

# La théorie de la décision et la psychologie du sens commun

Social Science Information

50(3–4) 351–374

© The Author(s) 2011

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0539018411411019

ssi.sagepub.com



**Philippe Mongin**

Centre National de la Recherche Scientifique & Ecole des Hautes Etudes Commerciales, France

## Résumé

L'article compare philosophiquement la théorie mathématique de la décision individuelle, d'une part, et la conception psychologique ordinaire de l'action, du désir et de la croyance, d'autre part. Il délimite plus strictement son objet en étudiant, sous cet angle comparatif, le système de Savage et son concept technique de probabilité subjective, rapporté, comme celui de Ramsey, au modèle élémentaire du pari. L'examen est scandé par trois thèses philosophiques: (i) la théorie de la décision n'est que la psychologie commune mise en langage formel (Lewis), (ii) la première améliore substantiellement la seconde, mais ne s'affranchit pas des limites caractéristiques de celle-ci, en particulier de son inaptitude à séparer empiriquement le désir et la croyance (Davidson), (iii) la première améliore substantiellement la seconde, et par les innovations qu'elle comporte, s'affranchit de certaines de ses limitations. On s'est donné pour but d'établir la thèse (iii) à la fois contre la thèse trop simple (i) et contre la thèse subtile (ii).

## Mots-clefs

croyance, Davidson, désir, Lewis, probabilité subjective, psychologie du sens commun, Ramsey, Savage, théorie de la décision, utilité dépendante des états

## Abstract

Taking the philosophical standpoint, this article compares the mathematical theory of individual decision-making with the folk psychology conception of action, desire and belief. It narrows down its topic by carrying the comparison for Savage's system and his technical concept of subjective probability, which, like Ramsey's, is referred to the basic model of betting. The argument is organized around three philosophical theses: (i) decision theory is nothing but folk psychology stated in formal language (Lewis), (ii)

---

## Corresponding author:

Philippe Mongin, HEC, 1 rue de la Libération, F-78350 Jouy-en-Josas, France

Email: mongin@greg-hec.com

the former substantially improves on the latter, but is unable to overcome its typical limitations, especially its failure to separate desire and belief empirically (Davidson), (iii) the former substantially improves on the latter and, through these innovations, overcomes some of the limitations. The aim of the article is to establish (iii) not only against the all too simple thesis (i), but also against the subtle thesis (ii).

### Keywords

belief, Davidson, decision theory, desire, folk psychology, Lewis, Ramsey, Savage, state-dependent utility, subjective probability

‘All our lives we are in a sense betting’ (FP Ramsey, 1931)

## I La théorie de la décision et ses problèmes philosophiques

La théorie de la décision est une discipline mathématisée qui étudie les décisions d’un agent, le plus souvent d’un individu, qu’elle suppose rationnel et considère isolément des autres. Avec la théorie des jeux et la théorie des choix collectifs, elle fait partie des spécialités mathématiques du choix rationnel, qui servent d’auxiliaires techniques à l’économie et, beaucoup plus rarement, à d’autres sciences sociales. Quand, par exemple, la sociologie ou la science politique se réclament des *théories du choix rationnel*, il s’agit le plus souvent de conceptions informelles et strictement qualitatives, dont la proximité avec le sens commun se révèle spontanément. Nous ne traiterons pas ici de ce corpus bien qu’il aborde les mêmes objets que les trois théories précédentes.<sup>1</sup>

Le groupe qu’elles forment est structuré, mais n’est pas entièrement unifié, et il se découpe différemment selon le critère qu’on privilégie. En considérant d’abord *le nombre et l’interaction des agents*, on dira que la théorie de la décision porte sur les actions de l’individu isolé, la théorie des jeux sur les actions des individus multiples et qui interagissent, et la théorie des choix collectifs sur les actions des individus multiples, mais qui n’interagissent pas. La classification isole clairement la première théorie des deux autres, mais ainsi formulée, elle menace de la transformer en ‘robinsonnade’ futile, et comme, en fait, la théorie couvre des situations réelles où les agents ne sont pas isolés, il faut lui trouver un meilleur critère d’identification. Celui que nous allons considérer, *le savoir de l’agent et l’emploi qu’il en fait*, répond à la question tout en nous rapprochant du sujet de l’article.

Quand il est incertain, le savoir de l’agent peut se présenter de deux façons très différentes. *L’incertitude naturelle* a trait aux phénomènes dont l’agent ignore ou reconnaît imparfaitement l’occurrence, mais que, pour les besoins de la décision qu’il doit prendre, il peut supposer déterminés en dehors de celle-ci. L’hypothèse convient lorsque ces phénomènes relèvent de la nature extérieure; alors, ils ne résultent pas eux-mêmes d’une décision et, en particulier, ils ne reflètent pas la malveillance d’un adversaire supposé. *Raffiniert ist der Herrgott, aber boshaft ist Er nicht*. *L’incertitude stratégique* a trait aux phénomènes que l’agent doit au contraire supposer dépendants de sa propre décision s’il veut la prendre correctement. Il en va ainsi dans certains jeux de société et

en stratégie militaire, du moins lorsque celle-ci est bien comprise, ce qui est loin d'être toujours le cas. L'agent suppose alors que le phénomène dont il est incertain résulte des décisions d'autres agents prises à la manière des siennes, quoique avec d'autres objectifs, ce qui l'engage dans une séquence vertigineuse d'anticipations mutuelles. La théorie des jeux s'occupe à titre principal d'incertitude stratégique, en répondant par ses concepts d'équilibre au problème difficile soulevé par les anticipations mutuelles. Quant à la théorie de la décision, elle s'occupe à titre exclusif d'incertitude naturelle, en répondant par ses concepts d'optimalité au problème plus simple, mais nullement trivial, de décrire la meilleure décision que l'agent doit prendre à la lumière de son anticipation, la seule qui soit alors en cause.

La force de la distinction précédente est qu'elle ne dépend pas principalement de la nature des phénomènes sur lesquels porte l'anticipation. Certes, un météorologue qui s'interroge sur les canicules estivales est dans une tout autre disposition de pensée qu'un candidat aux élections qui examine ses chances en fonction de celles de ses concurrents. Mais le météorologue peut se muer en stratège s'il pense que les canicules résultent en partie de l'effet anthropique et que la prévision plus ou moins inquiétante qu'il fait d'elles est susceptible de contrarier cet effet. Et l'homme politique peut lier ses chances de succès moins aux actions des concurrents qu'à l'issue de la crise financière, dont il n'influence pas le cours et qu'il peut alors se représenter comme un phénomène naturel. Ainsi, la typologie des incertitudes ne porte pas tant sur les phénomènes que sur l'agent lui-même et sur la rationalité à laquelle il se conforme; elle est, pour le dire plus exactement, l'affaire du théoricien qui décide de représenter l'agent suivant tel ou tel idéal-type rationnel, étant donné le phénomène considéré. La circularité que la distinction manifeste entre les types d'incertitude et de rationalité renforce le privilège des choix théoriques sur les classifications empiriques. La théorie de la décision et celle des jeux se caractérisent finalement par leur manière propre d'envisager des situations qui peuvent être les mêmes ou très voisines, le choix entre elles dépendant des idéalizations que l'observateur juge préférables.

Maintenant adossée au concept d'incertitude, la séparation disciplinaire présente un inconvénient que lui épargnait l'autre critère: elle semble exclure les décisions prises en pleine connaissance de cause, sur lesquelles on attendrait un développement particulier. Mais il est possible de voir dans la certitude une forme-limite de l'incertitude naturelle plutôt qu'une situation effective. Il ne sera pas besoin de développer séparément la *théorie du choix certain*, comme on la nomme dans les manuels de base; elle tiendra dans un simple chapitre introductif du cours de théorie de la décision.

L'outil probabiliste facilite l'absorption de la certitude par l'incertitude naturelle, et il rend bien d'autres services encore, mais il serait très incorrect de réduire l'extension de celle-ci à ce que permet cet outil. A coup sûr, la théorie de la décision a un modèle de base qui est probabiliste, celui de *l'espérance d'utilité*: parmi toutes les actions disponibles, l'agent rationnel choisirait celle qui maximise la somme pondérée des valeurs d'utilité des conséquences, les valeurs de probabilités des états fournissant les coefficients de pondération. Le modèle est restrictif non seulement parce que la fonction d'évaluation qu'il emploie dépend linéairement des probabilités, mais, plus fondamentalement, parce qu'il suppose au départ des probabilités qui sont bien définies. L'appellation de *bayésien* que la théorie lui accole désigne peu rigoureusement ces deux caractères à la fois.<sup>2</sup> L'hypothèse

de valeurs de probabilités qui sont ou bien déjà données, ou bien, par extension, susceptibles d'être produites, se défend lorsqu'on observe un gestionnaire d'assurance-vie ou un expert en management des risques, mais elle peut devenir ailleurs douteuse et même sérieusement contre-intuitive. Que l'on pense au climatologue qui anticipe le réchauffement de la planète, au géopoliticien qui guette les vagues d'attentats, au financier qui attend le retour des bulles spéculatives. Ainsi, l'incertitude naturelle se subdivise en trois classes de situations: *déjà probabilisées*, *probabilisables* et *non probabilisables*. La théorie de la décision s'occupe du dernier cas non moins que des deux premiers, et au fil des ans, il est même devenu le centre de ses préoccupations. Seule une vision datée de ses résultats peut faire négliger cette tendance.<sup>3</sup>

