose connues les préférences de l'agent sur l'issue du jeu (l'agent préfère gagner que perdre). Si l'on accepte cependant cette hypothèse auxiliaire, semble-t-il, très raisonnable,¹⁹ alors on dispose bien d'une classe de situations où

les prédictions du modèle classique sont mises en défaut et où ces mises en défaut sont étroitement corrélées à la difficulté du jeu.

6. Monotonie des adéquations (P2)

La prémisse (P1) est insuffisante pour contrer l'objection anti-réaliste dans la mesure où elle ne donne aucun argument positif en faveur de l'idée qu'une modélisation explicite des processus délibératifs s'imposerait. Il faut fournir une justification séparée à une seconde prémisse, celle selon laquelle la mise en adéquation *cognitive* (la capacité à décrire correctement les états mentaux et les processus cognitifs des agents) est un *moyen* privilégié pour corriger ces inadéquations comportementales (P2).

Nous allons de nouveau nous appuyer sur la littérature expérimentale pour justifier cette seconde prémisse. Ce que nous voulons défendre, c'est qu'il existe une monotonie entre adéquation cognitive et adéquation comportementale: *ceteris paribus*, une meilleure adéquation comportementale fait progresser l'adéquation comportementale. Pour justifier une telle conjecture, il faut en savoir plus sur les processus cognitifs des agents. Ce n'est que récemment que l'on a développé une méthodologie appropriée en décision individuelle (Payne, Bettman & Johnson, 1993) et stratégique (Costa-Gomes, Crawford & Broseta, 2001; Johnson & *ali.*, 2002); Camerer & Johnson, 2004).²⁰ On pourra trouver dans Crawford (2008) un panorama assez complet des contributions qui concernent les situations stratégiques. La méthode principale consiste à présenter des problèmes de décision à des sujets sur un écran d'ordinateur de manière à ce que la structure du problème ne soit pas toute révélée simultanément, mais information par information. Cette démarche est illustrée par la Figure 2 'Présentation d'une situation de décision dans MouseLab'. On y voit deux options (caméras A et B), le sujet pouvant avoir des informations sur trois des caractéristiques de chacune des options. Dans l'écran reproduit, toutes ces caractéristiques sont cachées dans les différentes cellules. En faisant glisser la souris sur les cellules, les informations sont révélées une par une au sujet.

Les sujets doivent rechercher les caractéristiques de leur problème, et les expérimentateurs peuvent alors enregistrer la façon dont les sujets recherchent l'information. Bien sûr, les expérimentateurs enregistrent également les choix des sujets. L'expérimentateur obtient donc deux classes de données: d'une part des données d'un genre nouveau, que l'on peut qualifier de cognitives, d'autre part des données comportementales. L'analyse de ces données prend la forme suivante. Pour les différentes classes de jeux étudiées, on commence par spécifier un *ensemble de types de solutions* propres

| Camera A | Camera B |
|----------------------|----------------------|
| Price option A | Price Option B |
| Accessories option A | Accessories option B |
| Features option A | Features option B |
| Camera A | Camera B |

FIGURE 2
Présentation d'une situation de décision dans MouseLaB
 (<http://www.mouselabweb.org/demopage.html>)

aux jeux de la classe en question. Pour les jeux sous forme normale, cet ensemble contiendra bien sûr l'équilibre de Nash ('type nashien'), mais aussi, par exemple, le type *LI*, qui consiste à jouer une meilleure réponse à une distribution équiprobable sur les stratégies de son adversaire.

Chaque type de solutions – chaque modèle de décision interactive si l'on veut – a en premier lieu des implications *comportementales*: pour un jeu G et un type de solution k , il existe un certain nombre de stratégies qui sont compatibles avec le type de solution k . En général, les jeux proposés aux sujets sont choisis de telle sorte que ces stratégies soient en nombre restreint, voire uniques. Chaque type de solutions a également des implications *cognitives*: il existe notamment un ensemble de procédures – en matière de recherches d'information – qui permettent d'aboutir à la solution d'un jeu G à partir de sa description. Il peut exister différentes procédures pour une même solution. Afin qu'une série de données cognitives soit pertinente pour l'assignation d'un type de solutions, deux ensembles d'hypothèses sont

nécessaires: un premier ensemble d'hypothèses sur les procédures cognitives caractéristiques de chaque type, et un second sur les 'traces' attendues de ces procédures sur la recherche d'information.²¹