La théorie de la décision est de nature à solliciter la réflexion philosophique à plus d'un titre. Comme elle obéit à l'heuristique générale de la rationalité individuelle, et qu'on peut même voir dans ses modèles autant de tentatives faites pour définir axiomatiquement cette notion, elle est un point de passage obligé pour la philosophie de l'économie et elle peut servir de repère éventuel à la philosophie des autres sciences sociales. La controverse rebattue entre Rawls et Harsanyi sur la 'position originelle' montre quelle importance elle peut revêtir en philosophie morale et politique lorsque celle-ci croit bon d'invoquer la rationalité dans l'incertain. Par ailleurs, la méthodologie des sciences aurait lieu d'examiner les techniques spéciales de la théorie de la décision en les comparant à celles d'autres disciplines. Mathématisée, elle ne l'est pas comme n'importe quelle théorie formalisée; elle tire vers l'emploi d'un genre assez particulier d'axiomatisation, inauguré par von Neumann et Morgenstern (1944–1947), qui met en rapport des formalismes symboliques avec des formalismes numériques.<sup>4</sup> Une de ses énigmes philosophiques est que, partant de principes normatifs qu'elle développe mathématiquement, c'est-à-dire a priori, elle ait pu nourrir en hypothèses la psychologie et l'économie expérimentales de la décision; celles-ci ont littéralement grandi à son ombre. La nature des hypothèses qu'elle transmet ainsi à des disciplines empiriques requerrait une étude plus approfondie que les linéaments actuels.<sup>5</sup> Enfin, l'outil probabiliste fait, comme nous l'avons dit, l'objet d'une problématique interne à la théorie. Chez Savage (1954–1972), à qui on doit le premier traitement systématique de l'incertitude probabilisable, elle fonde une interprétation subjectiviste de la probabilité ainsi que de l'inférence statistique. Ainsi, au-delà des études méthodologiques précédentes, la théorie de la décision communique avec l'épistémologie fondamentale. Cette liaison-là, du moins, est reconnue de longue date par les philosophes, même s'ils se réfèrent moins volontiers à Savage qu'à ses prédécesseurs, comme Ramsey.<sup>6</sup>

Le présent travail manifestera une autre liaison philosophique encore. Son objectif est d'apprécier au mieux ce qui rapproche et ce qui éloigne la théorie de la décision et la psychologie du sens commun. Or la philosophie contemporaine des sciences cognitives s'est longuement interrogée sur le rôle que la psychologie populaire – *folk psychology* – occuperait dans ces disciplines si elles parvenaient à un stade plus élevé de développement. Il s'est ensuivi une controverse, qui peut sembler aujourd'hui quelque peu datée, mais qui ne s'est pas résolue pour autant, entre l'éliminationnisme et d'autres positions plus mesurées. Pour ne citer que deux auteurs emblématiques, Churchland (1981) a proclamé l'inéluctable disparition de la *folk psychology* au profit de la neuroscience, tandis que Fodor (1987) a maintenu qu'elle anticipait la psychologie scientifique de demain.<sup>7</sup> Nous n'avons pas de titre à réexaminer cette controverse, mais pouvons espérer faire un peu de lumière sur elle et ce qui l'a fait surgir.

Notre enquête peut conduire à serrer les liens de la théorie de la décision avec la psychologie du sens commun; si telle était la conclusion, il serait loisible de confronter, à propos de la première, les thèses de la controverse qui s'affrontent à propos de la seconde. Mais la conclusion peut être toute différente; il se peut qu'elle place la théorie de la décision du côté d'une psychologie savante d'un certain type. Dans cas, l'étude monographique de la théorie de la décision aurait pour effet d'ajouter un terme au découpage trop simple entre la psychologie populaire et une psychologie savante de type neuroscientifique. On agirait sur les termes de la controverse antérieure au lieu, simplement, de lui donner matière à se réexprimer. Quoi qu'il arrive finalement, la théorie aura manifesté sa pertinence pour la philosophie des sciences cognitives.<sup>8</sup>

Nous insistons sur le rapprochement avec cette branche de la philosophie, parce qu'elle est, contre toute attente, l'une de celles où la théorie de la décision est le moins bien connue. Il semble bien que la théorie n'informe la controverse, ni chez les partisans résolus ou modérés de la psychologie populaire, ni chez ses adversaires éliminativistes. La lacune est d'autant plus curieuse que deux philosophes qui passent pour avoir inspiré la position favorable se sont exprimés sur elle: Davidson, dans des parties développées des *Essays on Actions and Events* (1980), et Lewis, dans des passages elliptiques des *Philosophical Papers* (1983, 1986). Pour ces deux auteurs, qui n'emploient pas encore la locution de *folk psychology*, la théorie de la décision reproduit, sous une forme rigoureuse et systématisée, certains raisonnements psychologiques de la vie courante: *l'explication par les raisons* (*reason explanation*) si l'on en croit Davidson et *la théorie des personnes suivant le bon sens* (*commonsense theory of persons*) si l'on en croit Lewis.

Quelques extraits des deux philosophes, à la section suivante, nous aideront à cerner deux thèses qui ordonneront l'examen philosophique ultérieur.

- (T1) La théorie de la décision n'est que la mise en forme de la psychologie du sens commun.
- (T2) La théorie de la décision améliore substantiellement la psychologie du sens commun, mais ne s'affranchit pas des principes et des limites caractéristiques de celle-ci.

La première peut s'attribuer à Lewis; nous l'écartérons parce qu'elle ne décrit pas adéquatement la théorie. La deuxième est celle de Davidson. Plus subtile, elle est aussi plus fuyante, ce qui fait deux bonnes raisons de l'approfondir. Nous cheminerons en sa compagnie avant de l'abandonner finalement au profit d'une autre, qui est plus généreuse sans l'être tout à fait pour autant.

- (T3) La théorie de la décision améliore substantiellement la psychologie du sens commun, et en développant des concepts et des principes nouveaux par rapport à elle, s'affranchit de certaines de ses limitations.

Quels concepts et principes de la théorie élaborent la psychologie ordinaire, et quels autres innove et lui confèrent son indépendance, quelles limitations perdurent et quelles autres disparaissent, voilà ce que nous voudrions être en mesure d'expliquer

globalement. Mais cette brève étude sera moins ambitieuse. Il est possible de trancher entre (T1), (T2) et (T3) en échantillonnant les concepts et les principes, et c'est ainsi que nous procéderons.

On obtient des variantes des trois thèses en remplaçant chaque fois 'théorie de la décision' par 'modèle de l'espérance d'utilité'. Les représentants historiques des deux premières n'emploient jamais que ce modèle particulier dans leurs commentaires de la théorie. Ils avaient – surtout Davidson, moins nettement Lewis – l'excuse du moment où ils écrivaient et, même transportés aujourd'hui, ils se défendraient sans doute au motif qu'il est préférable de commenter les théories canoniques, donc anciennes, que celles qui sont en train de s'élaborer. Il est un peu trop commode, de la part des philosophes, d'aller répétant que l'oiseau de Minerve ne s'envole qu'au crépuscule, mais c'est une autre facilité, de la part des techniciens, que de gagner contre les philosophes des procès en ignorance. Parce que nous voulons dialoguer avec ceux qui ne connaissent de la théorie de la décision que son modèle premier, et aussi parce qu'il est plus adroit d'établir (T3) à moindres frais, nous nous en tiendrons à celui-ci, malgré la tendance récente signalée plus haut. Mais nous l'appréhenderons à son maximum de généralité, en relisant Savage plutôt que ses prédécesseurs moins satisfaisants – Ramsey (1931), de Finetti (1937) et von Neumann et Morgenstern (1944–1947).

## 2 La théorie de la décision rencontre la philosophie de l'action

On connaît la thèse de Davidson voulant que, sous certaines conditions qu'il s'efforce de caractériser, l'action intentionnelle résulte causalement, et donc tire son explication, de ce qui la justifie, c'est-à-dire de ses raisons. Par nature, celles-ci relèvent de deux ordres, le désir et la croyance, que l'auteur de l'explication tentera de démêler en s'appuyant sur les données empiriques à sa disposition. Comme les données se ramènent encore et toujours à des actions effectuées, la dualité des raisons subjectives débouche sur une indétermination caractéristique: des couples différents d'hypothèses faites sur les désirs et de croyances rendent compte également bien des mêmes actions. Il est concevable d'étendre les données au-delà du seul comportement. Mais si l'observateur accroît son explanandum de données verbales, par exemple de réponses à des questionnaires, il devra simultanément élargir son explanans; il lui faut articuler des principes d'interprétation aux principes de rationalité pratique dont il se servait déjà; ainsi, la course des hypothèses avec les données empiriques apparaît sans limite.<sup>9</sup>

Dans l'explication par les raisons, Davidson voit un phénomène spontané de psychologie sociale tout autant qu'une forme réfléchie du raisonnement hypothétique. Ce phénomène, qu'il ne cherche pas à distinguer de la simple compréhension d'autrui, est à ses yeux universel dans la provenance et dans l'application: d'une part, il est accessible à tout homme en situation normale; d'autre part, il vaut pour toutes les actions qui demandent à être comprises ou expliquées. Davidson ne veut pas nier qu'il soit possible et quelquefois souhaitable de raffiner l'attribution intentionnelle, mais il manifeste à l'égard de ces tentatives un scepticisme appuyé. Car pour lui, l'indétermination des réponses explicatives est inhérente au procédé d'explication par les raisons subjectives, et non pas seulement due aux variantes naïves qu'en manifeste la vie quotidienne.

Finalement, ces variantes obtiennent des résultats suffisants, par rapport aux faibles efforts consentis, pour qu'il n'y ait pas souvent lieu de les perfectionner. C'est à ce point, bien entendu, que Davidson préfigure la thèse favorable à la psychologie populaire dans la controverse qu'évoquait la section précédente.

La théorie de la décision lui sert justement à faire sentir les limitations des raisonnements plus techniques. Dans le chapitre 14 de *Actions and Events*, qui est le plus révélateur pour notre propos avec le chapitre 12, Davidson commence par définir l'explication par les raisons, puis il enchaîne ainsi vers la théorie de la décision:

The discussion has been hampered ... by my sticking to a particularly simple form of reason explanation, and this has prevented me from saying anything sensible about a number of problems, such as how an agent might be expected to choose among several competing actions, each of which is recommended by reasons he has. Similarly, no mention has been made of the effect of variations in the strength of desire, or degree of belief. The theory of decision under uncertainty is designed to cope with these matters. ... *it offers a more sophisticated way of dealing with reason explanation, and is still based on commonsense ideas of how actions are explained* .... (1980: 268; c'est nous qui soulignons)

Le passage affirme déjà que la théorie de la décision reste inscrite dans la psychologie du sens commun. La conclusion dernière sera – en substance, car l'écriture est plus nuancée – que les postulats de la théorie ne font pas mieux que faciliter le repérage des désirs et des croyances chez les agents dont on veut expliquer les choix. Par la cohérence qu'ils instaurent entre les notions, et la détermination quantitative qu'ils impliquent, ils affrontent mieux que la psychologie ordinaire l'indétermination du procédé qui est le sien, mais ils sont incapables de la surmonter; ils entament l'adversaire sans l'abattre. D'ailleurs, malgré leur apparence d'universalité, les postulats de la théorie de la décision ne constituent pas des lois authentiques des choix humains, ou même des seuls choix humains rationnels. *En bref, il faut y voir des principes qui renforceraient l'attribution explicative ordinaire, et non pas les prémisses d'une psychologie scientifique de l'action.*

Contrairement à Davidson, Lewis ne s'est jamais vraiment penché sur la théorie de la décision, mais il ne se prive pas de faire connaître son jugement:

Decision theory (at least if we omit the frills) is not esoteric science, however unfamiliar it may seem to an outsider. Rather, it is a systematic exposition of the consequences of certain platitudes about belief, desire, preference and choice. It is the very core of our common-sense theory of persons, dissected out and elegantly systematized ... The machinery derived from such systematizations is common property, whether or not the accompanying jargon is. (1983: 114)

L'interprétation paraît à la fois plus et moins générale que la précédente. D'une part, contrairement à Davidson, Lewis n'appréhende pas la psychologie du sens commun sous l'angle exclusif ou privilégié de la fonction explicative. Mais d'autre part, il la présente comme si elle constituait une *théorie*, ce que Davidson se gardait bien d'écrire. Sans doute, la langue philosophique anglo-saxonne prend-elle souvent *theory* dans le sens d'une simple hypothèse générale, sans élaboration théorique particulière, mais quelques

lignes plus haut, le mot venait de revêtir un sens fort avec *decision theory*, et il serait maladroît dans ce contexte de glisser vers le sens faible. A tort ou à raison, ce passage et un petit nombre d'autres aidant, on attribue à Lewis une défense de la psychologie populaire dans la version spéciale qu'on a désignée par la suite comme théorique (la "*theory theory*" de la controverse actuelle; voir par exemple Stich, 1983: ch. 2). Que la théorie de la décision s'occupe de la personne entière, plutôt que, seulement, de ses croyances et de ses désirs, c'est là une autre thèse forte et contestable que semble affirmer Lewis. Mais ces différents aspects de sa position nous importent moins que le réductionnisme flagrant qui l'accompagne: *la théorie de la décision ne se distinguerait des intuitions ordinaires que par la sélection qu'elle opère, les regroupements qu'elle effectue et la précision terminologique dont elle accompagne ces opérations*. Bien qu'on puisse la décrire aussi comme réductionniste, l'interprétation de Davidson est plus généreuse: elle n'invoque pas seulement la mise en forme, mais des perfectionnements substantiels, comme ceux qu'apportent la formulation rigoureuse des postulats et la précision quantitative des fonctions d'évaluation.

Il est d'autant plus instructif de comparer Lewis et Davidson qu'ils attribuent à la théorie de la décision le même contenu restrictif. Chez eux, nous l'avons dit, elle se confond avec le modèle de l'espérance d'utilité. Comme d'habitude, Lewis est le plus carré des deux: 'I think common sense psychology, systematized, should look a lot like that Bayesian decision theory' (1983: 148-149).

Cette phrase résume adéquatement la thèse (T1) dans la variante restreinte qu'annonçait l'introduction. Davidson ne parle ni de bayésianisme, ni d'espérance utilité, mais seulement de théorie de la décision, comme si elle était une. Il distingue cependant la variante de Ramsey, qui est pour lui fondamentale, et celle de von Neumann et Morgenstern, qu'il juge moins intéressante (1980: 234). De fait, l'article 'Truth and Probability' préfigure la conception explicative rappelée plus haut, à cette différence capitale près que Ramsey ne voyait pas dans la dualité des raisons un obstacle intangible au raffinement de cette conception. Tout au contraire, la révélation de la probabilité subjective par les paris, qui est l'idée fondamentale de son article, devait servir à dépasser l'indétermination que Davidson présentera ensuite comme irrémédiable. Les philosophes qui se réclament des deux auteurs à la fois n'en font pas assez ressortir la divergence; d'une certaine manière, elle recoupe celle de nos deux thèses (T2) et (T3). Pour la théorie de la décision contemporaine, Ramsey n'est qu'un précurseur de Savage, et le travail axiomatique du second l'emporte sur celui du premier à tous les égards: les postulats sont à une plus grande distance heuristique du modèle de l'espérance d'utilité, qui forme la conclusion désirée; ils sont moins nombreux, plus commodément formulés, et on peut en examiner l'indépendance logique, ce qui est presque impossible avec les postulats de Ramsey; enfin, la conclusion elle-même est plus ferme (à la différence de son prédécesseur, Savage distingue l'existence et l'unicité de la mesure de probabilité subjective qu'il dérive). Comme notre travail n'a pas à s'embarasser du détail historique, nous pouvons supposer que Davidson, ainsi que, plus généralement, les tenants de la thèse (T2), s'expriment par rapport au représentant canonique de la théorie.<sup>10</sup>

Il faut maintenant préciser l'autre terme de la comparaison. Par la *psychologie du sens commun*, nous entendrons désormais, restrictivement, celle qui traite de l'action intentionnelle ou, pour le dire mieux, de l'action qui se laisse envisager comme telle. Il paraît

difficile de retenir l'idée, particulière à Lewis, qu'une conception entière de la personne servirait de répondant à la théorie de la décision. Une personne est aussi un centre d'émotions, de sensations, d'activités réflexes, qu'on ne saurait pas rapporter à quoi que soit dans la théorie. Il faut admettre que celle-ci a des préoccupations disciplinaires étroites, et l'on se gardera de confondre la limitation de ses objets avec l'économie de concepts dont elle fait montre lorsqu'elle étudie les mêmes objets que la psychologie du sens commun. Ainsi, le langage ordinaire a une palette riche et nuancée de notions téléologiques – le but, la fin, l'objectif, la valeur, la fonction – alors que la théorie se contente pratiquement de définir des préférences. La concentration qu'elle réalise là est d'une autre nature que son refus d'envisager, par exemple, les émotions qui sous-tendent la prise de décision.

La psychologie ordinaire de l'action intentionnelle comporte deux aspects qu'il est important de distinguer. D'une part, l'observateur *attribue* au sujet les attitudes de désir et de croyance qui, d'après lui, structurent l'intention. D'autre part, il *exploite* l'attribution à des fins épistémiques diverses: l'explication, la prédiction, la justification, et d'autres encore. Il n'y a pas de raison particulière pour qu'il se fixe sur l'explication plutôt que sur une autre fin. De ce point de vue, Davidson et la philosophie de l'action qu'il a inspirée sont trop restrictifs, alors qu'en parlant de *théorie*, Lewis et ses successeurs évitent de préjuger de ce que sera l'exploitation; c'est un avantage certain de leur terme problématique. La restriction des fins épistémiques est d'autant plus contestable que l'observateur ne peut atteindre l'une que s'il en atteint d'autres simultanément. De fait, le principe même de l'explication par les raisons est d'utiliser dans un rôle explicatif ce qui se présente d'abord comme une justification; et une explication de type quelconque ne mérite peut-être pas d'être considérée comme telle si elle ne prédit pas. Ainsi, notre conception de la psychologie du sens commun se décale à la fois de celle de Davidson et de celle de Lewis, chaque fois dans le sens de la généralité. Elle vise à neutraliser les désaccords entre les partisans de (T1), (T2) et (T3) qui ne tiendraient pas à ces affirmations seules.

Une autre distinction s'impose, qui, pour le coup, semble reproduire celle de nos deux sources philosophiques. Pour comparer la théorie de la décision et la psychologie du sens commun, on peut faire porter l'examen sur des *concepts* qui se répondent, en particulier sur le couple probabilité–utilité, dans l'une, et le couple croyance–désir, dans l'autre; c'est dans l'ensemble ainsi que procède Davidson. Mais on peut aussi considérer des *propositions* remarquables, c'est-à-dire les principes et théorèmes, dans l'une, et les 'plattitudes', dans l'autre; on peut supposer que Lewis procéderait de cette manière s'il développait sa critique. Nous commencerons par examiner des concepts avant de mettre à l'épreuve un principe, appelé ici d'universalité des paris, mais nous privilégierons le point de vue conceptuel. Un tel ordre convient non seulement à la philosophie générale du jugement, mais aux partis pris stratégiques de cet article: nous écarterons la thèse de Lewis plus vite que celle de Davidson, et le désaccord avec celui-ci ressortira d'autant mieux que nous aurons généralement adopté sa manière.

Une étude plus poussée que celle-ci ferait d'abord le bilan de tous les concepts qui sont de nature à guider la comparaison des deux blocs. Ils se rangeraient en trois classes dont chacune a son importance: ceux qui, à l'évidence, n'appartiennent qu'à l'un; ceux qui, à l'évidence, n'appartiennent qu'à l'autre; et ceux qui semblent partagés. Plus

expéditive, notre analyse partira des concepts organisateurs chez Savage pour explorer leurs répondants psychologiques ordinaires. En substance, nous nous intéresserons à la dernière classe plutôt qu'aux deux premières, et nous chercherons à y accéder depuis la théorie de la décision, alors que le cheminement inverse est aussi défendable.

### 3 Les concepts la théorie de la décision et les concepts psychologiques ordinaires

Léonard veut se préparer une omelette. Il vient de casser cinq œufs et d'en mélanger le contenu dans une assiette. Or il dispose encore d'un œuf dont la fraîcheur est incertaine. Léonard doit affronter le problème de décision suivant. En cassant le sixième œuf directement dans l'assiette des cinq premiers, il aura une omelette de six œufs si cet œuf est frais et pas d'omelette du tout sinon. En cassant le sixième œuf dans une assiette séparée, il aura une omelette soit de six, soit de cinq œufs, et en tout cas, une assiette de plus à laver; en jetant l'œuf sans tarder, il se résigne à ne consommer que cinq œufs. On a reconnu le début peu héroïque du maître livre de Savage, *The Foundations of Statistics* (1972). Il ne faudra pas moins d'une théorie mathématique entière pour approfondir le problème de Léonard.