Prenons un exemple: dans un jeu sous forme normale, si un sujet est du type 'pessimiste', i.e. s'il joue une stratégie *Minimax* par rapport aux stratégies possibles de son adversaire, alors on doit s'attendre à ce qu'il cherche des informations sur ses propres paiements, mais peu (ou pas) sur ceux de son adversaire. Autre exemple: dans un jeu sous forme extensive, si un sujet est du type 'équilibre parfait en sous-jeux', on doit s'attendre à ce qu'il cherche à connaître ses paiements, et ceux de son adversaire, jusqu'aux nœuds terminaux de l'arbre de jeu – et éventuellement qu'il remonte des nœuds terminaux au nœud initial conformément à l'algorithme de *backward induction*.

Les données cognitives ainsi obtenues et analysées peuvent être utilisées et interprétées de différentes manières. On peut s'en servir (et c'est ce que font principalement Costa-Gomes, Crawford & Broseta (2001)) pour identifier plus finement le type des sujets que l'on ne pourrait le faire avec les données purement comportementales. Par exemple, du point de vue comportemental le type *L1* décrit plus haut est difficilement séparable du type *D1* (élimination des stratégies dominées, puis meilleure réponse à une distribution équiprobable sur les stratégies restantes de l'adversaire). Sur la base des seules données comportementales, ces deux types sont attribués au sujet assez fréquemment, et approximativement avec la même fréquence. En revanche, quand on prend en compte les données cognitives, la classification se trouve bouleversée, presque tous les sujets étant considérés comme relevant du type *L1*. Johnson & ali. (2002) se servent des données cognitives pour distinguer entre deux théories explicatives des déviations de l'équilibre dans des jeux de marchandage séquentiel: une théorie qui invoquerait la rationalité limitée des sujets, l'autre des préférences 'sociales' qui intégreraient des considérations d'équité. Ici encore, les données cognitives permettraient de trancher entre ces deux hypothèses (en faveur de l'hypothèse de rationalité limitée).

Cet usage des données cognitives n'est cependant pas le plus important pour notre argument. On veut en effet mettre en avant la *monotonie des adéquations*, autrement dit la corrélation (positive) entre la classification des sujets en fonction des données cognitives et la classification des sujets en fonction des données comportementales. Or, précisément, on observe une telle corrélation entre les données cognitives et les données comportementales. Un recensement de ces corrélations serait fastidieux ici,²² mais à titre d'exemple, on peut considérer l'étude Costa-Gomes & Crawford (2006)

qui suit la même méthode que celle que nous venons d'exposer et qui se consacre aux jeux de *guessing* à deux joueurs. Si l'on compare la classification comportementale et la classification cognitive des sujets, on se rend compte que ces classifications coïncident dans près de la moitié des cas. De la même façon, Camerer & Johnson (2004) interprètent leurs analyses comme montrant qu' 'il est possible de prédire statistiquement ce qu'un joueur offrira [il s'agit d'un jeu de marchandage] à partir des informations qu'il regarde le plus souvent'.^{2,3}

Conclusion

Gul & Pesendorfer (2005) sont les représentants les plus récents de l'objection anti-réaliste que nous avons discutée et selon laquelle, parce que l'objectif des modèles de choix est de prédire les comportements des agents, la méthodologie associée devrait être *mindless*. Notre thèse est en substance que, précisément parce que l'objectif est avant tout de prédire les comportements, la méthodologie associée devrait être *mindful*. Pour bien circonscrire la portée de l'argument qui a mené à cette thèse, nous admettons, d'une part, qu'il puisse être *en pratique* extrêmement difficile de fournir des modèles qui tiennent compte des données procédurales, et d'autre part, que l'objection fondée sur la variante collective de l'anti-réalisme reste à considérer. Mais, si l'argument est accepté, alors le projet de modélisation de la rationalité limitée, si radical qu'il puisse paraître, n'est pas le contre-sens méthodologique auquel les tenants de l'anti-réalisme l'assimilent parfois.

Mikaël Cozic est chercheur à l'Institut d'Histoire et de Philosophie des Sciences et des Techniques (Paris I-CNRS-ENS Ulm), au GREGHEC (CRNS-HEC) et au Département d'Etudes Cognitives (ENS Ulm). Ses thèmes de recherche actuels sont la méthodologie de l'économie (en particulier les rapports entre psychologie et économie) et les théories formelles de la rationalité. Il a tout récemment publié *Philosophie de la logique* (avec D. Bonnay, Paris: Vrin) et *Précis de Philosophie des Sciences* (avec A. Barberousse et D. Bonnay, Paris: Vuibert). *Adresse de l'auteur*: DEC, Ecole Normale Supérieure, 29, rue d'Ulm, F-75005 Paris. [*email*: mikael.cozic@ens.fr]

Notes

* L'auteur tient à remercier pour leurs commentaires et leurs critiques A. Barberousse, D. Bonnay, P. Egré, H. Galinon, B. Walliser. Nous avons une dette particulière envers Ph. Mongin pour ses encouragements et sa relecture minutieuse d'une version préliminaire.

1. Voir Camerer & Loewenstein (2003). Il faut préciser que nous ne soutenons pas que la qualification de 'révisionnisme empirique' vaut pour *toutes* les recherches de ce domaine

extrêmement hétérogène. Elle ne vaut par exemple probablement pas pour la neuroéconomie dans son état actuel: bien des contributions de la neuroéconomie ne se prononcent tout simplement pas sur la valeur des modèles classiques de décision. La question de savoir comment relier ces modèles de décision au langage et aux données des neurosciences est d'ailleurs particulièrement difficile.

2. Nous nous y sommes essayés dans le chapitre 3 de Cozic (2005).

3. C'est par exemple dans ce sens que le terme est employé dans le panorama influent de C. Camerer, 'Bounded rationality in decision making' (Camerer, 1998). Voir aussi Camerer & Fehr (2006: 47): 'Quand les individus manifestent des croyances systématiquement biaisées concernant les événements extérieurs ou le comportement d'autrui ou qu'ils dévient systématiquement de l'action qui satisfait au mieux leurs préférences, nous parlons de rationalité limitée.'

4. Supposons que C soit fini, que L et L' dénotent génériquement des loteries sur C et que $L(c)$ pour $c \in C$ dénote la probabilité assignée à la conséquence c par la loterie L . Le théorème de représentation garantit que si \geq satisfait les axiomes, il existe une fonction d'utilité $u(\cdot)$ telle que $L \geq L'$ ssi $\sum_{c \in C} [u(c) \cdot L(c)] \geq \sum_{c \in C} [u(c) \cdot L'(c)]$.

5. L'expression n'est pas de Friedman (1953) mais s'inscrit dans une retranscription logique de sa thèse. Mongin (1988) attire également l'attention sur l'importance de ce que nous appelons le domaine-cible. Selon lui, la position de Friedman 'paraît plutôt ressortir à une forme assez inédite d'instrumentalisme, celle qui *interprète des lois scientifiques réfutées comme vraies dans un certain domaine*' (p. 285).

6. Cette position est plus ou moins celle de l' 'empirisme constructif' de B. van Fraassen selon lequel 'le but de la science est de nous donner des théories qui sont empiriquement adéquates' (van Fraassen, 1980: 12).

7. Il faut noter le double déplacement: non seulement on passe de l'individuel au collectif, mais on inclut l'ensemble des *phénomènes* collectifs, et pas uniquement les *comportements*. Sous cette variante, en effet, on inclut non seulement les comportements individuels agrégés, mais également des variables collectives qui, dans les modèles du moins, ne se réduisent pas à des comportements.