L'exemple illustre les trois concepts d'*état de la nature*, de *conséquence* et d'*acte* qui, avec celui de *préférence*, discuté plus loin, forment les éléments primitifs de la théorie.<sup>11</sup> Loin d'être seulement illustratif, cet exemple joue le rôle d'un procédé sémantique intégral. En le détaillant avec goguenardise, Savage annonce en fait à ses lecteurs: 'Je n'ai rien de plus à dire sur l'interprétation des états de la nature, des conséquences et des actes que ce que vous savez déjà quand vous employez ces concepts'. Le sixième œuf est pourri, le sixième est frais: voilà pour les états de la nature. Léonard mangera une omelette avec tant d'œufs ou aucun, et il aura ou n'aura pas d'assiette supplémentaire à nettoyer: voilà pour les conséquences. Tout au plus Savage fait-il un pas d'abstraction quand il définit les actes comme des *applications*, au sens mathématique habituel, qui vont des états du monde vers les conséquences. Même cette modeste trouvaille s'inscrit dans la configuration du bon sens. Elle repose sur une idée de l'action qu'on peut dire *conséquentialiste*: deux actes qui, en chaque état de la nature, produisent les mêmes conséquences ne sont pas distincts l'un de l'autre. Les philosophes et les théoriciens de la décision eux-mêmes ont débattu ailleurs du conséquentialisme.<sup>12</sup> Au vu des objections, il faut sans doute lui dénier une validité générale, mais il apparaît aussi qu'il devient très acceptable si l'on s'occupe de décisions aussi triviales que celle de Léonard: jeter le sixième œuf, ce n'est rien d'autre, pour ce qui lui importe en fait, que de manger une omelette à cinq œufs, et ainsi de suite pour les autres actes.

Le début des *Foundations* incline donc vers la thèse (T1). La théorie de la décision a sélectionné un petit nombre de concepts parmi ceux que le langage ordinaire emploie comme synonymes imparfaits – par exemple, elle a retenu l'acte et non pas l'action, la conséquence et non pas le résultat, l'état de la nature et non pas la situation. Dans la confusion d'idées qui entoure chaque terme, elle fait à nouveau des choix sans ajouter beaucoup qui lui soit propre. Une fois qu'elle a décanté le matériau sémantique, elle entreprend de le symboliser et de le formaliser, et là réside son véritable apport.

Mais il suffit de faire un pas de plus dans les *Foundations* pour mettre (T1) en difficulté. Savage est parvenu au concept de préférence, qui va conditionner toute sa construction axiomatique. Il l'identifie pratiquement à celui de choix: préférer  $f$  à  $g$ , cela 'signifie' – c'est le terme qu'il emploie – que, si l'on a le choix entre  $f$  et  $g$  seulement, on se décide pour  $f$  (1972: 17). Même si l'on restreint l'objet de la préférence aux actes, ce que Savage fera tout au long, elle ne se ramène pas à l'opération du choix. Non seulement elle constitue un état mental, affectif et dispositionnel, alors que le choix relève aussi du comportement observable, de sorte que les significations diffèrent, mais il n'est pas même vrai qu'on puisse – extensionnellement – associer à toute préférence binaire un choix déterminé, et vice-versa. C'est une expérience banale que de choisir entre deux actes qui sont indifférents ou entre lesquels on n'arrive pas à fixer sa préférence; et il est tout aussi banal d'accomplir des actes auxquels on préfère un autre qui était aussi réalisable. En fait, Savage adopte une sémantique particulière, dite ailleurs des *préférences révélées*, qui est en conflit partiel avec la sémantique ordinaire.<sup>13</sup> Elle ne la contredit pas intégralement, car nul ne contesterait qu'une préférence entre des actions doive s'exprimer de quelque manière par des choix, mais elle la simplifie au point de la dénaturer. On ne peut pas prétendre, cette fois-ci, que la théorie se contente de retenir une signification parmi celles qui existent; elle emprunte à plusieurs pour en composer une qui lui soit propre. Le volontarisme de la sémantique justifie que Savage l'introduise par une stipulation explicite, quasi-formelle, et non par voie d'exemple, comme il le faisait avec les autres concepts.

Faisant un pas de plus dans l'énoncé des concepts, Savage introduit enfin celui qu'il a contribué par dessus tout à fonder axiomatiquement, la *probabilité subjective* (il préfère dire 'personnelle', mais l'appellation n'est pas restée). Comme Ramsey et de Finetti, dont il s'inspire, il conçoit la probabilité non pas comme la valeur-limite d'une fréquence, non pas comme une proportion dictée par l'analyse logique des cas, non pas comme la mesure d'une tendance physique à produire un effet, mais comme une *mesure des degrés de la croyance subjective*. Suivant toujours la démarche de ses prédécesseurs, il va rechercher dans l'observation des choix de l'agent le moyen d'attribuer des valeurs numériques de probabilité à ses croyances. La construction axiomatique entière, faite de huit postulats et d'un théorème de mathématique avancée qui en tire les implications, vise à préciser sous quelles hypothèses préalables de rationalité les choix de l'agent deviennent ainsi révélateurs. Mais avant de suivre Savage dans la réalisation de son objectif théorique, il vaut la peine de rappeler, à la manière familière qui est la sienne, l'heuristique fondamentale du subjectivisme en probabilité.

Supposons qu'un observateur veuille déterminer quelle est la probabilité que Léonard attache aux deux états de la nature 'le sixième œuf est pourri', 'le sixième œuf est frais', et que, avec bon sens, il se méfie des réponses à l'emporte-pièce que Léonard pourrait lui donner s'il l'interrogeait directement. Il lui reste un recours, qui est d'éprouver la croyance de Léonard *sur les paris que celui-ci accepterait de prendre*. La définition savagienne des actes, qui a pu sembler formelle, répond en fait à cette idée concrète et même parfaitement terre-à-terre. Dans le cas d'espèce, l'observateur sélectionnera des actes entre lesquels il étudiera les préférences de Léonard, par exemple ces deux-ci: Léonard touche \$1 si l'œuf est pourri et rien sinon; Léonard touche \$1 si l'œuf est frais, et rien sinon. On voit que ces actes, définis à la manière savagienne, représentent des paris et

que, si Léonard exprimait sa préférence pour le premier des deux, l'observateur aurait motif à conclure que Léonard juge l'état 'l'œuf est pourri' plus probable que 'l'œuf est frais' (une préférence opposée de la part Léonard menant à la conclusion opposée). Avant toute extrapolation vers des idées plus ambitieuses, le formalisme des états, des conséquences et des actes, chez Savage, doit se comprendre par rapport à cette heuristique élémentaire du subjectivisme en probabilité.

Reconsidérée à sa lumière, la sémantique de la préférence révélée heurte celle de la préférence ordinaire moins vivement qu'il ne semblait tout d'abord: quand les actes sont des paris, et plus précisément des paris dotés d'enjeux monétaires, les raisons qui poussent à disjoindre les préférences et les choix hypothétiques perdent certainement de leur force. Les choix sans préférence deviennent rares parce que les paris en général ont une structure transparente d'objets conditionnels et que leurs enjeux monétaires facilitent encore les comparaisons entre eux. L'indifférence, qui par nature ne se traduit pas en choix, n'est plus qu'un embarras secondaire, parce qu'il suffit de faire varier les enjeux pour qu'elle se transforme en préférence d'un côté ou de l'autre. Quant aux choix contre la préférence, ils se produisent souvent lorsque le conséquentialisme est déjoué, ce qui ne devrait pas se produire ici, tant l'acte paraît adéquatement décrit par ses conséquences en chaque état. En définitive, la sémantique de la préférence heurte non pas les idées communes sur les paris, mais la dissimilitude que celles-ci impliquent entre un cas d'application aussi favorable et les autres. Simple remarque à ce point, le diagnostic influencera l'analyse de la section 4.

La marche vers la probabilité subjective, chez Savage, se réalise en deux étapes qu'il distingue beaucoup plus nettement que ses prédécesseurs.<sup>14</sup> Supposons que Léonard préfère que le montant de \$1 porte sur 'l'œuf est pourri' plutôt que sur 'l'œuf est frais'. L'observation de cette préférence ne peut révéler dans la meilleure hypothèse qu'un *degré qualitatif de croyance*: Léonard juge le premier état plus probable que le second. Savage (1972: 28) semble prêt à considérer que cette étape relève encore du sens psychologique commun. Admettons cette interprétation et poursuivons le raisonnement subjectiviste, qui est encore incomplet. La deuxième et dernière étape consiste à *quantifier* le degré de croyance obtenu dans l'expérience précédente. Pour cela, il faut à la fois plus de données empiriques et plus de théorie. Supposons que Léonard se voie demander non plus sa préférence entre les paris, mais d'indiquer – c'est une autre forme de préférence exprimée – le prix supérieur qu'il verserait pour participer au premier pari; soit \$x ce montant. Si l'on désigne par  $p$  la probabilité que Léonard donne à l'état 'l'œuf est pourri' et par  $u(x)$ ,  $u(1)$ ,  $u(0)$  l'utilité ou la satisfaction que lui procurent, respectivement, les sommes de \$x, \$1 et \$0, le modèle de l'espérance d'utilité implique:

$$p.u(1) + (1 - p).u(0) = u(x),$$

d'où vient l'équation de la probabilité qu'on recherche:

$$(*) . p = \frac{u(x) - u(0)}{u(1) - u(0)} .$$

On voit que la probabilité dépend de la fonction d'utilité  $u(\cdot)$ , qui est elle-même inconnue de l'observateur. Le problème initial de révélation, portant sur les probabilités, s'est compliqué d'un autre, portant sur l'utilité. Ramsey avait mis en relief cette difficulté, variante formelle du problème explicatif de Davidson, et il tente de la surmonter par ses postulats; à son tour, Savage la reprendra pour lui donner une solution axiomatique. Mais à ce point, nous ne voulons que dégager l'heuristique la plus élémentaire du subjectivisme, et nous nous satisferons d'un dénouement brutal. Il apparaît que l'équation (\*) porte sur des *variations* de  $u(\cdot)$ . Nous pouvons supposer, dans le cas d'espèce, qu'elles sont assez faibles pour être approchées par les variations correspondantes des montants monétaires; d'où l'évaluation finale  $p \approx x$ . Le raisonnement vient d'atteindre son but, qui était de donner une traduction probabiliste à la croyance de Léonard.<sup>15</sup>