8. Voir Mongin (2000), section 'Friedman, Machlup et la querelle du "réalisme des hypothèses"'.

9. Ces deux ciblés sont également combinés par la description de l'instrumentalisme que R. Sugden fournit dans un article méthodologique récent (Sugden, 2000: sec. 5): 'Dans l'interprétation instrumentaliste, l'objet de la théorie néo-classique est de prédire les changements dans les prix et les quantités totales échangées des différents biens comme résultant de changements exogènes (comme des changements de technologie ou de taxe). De ce point de vue, les statistiques économiques agrégées jouent le même rôle à l'égard de l'économie que le mouvement des corps céleste à l'égard de l'astronomie ancienne: ce sont les seuls phénomènes que l'on veut prédire, et les seules données (ou les seules données acceptables).'

10. Pour un panorama récent en langue française, voir Cohen & Tallon (2000).

11. Nous remercions Ph. Mongin pour cette remarque.

12. Voir par exemple le manuel Jehle & Reny (2001): 'Pour le moment, nous avons abordé la théorie de la demande en supposant que les consommateurs ont des préférences qui satisfont certaines propriétés (complétude, transitivité, monotonie stricte). Nous avons ensuite tenté de déduire toutes les propriétés observables de la demande de marché qui en découlent (équilibre budgétaire, symétrie, matrice de Slutsky semi-définie négative). Ainsi, nous avons commencé par supposer quelque chose concernant des entités que nous ne pouvons pas observer – les préférences – pour aboutir à des prédictions sur des choses que nous pouvons observer – le comportement de demande du consommateur.' (pp. 7–8)

13. Les tests non-paramétriques sont en général construits à partir de l'axiome GARP (Generalized Axiom of Revealed Preference). L'axiome porte sur des ensembles de paires paniers de biens-prix (p_t, x_t) , qui sont donc considérés comme observables. On peut montrer qu'un ensemble fini $\{(p_t, x_t)\}$ pour $t = 1, \dots, T$ satisfait GARP s'il existe une fonction d'utilité (localement non saturée) sur les paniers de bien qui rationalise les choix.

14. Stiglitz & Walsh (2002: 123–4); Kreps (1996: 6). Voir aussi les commentaires de Gilboa & Schmeidler (2001: 26).

15. Les auteurs parlent de 'données de préférence révélée' (*revealed preference data*) ou de 'données de choix' (*choice evidence*).

16. Voir par exemple: 'Si un économiste propose une nouvelle théorie qui n'est pas fondée sur des données de choix, alors ou bien la nouvelle théorie conduit à de nouvelles prédictions comportementales, auquel cas elle peut être testée à partir de données de préférence révélée, ou bien elle ne conduit pas à de nouvelles prédictions comportementales, auquel cas la modification est vide. En économie standard, les implications testables d'un modèle sont son contenu; une fois qu'elles sont identifiées, les données étrangères aux choix qui l'ont motivée deviennent non pertinentes' (Gul & Pesendorfer, 2005: 6–7). Notons que l'argumentaire de Gul & Pesendorfer est plus riche que ce que nous en retenons pour notre discussion et qu'il contient nombre de remarques qui nous semblent parfaitement légitimes. Hausman (2008), dont nous avons pris connaissance pendant la rédaction de la présente version de cet article, s'est livré à une reconstruction méticuleuse des différentes lignes argumentatives de l'essai. Il considère également que l'une de ces lignes argumentatives est précisément la ligne friedmanienne qui nous intéresse ici.

17. La littérature sur les anomalies empiriques du modèle d'espérance d'utilité est bien connue. Il vaut la peine de souligner que, depuis une dizaine d'années (et seulement depuis une dizaine d'années), le modèle de choix certain (et, en particulier, la théorie du consommateur) fait l'objet de test expérimentaux individuels, dont les résultats sont relativement équivoques. Voir Sippel (1997), Février & Visser (2004), Harbaugh, Krause & Berry (2001), Andreoni & Miller (2002) et Choi & *ali.* (2007).

18. L'algorithme est expliqué dans la section 1.5 de Binmore (1992).

19. Il convient de préciser que, comme de coutume, les sujets sont motivés financièrement à remporter les parties qu'ils jouent.

20. Notons que l'un des principaux contributeurs actuels à ces recherches, V. Crawford, a lui-même établi un lien entre le débat méthodologique qui fait suite à la contribution de Gul & Pesendorfer (2005) et les données cognitives dans un essai, Crawford (2008), dont nous avons pris connaissance pendant la rédaction de la présente version de cet article.

21. Costa-Gomes, Crawford & Broseta (2001): 'Chacun de nos types est associé naturellement avec certains algorithmes qui décrivent la manière dont les informations sur les paiements sont exploitées pour la prise de décision. En se servant de ces algorithmes comme de modèles des processus cognitifs du sujet et en s'appuyant sur deux hypothèses conservatrices quant à la façon dont la cognition est reliée à la recherche, on dispose d'assez de structure sur l'espace des suites possibles de recherche pour obtenir une caractérisation manipulable des implications sur la recherche de chacun des types' (p.1196, traduction de l'auteur).

22. Voir Cozic (2005: 184–91) qui se livre systématiquement à l'exercice pour Costa-Gomes, Crawford & Broseta (2001), Costa-Gomes & Crawford (2006) et Johnson & *ali.* (2002).

23. Voir également Camerer (2003): 'L'idée positiviste selon laquelle les règles de décision servent uniquement à prédire les choix, plutôt qu'elles ne contraignent nécessairement les détails du processus cognitif, autorise une absence de corrélation entre les décisions prises par le joueur et les mesures de traitement de l'information. Mais les décisions et les structures de

recherches (*search patterns*) sont corrélées. Pour le dire autrement, si la classification avait été établie sur la base des structures de recherches, les données cognitives auraient aidé à prédire ce que les sujets choisissent effectivement' (pp. 251–2).

Références

- Allais, M. (1953) 'Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'école américaine', *Econometrica* 21: 503–46
- Andreoni, J. & Miller, J. (2002) 'Giving according to Garp: an experimental test of the consistency of preferences for altruism', *Econometrica* 70(2): 737–53
- Binmore, K. (1992) *Fun and games: a text on game theory*. Lexington, MA: D. C. Heath & Co.
- Camerer, C. (1998) 'Bounded rationality in individual decision making', *Experimental economics* 1: 163–83.
- Camerer, C. (2003) *Behavioral game theory*. Princeton, NJ: Princeton University Press.
- Camerer, C. & Fehr, E. (2006) 'When "economic man" dominates social behavior', *Science* 311: 47–52.
- Camerer, C. & Johnson, E. (2004) 'Thinking about attention in games: backward and forward induction', in I. Brocas & J. Carillo (eds) *The psychology of economic decisions*, vol. 2, pp. 111–30. Oxford: Oxford University Press.
- Camerer, C. & Loewenstein, G. (2003) 'Behavioral economics: past, present, future', in C. Camerer et al. (eds) *Advances in behavioral economics*, pp. 3–52. Princeton, NJ: Princeton University Press.
- Camerer, C., Loewenstein, G. & Prelec, D. (2005) 'Neuroeconomics: how neuroscience can inform economics', *Journal of economic literature* XLIII: 9–64.
- Chiappori, P.-A. (1990) 'La théorie du consommateur est-elle réfutable?', *Revue économique* 41(6): 1001–25.
- Choi, S., Fisman, R., Gale, D. & Kariv, S. (2007) 'Consistency and heterogeneity in individual behavior under uncertainty', *The American economic review* 97(5): 1921–38.
- Cohen, M. & Tallon, J.-M. (2000) 'Décision dans le risque et l'incertain: l'apport des modèles non-additifs', *Revue d'économie politique* 110(5): 631–81
- Costa-Gomes, M. & Crawford, V. (2006) 'Cognition and behavior in two-person guessing games: an experimental study', *The American economic review* 96 (December): 1737–68.
- Costa-Gomes, M., Crawford, V. & Broseta, B. (2001) 'Cognition and behavior in normal-form games: an experimental study', *Econometrica* 69(5): 1193–235.
- Cozic, M. (2005) *Fondements cognitifs du choix en rationalité limitée*. Thèse de doctorat, Université de Paris IV Sorbonne, Paris.
- Crawford, V. (2008) 'Look-ups as the windows of the strategic soul: studying cognition via information search in game experiments', in A. Caplin & A. Schotter (eds) *Handbook of economic methodology*, vol. 1, pp. 249–80. Oxford: Oxford University Press.
- Ellsberg, D. (1961) 'Risk, ambiguity and the savage axioms', *Quarterly journal of economics* 75: 643–69.
- Février, P. & Visser, M. (2004) 'A study of consumer behavior using laboratory data', *Experimental economics* 7: 93–114.
- van Fraassen, B. (1980) *The scientific image*. Oxford: Clarendon Press.
- Friedman, M. (1953) 'The methodology of positive economics', in *Essays in positive economics*, pp. 3–46. Chicago, IL: University of Chicago Press.