Quoi qu'il en soit de la première étape, celle qui fixe les degrés qualitatifs de croyance, la seconde étape, qui est numérique, est inassimilable par la thèse (T1). Le compte-rendu psychologique ordinaire, tel qu'on le rencontre par exemple dans la narration historique, invoque tantôt la croyance *absolue*, tantôt la croyance *relative*, mais la croyance *numérique* ne s'y rencontre pour ainsi dire jamais. On lira: 'Napoléon jugeait probable que Grouchy le rejoindrait sur le champ de bataille de Waterloo', ou encore: 'Lorsqu'il vit marcher sur lui une armée inconnue, Napoléon pensa que Grouchy, plus probablement que Blücher, rejoignait la bataille'. Mais il faudra scruter longtemps le corpus napoléonien pour y trouver: 'Napoléon accordait quatre chances à l'arrivée de Grouchy contre une seule à celle de Blücher'. Et même si l'on rencontrait cette affirmation, l'historien irait-il jusqu'à prétendre que le compte des chances par l'Empereur obéissait aux lois contraignantes du calcul des probabilités? Il y a beaucoup de raisons, y compris expérimentales, de penser que la psychologie du sens commun ne se sente pas liée par l'axiome d'addition des probabilités quand elle combine des attributions de croyance multiples.<sup>16</sup>

Qu'en est-il maintenant de la première étape du raisonnement subjectiviste? C'était peut-être une interprétation prématurée que de la concéder au sens commun. Certes, l'aspect comparatif de la croyance de Léonard – il croit à un état plus fortement qu'à l'autre – ne fait pas à lui seul une différence pertinente; on rencontre de tels énoncés dans la narration historique – nous en avons cité un – et, tout simplement, dans la vie courante. Mais il est original, d'une part, *de ne considérer jamais de croyances absolues*, et d'autre part, *de relier la croyance relative à la comparaison entre deux actes*. Quand il relie l'action et la croyance ou le désir, le raisonnement attributif ordinaire porte sur *une* action et *un* couple de croyance et de désir, couple qui se ramène à une composante unique lorsque l'autre est déjà connue ou trop évidente pour être mentionnée. Si l'on se demande pourquoi le conférencier n'utilise pas de présentation PowerPoint, on dira, par exemple, qu'il croit que les exposés philosophiques interdisent ce genre de moyens simplificateurs, ou qu'il désire épater son public par la lecture d'un texte incompréhensible, ou que les deux raisons, la croyance et le désir, ont joué à la fois. Il se peut qu'on pointe finalement la vraie raison, qui est son manque d'appétence pour la technologie; mais si l'on n'approfondit pas l'attribution, et le cas ne justifie pas qu'on le fasse, elle aura simplement relié une action avec une raison ou, au plus, deux raisons. La conception de la probabilité subjective ne procède pas ainsi, et ses attributions répondent toujours au type: le conférencier a choisi l'acte *a* plutôt que l'acte *b*, donc il accorde plus de probabilité à l'état *s* qu'à l'état *t*, sachant qu'il désire plus la conséquence *x* que la conséquence *y*.

Nous venons de défendre l'idée que la quantification probabiliste de la croyance et, déjà même, la détermination de ses rapports qualitatifs excèdent le sens psychologique commun. C'était une manière d'identifier les 'perfectionnements substantiels' dont (T2) et (T3) s'accordent à reconnaître l'existence, et par là, d'en finir avec (T1), décidément trop simpliste. L'analyse s'est déroulée au plan des concepts, à la manière de Davidson, mais elle se transporte si besoin est au plan des propositions, où Lewis plaçait (T1): la conception de la probabilité subjective consiste si peu en 'platitudes' qu'elle rompt dès sa première étape technique avec le mode d'expression ordinaire.

L'analyse précédente suffit au but critique, mais elle pourrait être élargie comme il était indiqué à la fin de la section 2. Ce qui a été dit pour la croyance et la probabilité a son répondant pour les deux concepts de *désir* et d'*utilité*. En première analyse, le second, qui porte sur les conséquences des actes, retraduit le premier en langage technique lorsque l'un et l'autre se disent d'un même objet. Mais cette analyse est grossière. Non seulement l'utilité numérique échappe au sens commun tout autant que la probabilité numérique, mais, suivant l'analyse reçue en théorie de la décision, la fonction d'utilité ne permet que des évaluations comparatives.<sup>17</sup> Ainsi, elle n'est pas en mesure de transcrire des phrases sur le désir aussi banales que 'je désire que tu viennes'. Les répondants théoriques les plus proches diraient en substance 'je désire que tu viennes plus que je ne désire que tu ne viennes pas' ou 'il y a beaucoup de choses que je désire moins que ta venue', ce qui distord la signification.

Certains concepts techniques de la théorie de la décision ne sont pas même liés à leurs répondants ordinaires de cette manière distendue que nous venons de manifester. Ceux de *risque* et d'*attitude à l'égard du risque* sont dans ce cas, même s'ils appellent une réserve d'ordre historique: parfois, le sens commun finit par recevoir des acceptions spécialisées qui s'opposaient initialement à lui. S'agissant de deux actes à conséquences monétaires, *a* et *b*, les probabilités subjectives étant déterminées, on dira que *b* est plus risqué que *a* si *b* met plus de masse de probabilité que *a* sur les conséquences extrêmes, favorables ou défavorables, tout en conservant l'espérance mathématique de *a*.<sup>18</sup> En dehors des milieux professionnels où s'exerce déjà l'emprise de la théorie, une telle définition heurte le jugement spontané, car celui-ci tend à se fixer sur les seules conséquences défavorables quand il apprécie le risque de deux options. De même, les définitions techniques de l'aversion ou du goût pour le risque se rattachent mal aux catégories de la psychologie morale, et il est d'autant plus difficile d'établir une correspondance entre ces deux groupes que le langage courant sépare mal le risque lui-même – propriété objective et de type statistique – des attitudes subjectives que l'on a par rapport à lui. Mais la stratégie consistant à pointer des concepts techniques ignorés par le sens commun n'est pas la nôtre dans cette étude. Nous en avons dit beaucoup déjà pour écarter (T1), et il est urgent de commencer à différencier (T2) et (T3).

#### 4 Un principe à la fois spécifique et perfectible

Pour faire progresser la distinction entre (T1) et (T2), nous chercherons à mieux cerner la conception subjectiviste de la probabilité. La section précédente a mis en exergue la classe des préférences qui s'expriment sur des paris; or pour les subjectivistes, cette classe apparemment spéciale représente l'ensemble des préférences relatives à l'incertain.

Telle est l'affirmation que nous voulons discuter maintenant. Le *principe d'universalité des paris*, comme nous le nommerons, apparaîtra finalement, d'une part, étranger au sens commun, d'autre part, inacceptable comme tel, mais susceptible d'être perfectionné.

Ramsey l'énonce avec la candeur admirable des pionniers:

[The present theory] is based fundamentally on betting, but this will not seem unreasonable when it is seen that all our lives we are in a sense betting. (1931, in Gärdenfors & Sahlin: 36)

L'énoncé critique *all our lives we are in a sense betting* est lâché aussi crûment que possible, et la tournure qui sert à l'introduire, *it is seen*, est d'un dogmatisme consommé: chacun de nous devrait se rendre à l'évidence de ce qu'est la vie humaine. A la décharge de Ramsey, on se demande bien quelle preuve il pouvait invoquer en faveur d'une affirmation aussi vague que la sienne. Il pouvait la comparer, la paraphraser, l'illustrer, mais s'il voulait à proprement parler des arguments, les meilleurs consistaient encore à la développer par ses conséquences. Le principe d'universalité des paris est de ces propositions fondatrices qu'on rencontre dans les disciplines théoriques, sources d'implications empiriques, mais non pas empiriques elles-mêmes, défendables par ce qu'elles permettent de faire et non par ce qu'elles sont, en un mot *métaphysiques* dans une certaine acception du concept.

On chercherait en vain à quoi rattacher le principe dans le formalisme de Ramsey. C'est que le fondateur de l'école subjectiviste accomplit moins, par ce formalisme, que ses commentateurs élogieux l'ont fait accroire par la suite. Il surmonte le problème, certainement délicat, de dégager un concept numérique de 'valeur' – l'utilité dans notre terminologie – qui permette ensuite de mesurer la croyance comme le veut le calcul des probabilités. L'équation (\*) de la section précédente, que nous avons résolue par un subterfuge grossier, trouve chez lui une solution théoriquement correcte; mais celle-ci n'est valable que si les états de la nature sont décrits par des propositions 'éthiquement neutres'. Sans entrer dans une définition technique, il nous suffira de dire que la restriction ne se justifie à ses yeux que par la commodité.<sup>19</sup> Or si le principe d'universalité des paris devait entrer dans son formalisme, ce serait au stade démonstratif où la probabilité subjective se révélerait à *propos de tous les états de la nature* – c'est-à-dire aussi pour les propositions qui ne sont pas 'éthiquement neutres'. Non seulement il ne procède pas à cette étape de généralisation, mais il n'en indique pas clairement la nécessité.

Sur ce point comme sur d'autres, Savage domine son prédécesseur. Chez lui non plus, le principe d'universalité des paris ne figure pas directement dans le formalisme, mais on peut lui rattacher deux postulats, (P3) et (P4), qui reviennent à le décomposer. Toute préférence entre des actes dont les conséquences dépendent d'un état incertain peut s'analyser comme un pari: tel serait le principe rendu en langage savagien. Mais comme les paris n'importent que pour autant qu'ils révèlent des probabilités subjectives, il vaut mieux que l'énoncé du principe ne les mentionne pas et les remplace par les conditions qui les rendent révélateurs. Tel est le rôle précisément dévolu à (P3) et (P4). Posés relativement à des actes quelconques, ces deux postulats reviennent bien à dire 'tout acte est un pari', mais de la manière instructive qui fait connaître, en même temps, pourquoi les paris sont des modèles à rechercher pour les actes.