- Gilboa, I. & Schmeidler, D. (2001) *A theory of case-based decisions*. Cambridge: Cambridge University Press.
- Guala, F. (2006) 'Has game theory been refuted?', *Journal of philosophy* 103: 239–63.
- Gul, F. & Pesendorfer, W. (2005) 'The case for mindless economics', *Mimeo*.
- Harbaugh, W., Krause, K. & Berry, T. (2001) 'GARP for kids: on the development of rational choice behavior', *The American economic review* 91(5): 1539–45.
- Hausman, D. (2008) 'Mindless of mindful economics: a methodological evaluation', in A. Caplin & A. Schotter (eds) *Handbook of economic methodology*, vol. 1, pp. 125–54. Oxford: Oxford University Press.
- Jehle, G. & Reny, Ph. (2001) *Advanced microeconomic theory*. Indianapolis, IN: Addison-Wesley.
- Johnson, E., Camerer, C., Sen, S. & Rymon, T. (2002) 'Detecting failures of backward induction: monitoring information search in sequential bargaining', *Journal of economic theory* 104(1): 16–47.
- Kahneman, D. (2003) 'Maps of bounded rationality: psychology for behavioral economics', *The American economic review* 93(5): 1449–75.
- Kahneman, D. & Tversky, A. (1979) 'Prospect theory: an analysis of decision under risk', *Econometrica* 47(2): 263–91.
- Kahneman, D. & Tversky, A., eds (2000) *Choice, values and frames*. Cambridge: Cambridge University Press.
- Kreps, D. (1996) *Leçons de théorie microéconomique*. Paris: PUF.
- Luce, R. & Raiffa, H. (1985) *Games and decisions: introduction and critical survey*. New York: Dover (2nd edn).
- Machlup, F. (1967) 'Theories of the firm: marginalist, behavioral, managerial', *The American economic review* 57(1): 1–33.
- McKinney, C. & Huyck, J. (2006) 'Does seeing more deeply into a game increase one's chances of winning?', *Experimental economics* 9: 297–303.
- McKinney, C. & Huyck, J. (2007) 'Estimating bounded rationality and pricing performance under uncertainty', *Journal of economic behavior and organization* 62: 625–39.
- Mongin, Ph. (1988) 'Le réalisme des hypothèses et la partial interpretation view', *Philosophy of the social sciences* 18: 281–325.
- Mongin, Ph. (2000) 'La méthodologie économique au XXème siècle. Les controverses en théorie de l'entreprise et la théorie des préférences révélées', in A. Béraud & G. Faccarello (eds) *Nouvelle histoire de la pensée économique*, t. 3, ch. 36, pp. 340–78. Paris: La Découverte.
- von Neumann, J. & Morgenstern, O. (1944) *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press
- Payne, J., Bettman, J. & Johnson, E. (1993) *The adaptive decision maker*. Cambridge: Cambridge University Press.
- Savage, L. (1954) *The foundations of statistics*. New York: Dover (2nd edn 1972).
- Sen, A. (1971) 'Choice functions and revealed preference', *Review of economic studies* 38: 307–17
- Simon, H. (1955) 'A behavioral model of rational choice', *The Quarterly journal of economics* 69: 129–38.
- Sippel, R. (1997) 'An experiment on the pure theory of consumer's behavior', *Economic journal* 107: 1431–44.
- Stiglitz, J. & Walsh, C. (2002) *Principles of microeconomics*. New York: W. W. Norton.
- Sugden, R. (2000) 'Credible worlds: the status of theoretical models in economics', *Journal of economic methodology* 7(1): 1–31.
- Varian, H. (1992) *Microeconomic analysis*. New York: W. W. Norton.