(P3) demande que les préférences de l'agent sur les conséquences d'un acte quelconque ne soient pas influencées par l'état de la nature qui se réalise. Il faut par exemple que, relativement aux autres conséquences, Léonard apprécie l'omelette à cinq œufs exactement de la même manière dans les deux états 'le sixième œuf est pourri' et 'le sixième œuf est frais'. L'exigence vise à exclure un obstacle possible à la révélation des probabilités par les choix d'actes: si Léonard, par exemple, déprécie l'omelette à cinq œufs quand il constate l'état 'le sixième œuf est pourri', *il peut très bien le juger plus probable que l'autre et cependant ne pas se prononcer en faveur de l'acte qui lui attribue la meilleure conséquence.*

On formalise une telle situation mentale en supposant que la fonction d'utilité a deux variables, la conséquence, d'une part, et l'état de la nature, de l'autre: elle s'écrit  $u(s,c)$  au lieu de  $u(c)$ . On parle dans ce cas d'une *fonction d'utilité dépendante des états (state-dependent utility function)*.<sup>20</sup> Pour aller au plus simple, supposons que Léonard ne soit sensible à la différence d'une omelette à cinq œufs, notée par 1, avec la perte intégrale de l'omelette, notée par 0, que si l'œuf est frais; un œuf pourri le jette dans un dégoût tel qu'il ne différencie plus les deux conséquences. En écrivant  $s_1 =$  l'œuf est pourri et  $s_2 =$  l'œuf est frais, on a donc:

$$u(s_1,1) = u(s_1,0) \text{ et } u(s_2,1) > u(s_2,0).$$

Comparons l'acte  $a$  qui attache 1 à  $s_1$ , 0 à  $s_2$ , et l'acte  $b$  qui attache 1 à  $s_2$ , 0 à  $s_1$ . Si  $p$  désigne la probabilité de  $s_1$ , l'évaluation par l'espérance d'utilité donne respectivement:

$$p.u(s_1,1) + (1-p).u(s_2,0) \text{ pour } a \text{ et } p.u(s_1,0) + (1-p).u(s_2,1) \text{ pour } b.$$

Sous l'hypothèse précédente, la comparaison ne dépend que des seconds termes, et  $b$  l'emportera toujours sur  $a$ , si grand que soit  $p$ , sauf dans le cas-limite  $p = 1$  où les deux paris deviennent équivalents. Si, donc, Léonard exprimait à l'observateur sa préférence pour  $b$ , celui-ci aurait bien tort d'en tirer la moindre indication de probabilité. (P3) écarte d'autorité toutes les configurations de ce genre.

Dans le même esprit, (P4) demande que les choix de l'agent entre deux actes dotés chacun de deux conséquences restent identiques lorsqu'on redéfinit les conséquences tout en conservant leur ordre de préférence antérieur. L'exigence vise à exclure un autre empêchement à la révélation: si Léonard se prononce en faveur de  $b$  contre  $a$ , lorsque les conséquences sont fixées comme précédemment, mais à l'opposé, en faveur de  $a$  'contre  $b$ ' quand on remplace les conséquences notées 1 et 0 par d'autres classées de la même manière, quelle conclusion l'observateur pourrait-il bien tirer sur l'état que Léonard juge le plus probable?

Comme nous l'avons annoncé, on peut lire de deux façons les postulats (P3) et (P4). D'une part, ils servent – modestement – à élucider les conditions faisant que les préférences entre des paris ordinaires indiquent des probabilités; d'autre part, ils affirment – de manière grandiloquente – que ces conditions favorables sont réalisées à propos de tous les actes, et c'est en quoi ils se rattachent au principe d'universalité des paris. Ce n'est pas le moindre intérêt de l'axiomatique de Savage qu'elle formalise non seulement le principe, mais les raisons de le rejeter: il suffit d'énoncer (P3) et (P4) pour que

des contre-exemples sautent à l'esprit! Comme on l'a vite remarqué, (P3) obligerait un agent qui prend un risque mortel à évaluer identiquement toute somme d'argent donnée, soit qu'il survive, soit qu'il disparaisse.<sup>21</sup> Et (P4) se heurte à l'évidence que l'on peut vouloir limiter les sommes qu'on risque sur la réalisation d'un événement incertain. Les deux objections n'ont pas la même force, car l'une pointe une conséquence manifestement irrationnelle du postulat considéré, tandis que l'autre dit que sa violation ne constitue pas une irrationalité manifeste. Mais prises ensemble, elles incitent à rejeter la théorie savagienne initiale au profit d'une généralisation plus appropriée. La théorie contemporaine de l'utilité dépendante des états s'est précisément donné pour tâche d'en produire une.

Mais avant d'aborder ce développement, il faut se montrer équitable à l'égard du principe d'universalité des paris, affirmation douteuse, mais utile et importante pour ce qu'il permet à la théorie d'accomplir. De fait, sous la forme de (P3) et (P4), il est, en présence des autres postulats, nécessaire et suffisant pour que les préférences de l'agent se représentent par une fonction d'évaluation qui a la forme ordinaire de l'espérance d'utilité et qui est, à quelque chose près, seule de son espèce. La mesure de probabilité qui intervient dans l'espérance est absolument unique, ce qui justifie de la traiter comme la probabilité subjective de l'agent. C'est le théorème de Savage, fondamental parce qu'il caractérise l'incertitude naturelle probabilisable comme si elle était déjà probabilisée.<sup>22</sup> Les postulats trouvent leur meilleure justification dans la réorganisation conceptuelle que le théorème provoque: un type d'incertitude se ramène à l'autre, et le modèle naïf des loteries, chez von Neumann et Morgenstern, qui place les probabilités dans l'objet même de la préférence, se voit enrichi d'un potentiel de généralité qu'il n'avait pas en lui-même. *Les économistes pourront désormais appliquer le modèle de l'espérance d'utilité au comportement d'agents quelconques sans avoir à se préoccuper de la nature des probabilités qu'ils leur attribuent.* Que ces probabilités fassent partie du choix ou qu'un observateur les ait révélées, il n'importe plus, et la théorie financière ou la théorie de l'assurance peuvent aller de l'avant. Avec cette analyse, le principe d'universalité des loteries apparaît plus clairement comme nous l'avions qualifié, c'est-à-dire proche des hypothèses métaphysiques dont les sciences ont besoin pour se développer. Nous ne prétendons pas que, à l'instar de certaines de ces hypothèses, il ne provienne pas du sens commun; mais comme dans leur cas, la source est assez lointaine pour qu'il devienne factice d'y insister.<sup>23</sup>

Après ce diagnostic, les thèses (T2) et (T3) commencent à se différencier. Il apparaît que les principes de la théorie de la décision ne tirent pas tous dans le même sens. Certains parlent sans doute en faveur de l'idée qu'elle approfondit, renforce, systématise, la psychologie du sens commun; mais d'autres parlent de leur divergence et – si l'on peut reprendre l'expression connotée – de rupture épistémologique. Nous regarderons comme établi à ce point que la théorie a développé des principes qui lui sont propres. Mais une partie de la thèse (T3) reste à défendre contre (T2): avec ces principes spécifiques, réussit-elle à dépasser les limitations du raisonnement attributif ordinaire? Et qu'en est-il, particulièrement, de la sous-détermination des raisons par les données, qui, d'après Davidson, affecte encore la théorie au point de compromettre son projet scientifique?

Léonard et le subjectiviste qui l'observe nous aideront à éprouver et, finalement, à rejeter cette objection célèbre. Nous avons compris que les préférences de Léonard

pouvaient être contaminées par les valeurs qu'il attribue aux conséquences en fonction des états de la nature. Cette éventualité aggrave l'indétermination première du résultat et, suivant les apparences, consacre le triomphe de Davidson. Mathématiquement, si l'on écrit la fonction d'évaluation sous la forme dépendante des états:

$$\sum_{s \in S} p(s) \cdot u(s, a(s))$$

au lieu de:

$$\sum_{s \in S} p(s) \cdot u(a(s)),$$

on accroît strictement l'ensemble des couples de fonctions  $(p, u)$  que l'on peut attribuer à Léonard au vu des préférences qu'il manifeste à propos des actes  $a$ . Les spécialistes post-savagiens ont tenté de répondre à ce nouveau problème en élargissant à la fois les données disponibles et le traitement théorique. L'une des constructions représentatives demande que Léonard exprime à son observateur des comparaisons nouvelles en sus de ses choix hypothétiques. Dans le questionnaire subjectiviste antérieur, il devait seulement dire, pour chaque couple d'actes  $a$  et  $b$  qu'on lui propose d'envisager, s'il choisirait  $a$  ou  $b$ . Dorénavant, il indiquera aussi, pour chaque couple de conséquences  $(c, d)$  et pour chaque couple d'états  $(s, t)$ , s'il préfère  $c$  en  $s$  ou  $d$  en  $t$ .

Pour illustrer quel rôle joue cette information, nous considérerons à nouveau les préférences de Léonard entre les actes  $a$  et  $b$  qui ont été définis plus haut, en faisant toujours l'hypothèse qu'il apprécie les conséquences différemment selon les états. Il ne suffit donc plus à l'observateur de lui poser la question familière:

Que choisiriez-vous, de  $a$  (qui attache 1 à  $s_1$  et 0 à  $s_2$ ), ou de  $b$  (qui attache 0 à  $s_1$  et 1 à  $s_2$ )?

Le subjectiviste ajoutera cette question:

Comment s'établissent vos préférences entre (0 en  $s_1$ ), (0 en  $s_2$ ), (1 en  $s_1$ ) et (1 en  $s_2$ )?

Un calcul très simple montre que, dans certaines configurations de réponse, il est possible de décider si Léonard attache plus de probabilité à  $s_1$  ou à  $s_2$ .<sup>24</sup> Cette indication fait comprendre que l'information supplémentaire est utile. La construction technique à laquelle nous pensons étend le questionnaire plus loin encore et parvient ainsi à déterminer totalement la probabilité subjective.<sup>25</sup>

Nous avons formulé la première question du subjectiviste dans le langage du choix hypothétique et la seconde dans celui de la préférence. La première est compatible avec les deux modes, mais il *n'y a que celui de la préférence qui convienne à la seconde*: on ne parle normalement pas de choix si les options à comparer ne sont pas réalisables simultanément, et c'est le cas, par exemple, de (1 en  $s_2$ ) et (1 en  $s_1$ ). Les théoriciens de la décision qui sont fidèles à la sémantique de la préférence révélée pensent que seule la première question est légitime, et récusant donc la seconde, ils s'interdisent de réviser les postulats de Savage pour qu'il soit possible d'en tenir compte.<sup>26</sup> Notre analyse est

différente: il nous semble que, malgré sa concordance acceptable dans le cas des paris, la sémantique des préférences révélées n'appartient pas à l'heuristique fondamentale du théoricien subjectiviste; elle n'appartient qu'à son heuristique *élémentaire*. Il doit viser à la révélation des probabilités coûte que coûte, et il n'y a pas de raison qu'il se dispense de l'information fournie par les préférences dès lors qu'elles ne s'expriment plus comme des choix hypothétiques. Si la théorie peut tirer parti de l'information que procurent certaines préférences interprétées à la manière ordinaire, qu'elle aille donc de l'avant! Le subjectiviste pourra ainsi réviser (P3) et (P4), c'est-à-dire, le principe d'universalité des paris, dont l'inadéquation s'était manifestée au grand jour avec la critique de ces postulats.

Le nouveau système de postulats sera complexe parce qu'il devra, en substance, couvrir deux classes d'informations tout en les coordonnant. D'une part, les préférences de Léonard sur les objets du type *c* en *s* ou *d* en *t* vérifieront des conditions de rationalité qui sont parallèles à celles qui règlent déjà les préférences entre les actes. D'autre part, les deux classes de préférences entretiendront certaines relations de cohérence entre elles. L'enjeu du travail axiomatique est de définir la cohérence de manière qu'elle ait la force exacte pour que les données de préférences considérées dans leur ensemble révèlent, intégralement et de manière unique, comme chez Savage donc, la probabilité subjective de l'agent.

Aux yeux de Davidson, la construction précédente illustrerait la course entre la théorie et les données qui est à ses yeux le mauvais lot de la discipline. Chaque fois que l'on s'autorise à recueillir de nouvelles données chez l'agent, il faut renforcer les liaisons qu'on suppose entre les opérations mentales qu'il effectue; en bref, *il faut lui supposer plus de rationalité*. C'est exactement ce qui l'on constate ici: Léonard doit savoir répondre sur les actes d'une manière qui s'accorde avec ses réponses sur les états et les conséquences, et l'exigence va plus loin que celle que Savage imposait au classement des actes seuls. Davidson souligne à juste titre qu'on n'abandonne pas le béhaviorisme sans acquitter une contrepartie méthodologique: si l'on s'autorise à exploiter les données verbales en sus des comportements, on s'oblige du même coup à importer une théorie sémantique dans la théorie de la décision. Dans l'analyse qui est faite ici, la condition de rationalité se renforce dans le registre même où elle se trouvait, celui de la préférence, et le saut que Davidson annonce vers la sémantique ne se produit pas. Mais substantiellement, les deux idées se rejoignent, car, une fois le béhaviorisme abandonné, il faut postuler de quelque façon que les paroles de l'agent fournissent une information véridique à l'observateur.

Le désaccord avec Davidson, que nous prenons comme le meilleur avocat de (T2), ne concerne pas l'observation que la théorie doit grandir en même temps que les données, mais seulement la conclusion qu'il en tire défavorablement à la théorie. Certes, en s'accroissant, les exigences de rationalité deviennent de moins en moins plausibles. Mais elles font gagner en plausibilité par ailleurs. La construction que nous venons d'évoquer se répercute sur le principe d'universalité des paris. Avec (P3) et (P4), Savage lui donnait une forme inadéquate; or le nouveau système, contraignant par ailleurs, tient compte des violations de ces deux postulats; en les affaiblissant, il rend moindres les conditions qu'un acte doit satisfaire pour compter comme l'analogie d'un pari; ainsi le principe devient beaucoup plus acceptable. Il était difficile de sauver (P3) et (P4) comme des postulats de rationalité à rapprocher, par exemple, de (P1), qui impose aux préférences

d'être totales et transitives.<sup>27</sup> Savage restreignait sans le dire le *contenu* des désirs, donc les situations mentales et les individus mêmes auxquels il destinait sa théorie. Cette restriction fait maintenant place à une autre, d'après laquelle une forme d'intégration régènerait entre différentes activités mentales de l'agent. La restriction initiale portait sur l'extension – le nombre de situations couvertes – et la restriction finale porte sur l'intension – la manière de comprendre la rationalité. Sans importer ici un appareil lakatosien, nous croyons pouvoir conclure que l'arbitrage est *théoriquement progressif*.

Ce qu'on lit chez Davidson n'irait pas nécessairement à l'encontre de cette conclusion; en revanche, il n'y a pas de doute qu'à ses yeux, les développements précédents ne seraient pas empiriquement progressifs. Peut-être ces travaux ne font-ils pas avancer la *mesure* de la croyance au-delà de ce qu'accomplissaient déjà Ramsey et Savage; mais la réussite de la mesure n'est pas le tout de l'appréciation empirique, et c'est précisément ce que Davidson omet de reconnaître. Dans le béhaviorisme et sa variante opérationnaliste, toute l'activité empirique de la science est subordonnée à l'objectif de mesurer les concepts puisque, suivant la thèse centrale de cette école, la mesure fixe les interprétations. Que Davidson rejoigne la conclusion méthodologique du béhaviorisme paraît d'autant plus incongru qu'il a des thèses sémantiques bien différentes à faire valoir.

Nous avons dit que Savage avait influencé les économistes en libérant chez eux l'utilisation de la formule classique de l'espérance d'utilité  $\sum_{s \in S} p(s)u(x(s))$ . On pourrait en dire autant, avec certaines précautions, des théoriciens de l'utilité dépendante des états et de leur formule plus satisfaisante  $\sum_{s \in S} p(s)u(s, x(s))$ . Ces formules peuvent avoir un contenu empirique en l'absence d'une mesure effective pour  $p(\cdot)$ . Les économistes sont passés maîtres dans l'exploitation logique des contraintes simplement structurelles posées par leurs équations. Ainsi, dans la théorie de l'assurance, où les deux formules se rencontrent, ils leur associent des conséquences testables qui dépendent seulement de l'existence et de l'unicité de  $p(\cdot)$ . Soit la proposition micro-économique, démontrée naguère par Arrow, d'après laquelle une compagnie d'assurance doit proposer des contrats à franchise positive et à remboursement intégral. Elle découle de la formule de l'utilité espérée classique en présence d'hypothèses auxiliaires diverses, mais sans qu'il faille donner la moindre valeur numérique aux probabilités.<sup>28</sup> Cet exemple fait sentir l'arbitraire de la restriction latente chez Davidson, et il s'adresse, comme une mise en garde, à tous les autres défenseurs de (T2): il se peut que le progrès théorique de la théorie de la décision doive se payer d'une stagnation empirique; mais pour fonder un diagnostic aussi pessimiste, il ne suffit pas de suivre la théorie à travers les réponses qu'elle donne à la vieille question de la mesure; il faut examiner les applications empiriques dans toute leur ampleur.

## 5 Envoi

A travers ses trois formes chez Ramsey, Savage et dans la théorie contemporaine, le principe d'universalité des paris, qui appartient spécifiquement à la théorie de la décision, nous est apparu en voie de perfectionnement régulier. Le pessimisme de Davidson, qui admet la spécificité, mais non pas la perfectibilité, nous paraît donc démenti. Nous

n'avons mené la démonstration que sur un seul développement technique; sans doute les théoriciens de la décision l'ont-ils investi d'une importance particulière, mais il n'est qu'un exemple parmi d'autres de la manière dont ils révisent graduellement leurs postulats pour les rendre plus acceptables. Cette souplesse évolutive est inaccessible au sens commun parce qu'elle suppose sur des allées et venues entre les formalismes et le langage ordinaire, alors que le sens commun ne marche que sur une seule jambe. Le caractère de spécificité préparait donc celui de perfectibilité, et (T3) nous semble une meilleure approximation que (T2) du général mouvement qui anime la théorie.

## Remerciements

Ce travail a fait l'objet d'une conférence invitée au Congrès de la Société de Philosophie des Sciences (Paris, 12–14 novembre 2009). L'auteur remercie D Andler, M Cozic, B Hill, P Jacob, A Rosenberg et P Weirich des observations qu'elle leur a suggérées, et il remercie J Baccelli d'avoir bien voulu relire et commenter en détail le manuscrit.

## Financement

La recherche n'a bénéficié d'aucun financement provenant d'organismes publics, privés ou sans but non lucratif.

## Notes

- 1 Une autre étude souligne la dissimilarité des deux groupes (Mongin, 2002).
- 2 Le mot 'bayésien' évoquerait plus naturellement la règle de révision des probabilités qu'on fait remonter à Thomas Bayes, et c'est d'ailleurs ainsi que le prennent les statisticiens et les philosophes des sciences. Cette acception intervient relativement peu en théorie de la décision.
- 3 Un des auteurs des nouveaux développements, Gilboa (2009), vient de les faire entrer dans un manuel, qui est l'un des premiers du genre. On peut se reporter aussi à la synthèse plus ancienne, mais très informée, de Cohen & Tallon (2000).
- 4 Les *théorèmes de représentation*, comme on les nomme, effectuent cette mise en rapport. D'ordinaire, le formalisme symbolique décrit la préférence et ses propriétés, et le formalisme numérique consiste en fonctions d'utilité ou de probabilité, mais les variantes se sont multipliées à partir de ce schéma. Il correspond à une forme singulière d'axiomatisation que Mongin (2003) s'efforce de caractériser.
- 5 Voir Guala (2000, 2005) et Mongin (2009).
- 6 Leurs suffrages vont aussi à Jeffrey (1983), qui continue Ramsey autrement que Savage. Les relations entre les systèmes des trois auteurs sont encore mal comprises. Bradley (2007) a fait l'effort d'unification le plus poussé.
- 7 Wilkes (1991) synthétise agréablement la première décennie de la controverse tout en prenant position contre l'éliminativisme.
- 8 A notre connaissance, Pettit (1991) est le seul qui ait déjà confronté la psychologie du sens commun et la théorie de la décision; si le propos comparatif est le même, les conclusions diffèrent. La question traitée par Satz & Ferejohn (1994), reprise par Hausman (1995), est d'une autre nature, car elle porte sur le contenu psychologique en général d'un corpus plus vaste (la 'théorie sociale'). C'est une autre comparaison encore qui a intéressé Rosenberg: il entend relier la psychologie du sens commun à la micro-économie, voire à toute la discipline économique, en fait pour en proclamer l'échec scientifique (le ch. 5 de son livre de 1992 résume la thèse, déjà en filigrane dans son livre de 1976).

- 9 L'indétermination dont s'occupe Davidson peut évoquer celle dont la philosophie des sciences traite classiquement sous l'intitulé de *problème de Duhem*. Mais il s'agit chez Davidson d'explication et chez Duhem de réfutation empirique, et les deux idées ne sont pas adjacentes dans toutes les conceptions. De plus et surtout, le problème de Davidson est plus fortement spécifié, donc plus informatif, que celui de Duhem.
- 10 Dokic et Engel (2004), ainsi que les auteurs réunis par Galavotti (2006), discutent le système de Ramsey et ses liaisons philosophiques anciennes et récentes. Bradley (2004) l'a formalisé dans un langage mathématique plus actuel, ce qui facilite grandement la comparaison avec Savage.
- 11 Savage dit en fait 'état du monde', *state of the world*, et non pas 'état de la nature', *state of nature*. Notre terminologie emprunte à la théorie des jeux, qui distingue les deux concepts en les rapportant à l'incertitude stratégique et à l'incertitude naturelle respectivement.
- 12 Les philosophes débattent du conséquentialisme en philosophie morale surtout, et ils le présentent alors comme une généralisation de l'utilitarisme. Pour les théoriciens de la décision, il constitue une thèse sur l'action individuelle, et plus précisément, une supposition antérieure aux postulats ordinaires de rationalité. La supposition peut être elle-même déployée axiomatiquement, comme chez Hammond (1988).
- 13 L'expression de *préférences révélées* vient de l'économie, où elle a fait couler beaucoup d'encre. Elle y désigne à la fois une théorie économique particulière, qui porte sur les choix du consommateur, et la conception sémantique de la préférence qui est traitée ici.
- 14 Il y a en fait des raisons mathématiques à cette distinction, outre les raisons conceptuelles qui figurent seules ici.
- 15 Le troisième fondateur du subjectivisme, de Finetti (1937), a pu s'accommoder de cette résolution brutale. Il définit la probabilité comme un rapport de prix mesurés en unités monétaires.
- 16 Les recherches amorcées par Kahneman, Slovic et Tversky (1982) et abondamment développées par la suite, fournissent les arguments expérimentaux, à supposer qu'ils soient nécessaires à la conclusion.
- 17 L'affirmation vaut encore pour les fonctions d'utilité qu'on appelle *cardinales* par opposition aux fonctions *ordinales*, car les unes et les autres se comprennent toujours par rapport à la préférence, même si elles livrent sur elle des informations différentes, et la préférence est par nature un concept relationnel ou comparatif.
- 18 La définition rigoureuse fait intervenir les fonctions de répartition induites par  $a$  et  $b$ , étant donné les probabilités subjectives.
- 19 Le passage décisif est à la p. 33 de l'article 'Truth and Probability'.
- 20 L'expression est légèrement équivoque parce que l'utilité dépend des états même dans la formulation canonique: si  $a$  est l'acte considéré, on a  $u(c) = u(a(s))$ . On veut évidemment dire que la réalisation de l'état influence *directement* l'évaluation.
- 21 Voir la discussion ancienne d'Aumann et Savage reproduite chez Drèze (1987) et les commentaires qu'ajoute cet auteur.
- 22 Des variantes plus générales, aujourd'hui, évitent de supposer le modèle de l'espérance d'utilité pour dériver une mesure de probabilité unique. En substance, elles reposent toujours sur (P3) et (P4) (voir Machina et Schmeidler, 1992).
- 23 Savage lui-même souligne les tensions que sa conception de la probabilité implique par rapport à la sémantique ordinaire de cette notion:

A particular interpretation of probability to a person is given in terms of the theory of consistent decisions in the face of uncertainty... Much as I hope that the notion of probability defined here is consistent with ordinary usage, it should be judged by the contribution it makes to the theory of decision, not by the accuracy with which it analyzes ordinary usage. (1972: 27)

- 24 Supposons que Léonard réponde par le pari  $a$ , puis qu'il indique les préférences:  $(1 \text{ en } s_2) > (1 \text{ en } s_1) > (0 \text{ en } s_1) > (0 \text{ en } s_2)$ . Si l'on calcule les espérances d'utilité, on vérifie que Léonard donne la probabilité la plus élevée à  $s_1$ . Cette configuration des préférences est avantageuse sur le plan des informations requises. Il en existe d'autres pour lesquelles l'observateur devrait connaître en outre l'intensité des préférences que Léonard forme sur les quatre objets précédents.
- 25 Karni a développé en grand détail cette construction néo-subjectiviste dont nous ne faisons que donner l'idée sommaire; voir notamment Karni (1985, 1996) et Karni et Mongin (2000). Elle n'a pas l'exclusivité des réponses apportées au problème de l'utilité dépendante des états; voir par exemple celle, tout récente, de Hill (2009).
- 26 Drèze (1987) a pris cette position, d'où est résulté une controverse avec Karni.
- 27 Nous laissons de côté la discussion, à la fois plus complexe et plus familière, du 'principe de la chose sûre' (*sure-thing principle*) affirmé par (P2).
- 28 Karni (1985) développe d'autres implications en matière d'assurance, en partant cette fois de la formule dépendante des états.

## Références

- Bradley R (2004) Ramsey's representation theorem. *Dialectica* 58: 483–498.
- Bradley R (2007) A unified Bayesian decision theory. *Theory and Decision* 63: 233–263.
- Churchland PM (1981) Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78: 67–90.
- Cohen M, Tallon JM (2000) Décision dans le risque et l'incertain: l'apport des modèles non-additifs. *Revue d'économie politique* 110: 631–681.
- Davidson D (1980) *Essays on Actions and Events*. Oxford: Clarendon Press.
- Dokic J, Engel P (2001) *Ramsey: Vérité et succès*. Paris: Presses Universitaires de France. (Trad. anglaise, *Frank Ramsey: Truth and Success*, London: Routledge 2001)
- Drèze J (1987) *Essays on Economic Decisions under Uncertainty*. Cambridge: Cambridge University Press.
- De Finetti B (1937) La prévision: ses lois logiques, ses sources subjectives. *Annales Henri Poincaré* 7: 1–68.
- Fodor F (1987) *Psychosemantics*. Cambridge, MA: MIT Press.
- Galavotti MC (ed.) (2006) *Frank P. Ramsey and the Vienna Circle*. Boston, MA: Springer.
- Gilboa I (2009) *Theory of Decision under Uncertainty*. Cambridge: Cambridge University Press.
- Guala F (2000) The logic of normative falsification: Rationality and experiments in decision theory. *Journal of Economic Methodology* 7: 59–93.
- Guala F (2005) *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.
- Hammond PJ (1988) Consequentialist foundations for expected utility. *Theory and Decision* 25: 25–78.
- Hausman DM (1995) Rational choice and social theory: A comment. *The Journal of Philosophy* 92: 96–102.
- Hill B (2009) Living with state-dependent utilities. *Theory and Decision* 67: 405–432.
- Jeffrey R (1983) *The Logic of Decision*. Chicago, IL: University of Chicago Press (1ère éd. 1954).
- Kahneman D, Slovic P & Tversky A (1982) *Judgment under Uncertainty: Heuristic and Biases*. Cambridge: Cambridge University Press.
- Karni E (1985) *Decision Making under Uncertainty: The Case of State-Dependent Preferences*. Cambridge, MA: Harvard University Press.
- Karni E (1996) Probabilities and beliefs. *Journal of Risk and Uncertainty* 13: 249–262.

- Karni E, Mongin P (2000) On the determination of subjective probability by choice. *Management Science* 46: 233–248.
- Lewis D (1983) *Philosophical Papers, I*. New York: Oxford University Press.
- Lewis D (1986) *Philosophical Papers, II*. New York: Oxford University Press.
- Machina MJ, Schmeidler D (1992) A more robust definition of subjective probability. *Econometrica* 60: 745–780.
- Mongin P (2002) Le principe de rationalité et l'unité des sciences sociales. *Revue économique*, 53: 301–323.
- Mongin P (2003) L'axiomatisation et les théories économiques. *Revue économique* 54: 99–138.
- Mongin P (2009) Duhemian themes in expected utility theory. In: Brenner A, Gayon J (eds) *French Studies in the Philosophy of Science*. New York: Springer, 303–357.
- Pettit P (1991) Decision theory and folk psychology. In: Bacharach M, Hurley S (eds) *Foundations of Decision Theory*. Oxford, Blackwell, ch. 4, 147–175.
- Ramsey FP (1931) Truth and probability. In: Braithwaite RB (ed.) *The Foundations of Mathematics*. London: Routledge & Kegan Paul, 156–198. Rééd. 1988 in: Gärdenfors P & Sahlin NE (eds) *Decision, Probability and Utility*. Cambridge: Cambridge University Press, ch. 1. (Trad. française dans Ramsey FP, *Logique, philosophie et probabilités*. Paris: Vrin, 2003)
- Rosenberg A (1976) *Microeconomic Laws*. Pittsburgh, PA: University of Pittsburgh Press.
- Rosenberg A (1992) *Economics – Mathematical Politics or Science of Diminishing Returns*. Chicago, IL: The University of Chicago Press.
- Satz D, Ferejohn J (1994) Rational choice and social theory. *Journal of Philosophy* 91: 71–87.
- Savage LJ (1972) *The Foundations of Statistics*. New York: Dover Publications (1ère éd. 1954).
- Stich S (1983) *From Folk Psychology to Cognitive Science: The Case against Belief*. Cambridge, MA: Bradford Books & The MIT Press.
- Von Neumann J, Morgenstern O (1947) *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press (1ère éd. 1944).
- Wilkes K (1991) The relationship between scientific psychology and common-sense psychology. *Synthèse* 89: 15–39.

### Biographie de l'auteur

Philippe Mongin est Directeur de recherche de classe exceptionnelle au Centre National de la Recherche Scientifique (CNRS) et Professeur affilié d'économie et de philosophie à l'École des Hautes Études Commerciales (HEC). Tout en travaillant de longue date au CNRS, il a enseigné la théorie de la décision individuelle et collective, ainsi que la philosophie de l'économie, ses deux spécialités principales, dans plusieurs institutions françaises et étrangères, notamment à l'Université de Montréal, à l'Université Catholique de Louvain, à la London School of Economics, au California Institute of Technology et à Brown University. Il a publié des ouvrages collectifs et de nombreux articles, en particulier dans les revues suivantes: *Journal of Economic Theory*, *Journal of Mathematical Economics*, *Revue économique*, *Games and Economic Behavior*, *Management Science*, *Social Choice and Welfare*, *Synthèse*, *Journal of Philosophical Logic*, *Economics and Philosophy*, *History of Political Economy*, *Annales*, *Histoire*, *sciences sociales*.